Kidney Failure Prediction Models: A Comprehensive External Validation Study in Patients with Advanced CKD

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] and Merel van Diepen,[1] on behalf of the EQUAL Study Investigators*

Due to the number of contributing authors, the affiliations are listed at the end of this article.

ABSTRACT

Background Various prediction models have been developed to predict the risk of kidney failure in patients with CKD. However, guideline-recommended models have yet to be compared head to head, their validation in patients with advanced CKD is lacking, and most do not account for competing risks.

Methods To externally validate 11 existing models of kidney failure, taking the competing risk of death into account, we included patients with advanced CKD from two large cohorts: the European Quality Study (EQUAL), an ongoing European prospective, multicenter cohort study of older patients with advanced CKD, and the Swedish Renal Registry (SRR), an ongoing registry of nephrology-referred patients with CKD in Sweden. The outcome of the models was kidney failure (defined as RRT-treated ESKD). We assessed model performance with discrimination and calibration.

Results The study included 1580 patients from EQUAL and 13,489 patients from SRR. The average *c* statistic over the 11 validated models was 0.74 in EQUAL and 0.80 in SRR, compared with 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 four- and eight-variable 2-year KFRE and the 4-year Grams model showed excellent calibration and good discrimination in both cohorts.

Conclusions Some existing models can accurately predict kidney failure in patients with advanced CKD. KFRE performed well for a shorter time frame (2 years), despite not accounting for competing events. Models predicting over a longer time frame (5 years) overestimated risk because of the competing risk of death. The Grams model, which accounts for the latter, is suitable for longer-term predictions (4 years).

JASN 32: 1174–1186, 2021. doi: https://doi.org/10.1681/ASN.2020071077

The worldwide burden of 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 toward 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 counseling 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 patients with CKD.[6–12] These existing models have been externally validated to various degrees and are recommended in multiple guidelines.[3,9,13,14] The decisional dilemma underlying the clinical use of such models varies depending on the care setting and disease severity of the patient. Although existing models have been shown to predict kidney failure with high discrimination, most were developed and validated on patients with CKD 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 (stages 4 and 5).[15,16]

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.[17] 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 who 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.[18] 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.[19] Although 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 the clinic.[19] 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 patients with advanced CKD while taking the competing risk of death into account in the assessment of predictive performance. Models that can be used in patients with advanced CKD for timely RRT preparation and informing patients of their expected prognosis were included.

Significance Statement

Most kidney failure prediction models have been developed and validated in cohorts of patients with a wide range of disease severity, without accounting for the competing risk of death. Models recommended by guidelines, currently used in the clinic, have not undergone a head-to-head comparison. This study provides a comprehensive external validation of kidney failure prediction tools in two cohorts of patients with advanced CKD, taking the competing risk of death into account. Models that predict over a longer time frame of 5 years overestimate risk due to the competing risk of death. In patients with advanced CKD, the eight-variable 2-year Kidney Failure Risk Equation is recommended for short-term predictions surrounding preparation for RRT. The 4-year Grams model, which accounts for competing risk, is most suitable for longer-term predictions.

Received July 26, 2020. Accepted December 26, 2020.

*The EQUAL study investigators include Adamasco Cupisti, Adelia Sagliocca, Alberto Ferraro, Aleksandra Musiała, Alessandra Mele, Alessandro Naticchia, Alex Còsaro, Alistair Woodman, Andrea Ranghino, Andrea Stucchi, Andreas Jonsson, Andreas Schneider, Angelo Pignataro, Anita Schrander, Anke Torp, Anna McKeever, Anna Szymczak, Anna-Lena Blom, Antonella De Blasio, Antonello Pani, Aris Tsalouichos, Asad Ullah, Barbara McLaren, Bastiaan van Dam, Beate Iwig, Bellasi Antonio, Biagio Raffaele Di Iorio, Björn Rogland, Boris Perras, Butti Alessandra, Camille Harron, Carin Wallquist, Carl Siegert, Carla Barrett, Carlo Gaillard, Carlo Garofalo, Cataldo Abaterusso, Charles Beerenhout, Charlotte O'Toole, Chiara Somma, Christian Marx, Christina Summersgill, Christof Blaser, Claudia D'alessandro, Claudia Emde, Claudia Zullo, Claudio Pozzi, Colin Geddes, Cornelis Verburgh, Daniela Bergamo, Daniele Ciurlino, Daria Motta, Deborah Glowski, Deborah McGlynn, Denes Vargas, Detlef Krieter, Domenico Russo, Dunja Fuchs, Dympna Sands, Ellen Hoogeveen, Ellen Irmler, Emöke Dimény, Enrico Favaro, Eva Platen, Ewelina Olczyk, Ewout Hoorn, Federica Vigotti, Ferruccio Ansali, Ferruccio Conte, Francesca Cianciotta, Francesca Giacchino, Francesco Cappellaio, Francesco Pizzarelli, Fredrik Sundelin, Fredrik Uhlin, Gaetano Greco, Geena Roy, Gaetana Porto, Giada Bigatti, Giancarlo Marinangeli, Gianfranca Cabiddu, Gillian Hirst, Giordano Fumagalli, Giorgia Caloro, Giorgina Piccoli, Giovanbattista Capasso, Giovanni Gambaro, Giuliana Tognarelli, Giuseppe Bonforte, Giuseppe Conte, Giuseppe Toscano, Goffredo Del Rosso, Gunilla Welander, Hanna Augustyniak-Bartosik, Hans Boots, Hans Schmidt-Gürtler, Hayley King, Helen McNally, Hendrik Schlee, Henk Boom, Holger Naujoks, Houda Masri-Senghor, Hugh Murtagh, Hugh Rayner, Ilona Miśkowiec-Wiśniewska, Ines Schlee, Irene Capizzi, Isabel Bascaran Hernandez, Ivano Baragetti, Jacek Manitius, Jane Turner, Jan-Willem Eijgenraam, Jeroen Kooman, Joachim Beige, Joanna Pondel, Joanne Wilcox, Jocelyn Berdeprado, Jochen Röthele, Jonathan Wong, Joris Rotmans, Joyce Banda, Justyna Mazur, Kai Hahn, Kamila Jędrzejak, Katarzyna Nowańska, Katja Blouin, Katrin Neumeier, Kirsteen Jones, Kirsten Anding-Rost, Knut-Christian Gröntoft, Lamberto Oldrizzi, Lesley Haydock, Liffert Vogt, Lily Wilkinson, Loreto Gesualdo, Lothar Schramm, Luigi Biancone, Łukasz Nowak, Maarten Raasveld, Magdalena Durlik, Manuela Magnano, Marc Vervloet, Marco Ricardi, Margaret Carmody, Maria Di Bari, Maria Laudato, Maria Luisa Sirico, Maria Stendahl, Maria Svensson, Maria Weetman, Marjolijn van Buren, Martin Joinson, Martina Ferraresi, Mary Dutton, Merel van Diepen, Michael Matthews, Michele Provenzano, Monika Hopf, Moreno Malaguti, Nadja Wuttke, Neal Morgan, Nicola Palmieri, Nikolaus Frischmuth, Nina Bleakley, Paola Murrone, Paul Cockwell, Paul Leurs, Paul Roderick, Pauline Voskamp, Pavlos Kashioulis, Pawlos Ichtiaris, Peter Blankestijn, Petra Kirste, Petra Schulz, Phil Mason, Philip Kalra, Pietro Cirillo, Pietro Dattolo, Pina Acampora, Rincy Sajith, Rita Nigro, Roberto Boero, Roberto Scarpioni, Rosa Sicoli, Rosella Malandra, Sabine Aign, Sabine Cäsar, Sadie van Esch, Sally Chapman, Sandra Biribauer, Santee Navjee, Sarah Crosbie, Sharon Brown, Sheila Tickle, Sherin Manan, Silke Röser, Silvana Savoldi, Silvio Bertoli, Silvio Borrelli, Siska Boorsma, Stefan Heidenreich, Stefan Melander, Stefania Maxia, Stefano Maffei, Stefano Mangano, Stephanie Palm, Stijn Konings, Suresh Mathavakkannan, Susanne Schwedler, Sylke Delrieux, Sylvia Renker, Sylvia Schättel, Szyszkowska Dorota, Teresa Cicchetti, Teresa Nieszporek, Theresa Stephan, Thomas Schmiedeke, Thomas Weinreich, Til Leimbach, Tiziana Rappa, Tora Almquist, Torsten Stövesand, Udo Bahner, Ulrika Jensen, Valentina Palazzo, Walter De Simone, Wolfgang Seeger, Ying Kuan, Zbigniew Heleniak, and Zeynep Aydin.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. Chava L. Ramspek, Department of Clinical Epidemiology, Leiden University Medical Center, Albinusdreef 2, 2333ZA Leiden, the Netherlands. Email: c.l.ramspek@lumc.nl

