Kidney failure prediction models: a comprehensive external validation study in patients with advanced CKD

Ramspek, C.L.; Evans, M.; Wanner, C.; Drechsler, C.; Chesnaye, N.C.; Szymczak, M.; ... ; EQUAL Study Investigators

Which prognostic model predicts kidney failure best? A comprehensive external validation study in patients with advanced chronic kidney disease, accounting for the competing risk of mortality

Running title: kidney failure prediction models

Chava L. Ramspek[1], Marie Evans[2], Christoph Wanner[3], Christiane Drechsler[4], Nicholas C. Chesnaye[5], Maciej Szymczak[6], Magdalena Krajewska[6], Claudia Torino[7], Gaetana Porto[7], Samantha Hayward[8,9], Fergus Caskey[10], Friedo W. Dekker[1], Kitty J. Jager[5], Merel van Diepen[1] and the EQUAL study investigators*

[1] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
[2] Division of Renal Medicine, Department of Clinical Science, Intervention and Technology, Karolinska Institutet and Karolinska University hospital, Stockholm, Sweden
[3]Division of Nephrology, University Hospital of Würzburg, Würzburg, Germany.
[4] University Hospital Würzburg, Dept of Internal Medicine 1, Div of Nephrology, Oberdürrbacher Str. 6, 97080 Würzburg, Germany.
[5]ERA-EDTA Registry, Amsterdam UMC, University of Amsterdam, Department of Medical Informatics, Amsterdam Public Health Institute, Meibergdreef 9, Amsterdam, The Netherlands
[6]Department of Nephrology and Transplantation Medicine, Wroclaw Medical University, Wroclaw, Poland.
[7]CNR-IFC, Clinical Epidemiology of Renal Diseases and Hypertension, Reggio Calabria, Italy
[8]Translational health sciences, Bristol medical school, University of Bristol, UK BS2 8DZ
[9]UK Renal Registry
[10]Population Health Sciences, Bristol medical school, University of Bristol, UK BS8 2PL.

*The list of EQUAL investigators

Key words: comprehensive external validation, chronic kidney disease, kidney failure, prediction model, prognostic.

Corresponding author: C.L. Ramspek, c.l.ramspek@lumc.nl

Date: 23 December 2020

Abstract

Background

Various prediction models have been developed to predict the risk of kidney failure in patients with chronic kidney disease. However, guideline recommended models have yet to be compared head-to-head, validation in advanced chronic kidney disease (CKD) patients is lacking, and most models don't account for competing risks. The aim of the current study is to externally validate 11 existing models of kidney failure in two large cohorts of advanced CKD patients, whilst taking the competing risk of death into account.

Methods

The models were validated in EQUAL, a European prospective cohort of older advanced CKD patients and the Swedish Renal Registry (SRR) of nephrology-referred CKD patients. Model performance was assessed with discrimination and calibration.

Results

1580 patients from EQUAL and 13489 patients from the SRR were included. The average C-statistic over the 11 validated models was 0.74 in EQUAL and 0.80 in the SRR compared to 0.89 in previous validations. Most models with longer prediction-horizons overestimated the risk of kidney failure considerably. The 5-year Kidney Failure Risk Equation (KFRE) overpredicted risk by 10% -18%. The 4 and 8 variable 2-year KFRE and the 4-year Grams model showed excellent calibration and good discrimination in both cohorts.

Conclusion

Existing models can accurately predict kidney failure in patients with advanced CKD. For a shorter time-frame of 2 years, the KFRE had a good performance despite the fact that this model does not account for competing events. However, models that predicted over a longer time-frame of 5 years overestimated risk due to the competing risk of death.

Significance statement

Most kidney failure prediction models have been developed and validated on cohorts with a wide range of disease severity, without accounting for the competing risk of death. A head-to-head comparison is lacking for guideline recommended models, currently used in clinic. Therefore, the current study provides a comprehensive external validation of kidney failure prediction tools in two advanced CKD cohorts, taking the competing risk of death into account. Models that predicted over a longer time-frame of 5 years overestimated risk due to the competing risk of death. In advanced CKD patients the 8 variable 2-year kidney failure risk equation is recommended for short-term predictions surrounding RRT preparation. The 4-year Grams model, which accounts for competing risk, is most suitable for longer-term predictions.

Introduction

The worldwide burden of chronic kidney disease (CKD) on public health is large and increasing, with an estimated worldwide prevalence of 844 million people.[1] As CKD can lead to kidney failure, striving towards the most optimal treatment and decision-making is of high importance.[2] Obtaining individualized risk-based information is key as rates of progression vary highly between individuals.[3] Risk assessment is important to inform patients, guide treatment decisions and provide information for planning and prioritization of resources.[3 4] Specifically for nephrologists and other advanced CKD care providers risk assessment is central to individualized management and can be used for decisions regarding vascular access placement, other dialysis preparations and counselling on kidney transplant options. For such outcomes a short-term prediction (over 1 or 2 years) is most informative.[3 4] In addition risk assessment can guide referral back to primary care for CKD treatment, this calls for a long-term prediction (over 4 or 5 years).[3 4] Finally, receiving information on prognosis can relieve uncertainty and distress on disease progression, for patients with advanced CKD.[5]

Multiple prediction models have been developed that provide individualized information on the risk of kidney failure in CKD patients.[6-12] These existing models have been externally validated to various degrees and are recommended in multiple guidelines.[13-15] [3] The decisional dilemma underlying the clinical use of such models varies depending on the care setting and disease severity of the patient. Though existing models have shown to predict kidney failure with high discrimination, most were developed and validated on CKD patients with a wide range of disease severity from various care settings. Head-to-head comparison of multiple models is lacking, particularly in patients with advanced CKD (stage 4/5).[16] [17]

In patients with advanced CKD the competing risk of death plays an important role in risk assessment. Most existing models do not consider this competing event in the risk estimation.[18] Competing risk is more important to consider in frail, older populations in which the competing event occurs frequently, and when predicting over long time frames. Most existing kidney failure prediction models censor patients that die. As this censoring is assumed to be uninformative (e.g. unrelated to the risk of kidney failure) the resulting prognosis should be interpreted as the risk of kidney failure in a hypothetical setting in which patients do not die. This risk is an overestimation of the true risk of kidney failure.[19] For patients with a high risk of dying prior to kidney failure, a less aggressive treatment may be in their best interest. If the competing risk of death is disregarded, these patients may undergo unnecessary dialysis preparation, including a vascular access surgery.[20] Though a recent publication recommends that kidney failure calculators should account for death as a competing risk, many of these calculators (which do not account for competing risks) are already used in clinic.[20] As these prediction models are used to predict risk of kidney failure (and not the hypothetical risk of kidney failure given that no patient dies), we deem external validation in which the observed risks are calculated taking competing risks into account of paramount importance.

Therefore, the aim of this study is to externally validate published models that predict kidney failure in two large cohorts of advanced CKD patients whilst taking the competing risk of death into account in the assessment of predictive performance. Models that can be employed in patients with advanced CKD for timely RRT preparation and informing patients on their expected prognosis were included.