METHODS

Selecting Prediction Models for Validation

A recent systematic review, conducted by our research group, identified prediction models for RRT initiation in patients with CKD.[16] 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 this study, we formulated a number of inclusion criteria. First, the model must have been developed for a general CKD population. Second, only models that predict initiation of RRT within a specified time frame were considered for validation. Third, 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 patients with advanced CKD as part of the development population, as our goal was to validate models that were applicable for use in patients with advanced CKD. For RRT preparation, a short-term model is more relevant, whereas 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.[20]

Validation Cohorts

This study follows the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis guidelines (TRIPOD checklist is given in Supplemental Material).[21,22] All included prediction models were validated in two cohorts of patients with CKD: the European Quality Study (EQUAL) and the Swedish Renal Registry (SRR). EQUAL is an ongoing international European prospective, multicenter cohort study of older nephrology-referred patients with CKD.[23] 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 eGFR first dropped below 20 ml/min per 1.73 m^2, and each patient was followed for 4–8 years. Patients with AKI or previous RRT were excluded. Clinical characteristics and laboratory values are registered every 6 months. Patients were included between March 2012 and December 2018. Some patients' kidney function increased above 20 ml/min per 1.73 m^2 at study baseline, as eligibility assessment took place earlier. Thus, for the main analysis, we restricted to patients with an eGFR between 8 and 30 ml/min per 1.73 m^2 at baseline.

SRR is an ongoing registry of patients with CKD 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 per 1.73 m^2 or when the eGFR first drops below 30 ml/min per 1.73 m^2, with an option for the clinics to include patients earlier when their eGFR drops below 45 ml/min per 1.73 m^2. Although the registry started in 2005, this study was restricted to patients included from January 1st, 2012 to June 30th, 2018. This was done to include only incident patients and because patients in SRR 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 per 1.73 m^2 at the time of registry.

Ethical Approval and Consent to Participate

EQUAL 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.

Predictors

All predictors were measured at baseline; this was the first visit after the patient was included in either EQUAL or SRR. For EQUAL, this was within 6 months of recruitment, when the patient first had an eGFR below 20 ml/min per 1.73 m^2. For SRR, this is at the first registered visit at a nephrology clinic with an eGFR under 30 ml/min per 1.73 m^2. Both these baseline time points 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 per 1.73 m^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; they were the Chronic Kidney Disease Epidemiology Collaboration equation for KFRE, the Grams model, and KPNW score and the Modification of Diet in Renal Disease equation for all other models and risk scores. For predictors not available in the validation cohorts, proxies were used (Supplemental Material).