Methods

Selecting prediction models for validation

A recent systematic review, conducted by our research group, identified prediction models for renal replacement therapy (RRT) initiation in CKD patients.[17] As the review included articles published up to December 31st 2017, we updated the search to include articles published up to December 31st 2018. For the current study we formulated a number of inclusion criteria. Firstly, the model must have been developed for a general CKD population. Secondly, only models that predict initiation of renal replacement therapy within a specified time-frame were considered for validation. Thirdly, models were only validated if they provided calculation options to determine an individual's risk of RRT. For studies that did not provide this the authors were contacted via email and requested to provide a calculation option. Finally, we only included models that included advanced CKD patients as part of the development population, as our goal was to validate models that were applicable for use in advanced CKD patients. For RRT preparation, a short-term model is more relevant whilst for opting for less aggressive treatment regiments or referral back to primary care for CKD follow-up, longer term predictions might be preferred. For each included model, the risk of bias and applicability to our prognostic question was assessed using the PROBAST tool.[21]

Validation cohorts

The current study follows the TRIPOD (transparent reporting of a multivariable prediction model for

individual prognosis or diagnosis) guidelines (TRIPOD-checklist given in supplement).[22 23] All included prediction models were validated in two cohorts of CKD patients, the European Quality Study (EQUAL) and Swedish Renal Registry (SRR). EQUAL is an ongoing international European prospective multi-centre cohort study of older nephrology-referred CKD patients.[24] Patients ≥65 years were included in Germany, Italy, the Netherlands, Poland, Sweden, and the United Kingdom. Patients were recruited at the nephrology clinic when their estimated glomerular filtration rate (eGFR) first dropped below 20 ml/min/1.73m^2 and each patient is followed for 4 to 8 years. Patients with acute kidney injury or previous RRT were excluded. Clinical characteristics and lab values are registered every 6 months. Patients were included between March 2012 and December 2018. Some patients' kidney function increased above 20 ml/min/1.73m^2 at study baseline, as eligibility assessment took place earlier. Thus, for the main analysis we restricted to patients with an eGFR between 8-30 ml/min/1.73m^2 at baseline.

The SRR is an ongoing registry of CKD patients from 98% of the nephrology clinics in Sweden. Patients are registered when they are first referred to the nephrologist with an eGFR below 30 ml/min/1.73m^2 or when the eGFR first drops below 30 ml/min/1.73m^2, with an option for the clinics to include patients earlier, when their eGFR drops below 45 ml/min/1.73m^2. Though the registry started in 2005 the current study restricted to patients included from January 1st 2012 till June 30th 2018. This was done to include only incident patients and because SRR patients included between 2005 and 2011 comprise 1-2% of the CKD-prognosis consortium population used for the development of the updated kidney failure risk equation (KFRE) and CKD G4+ risk calculator (referred to as the Grams model). The main analysis was restricted to patients 18 years and older with an eGFR between 8 and 30 ml/min/1.73m^2 at the time of registry.

Predictors

All predictors were measured at baseline, this was the first visit after the patient was included in

either EQUAL or the SRR. For EQUAL this was within 6 months of recruitment, when the patient first had an eGFR below 20 ml/min/1.73m^2. For the SRR this is at the first registered visit at a nephrology clinic with an eGFR under 30 ml/min/1.73m^2. Both these baseline timepoints were considered clinically relevant moments for RRT prediction specifically for managing expectations on prognosis and preparing for RRT. Patients with an eGFR under 8 ml/min/1.73m^2 at baseline were excluded as their late presentation makes RRT prediction less meaningful. The predictors are shown in table 1. The eGFR equation that was used in each original prediction model was used in our model validation, this was the CKD-EPI equation for the KFRE, Grams model and KPNW score, and the MDRD equation for all other models and risk-scores. For predictors not available in the validation cohorts, proxies were used (see supplement).

Outcome

The outcome of all validated models was kidney failure defined as RRT-treated end stage kidney disease, which comprises start of haemodialysis, peritoneal dialysis, or pre-emptive kidney transplantation. To calculate the observed risk of the outcome RRT, cumulative incidence functions were used in which the competing event was death before RRT. Patients that completed the study without death or RRT and patients that were lost to follow-up were right-censored.[25]

Statistical analysis

Continuous baseline characteristics are presented as mean values with standard deviations or median values with interquartile ranges when not normally distributed. Categorical variables are presented as valid percentages. Missing data were assumed to be largely missing at random. Therefore, 10-fold multiple imputation with fully conditional specification was performed separately in both validation cohorts using the R-package 'mice'. All predictors, various patient characteristics, outcome and death were included in the imputation models.[26 27]

For each model the probabilities of RRT were calculated per individual (prediction formulas given in supplement). The performance of each model was then assessed in both validation cohorts based on discrimination and calibration. Discrimination is a relative measure of how well a model can discriminate between people with and without the event of interest. To assess discrimination time-to-event C-statistics were computed for each validated model. The competing event of death was taken into account by censoring patients who die at infinity, thereby indicating that these patients cannot experience RRT after death.[28] C-statistics were pooled over the ten imputation datasets according to Rubin's Rules.[29] A C-statistic of 1 is perfect, 0.5 is equal to chance and ≥0.8 is generally considered good for prognostic models.[30] Importantly, the C-statistic of the same model can vary highly depending on the validation population. A more homogeneous population will make it difficult to distinguish low and high-risk patients and will result in a lower C-statistic.[31]

Calibration determines whether the absolute predicted risks are similar to the observed risks. First, the predicted probabilities were combined over the ten imputation datasets by calculating the mean probability per patient. The observed risk of kidney failure was calculated using crude cumulative incidence functions, this allowed us to take the substantial competing risk of death before RRT into account.[25] The calibration-in-the-large is the overall observed risk of RRT compared to the predicted risk. A calibration plot presents the predicted risk and the observed risk, such that the 45˚ line indicates perfect agreement between predicted and observed. These plots were computed using a smoothed (lowess) regression line and patient deciles grouped by predicted risk.[32] The distribution of predicted probabilities is shown in histograms per validated model and separate calibration plots were computed per model.

To assess the impact of taking competing risk into account, each model's predicted probabilities were also compared to observed risks in which the competing risk was not accounted for. For Cox prediction models the observed risk was assessed by censoring patients who died before RRT and

calculating a Harrel's C-statistic. For (multinomial) logistic models this was done by assuming censored/died patients did not have the outcome and calculating an AUC. To explore the influence of eGFR at baseline three sensitivity analyses were performed in which the SRR was restricted to patients with an eGFR of 8-20, 20-30 and 8-45 ml/min/1.73m^2. Additionally, all analyses in EQUAL were repeated excluding Swedish patients, as these patients are most likely also included in the SRR. All analyses were performed in R version 3.5.1.

Results

Models selected for validation

In our previous systematic review 20 studies were identified that developed and/or validated prediction models for RRT in a general CKD population and the update of our search strategy identified an additional 5 studies. [17] A flowchart of the model selection process is given in Figure S1 (supplement). Many studies did not provide calculation options for absolute risks and a total of 7 studies containing 11 prediction models were finally included for external validation. The characteristics of these 11 models are shown in Table 1. In general, the models use similar predictors, with age, eGFR and sex being the most commonly used. The majority was developed in patients with an eGFR between 0 and 60 ml/min/1.73m^2. Only one study (by Grams et al.) took the competing risk of death into account during model development, the majority were Cox prediction models which censored patients who died. The prediction horizon ranged from 1 to 5 years. The risk of bias and applicability per model is shown in supplemental Table S1. All studies were scored as having an overall high risk of bias mainly due to competing risks not being accounted for, missing data not being handled appropriately and it being unclear at what time-point predictors were assessed. When assessing applicability, all models were applicable to our research question concerning the included predictors and outcome. However, the models had a varying degree of applicability to our patient population. Though each model included patients with advanced CKD in

the development population, only the VA model and Grams model were developed on exclusively advanced CKD patients. The development population of these models resembles our validation cohorts much closer than some of the other development populations. For instance, in the KPNW cohort only 7% of patients had CKD stage 4, and none of the included patients had CKD stage 5. This marked difference in populations can heavily influence external validation results.

Baseline characteristics

Baseline characteristics for EQUAL and the SRR are shown in Table 2. In general, EQUAL patients are slightly older, have a slightly lower kidney function and substantially more comorbidities. The SRR patients have more heterogeneity in the continuous predictors and are more similar to the derivation cohorts of the validated models than the EQUAL patients. This is most apparent for the important predictors age and eGFR (see supplement Figure S2). Extensive baseline tables of EQUAL and the SRR including number of missing values are given in the supplement (Table S2 and S3). For most predictors the proportion of missing values was low. Laboratory values had the highest amount of missings, as the time of measurement sometimes didn't coincide with study baseline and only routinely collected lab data was used. The two subsequent lab measurements (at 6 and 12 months) were included in the imputation models to estimate these missing values. Smoking, ethnicity and mean corpuscular volume were not collected in the SRR.

Outcome assessment

In total, 1580 patients from EQUAL were included. Of these patients 458 started RRT within 5 years of study inclusion. Of the RRT initiators 74% started on haemodialysis, 23% on peritoneal dialysis and 3% received a pre-emptive kidney transplant. The median observation time was 24 months. A total of 330 patients died before RRT initiation and 215 patients withdrew or were lost to follow-up. A total of 13489 patients were included from the SRR, of which 2764 started RRT within 5 years. Of these patients 58% started on haemodialysis, 35% on peritoneal dialysis and 6% received a pre-

emptive kidney transplant. The median observation time was 21 months. A total of 3357 patients died before RRT start and no patients were lost to follow up.

Predictive performance of validated models

In general, the models had a good discrimination, see Table 3. The average validated C-statistic reported in the original papers was 0.89. In EQUAL the average C-statistic was 0.74 and in the SRR it was 0.80. The C-statistics in EQUAL ranged from 0.611 (Johnson score) to 0.807 (VA model) and in the SRR they ranged from 0.662 (Johnson score) to 0.835 (2-year Grams model). For short-term prediction the VA model showed the best discriminatory performance. For long-term prediction the Landray model had the highest C-statistics. In the sensitivity analysis where patients who died were censored and the competing risk therefor not accounted for, the average C-statistic was slightly higher (0.75 in EQUAL and 0.82 in the SRR, see Table S4). Increasing and decreasing the eGFR range of included SRR patients moderately increased and decreased the C-statistics, respectively (Table S6).

The calibration-in-the large (shown in Table 4) was reasonably accurate for the Grams models and 2-year KFREs, but most models predicting over a longer horizon overestimated the risk of RRT. In Figure 1 and 2 each model's calibration is plotted per validation cohort. In both EQUAL and the SRR, the 4 and 8 variable 2-year KFRE and 4-year Grams model are most accurate. The sensitivity analysis in which competing risks were not accounted for in the observed risks, showed markedly different calibration results (Table S5, Figure S5 & S6). When censoring for patients who die, the 5-year KFREs have an almost perfect calibration in the SRR. The 4 variable 5-year KFRE predicts an average RRT risk of 40.9% in the SRR, the observed risk when censoring for death is 40.8% but the observed risk calculated whilst taking competing events into account is 31.0%. This discrepancy is further exaggerated in high-risk patients, who not only have a high risk of kidney failure but also of dying. The 8 variable 5-year KFRE on average overpredicted risk of RRT by 18% in EQUAL and by 17% in the

SRR. The distribution of predicted probabilities is shown in Figure S3 and S4. Calibration remained similar when varying SRR eGFR exclusion criteria and when excluding Swedish patients from EQUAL (see supplement, Table S7 & S8, figure S7-S10).

Discussion

Main findings

The current study externally validates 11 prediction tools that predict the risk of kidney failure treated with RRT within one to five years. The discrimination and calibration of these models were assessed within two different cohorts of advanced CKD patients, taking into account the competing risk of death. In general, the C-statistics showed reasonable to good discrimination, though considerably lower than reported in previous studies which were performed on CKD patients with a wider range in disease severity. The apparent decline in discrimination may be explained by the narrower case-mix of our validation cohorts, compared to the development populations. The agreement between observed and predicted risks varied greatly per model. By accounting for death before kidney failure in the observed risks, it became apparent that models predicting over a longer time-frame overestimated the risk of RRT. This was most extreme for high-risk patients.

Comparison with other studies

In recent years many prediction models have been developed and compared to the KFRE. However, no previous study has validated multiple independent prediction models in the same external cohort. Additionally, the KFRE has not been externally validated while taking competing risks into account. As different countries show considerable variation in CKD progression rates and mortality, it is important to validate these models in various settings.[33] The C-statistic is highly dependent on the study population and its heterogeneity in predictor values.[34] In our cohorts of advanced CKD patients these predictor values are more homogeneous than in many of the development cohorts,

which included patients with a wide range in disease severity. This explains the lower C-statistics observed in the current study. Our findings that C-statistics decreased as we restricted the population to smaller eGFR ranges further exemplify this and the importance of selecting validation populations that correspond to the proposed clinical use of the prediction model.[35]

Competing risk

The failure to consider the competing risk of death can bias prediction models and result in predicted risks that are too high. This bias is more extreme in frail patient populations and for long follow-up durations, as the competing event of death is more frequent in such settings.[20 36] After accounting for the competing risk of death, we found that models with shorter prediction horizons (of 1 or 2 years) were not biased much; our main results were very similar to our sensitivity analyses in which competing risks were ignored. The 2-year KFRE specifically showed an accurate calibration; it seems taking the competing risk of death into account is not necessary for these short-term predictions in patients with advanced CKD. However, models predicting over 5 years significantly overestimated the risk of RRT in our advanced CKD population. The 5-year KFRE showed a structural over-prediction of RRT risk which can be fully attributed to the competing risk of death (as shown in our sensitivity analyses). The failure to consider the competing risk of death can result in incorrect predicted 5-year risks, this in turn may lead to poor treatment decisions. To our knowledge the current study is the first to externally validate existing logistic and Cox models for a competing risk scenario.