Outcome

The outcome of all validated models was kidney failure defined as RRT-treated ESKD, which comprises start of hemodialysis, peritoneal dialysis, or preemptive 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 who completed the study without death or RRT and patients who were lost to follow-up were right censored.[24]

Statistical Analyses

Continuous baseline characteristics are presented as mean values with SDs or median values with interquartile ranges when

Table 1. Characteristics of validated prediction models

Model	Prediction Horizon, yr	Predictors	Type of Prediction Tool	Competing Risk Model	Derivation Population	Country	Mean Age, yr	Mean eGFR	Sample Size	Previous External Validations, Country, and *c* Statistic[a]
VA model[47]	1	Age, eGFR, congestive heart failure, SBP, s. potassium, s. albumin	Cox formula	No	eGFR<30; age ≥65 yr	USA	78	25	1866	1×, USA, *c* statistic: 0.82
Grams *et al.*[13] model	2	Age, eGFR, sex, race, CVD, diabetes, SBP, uACR, smoking	Multinomial formula and web tool	Yes	eGFR<30	29 cohorts from five continents	72	24	264,296	—
KFRE 4v model[8]	2	Age, eGFR, sex, uACR	Cox formula and web tool	No	eGFR<60	31 cohorts from four continents	74	46	267,479[b]	—
KFRE 8v model[8]	2	Age, eGFR, sex, uACR, s. albumin, s. phosphate, s. bicarbonate, s. calcium	Cox formula and web tool	No	eGFR<60	31 cohorts from four continents	74	46	40,221[b]	—
Grams *et al.*[13] model	4	Age, eGFR, sex, race, CVD, diabetes, SBP, uACR, smoking	Multinomial formula and web tool	Yes	eGFR<30	29 cohorts from five continents	72	24	234,286[c]	—
KFRE 4v model[8]	5	Age, eGFR, sex, uACR	Cox formula and web tool	No	eGFR<60	31 cohorts from four continents	74	46	267,479[b]	3×, USA, UK, The Netherlands, *c* statistic: 0.83, 0.88, 0.95
KFRE 8v model[8]	5	Age, eGFR, sex, uACR, s. albumin, s. phosphate, s. bicarbonate, s. calcium	Cox formula and web tool	No	eGFR<60	31 cohorts from four continents	74	46	40,221[b]	1×, The Netherlands, *c* statistic: 0.89
Landray *et al.*[10] model	5	S. creatinine, sex, uACR, s. phosphate	Cox formula[c]	No	eGFR<60	UK	62	22	382	1×, UK, *c* statistic: 0.91
Marks *et al.*[48] model	5	Age, eGFR, sex, microalbuminuria, macroalbuminuria	Logistic formula	No	eGFR<60	UK	79	33	3396	1×, UK, *c* statistic: 0.96
KPNW score[6]	5	Age, eGFR, sex, diabetes, diabetic complications, antihypertensive medication, SBP, hemoglobin, albuminuria	Risk score	No	eGFR 15–60	USA	75	47	22,460	1×, USA, *c* statistic: 0.95
Johnson *et al.*[12] score	5	Age, eGFR, sex, diabetes, hypertension, anemia	Risk score	No	eGFR 15–60	USA	73	46	9782	—

VA, Veterans Affairs; SBP, systolic BP; s., serum; USA, United States of America; CVD, cardiovascular disease; uACR, urinary albumin-creatinine ratio; —, non-existent; 4v, four variable; 8v, eight variable; UK, United Kingdom; KPNW, Kaiser Permanente Northwest.

[a]Data on the basis of previously conducted systematic review of these studies.[16]

[b]Sample size of external validation and recalibration meta-analysis. These recalibrated models are validated in this study. The original KFRE model was developed on 3449 patients.[9]

[c]Made available through personal communication with M. Landray.

not normally distributed. Categorical variables are presented as valid percentages. Missing data were assumed to be largely missing at random. Therefore, ten-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.[25,26]

For each model, the probabilities of RRT were calculated per individual (prediction formulas are given in Supplemental Material). The performance of each model was then assessed in both validation cohorts on the basis of 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.[27] The *c* statistics were pooled over the ten imputation datasets according to the Rubin rules.[28] A *c* statistic of 1.0 is perfect, 0.5 is equal to chance, and ≥0.8 is generally considered good for prognostic models.[29] 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.[30]

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.[24] The calibration in the large is the overall observed risk of RRT compared with 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.[31] The distribution of predicted probabilities is shown in histograms per validated model, and separate calibration plots were computed per model.

To assess the effect of taking competing risk into account, each model's predicted probabilities were also compared with 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 *c* statistic. For (multinomial) logistic models, this was done by assuming patients who were censored or died did not have the outcome and calculating an AUC. To explore the influence of eGFR at baseline, three sensitivity analyses were performed in which SRR was restricted to patients with eGFR of 8–20, 20–30, and 8–45 ml/min per 1.73 m^2. Additionally, all analyses in EQUAL were repeated excluding Swedish patients, as these patients are most likely also included in 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 five studies.[16] A flowchart of the model selection process is given in Supplemental Figure 1. Many studies did not provide calculation options for absolute risks, and a total of seven 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 were developed in patients with an eGFR between 0 and 60 ml/min per 1.73 m^2. Only one study (by Grams *et al.*[13]) 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 are shown in Supplemental Table 1. 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. Although each model included patients with advanced CKD in the development population, only the VA model and the Grams model were developed exclusively on patients with advanced CKD. 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 SRR are shown in Table 2. In general, patients in EQUAL are slightly older, have a slightly lower kidney function, and have substantially more comorbidities. The patients in SRR have more heterogeneity in the continuous predictors and are more similar to the derivation cohorts of the validated models than the patients in EQUAL. This is most apparent for the important predictors age and eGFR (Supplemental Figure 2). Extensive baseline tables of EQUAL and SRR including number of missing values are given in Supplemental Tables 2 and 3. For most predictors, the proportion of missing values was low. Laboratory values had the highest amount of missing data, as the time of measurement sometimes did not coincide with study baseline and only routinely collected laboratory data were used. The two subsequent laboratory measurements (at 6 and 12 months) were included in the imputation models to estimate

Table 2. Baseline characteristics of the EQUAL cohort and SRR

Characteristic	EQUAL Cohort, *n*=1580	SRR, *n*=13,489
Age, yr	76.2 (70.7–81.5)	74.3 (65.7–81.2)
Men, %	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 by MDRD, ml/min per 1.73 m^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)

Laboratory values are shown in the International System of Units and can be converted to conventional units as follows: urinary ACR in milligrams per gram: multiply by 8.85; calcium in milligrams per deciliter: multiply by 4.0. MDRD, Modification of Diet in Renal Disease; ACR, albumin-creatinine ratio.

these missing values. Smoking, ethnicity, and mean corpuscular volume were not collected in 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 hemodialysis, 23% started on peritoneal dialysis, and 3% received a preemptive 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 13,489 patients were included from SRR, of which 2764 started RRT within 5 years. Of these patients, 58% started on hemodialysis, 35% started on peritoneal dialysis, and 6% received a preemptive 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 good discrimination (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 SRR, it was 0.80. The *c* statistics in EQUAL ranged from 0.61 (Johnson score) to 0.81 (VA model), and in SRR, they ranged from 0.66 (Johnson score) to 0.84 (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, therefore, was not accounted for, the average *c* statistic was slightly higher (0.75 in EQUAL and 0.82 in SRR) (Supplemental Tables 4 and 5). Increasing and decreasing the eGFR range of included patients from SRR moderately increased and decreased the *c* statistics, respectively (Supplemental Table 6).