Strengths and limitations

The current study has a number of strengths. It provides a comparison of multiple prediction models in the same cohorts in a structured, comprehensive and methodologically sound fashion. Specifically, the first external validation of the Grams model and the comparison of this with the KFRE is critical for evidence-based-medicine as both have been recommended in guidelines.[3 15] This

study does employ somewhat unconventional statistical methods. From a statistical point of view, it seems inconsistent to validate a logistic or Cox model as if it were a competing risk model (e.g. Fine & Gray or Markov models), it is therefore common practice to validate Cox prediction models as developed (by censoring patients who die before RRT). Though this approach was considered, we decided against it as the observed risk is then the risk of RRT in the hypothetical scenario in which no patients would die. If we validate Cox models as such, our external validation might show a perfect prediction, though this risk is not interpretable or of use in clinical practice. By taking death into account in the observed risks, we give a better representation of the true RRT risks and the ability of these models to predict this. Furthermore, it is unique that the two validation cohorts are contemporary European nephrologist-referred patients. However, our findings should be placed in light of a number of limitations. Firstly, not all predictors were available in our cohorts and the use of proxies might have influenced model performance. Secondly, it is a major limitation that patients who chose to forgo RRT and opted for conservative care are not included in our kidney failure outcome, due to limitations of the data. As conservative care is becoming a more frequent approach in many European countries, particularly in older patients, this may have resulted in an underestimation of kidney failure incidence. Thirdly, both cohorts contain routinely collected clinical data, though this can be perceived as a strength because it mirrors routine clinical nephrology care, it is a limitation concerning the completeness of laboratory data. To deal with this missingness as best as possible multiple imputation was used. In addition, almost 14% of patients in EQUAL were lost to follow-up, if this dropout is related to kidney failure or death, this may have led to some form of selection bias which in turn may lead to miscalibration. Fourthly, this external validation study cannot ascertain the best model for non-European countries or different patient populations such as primary care cohorts.[37] Model performance was tested in only two advanced CKD cohorts, one which included only older patients; validation in other cohorts may show different model performance. And finally, this study does not provide evidence on how to use these models to guide binary clinical decisions in individual patients.

Clinical implications

When selecting a prediction model the intended use as well as discrimination and calibration should be considered. [38-40] Good discrimination allows for a large range of predicted risks[38] and calibration is important for accurate absolute risk prediction. A predicted risk that is too high or too low may result in wrong treatment decisions. In the nephrology clinic, short-term risk predictions are probably most relevant, particularly when considering that these predictions can be updated at every follow-up visit. RRT prediction could improve the timing of providing treatment option information and counseling, timing of referral for vascular access, initiating transplant investigation and guide referral back to primary care for CKD treatment and follow-up. This would allocate more valuable specialist resources to patients with high risk of disease progression. For predicting short-term kidney failure risk in advanced CKD patients, we would recommend the 4 or 8 variable 2-year KFRE. These models would be suitable for the timing of RRT preparation. For longer prediction horizons the 4-year Grams model is recommended. These recommendations are based on consistently good discrimination and calibration results in both validation cohorts, the robust development data underlying these models, and the availability of an easy to use web-calculator. When validating these models in a competing risk scenario they remained accurate. The 2-year Grams model under-estimated the risk of kidney failure in high risk patients considerably, if this model is used and predicts risks >40% these are most likely underestimations of the actual risk. As both Grams models predict the risk of multiple adverse outcomes including cardiovascular disease and death before and after RRT start, these models are more informative and conducive for decision making. We therefore agree with the KDIGO conference report and recommend the use of the Grams models in advanced CKD patients for predicting RRT, with a preference for the 4-year Grams model.[3] Further external validation of the other outcomes predicted by the Grams models is advised. The 4 and 8 variable 5-year KFREs substantially overpredicted the risk of RRT in both cohorts when considering the competing risk of death, these are therefore not recommended for

use in the nephrology clinic. The Landray model performed reasonably well but overpredicted in higher risk patients and the lack of a web application makes use more difficult. The VA model overestimated risks greatly and Marks model showed mediocre performance, these are not recommended. The two categorical risk scores (KPNW and Johnson) performed poorly in our validation and their use is discouraged in nephrology referred patients with CKD stage 4+; these scores appear to be inapplicable to this population.

Future studies

We would advise against future development of similar prediction models of RRT, as existing models have shown consistently good results. It would be valuable to evaluate these models in other clinically relevant settings and populations, including the calculation of the model-based concordance measure, which allows quantification of how case-mix heterogeneity influences each model's discriminative capability in validation.[41] Furthermore, these models might be recalibrated to various settings and to correct for the competing risk of death. Additionally, studies should look into optimal risk thresholds to base specific clinical decisions on and assess the impact of using such models in clinical practice. This would preferably be done in a clinical impact trial to assess whether using such models will benefit patients.[3 42-44] If these models are integrated in clinical practice, they would be updated at every visit, this should also be considered in future studies. Further work on competing risk and dynamic prediction models is warranted. For prediction models that are used on chronically ill patients in clinical practice, we encourage researchers to externally validate such existing (logistic and Cox) prediction models whilst taking competing risks into account. Finally, future studies might focus further on predicting other quality of life related outcomes such as symptom burden, functional and cognitive status and hospitalization as these are highly relevant to patients.[45-48]

Conclusions

This study is the first to provide a comprehensive validation of all available models that predict kidney failure in patients with CKD. The validation has been performed in two cohorts of advanced CKD patients. We found that for short-term predictions the 4 and 8 variable 2-year KFRE are most suitable for predicting the risk of kidney failure. For this 2-year time-frame the predictions were accurate, despite the model not accounting for the competing risk of death. However, when predicting over a longer time frame, the 5-year KFREs overestimated the actual risk of RRT considerably due to the competing risk of death. Use of these models should be reconsidered in patients with advanced CKD (stage 4/5) and instead the 4-year Grams model is recommended.

Ethical approval and consent to participate

The EQUAL study was approved by the medical ethics committee or institutional review boards (as appropriate) of all participating centers (main medical ethical committee approval obtained in the Amsterdam Medical Center, NL38874.018.11). Written informed consent was obtained from all patients. This study was approved by the Regional Ethical Review Board in Stockholm, Sweden (Dnr 2018/1591-31/2). According to Swedish law, health care quality registries can be used for research. Patients have the right to opt out, but no additional individual consent is required for specific research projects.

Consent for publication

Not applicable.

Availability of data

Data are not publicly available. Data from the EQUAL study may be requested with protocol and statistical analysis plan at the EQUAL publication committee (contact: n.c.chesnaye@amsterdamumc.nl). Data from the SRR may be requested with protocol and statistical analysis plan and reviewed by the regional ethical review board in Stockholm (contact: marie.evans@ki.se).

Competing interests

All authors declare that they have no competing interests..

Funding

The work on this study by MvD was supported by a grant from the Dutch Kidney Foundation (16OKG12).

Authors' Contributions

Contributors: CLR, ME, FWD and MvD conceived the current study. ME, CW, MS, CT, FC, FWD and KJJ oversaw design and data collection for the EQUAL study and ME oversaw data collection for the SRR. NCC and KJJ oversaw data management and data quality assurance for the EQUAL study. CD, MK, CT, GP and SH contributed to data collection. CLR and MvD performed the analysis. CLR prepared the first draft. All authors reviewed the whole draft and approved the final manuscript. CLR and MvD are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Acknowledgements

We would like to thank all the patients and health professionals involved in the EQUAL study and the Swedish Renal Registry. ME was funded by a grant from Centre of Innovative Medicine (CIMED), Karolinska Institutet and Stockholm City Council. Funding for the EQUAL-study was received from the European Renal Association – European Dialysis and Transplant Associations, the Swedish Medical Association, the Stockholm County Council ALF, Njurfonden, the Italian Society of Nephrology (SIN – Reni), the Dutch Kidney Foundation (SB142), the Young Investigators grant in Germany and the National Institute for Health research in the UK.

Supplemental material – table of contents

Tables & Figures

Table 1. Characteristics of validated prediction models.

Model	Prediction horizon	Predictors	Type of prediction tool.	Competing risk model	Derivation population	Country	Mean age	Mean eGFR	Sample size	Previous external validations, country and C-statistic.*
VA model[49]	1 year	Age, eGFR, congestive heart failure, SBP, s. potassium, s. albumin	Cox formula	No	eGFR<30 age ≥ 65 years	USA	78	25	1866	1x USA C-stat: 0.82
Grams model[14]	2 years	Age, eGFR, sex, race, CVD, diabetes, SBP, uACR, smoking	Multinomial formula & webtool	Yes	eGFR <30	29 cohorts from 5 continents	72	24	264,296	-
KFRE 4v model[50]	2 years	Age, eGFR, sex, uACR	Cox formula & webtool	No	eGFR<60	31 cohorts from 4 continents	74	46	267,479[a]	-
KFRE 8v model[50]	2 years	Age, eGFR, sex, uACR, s. albumin, s. phosphate, s. bicarbonate, s. calcium	Cox formula & webtool	No	eGFR<60	31 cohorts from 4 continents	74	46	40,221[a]	-
Grams model[14]	4 years	Age, eGFR, sex, race, CVD, diabetes, SBP, uACR, smoking	Multinomial formula & webtool	Yes	eGFR <30	29 cohorts from 5 continents	72	24	234,286[b]	-
KFRE 4v model[50]	5 years	Age, eGFR, sex, uACR	Cox formula & webtool	No	eGFR<60	31 cohorts from 4 continents	74	46	267,479[a]	3x, USA, UK, Netherlands C-stat: 0.83, 0.88, 0.95
KFRE 8v model[50]	5 years	Age, eGFR, sex, uACR, s. albumin, s. phosphate, s. bicarbonate, s. calcium	Cox formula & webtool	No	eGFR<60	31 cohorts from 4 continents	74	46	40,221[a]	1x Netherlands C-stat: 0.89
Landray model[51]	5 years	s. creatinine, sex, uACR, s. phosphate	Cox formula[b]	No	eGFR<60	UK	62	22	382	1x, UK C-stat: 0.91
Marks model[52]	5 years	Age, eGFR, sex, micro-albuminuria, macro-albuminuria	Logistic formula	No	eGFR<60	UK	79	33	3396	1x UK C-stat: 0.96
KPNW score[6]	5 years	Age, eGFR, sex, diabetes, diabetic complications, antihypertensive medication, SBP, haemoglobin, albuminuria	Risk score	No	eGFR 15-60	USA	75	47	22,460	1x USA C-stat:0.95
Johnson score[53]	5 years	Age, eGFR, sex, diabetes, hypertension, anemia	Risk score	No	eGFR 15-60	USA	73	46	9782	-

Abbreviations: SBP: systolic blood pressure, eGFR: estimated glomerular filtration rate, CVD: cardiovascular disease, uACR: urinary albumin creatinine ratio, s.: serum, 4v: 4 variable, 8v: 8 variable, USA: United States of America, UK: United Kingdom. *Data based on previously conducted systematic review of these studies.[17]
[a]sample size of external validation and recalibration meta-analysis, these recalibrated models are validated in the current study. The original KFRE model was developed on 3449 patients. [13]
[b]Made available through personal communication with author.

Table 2. Baseline characteristics of the EQUAL cohort and Swedish renal registry (SRR)

	EQUAL cohort n = 1580	**SRR n=13489**
Age (year)	76.2 (70.7-81.5)	74.3 (65.7-81.2)
Sex (% male)	65.5%	61.3%
Current smoker (%)	9.1%	-
Country of residence (%)		
Germany	8.5%	0%
Italy	24.3%	0%
The Netherlands	15.0%	0%
Poland	4.2%	0%
Sweden	18.1%	100%
United Kingdom	29.9%	0%
Primary Kidney Disease (%)		
Diabetes mellitus	20.3%	21.5%
Glomerular disease	9.2%	6.9%
Hypertension	36.4%	30.2%
Other	34.2%	41.4%
Comorbidities (%)		
Cardiovascular disease	62.2%	33.1%
Hypertension	91.7%	73.2%
Diabetes mellitus	42.1%	36.4%
Laboratory parameters		
eGFR (MDRD, ml/min/1.73m^2)	18.5 (4.7)	21.9 (5.7)
Urinary ACR (mg/mmol)	40 (8 - 165)	36 (7 - 155)
Serum Calcium (mmol/L)	2.24 (0.32)	2.29 (0.29)

Abbreviations: RRT: renal replacement therapy, eGFR: estimated Glomerular Filtration Rate, MDRD: Modification of Diet in Renal Disease formula, ACR: albumin creatinine ratio. Lab values are shown SI units and can be converted to conventional units as follows: urinary ACR in mg/g: multiply by 8.85, calcium in mg/dL: multiply by 4.0.

Table 3. Discrimination of validated models in EQUAL and the SRR

	Time-frame	Original C-statistic (95%CI)	C-statistic (95%CI) EQUAL	C-statistic (95%CI) SRR
VA model	1 year	0.823[a]	0.807 (0.778-0.836)	0.835 (0.823-0.847)
Grams model	2 years	0.814 (IQR: 0.755-0.850)[b]	0.763 (0.730-0.797)	0.842 (0.833-0.851)
KFRE 4v model	2 years	0.90 (0.89-0.92)[a]	0.757 (0.718-0.795)	0.838 (0.828-0.847)
KFRE 8v model	2 years	0.89 (0.88-0.91)[a]	0.777 (0.747-0.808)	0.838 (0.828-0.848)
Grams model	4 years	0.784 (IQR: 0.745-0.852)[b]	0.742 (0.713-0.771)	0.826 (0.818-0.834)
KFRE 4v model	5 years	0.88 (0.86-0.90)[a]	0.745 (0.712-0.779)	0.812 (0.804-0.820)
KFRE 8v model	5 years	0.86 (0.85-0.88)[a]	0.760 (0.732-0.787)	0.807 (0.798-0.816)
Landray model	5 years	0.91 (0.87-0.96)[a]	0.778 (0.753-0.802)	0.805 (0.797-0.813)
Marks model	5 years	0.960 (0.947-0.974)[c]	0.705 (0.682-0.728)	0.779 (0.771-0.787)
KPNW score	5 years	0.95 (0.94-0.97)[a]	0.656 (0.637-0.675)	0.762 (0.754-0.769)
Johnson score	5 years	0.89[d]	0.611 (0.591-0.630)	0.662 (0.654-0.671)

Abbreviations: 4v: 4 variable, 8v: 8 variable.

[a]External validation results. [b]apparent C-statistic, received via email from M. Grams. [c]Temporal validation result (development cohort was nested in external validation cohort). [d] Internal validation result (bootstrapped).