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 Figures 1 and 2, each model's calibration is plotted per validation cohort. In both EQUAL and SRR, the four- and eight-variable 2-year KFRE and 4-year Grams models are most accurate. The sensitivity analysis in which competing risks were not accounted for in the observed risks showed markedly different calibration results (Supplemental Figures 3–6, Supplemental Table 5). When censoring for patients who die, the 5-year KFREs have an almost perfect calibration in SRR. The four-variable 5-year KFRE predicts an average RRT risk of 41% in SRR; the observed risk when censoring for death is 41%, but the observed risk calculated while taking competing events into account is 31%. This discrepancy is further exaggerated in high-risk patients, who not only have a high risk of kidney failure but also, a high risk of dying. The eight-variable 5-year KFRE on average overpredicted risk of RRT by 18% in EQUAL and by 17% in SRR. The distribution of predicted probabilities is shown in Supplemental Figures 3 and 4. Calibration remained similar when varying SRR eGFR exclusion criteria and when excluding Swedish patients from EQUAL (Supplemental Figures 7–10, Supplemental Material, Supplemental Tables 7 and 8).

DISCUSSION

Main Findings

This study externally validates 11 prediction tools that predict the risk of kidney failure treated with RRT within 1–5 years. The discrimination and calibration of these models were assessed within two different cohorts of patients with advanced CKD, taking into account the competing risk of death. In general, the *c* statistics showed reasonable to good discrimination, although considerably lower than reported in previous studies that were performed on patients with CKD with a wider range in disease severity. The apparent decline in discrimination may be explained by the narrower patient mix of our validation cohorts, compared with 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.

Table 3. Discrimination of validated models in EQUAL and SRR

Validated Model	Time Frame, yr	Original *c* Statistic (95% CI)	*c* Statistic (95% CI) EQUAL	*c* Statistic (95% CI) SRR
VA model	1	0.82[a]	0.81 (0.78 to 0.84)	0.84 (0.82 to 0.85)
Grams model	2	0.81 (IQR, 0.755–0.850)[b]	0.76 (0.73 to 0.80)	0.84 (0.83 to 0.85)
KFRE 4v model	2	0.90 (0.89 to 0.92)[a]	0.76 (0.72 to 0.80)	0.84 (0.83 to 0.85)
KFRE 8v model	2	0.89 (0.88 to 0.91)[a]	0.78 (0.75 to 0.81)	0.84 (0.83 to 0.85)
Grams model	4	0.78 (IQR, 0.745–0.852)[b]	0.74 (0.71 to 0.77)	0.83 (0.82 to 0.83)
KFRE 4v model	5	0.88 (0.86 to 0.90)[a]	0.75 (0.71 to 0.78)	0.81 (0.80 to 0.82)
KFRE 8v model	5	0.86 (0.85 to 0.88)[a]	0.76 (0.73 to 0.79)	0.81 (0.80 to 0.82)
Landray model	5	0.91 (0.87 to 0.96)[a]	0.78 (0.75 to 0.80)	0.81 (0.80 to 0.81)
Marks model	5	0.96 (0.95 to 0.97)[c]	0.71 (0.68 to 0.73)	0.78 (0.77 to 0.79)
KPNW score	5	0.95 (0.94 to 0.97)[a]	0.66 (0.64 to 0.68)	0.76 (0.75 to 0.77)
Johnson score	5	0.89[d]	0.61 (0.59 to 0.63)	0.66 (0.65 to 0.67)

95% CI, 95% confidence interval; VA, Veterans Affairs; IQR, interquartile range; 4v, four variable; 8v, eight variable; KPNW, Kaiser Permanente Northwest.
[a]External validation results.
[b]Apparent *c* statistic
(received *via* email; M. Grams, personal communication).
[c]Temporal validation result
(development cohort was nested in external validation cohort).
[d]Internal validation result
(bootstrapped).

Comparison with Other Studies

In recent years, many prediction models have been developed and compared with KFRE. However, no previous study has validated multiple independent prediction models in the same external cohort. Additionally, 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.[32] The *c* statistic is highly dependent on the study population and its heterogeneity in predictor values.[33] In our cohorts of patients with advanced CKD, these predictor values are more homogeneous than in many of the

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

Validated Model	Time Frame, yr	Predicted versus EQUAL, %	Observed EQUAL, %	Predicted versus EQUAL, %	Observed SRR, %
VA model	1	18	13	16	8
Grams model	2	20	24	16	16
KFRE 4v model	2	22	24	17	16
KFRE 8v model	2	25	24	20	16
Grams model	4	30	35	26	27
KFRE 4v model	5	51	37	41	31
KFRE 8v model	5	55	37	48	31
Landray model	5	42	37	32	31
Marks model	5	25	37	22	31
KPNW score	5	56	37	38	31
Johnson score	5	42	37	32	31

VA, Veterans Affairs; 4v, four variable; 8v, eight variable; KPNW, Kaiser Permanente Northwest.