Table 4. Calibration-in-the-large of validated models in EQUAL and the SRR

	Time-frame	Predicted vs. observed EQUAL	Predicted vs. observed SRR
VA model	1 year	17.5% vs 12.7%	15.5% vs 7.9%
Grams model	2 years	20.0% vs 23.7%	15.7% vs 16.1%
KFRE 4v model	2 years	22.3% vs. 23.7%	17.1% vs 16.1%
KFRE 8v model	2 years	24.5% vs. 23.7%	20.3% vs 16.1%
Grams model	4 years	30.3% vs 35.2%	25.8% vs 26.8%
KFRE 4v model	5 years	51.0% vs 36.8%	40.9% vs 31.0%
KFRE 8v model	5 years	55.2% vs 36.8%	47.7% vs 31.0%
Landray model	5 years	42.1% vs 36.8%	31.8% vs 31.0%
Marks model	5 years	24.6% vs 36.8%	21.5% vs 31.0%
KPNW score	5 years	55.8% vs 36.8%	38.3% vs 31.0%
Johnson score	5 years	42.2% vs 36.8%	31.5% vs 31.0%

Abbreviations: 4v: 4 variable, 8v: 8 variable.

Figure 1: Calibration plots per validated model in EQUAL. The predicted probability is shown on the x-axis and the observed kidney failure rate given on the y-axis. The dotted 45 degree line represents perfect agreement between predicted and observed probability. The smoothed line is a lowess line through all predicted risks and corresponding observed risks. The dots represent a decile of the validation population (10%), ranked by predicted probability. For the KPNW score and Johnson score, each dot represents a risk group category, which corresponds to the risk score categories. The observed probability was calculated with cumulative incidence functions.

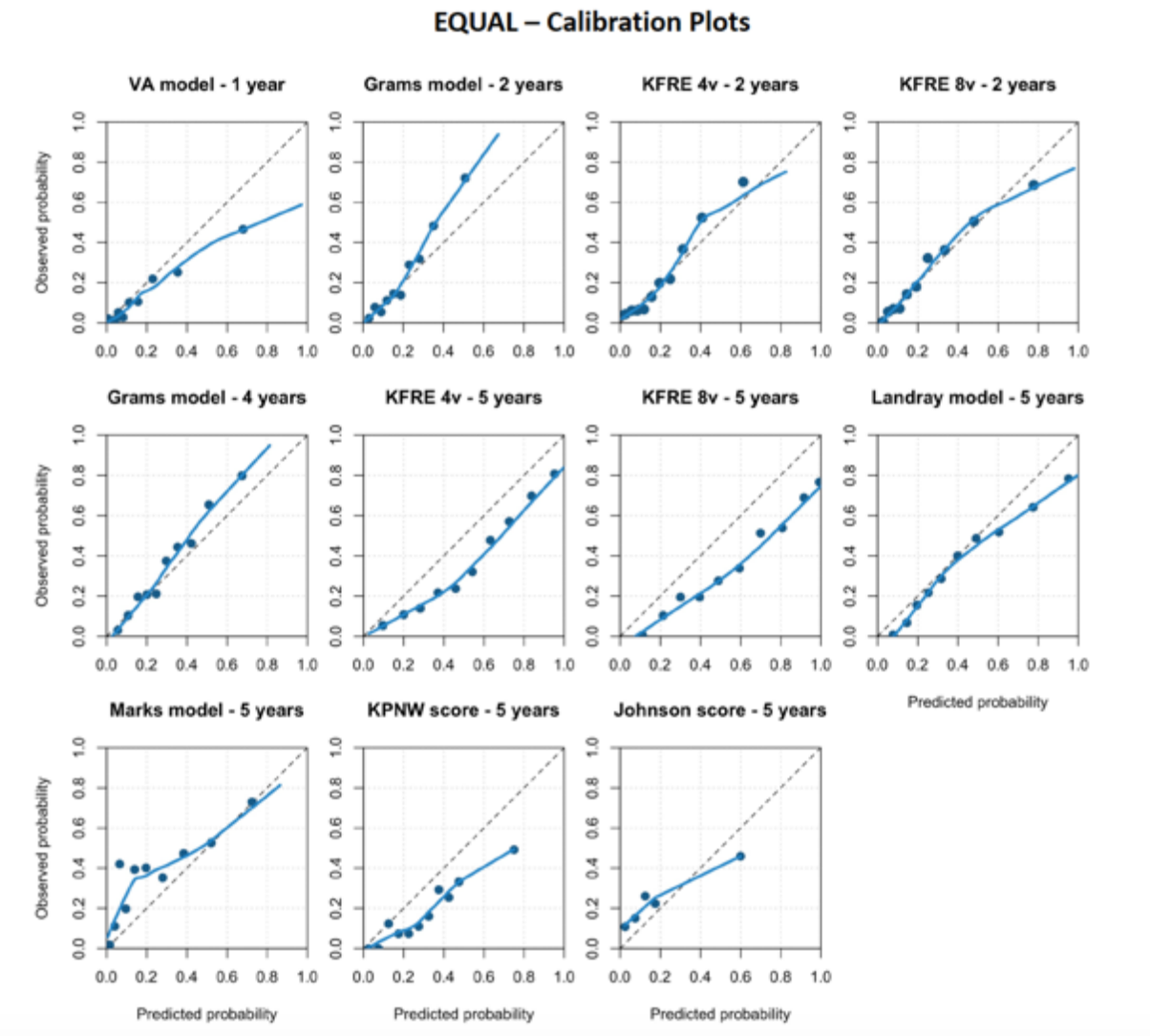
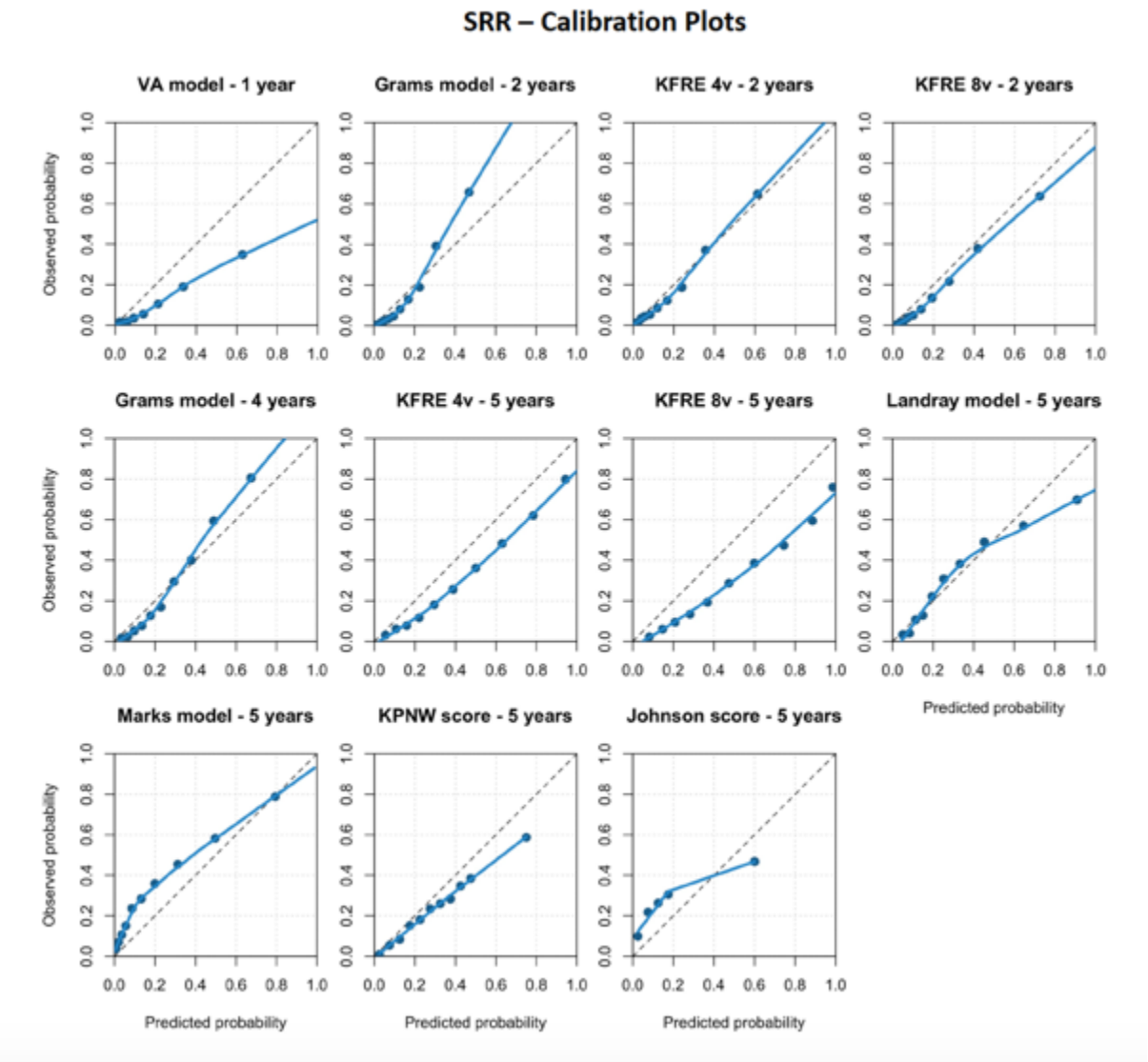