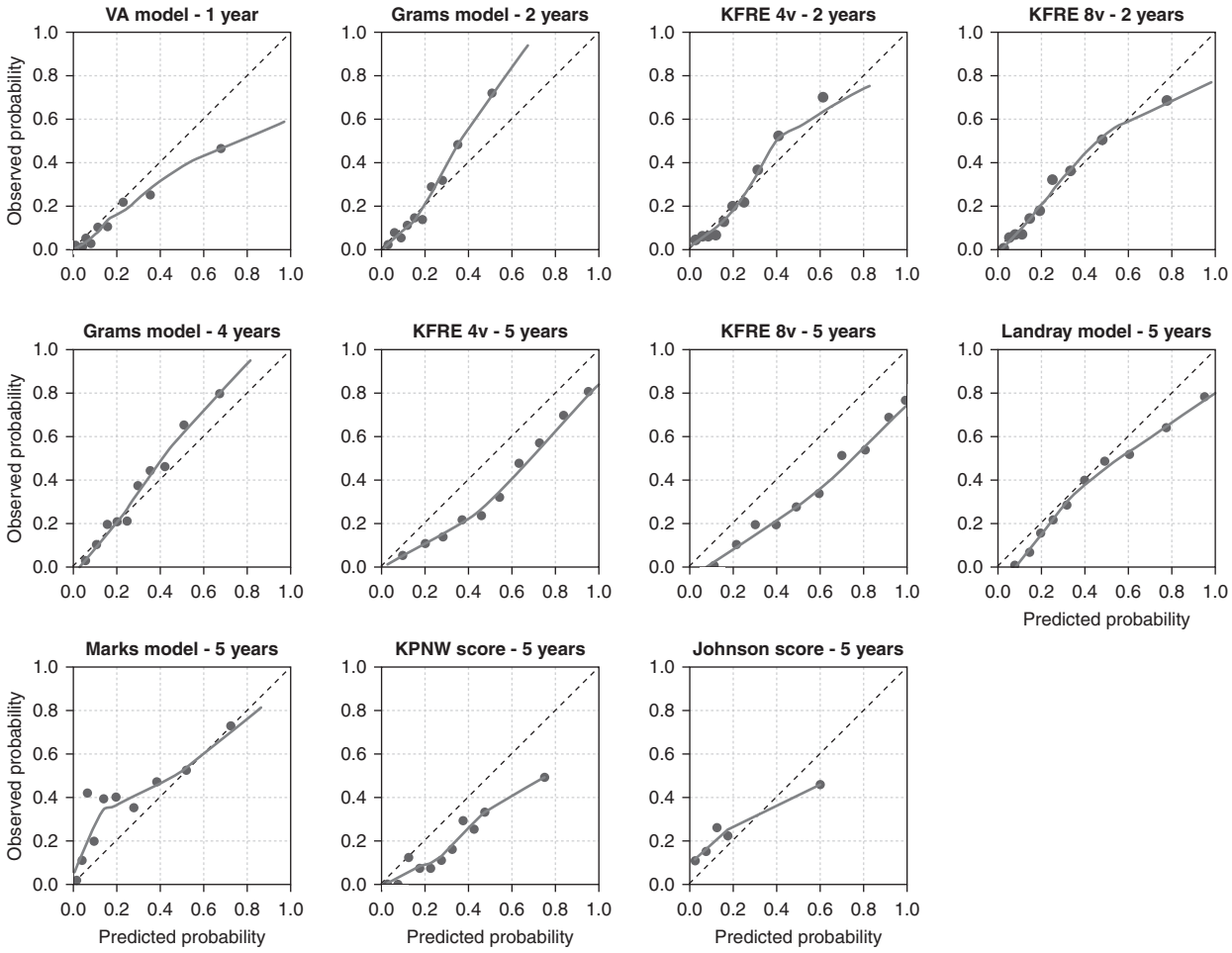

Figure 1. Calibration plots for each validated model in EQUAL. The predicted probability is shown on the *x* axis, and the observed kidney failure rate is given on the *y* axis. The dotted 45° line represents perfect agreement between predicted and observed probabilities. The smoothed line is a lowess 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 the 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. KPNW, Kaiser Permanente Northwest; VA, Veterans Affairs; 4v, four variable; 8v, eight variable.

development cohorts, which included patients with a wide range in disease severity. This explains the lower *c* statistics observed in this 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.[34]

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.[19,35] 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 that 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 overprediction 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, this study is the first to externally validate existing logistic and Cox models for a competing risk scenario.

Strengths and Limitations

This study has a number of strengths. It provides a comparison of multiple prediction models in the same cohorts in a structured,

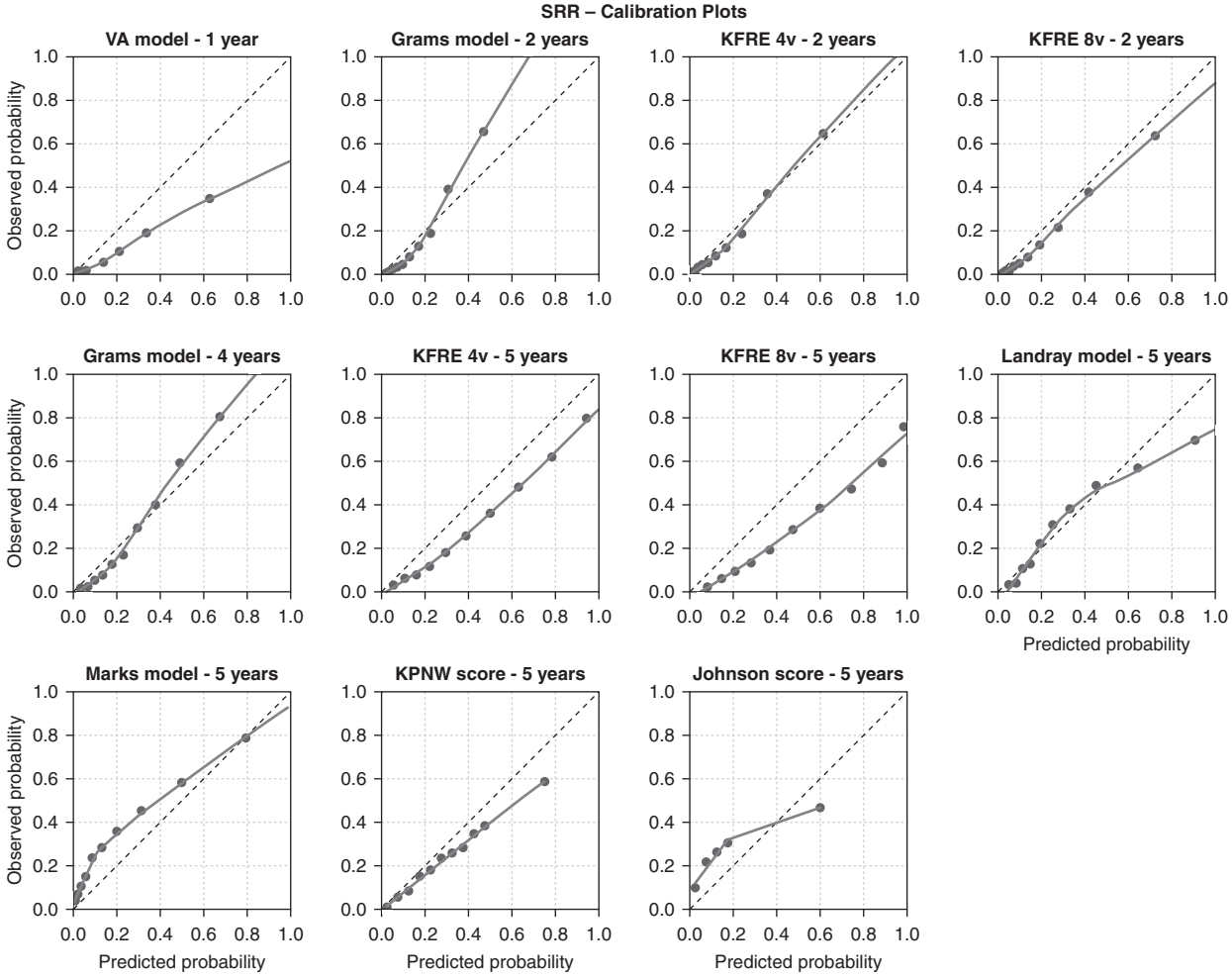

Figure 2. Calibration plots for each validated model in the SRR. The predicted probability is shown on the *x* axis, and the observed kidney failure rate is given on the *y* axis. The dotted 45° line represents perfect agreement between predicted and observed probabilities. The smoothed line is a lowess 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 the 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. KPNW, Kaiser Permanente Northwest; VA, Veterans Affairs; 4v, four variable; 8v, eight variable.

comprehensive, and methodologically sound fashion. Specifically, the first external validation of the Grams model and the comparison of this with KFRE are critical for evidence-based medicine as both have been recommended in guidelines.[3,14] This study does use 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 and Gray or Markov models); it is, therefore, common practice to validate Cox prediction models as developed (by censoring patients who die before RRT). Although 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, although 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. First, not all predictors were available in our cohorts, and the use of proxies might have influenced model performance. Second, 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. Third, both cohorts contain routinely collected clinical data, although this can be perceived as a strength because it mirrors routine clinical nephrology care; however, 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. Fourth, this external validation study cannot ascertain the best model for non-European countries or different patient populations, such as primary care cohorts.[36] Model performance was tested in only two advanced CKD cohorts, one of which included only older patients; validation in other cohorts may show different model performance. 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.[23,37,38] Good discrimination allows for a large range of predicted risks,[37] 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 patient counseling, timing of vascular access placement, transplant preparation and 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 patients with advanced CKD, we would recommend the four- or eight-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 on the basis of 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 underestimated 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 Kidney Disease Improving Global Outcomes conference report and recommend the use of the Grams models in patients with advanced CKD 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 four- and eight-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 the 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 seem 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 patient mix heterogeneity influences each model's discriminative capability in validation.[39] 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 effect of using such models in clinical practice. This would preferably be done in a clinical effect trial to assess whether using such models will benefit patients.[3,40–42] 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 while 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.[43–46]

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 patients with advanced CKD. We found that for short-term predictions, the four- and eight-variable 2-year KFREs 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 KFRE 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 (stages 4 and 5), and instead, the 4-year Grams model is recommended.

DISCLOSURES

F. Caskey reports research funding from Kidney Research UK and National Institute for Health Research and honoraria from Baxter. F.W. Dekker reports research funding from Astellas and Chiesi; being a scientific advisor

or member as Nephrology Dialysis Transplantation editorial board; and other interests/relationships *via* collaboration with Dutch Kidney Patients Association and collaboration with Dutch Quality Institute for Renal Care (Nefrovisie). C. Drechsler reports research funding from Genzyme. M. Evans reports honoraria from payment for lectures by Astellas, AstraZeneca, and Vifor Pharma; being a scientific advisor or member with Astellas, AstraZeneca, and Vifor Pharma advisory board; other interests/relationships as a member of the steering committee of SRR and the European Renal Association–European Dialysis and Transplant Association (ERA-EDTA) Registry Committee. M. Evans was funded by a grant from the Centre of Innovative Medicine, Karolinska Institutet and Stockholm City Council. K.J. Jager reports honoraria from Fresenius and being a scientific advisor or member on the editorial board of *Nephrology Dialysis Transplantation*, the editorial board of *Kidney International Reports*, and the editorial board of *Journal of Renal Nutrition*. M. Krajewska reports other interests/relationships as ERA-EDTA member and with the Polish Society of Nephrology and the Polish Society of Transplantology. C. Wanner reports consultancy agreements with Akebia, Bayer, Boehringer-Ingelheim, Gilead, GSK, MSD, Sanofi-Genzyme, Triceda, and Vifor; research funding from an Idorsia grant to the institution and Sanofi-Genyzme from a grant to the institution; honoraria from Astellas, AstraZeneca, Bayer, Boehringer-Ingelheim, Eli-Lilly, FMC, Sanofi-Genzyme, and Shire-Takeda; and other interests/relationships with ERA-EDTA. Funding for EQUAL was received from the Dutch Kidney Foundation (grant: SB142), ERA-EDTA, the Italian Society of Nephrology (Reni), the National Institute for Health Research in the United Kingdom, Njurfonden, the Stockholm County Council ALF, the Swedish Medical Association, and the Young Investigators grant in Germany. All remaining authors have nothing to disclose.

FUNDING

The work on this study by M. van Diepen was supported by Nierstichting grant 16OKG12.

ACKNOWLEDGMENTS

The authors thank all of the patients and health professionals involved in EQUAL and SRR. The authors thank Morgan Grams, Eric Johnson and Martin Landray for their correspondence on the validated models.