Figure 2: Calibration plots per validated model in the SRR. The predicted probability is shown on the x-axis and the observed kidney failure rate given on the y-axis. The dotted 45 degree line represents perfect agreement between predicted and observed probability. The smoothed line is a lowess line through all predicted risks and corresponding observed risks. The dots represent a decile of the validation population (10%), ranked by predicted probability. For the KPNW score and Johnson score, each dot represents a risk group category, which corresponds to the risk score categories. The observed probability was calculated with cumulative incidence functions.



References

1. Jager KJ, Kovesdy C, Langham R, et al. A single number for advocacy and communication—worldwide more than 850 million individuals have kidney diseases. *Kidney international* 2019;96(5):1048-50. doi: 10.1016/j.kint.2019.07.012
2. Liyanage T, Ninomiya T, Jha V, et al. Worldwide access to treatment for end-stage kidney disease: a systematic review. *Lancet (London, England)* 2015;385(9981):1975-82. doi: 10.1016/s0140-6736(14)61601-9 [published Online First: 2015/03/18]
3. Eckardt KU, Bansal N, Coresh J, et al. Improving the prognosis of patients with severely decreased glomerular filtration rate (CKD G4+): conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney Int* 2018;93(6):1281-92. doi: 10.1016/j.kint.2018.02.006 [published Online First: 2018/04/17]
4. Chapter 2: Definition, identification, and prediction of CKD progression. *Kidney international supplements* 2013;3(1):63-72. doi: 10.1038/kisup.2012.65 [published Online First: 2013/01/01]
5. Schell JO, Patel UD, Steinhauser KE, et al. Discussions of the kidney disease trajectory by elderly patients and nephrologists: a qualitative study. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2012;59(4):495-503. doi: 10.1053/j.ajkd.2011.11.023 [published Online First: 2012/01/10]
6. Schroeder EB, Yang X, Thorp ML, 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* 2017;12(1):87-94. doi: 10.2215/cjn.01290216 [published Online First: 2016/12/29]
7. Tangri N, Inker LA, Hiebert B, et al. A Dynamic Predictive Model for Progression of CKD. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2017;69(4):514-20. doi: 10.1053/j.ajkd.2016.07.030 [published Online First: 2016/10/04]
8. Tangri N, Grams ME, Levey AS, et al. Multinational Assessment of Accuracy of Equations for Predicting Risk of Kidney Failure: A Meta-analysis. *Jama* 2016;315(2):164-74. doi: 10.1001/jama.2015.18202 [published Online First: 2016/01/13]
9. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *Jama* 2011;305(15):1553-9. doi: 10.1001/jama.2011.451 [published Online First: 2011/04/13]
10. 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. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2010;56(6):1082-94. doi: 10.1053/j.ajkd.2010.07.016 [published Online First: 2010/11/03]
11. Day CJ, Howie AJ, Nightingale P, et al. Prediction of ESRD in pauci-immune necrotizing glomerulonephritis: quantitative histomorphometric assessment and serum creatinine. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2010;55(2):250-8. doi: 10.1053/j.ajkd.2009.10.047 [published Online First: 2010/01/05]
12. Johnson ES, Thorp ML, Platt RW, et al. 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* 2008;52(4):653-60. doi: 10.1053/j.ajkd.2008.04.026 [published Online First: 2008/07/01]

13. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *Jama* 2011;305(15):1553-9. doi: 10.1001/jama.2011.451 [published Online First: 2011/04/13]
14. Grams ME, Sang Y, Ballew SH, et al. Predicting timing of clinical outcomes in patients with chronic kidney disease and severely decreased glomerular filtration rate. *Kidney Int* 2018;93(6):1442-51. doi: 10.1016/j.kint.2018.01.009 [published Online First: 2018/04/02]
15. Findlay A, Farrington K, Joosten H, et al. Clinical Practice Guideline on management of older patients with chronic kidney disease stage 3b or higher (eGFR<45 mL/min/1.73 m2): a summary document from the European Renal Best Practice Group. *Nephrology Dialysis Transplantation* 2017;32(1):9-16. doi: 10.1093/ndt/gfw411 %J Nephrology Dialysis Transplantation
16. Major RW, Shepherd D, Medcalf JF, et al. The Kidney Failure Risk Equation for prediction of end stage renal disease in UK primary care: An external validation and clinical impact projection cohort study. *PLoS medicine* 2019;16(11):e1002955-e55. doi: 10.1371/journal.pmed.1002955
17. Ramspek CL, de Jong Y, Dekker FW, et al. Towards the best kidney failure prediction tool: a systematic review and selection aid. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2019 doi: 10.1093/ndt/gfz018 [published Online First: 2019/03/05]
18. Li L, Yang W, Astor BC, et al. Competing Risk Modeling: Time to Put it in Our Standard Analytical Toolbox. 2019;30(12):2284-86. doi: 10.1681/ASN.2019101011 %J Journal of the American Society of Nephrology
19. Wolkewitz M, Cooper BS, Bonten MJ, et al. Interpreting and comparing risks in the presence of competing events. *BMJ (Clinical research ed)* 2014;349:g5060. doi: 10.1136/bmj.g5060 [published Online First: 2014/08/26]
20. Ravani P, Fiocco M, Liu P, et al. Influence of Mortality on Estimating the Risk of Kidney Failure in People with Stage 4 CKD. *Journal of the American Society of Nephrology : JASN* 2019;30(11):2219-27. doi: 10.1681/asn.2019060640 [published Online First: 2019/09/22]
21. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170(1):51-58. doi: 10.7326/m18-1376 [published Online First: 2019/01/01]
22. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ (Clinical research ed)* 2015;350:g7594. doi: 10.1136/bmj.g7594 [published Online First: 2015/01/09]
23. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine* 2015;162(1):W1-73. doi: 10.7326/m14-0698 [published Online First: 2015/01/07]
24. Jager KJ, Dekker FW, Ocak G, et al. The EQUAL study: a European study in chronic kidney disease stage 4 patients. *Nephrology Dialysis Transplantation* 2012;27(suppl_3):iii27-iii31. doi: 10.1093/ndt/gfs277 %J Nephrology Dialysis Transplantation
25. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation* 2016;133(6):601-9. doi: 10.1161/circulationaha.115.017719 [published Online First: 2016/02/10]
26. de Goeij MC, van Diepen M, Jager KJ, et al. Multiple imputation: dealing with missing data. *Nephrology, dialysis, transplantation : official publication of the European*

Dialysis and Transplant Association - European Renal Association 2013;28(10):2415-20. doi: 10.1093/ndt/gft221 [published Online First: 2013/06/05]
27. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research* 2007;16(3):219-42. doi: 10.1177/0962280206074463 [published Online First: 2007/07/11]
28. Wolbers M, Koller MT, Witteman JC, et al. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology (Cambridge, Mass)* 2009;20(4):555-61. doi: 10.1097/EDE.0b013e3181a39056 [published Online First: 2009/04/16]
29. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley 1987.
30. Royston P, Moons KG, Altman DG, et al. Prognosis and prognostic research: Developing a prognostic model. *BMJ (Clinical research ed)* 2009;338:b604. doi: 10.1136/bmj.b604 [published Online First: 2009/04/02]
31. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *American journal of epidemiology* 2010;172(8):971-80. doi: 10.1093/aje/kwq223 [published Online First: 2010/09/03]
32. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Statistics in medicine* 2014;33(3):517-35. doi: 10.1002/sim.5941 [published Online First: 2013/09/05]
33. Bruck K, Jager KJ, Zoccali C, et al. Different rates of progression and mortality in patients with chronic kidney disease at outpatient nephrology clinics across Europe. *Kidney international* 2018;93(6):1432-41. doi: 10.1016/j.kint.2018.01.008 [published Online First: 2018/04/17]
34. Pencina MJ, D'Agostino RB, Sr. Evaluating Discrimination of Risk Prediction Models: The C Statistic. *Jama* 2015;314(10):1063-4. doi: 10.1001/jama.2015.11082 [published Online First: 2015/09/09]
35. Peeters MJ, van Zuilen AD, van den Brand JA, et al. Validation of the kidney failure risk equation in European CKD patients. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2013;28(7):1773-9. doi: 10.1093/ndt/gft063 [published Online First: 2013/05/07]
36. Riphagen IJ, Kleefstra N, Drion I, et al. Comparison of methods for renal risk prediction in patients with type 2 diabetes (ZODIAC-36). *PLoS One* 2015;10(3):e0120477-e77. doi: 10.1371/journal.pone.0120477
37. Tangri N, Ferguson T, Komenda P. Pro: Risk scores for chronic kidney disease progression are robust, powerful and ready for implementation. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2017;32(5):748-51. doi: 10.1093/ndt/gfx067 [published Online First: 2017/05/13]
38. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical chemistry* 2008;54(1):17-23. doi: 10.1373/clinchem.2007.096529 [published Online First: 2007/11/21]
39. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115(7):928-35. doi: 10.1161/circulationaha.106.672402 [published Online First: 2007/02/21]
40. Moons KGM, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;162(1):W1-W73. doi: 10.7326/m14-0698

41. van Klaveren D, Gönen M, Steyerberg EW, et al. A new concordance measure for risk prediction models in external validation settings. *Statistics in medicine* 2016;35(23):4136-52. doi: 10.1002/sim.6997 [published Online First: 2016/06/01]
42. Chiu HH, Tangri N, Djurdjev O, et al. Perceptions of prognostic risks in chronic kidney disease: a national survey. *Canadian journal of kidney health and disease* 2015;2:53. doi: 10.1186/s40697-015-0088-z [published Online First: 2015/12/22]
43. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European heart journal* 2014;35(29):1925-31. doi: 10.1093/eurheartj/ehu207 [published Online First: 2014/06/06]
44. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass)* 2010;21(1):128-38. doi: 10.1097/EDE.0b013e3181c30fb2 [published Online First: 2009/12/17]
45. Ramspek CL, Voskamp PW, van Ittersum FJ, et al. Prediction models for the mortality risk in chronic dialysis patients: a systematic review and independent external validation study. *Clinical epidemiology* 2017;9:451-64. doi: 10.2147/clep.S139748 [published Online First: 2017/09/19]
46. Ramer SJ, McCall NN, Robinson-Cohen C, et al. Health Outcome Priorities of Older Adults with Advanced CKD and Concordance with Their Nephrology Providers' Perceptions. *Journal of the American Society of Nephrology : JASN* 2018;29(12):2870-78. doi: 10.1681/asn.2018060657 [published Online First: 2018/11/06]
47. Couchoud C, Hemmelgarn B, Kotanko P, et al. Supportive Care: Time to Change Our Prognostic Tools and Their Use in CKD. *Clinical journal of the American Society of Nephrology : CJASN* 2016;11(10):1892-901. doi: 10.2215/cjn.12631115 [published Online First: 2016/08/12]
48. Verberne WR, Das-Gupta Z, Allegretti AS, et al. Development of an International Standard Set of Value-Based Outcome Measures for Patients With Chronic Kidney Disease: A Report of the International Consortium for Health Outcomes Measurement (ICHOM) CKD Working Group. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2019;73(3):372-84. doi: 10.1053/j.ajkd.2018.10.007 [published Online First: 2018/12/24]
49. Drawz PE, Goswami P, Azem R, et al. A simple tool to predict end-stage renal disease within 1 year in elderly adults with advanced chronic kidney disease. *Journal of the American Geriatrics Society* 2013;61(5):762-8. doi: 10.1111/jgs.12223 [published Online First: 2013/04/27]
50. Tangri N, Grams ME, Levey AS, et al. Multinational Assessment of Accuracy of Equations for Predicting Risk of Kidney Failure: A Meta-analysis. *Jama* 2016;315(2):164-74. doi: 10.1001/jama.2015.18202 [published Online First: 2016/01/13]
51. 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. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2010;56(6):1082-94. doi: 10.1053/j.ajkd.2010.07.016 [published Online First: 2010/11/03]
52. Marks A, Fluck N, Prescott GJ, 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* 2015;30(9):1507-17. doi: 10.1093/ndt/gfv089 [published Online First: 2015/05/07]

53. Johnson ES, Thorp ML, Platt RW, et al. 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* 2008;52(4):653-60. doi: 10.1053/j.ajkd.2008.04.026 [published Online First: 2008/07/01]