F.W. Dekker, M. Evans, C.L. Ramspek, and M. van Diepen conceived the study; F. Caskey, F.W. Dekker, M. Evans, K.J. Jager, M. Szymczak, C. Torino, and C. Wanner oversaw design and data collection for EQUAL; M. Evans oversaw data collection for SRR; N. C. Chesnaye and K.J. Jager oversaw data management and data quality assurance for EQUAL; C. Drechsler, S. Hayward, M. Krajewska, G. Porto, and C. Torino contributed to data collection; C.L. Ramspek and M. van Diepen performed the analysis; C.L. Ramspek prepared the first draft; all authors reviewed the whole draft and approved the final manuscript; C.L. Ramspek and M. van Diepen are the guarantors; and C.L. Ramspek attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

DATA SHARING STATEMENT

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

SUPPLEMENTAL MATERIAL

This article contains the following supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2020071077/-/DCSupplemental.

Supplemental Material. Formulas of validated models, use of proxies, and TRIPOD checklist.

Supplemental Table 1. PROBAST risk of bias and applicability table.

Supplemental Table 2. Full baseline table EQUAL.

Supplemental Table 3. Full baseline table SRR.

Supplemental Table 4. Discrimination disregarding competing risk.

Supplemental Table 5. Calibration disregarding competing risk.

Supplemental Table 6. SRR discrimination stratified by kidney function.

Supplemental Table 7. SRR calibration stratified by kidney function.

Supplemental Table 8. EQUAL results excluding Swedish patients.

Supplemental Figure 1. Flowchart of prediction model selection.

Supplemental Figure 2. Box plots of predictor distributions.

Supplemental Figure 3. Distribution of predicted probabilities—EQUAL.

Supplemental Figure 4. Distribution of predicted probabilities—SRR.

Supplemental Figure 5. Calibration plot disregarding competing risk—EQUAL.

Supplemental Figure 6. Calibration plot disregarding competing risk—SRR.

Supplemental Figure 7. Calibration plot SRR: eGFR 8–45.

Supplemental Figure 8. Calibration plot SRR: eGFR 20–30.

Supplemental Figure 9. Calibration plot SRR: eGFR 8–20.

Supplemental Figure 10. Calibration plot EQUAL excluding Swedish patients.

REFERENCES

1. Jager KJ, Kovesdy C, Langham R, Rosenberg M, Jha V, Zoccali C: A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases. *Kidney Int* 96: 1048–1050, 2019
2. Liyanage T, Ninomiya T, Jha V, Neal B, Patrice HM, Okpechi I, et al.: Worldwide access to treatment for end-stage kidney disease: A systematic review. *Lancet* 385: 1975–1982, 2015
3. Eckardt KU, Bansal N, Coresh J, Evans M, Grams ME, Herzog CA, et al.; Conference Participants: 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* 93: 1281–1292, 2018
4. Chapter 2: Definition, identification, and prediction of CKD progression. *Kidney Int Suppl (2011)* 3: 63–72, 2013
5. Schell JO, Patel UD, Steinhauser KE, Ammarell N, Tulsky JA: Discussions of the kidney disease trajectory by elderly patients and nephrologists: A qualitative study. *Am J Kidney Dis* 59: 495–503, 2012
6. Schroeder EB, Yang X, Thorp ML, Arnold BM, Tabano DC, Petrik AF, et al.: Predicting 5-year risk of RRT in stage 3 or 4 CKD: Development and external validation. *Clin J Am Soc Nephrol* 12: 87–94, 2017
7. Tangri N, Inker LA, Hiebert B, Wong J, Naimark D, Kent D, et al.: A dynamic predictive model for progression of CKD. *Am J Kidney Dis* 69: 514–520, 2017
8. Tangri N, Grams ME, Levey AS, Coresh J, Appel LJ, Astor BC, et al.; CKD Prognosis Consortium: Multinational assessment of accuracy of equations for predicting risk of kidney failure: A meta-analysis [published correction appears in *JAMA* 315: 822, 2016 10.1001/jama.2016.0342]. *JAMA* 315: 164–174, 2016
9. Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al.: A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* 305: 1553–1559, 2011

10. Landray MJ, Emberson JR, Blackwell L, Dasgupta T, Zakeri R, Morgan MD, 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* 56: 1082–1094, 2010
11. Day CJ, Howie AJ, Nightingale P, Shabir S, Adu D, Savage CO, et al.: Prediction of ESRD in pauci-immune necrotizing glomerulonephritis: Quantitative histomorphometric assessment and serum creatinine. *Am J Kidney Dis* 55: 250–258, 2010
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* 52: 653–660, 2008
13. Grams ME, Sang Y, Ballew SH, Carrero JJ, Djurdjev O, Heerspink HJL, et al.: Predicting timing of clinical outcomes in patients with chronic kidney disease and severely decreased glomerular filtration rate. *Kidney Int* 93: 1442–1451, 2018
14. Farrington K, Covic A, Nistor I, Aucella F, Clyne N, De Vos L, 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. *Nephrol Dial Transplant* 32: 9–16, 2017
15. Major RW, Shepherd D, Medcalf JF, Xu G, Gray LJ, Brunskill NJ: 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 [published correction appears in *PLoS Med* 17: e1003313, 2020 10.1371/journal.pmed.1003313]. *PLoS Med* 16: e1002955, 2019
16. Ramspek CL, de Jong Y, Dekker FW, van Diepen M: Towards the best kidney failure prediction tool: A systematic review and selection aid. *Nephrol Dial Transplant* 35: 1527–1538, 2020
17. Li L, Yang W, Astor BC, Greene T: Competing risk modeling: Time to put it in our standard analytical toolbox. *J Am Soc Nephrol* 30: 2284–2286, 2019
18. Wolkewitz M, Cooper BS, Bonten MJ, Barnett AG, Schumacher M: Interpreting and comparing risks in the presence of competing events. *BMJ* 349: g5060, 2014
19. Ravani P, Fiocco M, Liu P, Quinn RR, Hemmelgarn B, James M, et al.: Influence of mortality on estimating the risk of kidney failure in people with stage 4 CKD. *J Am Soc Nephrol* 30: 2219–2227, 2019
20. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al.; PROBAST Group†: PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 170: 51–58, 2019
21. Collins GS, Reitsma JB, Altman DG, Moons KG: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ* 350: g7594, 2015
22. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al.: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Ann Intern Med* 162: W1–W73, 2015
23. Jager KJ, Ocak G, Drechsler C, Caskey FJ, Evans M, Postorino M, et al.: The EQUAL study: A European study in chronic kidney disease stage 4 patients. *Nephrol Dial Transplant* 27[Suppl 3]: iii27–iii31, 2012 J
24. Austin PC, Lee DS, Fine JP: Introduction to the analysis of survival data in the presence of competing risks. *Circulation* 133: 601–609, 2016
25. de Goeij MC, van Diepen M, Jager KJ, Tripepi G, Zoccali C, Dekker FW: Multiple imputation: Dealing with missing data. *Nephrol Dial Transplant* 28: 2415–2420, 2013
26. van Buuren S: Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 16: 219–242, 2007
27. Wolbers M, Koller MT, Witteman JC, Steyerberg EW: Prognostic models with competing risks: Methods and application to coronary risk prediction. *Epidemiology* 20: 555–561, 2009
28. Rubin DB: *Multiple Imputation for Nonresponse in Surveys*, New York, Wiley, 1987
29. Royston P, Moons KG, Altman DG, Vergouwe Y: Prognosis and prognostic research: Developing a prognostic model. *BMJ* 338: b604, 2009
30. Vergouwe Y, Moons KG, Steyerberg EW: External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol* 172: 971–980, 2010
31. Austin PC, Steyerberg EW: Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med* 33: 517–535, 2014
32. Brück K, Jager KJ, Zoccali C, Bello AK, Minutolo R, Ioannou K, et al.; European CKD Burden Consortium: Different rates of progression and mortality in patients with chronic kidney disease at outpatient nephrology clinics across Europe. *Kidney Int* 93: 1432–1441, 2018
33. Pencina MJ, D'Agostino RB Sr: Evaluating discrimination of risk prediction models: The C statistic. *JAMA* 314: 1063–1064, 2015
34. Peeters MJ, van Zuilen AD, van den Brand JA, Bots ML, Blankestijn PJ, Wetzels JF; MASTERPLAN Study Group: Validation of the kidney failure risk equation in European CKD patients. *Nephrol Dial Transplant* 28: 1773–1779, 2013
35. Riphagen IJ, Kleefstra N, Drion I, Alkhalaf A, van Diepen M, Cao Q, et al.: Comparison of methods for renal risk prediction in patients with type 2 diabetes (ZODIAC-36). *PLoS One* 10: e0120477, 2015
36. Tangri N, Ferguson T, Komenda P: Pro: Risk scores for chronic kidney disease progression are robust, powerful and ready for implementation. *Nephrol Dial Transplant* 32: 748–751, 2017
37. Cook NR: Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. *Clin Chem* 54: 17–23, 2008
38. Cook NR: Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115: 928–935, 2007
39. van Klaveren D, Gönen M, Steyerberg EW, Vergouwe Y: A new concordance measure for risk prediction models in external validation settings. *Stat Med* 35: 4136–4152, 2016
40. Chiu HH, Tangri N, Djurdjev O, Barrett BJ, Hemmelgarn BR, Madore F, et al.: Perceptions of prognostic risks in chronic kidney disease: A national survey. *Can J Kidney Health Dis* 2: 53, 2015
41. Steyerberg EW, Vergouwe Y: Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur Heart J* 35: 1925–1931, 2014
42. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al.: Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* 21: 128–138, 2010
43. Ramspek CL, Voskamp PW, van Ittersum FJ, Krediet RT, Dekker FW, van Diepen M: Prediction models for the mortality risk in chronic dialysis patients: A systematic review and independent external validation study. *Clin Epidemiol* 9: 451–464, 2017
44. Ramer SJ, McCall NN, Robinson-Cohen C, Siew ED, Salat H, Bian A, et al.: Health outcome priorities of older adults with advanced CKD and concordance with their nephrology providers' perceptions. *J Am Soc Nephrol* 29: 2870–2878, 2018
45. Couchoud C, Hemmelgarn B, Kotanko P, Germain MJ, Moranne O, Davison SN: Supportive care: Time to change our prognostic tools and their use in CKD. *Clin J Am Soc Nephrol* 11: 1892–1901, 2016
46. Verberne WR, Das-Gupta Z, Allegretti AS, Bart HAJ, van Biesen W, García-García G, 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. *Am J Kidney Dis* 73: 372–384, 2019
47. Drawz PE, Goswami P, Azem R, Babineau DC, Rahman M: A simple tool to predict end-stage renal disease within 1 year in elderly adults with advanced chronic kidney disease. *J Am Geriatr Soc* 61: 762–768, 2013
48. Marks A, Fluck N, Prescott GJ, Robertson L, Simpson WG, Cairns Smith W, et al.: Looking to the future: Predicting renal replacement outcomes in a large community cohort with chronic kidney disease. *Nephrol Dial Transplant* 30: 1507–1517, 2015

AFFILIATIONS

[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 Institute and Karolinska University Hospital, Stockholm, Sweden
[3]Division of Nephrology, University Hospital of Wurzburg, Wurzburg, Germany
[4]Division of Nephrology, Department of Internal Medicine 1, University Hospital Wurzburg, Wurzburg, Germany
[5]Department of Medical Informatics, European Renal Association–European Dialysis and Transplant Association Registry, Amsterdam University Medical Center, University of Amsterdam, Amsterdam Public Health Institute, Amsterdam, The Netherlands
[6]Department of Nephrology and Transplantation Medicine, Wroclaw Medical University, Wroclaw, Poland
[7]Department of Clinical Epidemiology of Renal Diseases and Hypertension, Consiglio Nazionale della Ricerche - Istituto di fisiologia clinica, Reggio Calabria, Italy
[8]Department of Translational Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
[9]United Kingdom Renal Registry, Bristol, United Kingdom
[10]Departmen of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom