

Prediction models for the mortality risk in chronic dialysis patients: a systematic review and independent external validation study

Chava L Ramspek¹
Pauline WM Voskamp¹
Frans J van Ittersum²
Raymond T Krediet³
Friedo W Dekker¹
Merel van Diepen¹

On behalf of the
NECOSAD study group

¹Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, ²Department of Nephrology, VU University Medical Center, ³Department of Nephrology, Academic Medical Center, Amsterdam, The Netherlands

Objective: In medicine, many more prediction models have been developed than are implemented or used in clinical practice. These models cannot be recommended for clinical use before external validity is established. Though various models to predict mortality in dialysis patients have been published, very few have been validated and none are used in routine clinical practice. The aim of the current study was to identify existing models for predicting mortality in dialysis patients through a review and subsequently to externally validate these models in the same large independent patient cohort, in order to assess and compare their predictive capacities.

Methods: A systematic review was performed following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines. To account for missing data, multiple imputation was performed. The original prediction formulae were extracted from selected studies. The probability of death per model was calculated for each individual within the Netherlands Cooperative Study on the Adequacy of Dialysis (NECOSAD). The predictive performance of the models was assessed based on their discrimination and calibration.

Results: In total, 16 articles were included in the systematic review. External validation was performed in 1,943 dialysis patients from NECOSAD for a total of seven models. The models performed moderately to well in terms of discrimination, with *C*-statistics ranging from 0.710 (interquartile range 0.708–0.711) to 0.752 (interquartile range 0.750–0.753) for a time frame of 1 year. According to the calibration, most models overestimated the probability of death.

Conclusion: Overall, the performance of the models was poorer in the external validation than in the original population, affirming the importance of external validation. Floege et al's models showed the highest predictive performance. The present study is a step forward in the use of a prediction model as a useful tool for nephrologists, using evidence-based medicine that combines individual clinical expertise, patients' choices, and the best available external evidence.

Keywords: external validation, prediction, nephrology, dialysis, mortality

Introduction

End-stage renal disease (ESRD) is a large public health problem. Treatment options for ESRD consist of transplantation, conservative treatment, or dialysis, of which there are two main forms: hemodialysis (HD) and peritoneal dialysis (PD). Dialysis patients are a very vulnerable population, with 3-year cumulative survival of approximately 50%, considerably lower than survival rates for cancer and heart-failure patients.^{1,2} Many different patient characteristics have been reported to influence this poor survival, and there is wide interindividual variability in prognosis.^{3–5}

Identifying dialysis patients with an increased risk of dying through prediction models could offer valuable information on prognosis to patients, as well as help clinicians

Correspondence: Chava L Ramspek
Department of Clinical Epidemiology,
Leiden University Medical Center,
2 Albinusdreef, Leiden 2300, The
Netherlands
Tel +31 71 526 4037
Fax +31 71 526 6994
Email c.l.ramspek@lumc.nl

recognize these high-risk patients and apply more intensive care and supportive therapies when needed.^{3,6,7} Furthermore, a risk-stratification model could be used as an important research tool when adjusting for baseline risk in comparative studies, targeting specific interventions to high- or low-risk patients or even provide tools to compare mortality risks among different centers, regions, or countries.^{3,7,8}

Various models to predict mortality in dialysis patients have been published.^{7–13} However, no review of these articles exists, and only a fraction of these prediction models or risk scores have been externally validated.⁷ The few studies that have externally validated models to predict mortality in dialysis patients did so for not more than one or two models at a time.^{7,12,13} Differences in cohorts used to validate various models make it difficult to compare predictive performances fairly; ideally, one would externally validate all available models in the same cohort. External validation entails that a model is tested in other data than that in which the model was derived, in order to test the generalizability of the tool on the targeted patient population.¹⁴ Since the performance of a prediction model is generally lower in new individuals than in the population in which the model was developed,¹⁵ these models cannot be recommended for clinical use before external validity is established, as is also stressed in the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement for reporting prediction models.^{14–20} Though the importance of combining evidence is recognized, many more models have been developed in the medical research field than are used in clinical practice or even tested on different populations.^{17,21,22} To combat this research waste, the realization of more external validation studies is imperative and increasingly in demand.^{14,17,20,23–25}

A direct comparison of the predictive performance of existing models in the same external-validation cohort could help bridge the gap between the development of models and their clinical application. The aim of the current study was to perform a systematic literature review to identify existing models for predicting the risk of death in dialysis patients and subsequently to externally validate these models in the same large independent patient cohort, in order to assess and compare their predictive capacities.

Methods

Systematic review

The PRISMA (preferred reporting items for systematic reviews and meta-analyses) guidelines were followed for the review process.²⁶ A systematic literature search was performed in PubMed to identify articles that described prediction models for risk of death in dialysis patients.

PubMed was searched on October 28, 2016 using the search string presented in the [Supplementary materials](#). Additional relevant articles were identified by checking the references of the selected articles from the PubMed search. Titles and abstracts were screened, and full texts of potentially eligible articles were read carefully.

Studies were included in the systematic review if they met the following criteria: 1) the study contained at least one formal, original prediction model from which individual risk estimates could be calculated; 2) the study population consisted of chronic dialysis patients or patients with ESRD; 3) the end point of the prediction was all-cause mortality for a defined time frame, studied in a longitudinal design; and 4) the goal of the study was to predict mortality to the best of the authors' ability, and not to research whether a particular comorbidity score or single variable was predictive of mortality.

Titles, abstracts, and full papers were reviewed by two authors independently (CLR and PWMV), and any discrepancies in the full-text review were solved by consulting a third party (MvD). For the articles included, two investigators (CLR and PWMV) independently extracted the main characteristics and necessary data from the main paper and any accompanying supplemental material, as recommended by the PRISMA guidelines.²⁶

Validation cohort

NECOSAD (Netherlands Cooperative Study on the Adequacy of Dialysis) was a multicenter, prospective cohort study, in which 38 dialysis centers throughout the Netherlands participated. Incident dialysis patients were included at the start of dialysis treatment if they were aged ≥ 18 years and had no history of previous renal replacement therapy. Patient characteristics on dialysis modality, comorbidities, laboratory investigations, and clinical parameters were recorded. High-sensitivity CRP was not assessed in standard fashion, but retrospectively determined in 2002 from frozen blood samples for all patients included till then. Patients were followed till time of death or censored due to kidney transplantation or loss to follow-up. Inclusion of patients took place between 1997 and 2007, and follow-up data on death were available until February 1, 2015. For more information on study design, please refer to one of the hallmark NECOSAD articles published in the *Lancet*.²⁷ The study protocol was fully approved by the Amsterdam Medical Centre ethics committee, and subsequently all local medical ethics committees gave a declaration of agreement. All patients gave written informed consent. The NECOSAD data set will be made available upon request.

For the current study, baseline measurements were taken 3 months after dialysis initiation, with the exception of predialysis serum creatinine. This was primarily done to ascertain that the dialysis patients included had chronic renal failure and the choice of treatment modality would be more definite. Furthermore, laboratory investigations were collected at this time, since the clinical condition of the patient was more stabilized. Patients who had a follow-up within NECOSAD of less than 90 days were thus excluded from the current study.

Studies selected for validation

Predictors from the studies included in the review were assessed and matched to variables available in NECOSAD. For predictors that were not recorded in NECOSAD, a panel of researchers and nephrologists was consulted on the possible use of proxies. When the panel considered it appropriate, original predictors were replaced with proxies, in order to allow validation of the model. If these studies presented more than one formal prediction model, their final model or the model with the best predictive performance was selected. The original prediction models with regression coefficients per predictor and intercept and/or baseline hazard were extracted from the selected studies and used as published to evaluate predictive performance within our external population. If the full prediction formula was not provided in the article or its supplementary material, the authors were approached multiple times by email or phone with a request to supply this formula.

Statistical analysis

Baseline characteristics of the patients in NECOSAD are presented as mean values with standard deviations (SD) or as median values with interquartile ranges (IQRs) for continuous variables (depending on the normality of the distribution), and as percentages for categorical variables. Baseline characteristics included all the predictors from the validated studies, and were stratified per treatment modality.

To account for missing data on predictors, multiple imputation was performed using the fully conditional specification.^{28–31} The imputation model contained all predictor variables, outcome variables, and any variables that could help impute the missing data.³⁰ Normally distributed variables were log-transformed. Outcomes from the ten imputed data sets were combined into one overall estimate, with associated variance that incorporated both within- and between-imputation variability. This was done through applying the rules described by Marshall et al, which are an extension of Rubin's rules aimed specifically at imputation in prognostic modeling studies.³⁰

Probabilities of death predicted by each validated model were calculated for each individual within the validation cohort, using the original model formulae and the time frame used in the original papers. The predictive performance of the models was assessed based on outcomes for discrimination and calibration, and these measures are described later (see [Supplementary materials](#) for a full description of the model formulae and predictors/proxies used).

Discrimination is a measure for how well a model is able to distinguish between people who will or will not have the event (death) within the given time frame.³² It can be assessed with the *C*-statistic, which indicates the probability that the model will assign a higher risk to a random case than to a random non-case.³³ The *C*-statistic lies between 0.5 and 1, where 1 is perfect, 0.7 is considered reasonable, and 0.8 is considered good discrimination. For a logistic regression model, the *C*-statistic is equivalent to the area under the receiver-operating curve. For Cox proportional-hazard models, Harrell's *C*-index of concordance is presented. This outcome measure takes censoring into account.³⁴ The *C*-statistic was assessed per model for multiple prediction time frames and stratified by dialysis modality. Additionally, we computed Kaplan–Meier survival curves for four risk groups, based on quartiles of prognostic indices. These survival curves can provide informal evidence of discrimination, and allow for a visual interpretation of model performance over time: the further the survival curves lie from one another, the better the discrimination.³⁴

Calibration describes how accurately the predicted risks from a model reflect the mortality risks in the observed data.³⁴ This is best represented visually in a calibration plot, in which observed mean mortality risk is plotted against predicted mean mortality risk by a tenth of the predicted risk.^{22,33,35} The calibration plot can be characterized by a slope and calibration-in-the-large. The calibration-in-the-large indicates to what extent the model may systematically over- or underestimate the probability of death. For a perfectly calibrated risk score, the calibration plot would lie on the 45° line and have a slope of 1, and the calibration-in-the-large would be identical to the observed risk.^{24,34,35} In the current study, calibration was assessed per model by presenting a calibration plot, calibration slope, and calibration-in-the-large for multiple prediction time frames (if possible) and stratified by dialysis modality.

To test the additional benefit of recalibration, we updated the models in a sensitivity analysis. For the Cox models, this was done by adjusting the baseline hazard of death to the hazard of death observed within the NECOSAD population in the model.^{15,25,36} For the logistic model, this was done by

calculating a correction factor for the intercept based on the incidence of death in the validation cohort, as described by Janssen et al.³⁶ Recalibration can improve calibration measures, but does not affect the discrimination (*C*-statistic) of the model performance. All statistical analyses were performed in SPSS version 23.

Results

Systematic review

Using the PubMed search strategy ([Supplementary materials](#)), 508 articles were identified and their titles screened. From these articles, 112 were selected for abstract review. From the reviewed abstracts, 42 articles were selected for

full-text review and three that had been identified through references of other articles were added to the full-text review selection. After the full review 16 articles were included in the systematic review, based on the predefined inclusion criteria as described in the “Methods” section. The study-inclusion flowchart is presented in Figure 1. The main reasons for exclusion were that no formal prediction model was presented or that the goal of the model was not to predict mortality in individual patients. Characteristics from the selected studies were collected by two independent researchers, and are summarized in Table 1. Most of the studies were performed on HD patients, as opposed to PD patients, and made use of Cox models. Many studies used predictors not routinely

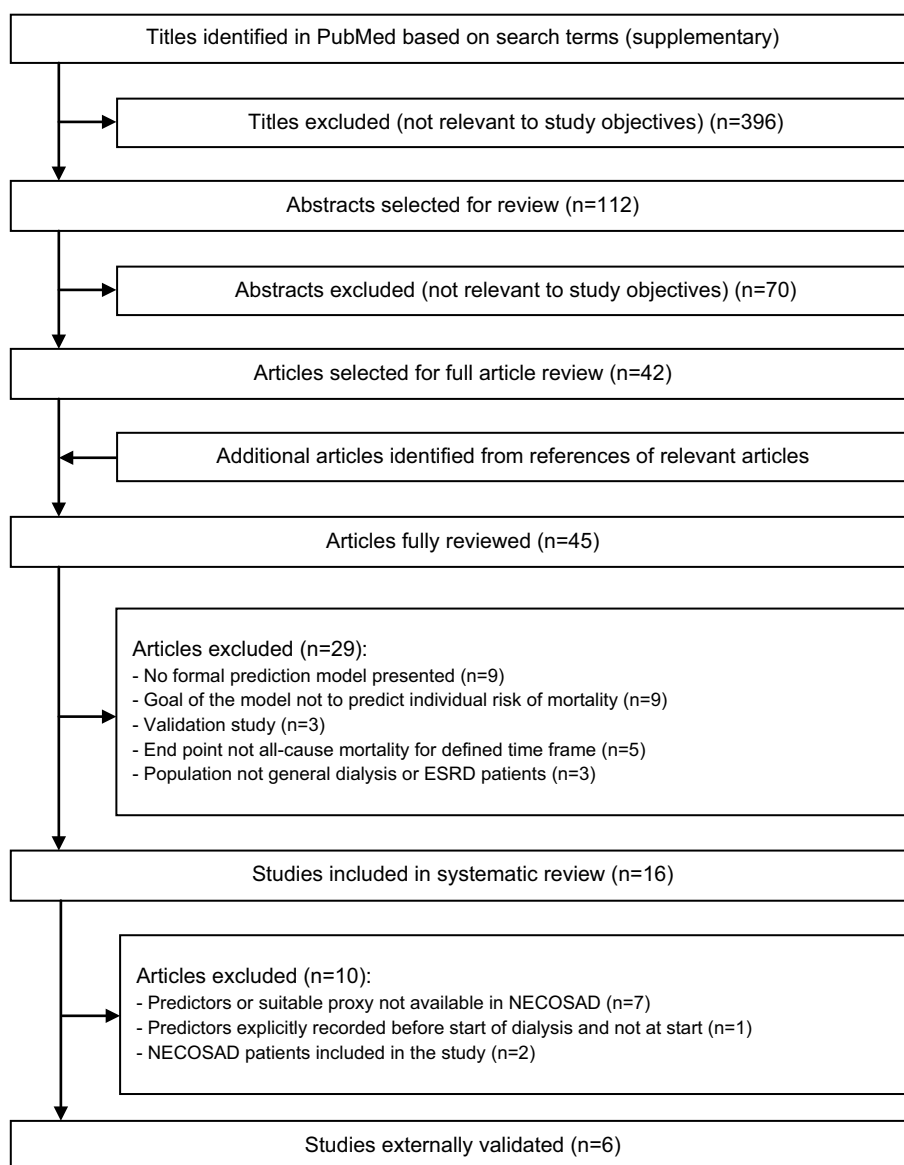


Figure 1 Flowchart of study selection for systematic review and external validation.

Abbreviations: ESRD, end-stage renal disease; NECOSAD, Netherlands Cooperative Study on the Adequacy of Dialysis.

collected by nephrologists; these were: socioeconomic status, fasting blood glucose, left ventricular ejection fraction, surprise question, plasma S100A12, peritoneal equilibration test, and ventilation.^{10,11,37–41}

As shown in Table 1, a discrimination value was presented for 13 of the 16 models, and ranged from 0.71 to 0.87 in the derivation cohorts. For 12 models, information on calibration in the form of a plot or table was present. A full prediction formula, including baseline hazard and/or intercept and regression coefficients, was available in ten of 16 studies. The baseline hazard function, with which one can calculate predicted probabilities for any time frame in a Cox model, was only offered in one study by Geddes et al.³ Multiple studies offered their estimated baseline hazard for multiple time frames, allowing the model to be used to predict mortality over different prediction horizons.

Selection for external validation

All studies included in the systematic review were considered carefully for validation. A number of studies used very specific predictors that could not be replaced by a proxy. These studies (n=7) were excluded from the validation (see [Supplementary material](#) for the specific variables).^{10,11,37–41} Two studies were excluded because they were (partly) developed on NECOSAD patients.^{42,43} Finally, one study was excluded because the authors explicitly stated that they had recorded all their predictors before the initiation of dialysis and the values of these predictors change notably after dialysis initiation.⁴⁴ As shown in Figure 1, a total of six studies presenting seven final-prediction models were finally selected for external validation. Of these studies, four were developed in HD patients and two in dialysis patients in general (HD and PD). The statistical model used was a Cox model in all but one of the validated models, the exception being Mauri et al.⁴⁵ Upon request, the full prediction formula was made available for all studies.

Baseline characteristics of the validation cohort

Baseline patient characteristics of the NECOSAD patients included at 3 months after dialysis initiation stratified for dialysis modality are summarized in Table 2. The mean time between dialysis initiation and baseline measurement was 93 (SD 13) days. As expected, some marked differences between HD and PD patients were present. For example, the median age was much lower in PD patients. In [Table S1](#), baseline characteristics are given for each predictor/proxy used for validation. In [Table S2](#), baseline characteristics from the

derivation cohorts of the validated models are shown. In the first year after baseline (at approximately 15 months after dialysis initiation), 12.6% of the patients (n=244) died. After 2 years 23.3% (n=452) had passed away, and this was 30.2% (n=586) after 3 years and 40.5% (n=787) after 5 years.

Performance of validated models

Table 3 shows the discrimination of the validated models presented as the pooled mean Harrell's C-index from the ten imputation sets with IQRs. In Table 4, the discrimination is presented when stratified for dialysis modality. The models performed moderately to well in terms of discrimination, with C-statistics ranging from 0.71 (IQR 0.708–0.711) to 0.752 (IQR 0.750–0.753) for a time frame of 1 year. Models tested for different time frames performed consistently more poorly when the prediction horizon increased. All models performed better in the PD population than in the HD population. The model of Wagner et al⁸ showed the highest discriminatory ability for 1 year, with a C-statistic of 0.752 (IQR 0.750–0.753), while models by Floege et al⁷ showed similarly high discrimination, with highest discriminatory ability for a 2-year prediction time frame with a C-statistic of 0.740 (IQR 0.737–0.742). To assist comparison, C-statistics found through current external validation were plotted against original C-statistics (if available; Figure 2). As one can tell from this figure, all studies that presented a C-statistic performed more poorly in external validation. In Figure 3, Kaplan–Meier curves are shown per model for quartiles of the prognostic indices. This clarifies the models' capability of distinguishing between patients with high or low mortality rates, and gives insight on how the discrimination capacity changes over time.

The overall calibration of the models can best be judged by examining the calibration plots presented in Figure 4 and taking the calibration slope and calibration-in-the-large into account, as shown in Tables 5 and 6. Calibration plots are presented for single time frames and combined for HD and PD patients (see [Figure S1](#) for calibration plots stratified by dialysis modality and for various prediction time frames). Though the model tended to overestimate slightly, the calibration of Holme et al's⁹ model was fairly good, as can be seen in the plot, and was exemplified by a calibration slope close to 1. Floege et al's⁷ 1- and 2-year models also showed fairly good calibration. The models of Geddes et al³ and Hutchinson et al⁵⁵ both overestimated considerably, especially in the higher risk ranges. The model of Wagner et al⁸ significantly underestimated the probability of mortality, and the model of Mauri et al⁴⁵ slightly underestimated and had a poor

Table 1 General characteristics of models to predict risk of all-cause mortality in dialysis patients

Reference	Events:sample size	Study population	Patient inclusion time span	Outcome, prediction horizon	Statistical model
Hemke et al, ⁴³ The Netherlands, "easy" model	1,225	Incident dialysis patients	1997–2007	3-, 5-, 10-year mortality	Cox
Cao et al, ⁴⁰ China	371:5,405	Incident CAPD patients	2012	1-year mortality	Cox
Doi et al, ⁴⁴ Japan	62:688	Incident HD patients	2006–2011	1-year mortality	Logistic
Floege et al, ⁷ Germany*	1,060, 1,714:9,722 ^b	Incident and prevalent HD patients	2007–2009	1- and 2-year mortality	Cox
Chua et al, ³⁷ Singapore	169:983	Incident dialysis patients	2005–2010	1-year mortality	Logistic and linear
Zhao et al, ¹¹ China, model 3	175:903	Incident CAPD patients	2006–2011	2-year mortality	Cox
Hemke et al, ⁴² The Netherlands	–, 3502,** 4549**,:6934 ^c	Incident RRT patients	1995–2005	3-, 5-, 10-year mortality	Cox
Shiotsu et al, ³⁹ Japan	50:550	Prevalent HD patients	2009	2-year mortality	Cox
Holme et al, ⁹ Norway*	880:1,868	Prevalent HD patients	2003–2004	3-year mortality	Cox
Quinn et al, ¹⁰ Canada	1,326:8,103	Incident dialysis patients	1998–2005	1-year mortality	Logistic
Wagner et al, ⁸ UK, model 3*	1,078:3,631	Incident dialysis patients	2002–2004	3-year mortality	Cox
Cohen et al, ³⁸ USA	43:512	Prevalent HD patients	2006–2007	6-month mortality	Cox
Mauri et al, ⁴⁵ Spain*	558:3,455	Incident HD patients	1997–2003	1-year mortality	Logistic
Geddes et al, ³ UK, MVR model*	351,** 837**,:1,139 ^d	Incident RRT patients	1997	1- and 5-year mortality	Cox
Foley, ⁴¹ Canada	73:325	Incident dialysis patients	1980–1991	6-month mortality (logistic)	Logistic and Cox model
Hutchinson et al, ⁵⁵ Canada*	101**,:220	Incident HD patients	1970–1975	5-year mortality	Cox

Notes: *Externally validated in the current independent study; **deducted/calculated based on information available in article; ***baseline hazard provided upon request; ****shrinkage factor provided upon request. ^a3, 5, 10-year mortality, respectively; ^b1 and 2-year mortality events respectively; ^c3, 5 and 10 year mortality events, respectively; ^d1, and 5-year mortality events, respectively. "–" not provided.

Abbreviations: CAPD, continuous ambulatory peritoneal dialysis; HD, hemodialysis; HR, hazard ratio; MVR, multivariate regression; RRT, renal replacement therapy.

slope of 0.27. After adjustment for differences in mortality rates between NECOSAD and the development cohorts as a sensitivity analysis, overall the calibration plots improved (Table S3 and Figure S2).

Discussion

In this study, which included a systematic review and external validation, it was shown that there are 16 published models predicting death in dialysis patients, which previously have barely been validated and have not been implemented in clinical practice on a large scale. Validating them in a large independent cohort showed that they all performed more

poorly on new patients, demonstrating the importance of external validation before clinical implementation in the shared decision-making process.

The systematic review exemplified that many articles describing the development of a prediction model used rare predictors that are difficult to validate. Besides complicating validation, the use of rare predictors makes these tools extremely impractical for nephrologists and large-scale clinical implementation highly improbable. The systematic review also showed that many studies were incomplete in providing all information necessary for complete external validation or described outcome-assessment measures differently from

Validation	Predictors, n	Discrimination, C-statistic (CI) (of derivation)	Discrimination, C-statistic (CI) (of validation)	Calibration plot/table, slope	Presentation of model
Internal split-sample validation (610 patients)	10	–	0.79, 0.79, 0.78 ^a	Yes	Full prediction formula provided
External validation (1,764 patients)	6	0.71 (0.6–0.83)	0.72 (0.63–0.81)	–	β -value per predictor given
Internal bootstrap validation	6	0.83	0.83 (0.79–0.89)	Yes	β -value per predictor given and score chart
External validation (10,615 patients)	12, 13 ^b	–	0.73, 0.72 ^b	Yes	Full prediction formulae provided
–	7	0.74, 0.74, 0.75 (logistic, linear, score)	–	–	Logistic, linear model, and score chart presented
Internal split-sample validation (451 patients)	10	0.79	0.76 (0.68–0.84)	Yes	Full prediction formula provided
Internal split-sample validation (6,934 patients)	4	–	0.71 (0.7–0.72), 0.72 (0.7–0.72), 0.72 (0.71–0.73) ^c	Yes	Full prediction formula provided
Bootstrapping and external validation (303 patients)	4	0.73 (0.66–0.8)	0.72 (0.63–0.82)	Yes	β -value per predictor given and score chart
Internal split-sample validation (905 patients)	5	0.73 (0.71–0.75)	0.72 (0.69–0.76)	Yes	Full prediction formula provided
Internal split-sample validation, twice (4,052 and 4,051 patients) and bootstrapping	15	0.77	0.76 and 0.76	Yes	Full prediction formula provided and score chart
Internal split-sample validation (1,816 patients)	11	0.75 (0.73–0.77)	0.73 (0.71–0.76)	Yes	HR per predictor given***
External validation (514 patients)	5	0.87 (0.82–0.92)	0.8 (0.73–0.88)	–	HR per predictor given
Internal split-sample validation (2,283 patients)	10	0.78	0.78	Yes	Full prediction formula provided
Internal split-sample validation (1,171 patients)	17	–	–	Yes	Full prediction formula provided****
–	14	–	–	Yes	Full prediction formula (logistic), HRs for Cox and score chart
–	3	–	–	Yes	Full prediction formula and score chart

standard discrimination and calibration, thus complicating comparison of different models and comparison of model performance in different populations. Similar problems gave rise to the development of the TRIPOD guidelines and were also described in a review by Bouwmeester et al, in which it was concluded that the majority of prediction studies do not follow current methodological recommendations.^{20,46}

The current study externally validated seven models (presented in six studies) on the same patient population for the same outcome measures. Depending on the intended use of the models and the available predictors, the preferred model

to implement would differ. Considering discrimination, the model proposed by Wagner et al⁸ seems to be best suited for distinguishing high- and low-mortality-risk patients in a population of HD and PD patients over a time frame of 1 year in this validation cohort. The models of Floege et al⁷ also showed high discriminatory ability, with the highest *C*-statistics for a 2-year time frame. For predicting 3-year probability of mortality, Holme et al's⁹ model performed best on discrimination. As far as calibration goes, most models showed poor calibration when eyeing the calibration plots, and calibration seemed to be best in the models of Floege et al⁷ and Holme et al⁹.

Table 2 Baseline characteristics of external validation cohort (NECOSAD) at 3 months after dialysis initiation

Characteristics	Missing	Total (n=1,943)	HD patients (n=1,266)	PD patients (n=677)
Age, years	0.2%	62.8 (49.8–72.2)	66.7 (55.6–74.4)	54.3 (43.0–65.1)
Sex (% male)	0.2%	62.3%	59.4%	67.8%
Current smokers	10.2%	22.5%	21.2%	24.8%
Ethnicity (% Caucasian)	0.2%	91.9%	92.0%	91.6%
Primary kidney disease	0.2%			
Diabetes mellitus		14.4%	14.1%	15.1%
Glomerulonephritis		12.4%	9%	18.8%
Renal vascular disease		16.9%	19.5%	11.8%
Other		56.3%	57.4%	54.3%
Comorbidities				
Cardiovascular disease	9.8%	35.3%	40.9%	25.3%
Diabetes mellitus	0.2%	5.6%	6.8%	3.3%
Cancer	10.8%	9.6%	11.9%	5.5%
Medication				
ACE inhibitors	0.6%	24.5%	22.4%	28.6%
Calcium antagonists	0.6%	34.1%	32.4%	37.3%
β-blockers	0.6%	31.7%	30.3%	34.2%
Diuretics	0.6%	27.3%	27.0%	28.0%
Erythropoietin	1.5%	85.3%	91.7%	73.2%
Laboratory				
Albumin, g/L	5.8%	36.0 (5.3)	35.8 (5.3)	36.2 (5.3)
mGFR, mL/min/1.73 m ²	21.0%	3.4 (1.8–5.4)	3.1 (1.6–4.9)	4.1 (2.2–6.3)
Hemoglobin, mmol/L	1.6%	6.9 (1.0)	6.7 (0.9)	7.4 (1.0)
HS-CRP mg/L	56.6%	5.0 (3.0–14.0)	6.0 (3.0–16.0)	4.0 (3.0–9.0)
BMI (kg/m²)	10.1%	24.7 (4.2)	24.7 (4.4)	24.8 (3.9)
Vascular access mode	4.6%	NA		NA
Catheter		NA	21.0%	NA
Graft/fistula		NA	79.0%	NA

Notes: Continuous variables presented as mean values with standard deviations for normally distributed variables. Continuous variables presented as median values with interquartile ranges for non-normal distributed variables. Categorical variables presented as valid percentages.

Abbreviations: NECOSAD, Netherlands Cooperative Study on the Adequacy of Dialysis; HD, hemodialysis; PD, peritoneal dialysis; mGFR, measured glomerular filtration rate (combined creatinine and urea clearance in 24-hour urine samples at 3 months); BMI, body-mass index; NA, not applicable; HS-CROP, high sensitivity C-reactive protein.

Table 3 Discrimination results of external validation for HD and PD patients

Reference	Time frame	Original population	Discrimination: C-statistic,* HD and PD
Floege et al ⁷	1 year	HD	0.740 (0.738–0.742)
Floege et al ⁷	2 years	HD	0.740 (0.737–0.742)
Holme et al ⁹	3 years	HD	0.734 (0.730–0.737)
Wagner et al ⁸	1 year	HD and PD	0.752 (0.750–0.753)
	2 years		0.738 (0.737–0.738)
	3 years		0.730 (0.729–0.731)
Mauri et al ⁴⁵	1 year	HD	0.728 (0.724–0.730)
Geddes et al ³	1 year	HD and PD	0.721 (0.717–0.723)
	2 years		0.706 (0.704–0.708)
	3 years		0.705 (0.702–0.707)
	5 years		0.698 (0.695–0.700)
Hutchinson et al ⁵⁵	1 year	HD	0.710 (0.708–0.711)
	2 years		0.702 (0.701–0.706)
	3 years		0.700 (0.699–0.702)
	5 years		0.697 (0.696–0.699)

Note: *C-statistics presented as medians and interquartile ranges from pooled imputation results.

Abbreviations: HD, hemodialysis; PD, peritoneal dialysis.

When comparing the predictive performances found in this external validation to the performance reported in the original articles (Figure 2), we can state that overall performance was poorer in the validation (as expected), especially when looking at performance in only HD patients. In conclusion, the models presented by Floege et al⁷ seem to be best suited to their goal, and showed results in validation that were very consistent with their original findings. This might be due to the fact that they presented predictive performances that were already based on external validation (as well as internal validation), which they had performed on their model in the original publication, but also could have been influenced by the high resemblance among the Western European patient populations in which the model was derived and validated. In addition, parsimony of a model may weigh against a slight loss in performance. Since the models compared differed greatly in number of predictors included, one could imagine that a model with fewer predictors, such as that of Holme et al⁹, which uses only five variables, might be preferred in

Table 4 Discrimination results of external validation stratified for dialysis modality

Reference	Time frame	Original population	Discrimination: C-statistic*	
			HD	PD
Floege et al ⁷	1 year	HD	0.701 (0.699–0.709)	0.778 (0.768–0.780)
Floege et al ⁷	2 years	HD	0.711 (0.705–0.718)	0.775 (0.767–0.777)
Holme et al ⁹	3 years	HD	0.705 (0.700–0.708)	0.767 (0.765–0.768)
Wagner et al ⁸	1 year	HD and PD	0.712 (0.711–0.715)	0.709 (0.789–0.793)
	2 years		0.712 (0.709–0.713)	0.770 (0.770–0.775)
	3 years		0.703 (0.702–0.706)	0.763 (0.762–0.767)
Mauri et al ⁴⁵	1 year	HD	0.670 (0.668–0.675)	0.800 (0.791–0.808)
Geddes et al ³	1 year	HD and PD	0.676 (0.671–0.677)	0.772 (0.769–0.774)
	2 years		0.671 (0.668–0.674)	0.748 (0.746–0.749)
	3 years		0.667 (0.664–0.669)	0.750 (0.748–0.753)
	5 years		0.663 (0.660–0.665)	0.737 (0.735–0.740)
Hutchinson et al ⁵⁵	1 year	HD	0.667 (0.666–0.671)	0.767 (0.763–0.771)
	2 years		0.667 (0.666–0.669)	0.752 (0.748–0.754)
	3 years		0.663 (0.661–0.665)	0.749 (0.748–0.753)
	5 years		0.663 (0.662–0.665)	0.742 (0.741–0.745)

Note: *C-statistics presented as medians and interquartile ranges from pooled imputation results.

Abbreviations: HD, hemodialysis; PD, peritoneal dialysis.

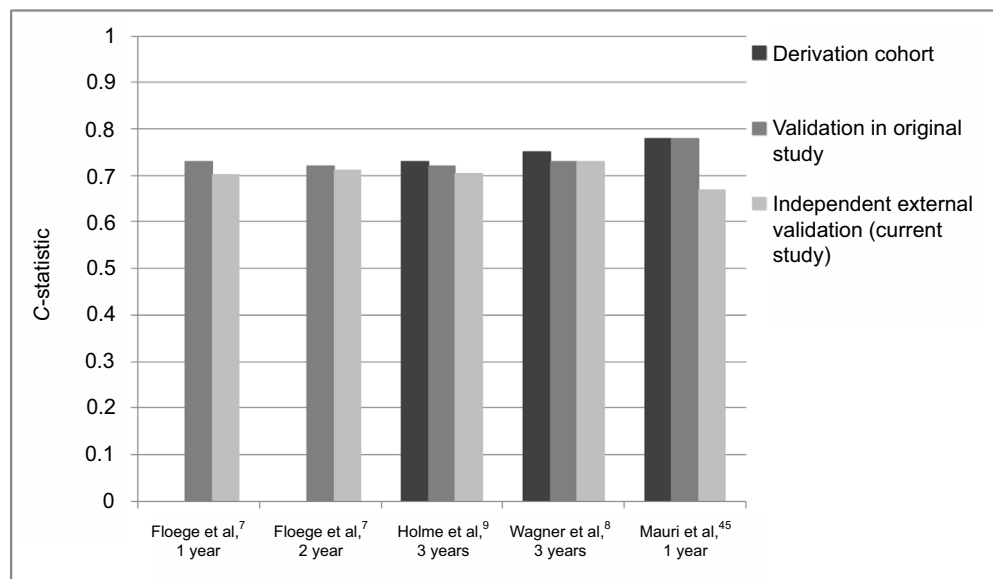


Figure 2 C-statistics reported in the original study compared to those found through external validation (for the original time frame and original dialysis modality).

Note: The validation in the original study of Floege et al⁷ was an external validation also.

a fast-paced clinical setting to Floege et al's⁷ models, which both include over ten predictors. One might also take into account that certain predictors might not be readily available in all settings. The simplicity of a model partly determines its clinical usefulness and reliability.¹⁴

By evaluating the different models in the same data set, this study provides a far greater amount of information than separate validation studies would, as attested by Collins and Moons and recently demonstrated in a study by Abbasi et al, who systematically validated prediction models for the risk of developing diabetes.^{47,48} Unfortunately,

a lack of well-conducted external-validation studies of prognostic models was found, and many validation studies handle missing data inadequately or do not report measures of model calibration.^{18,22} These are issues that we have accounted for. Still, there are a number of limitations to our study.

First, our results were influenced by differences in case-mix and mortality rates between the derivation cohorts and our validation cohort. Inherently, some of the derivation cohorts will resemble our validation cohort more than others, since patient characteristics can differ strongly per country,

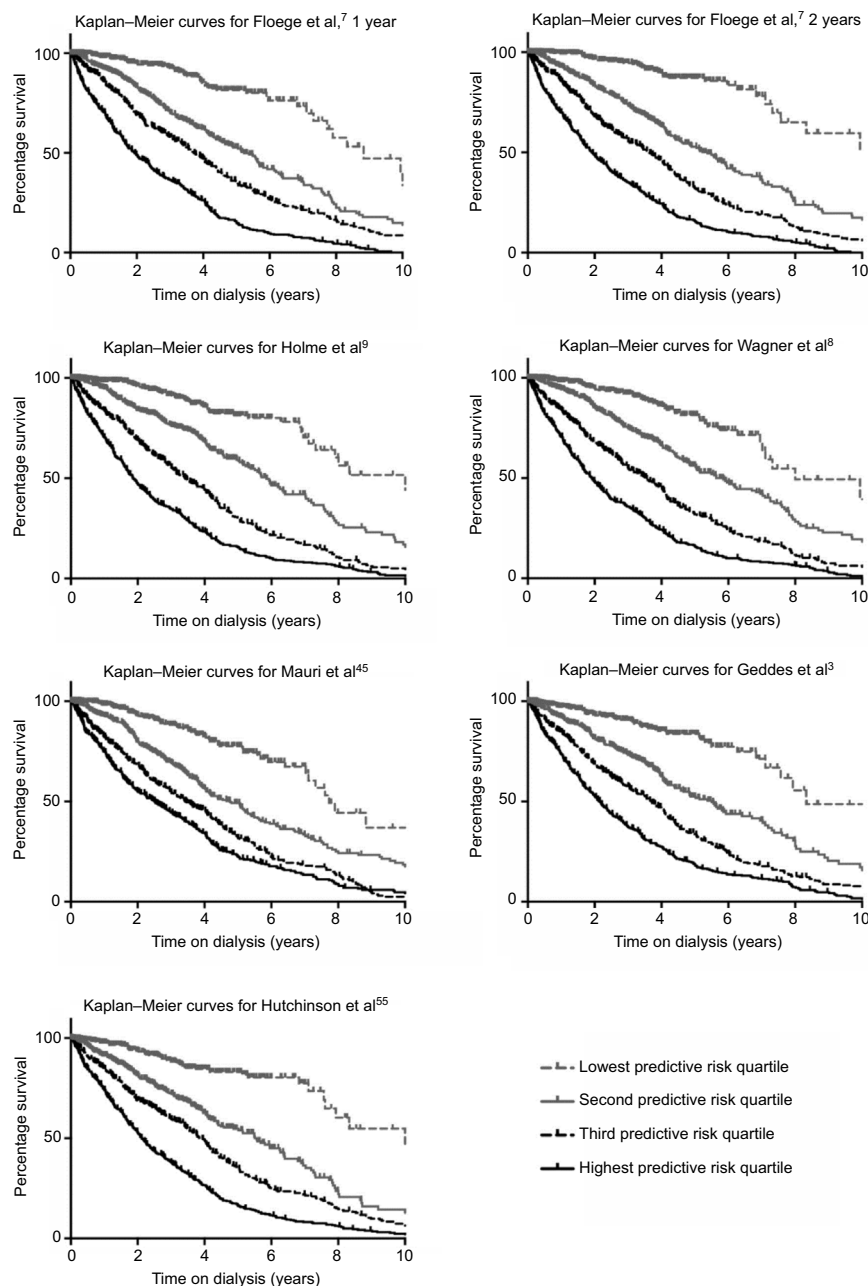


Figure 3 Kaplan-Meier survival curves by quartile of the prognostic index.

per time span of patient inclusion, or per chosen method for patient inclusion.

Furthermore, percentage of missing data of predictors can differ between derivation and validation, especially for CRP, which had a large missing rate in our cohort. However, this was dealt with by using multiple imputation, under the assumption that these data were missing at random. Baseline mortality, which can largely influence calibration results, differs strongly per time period and also per country and setting, such as an inpatient clinic, hospital, or outpatient clinic. The marked overestimation of mortality in the model of

Hutchinson et al can be accounted for by the dramatic change in mortality rates of dialysis patients between the 1970s and late 1990s,⁴⁹ and exemplifies the fact that recalibration over time is important to adjust models to new event rates.

The importance of differences in patient characteristics was also seen in the large difference between results in HD and PD patients. It seems counterintuitive that all models showed better discrimination for PD patients, even though they were not designed for this group of patients. However, this can be accounted for by a larger case-mix heterogeneity within the PD-patient population, specifically in age, allowing

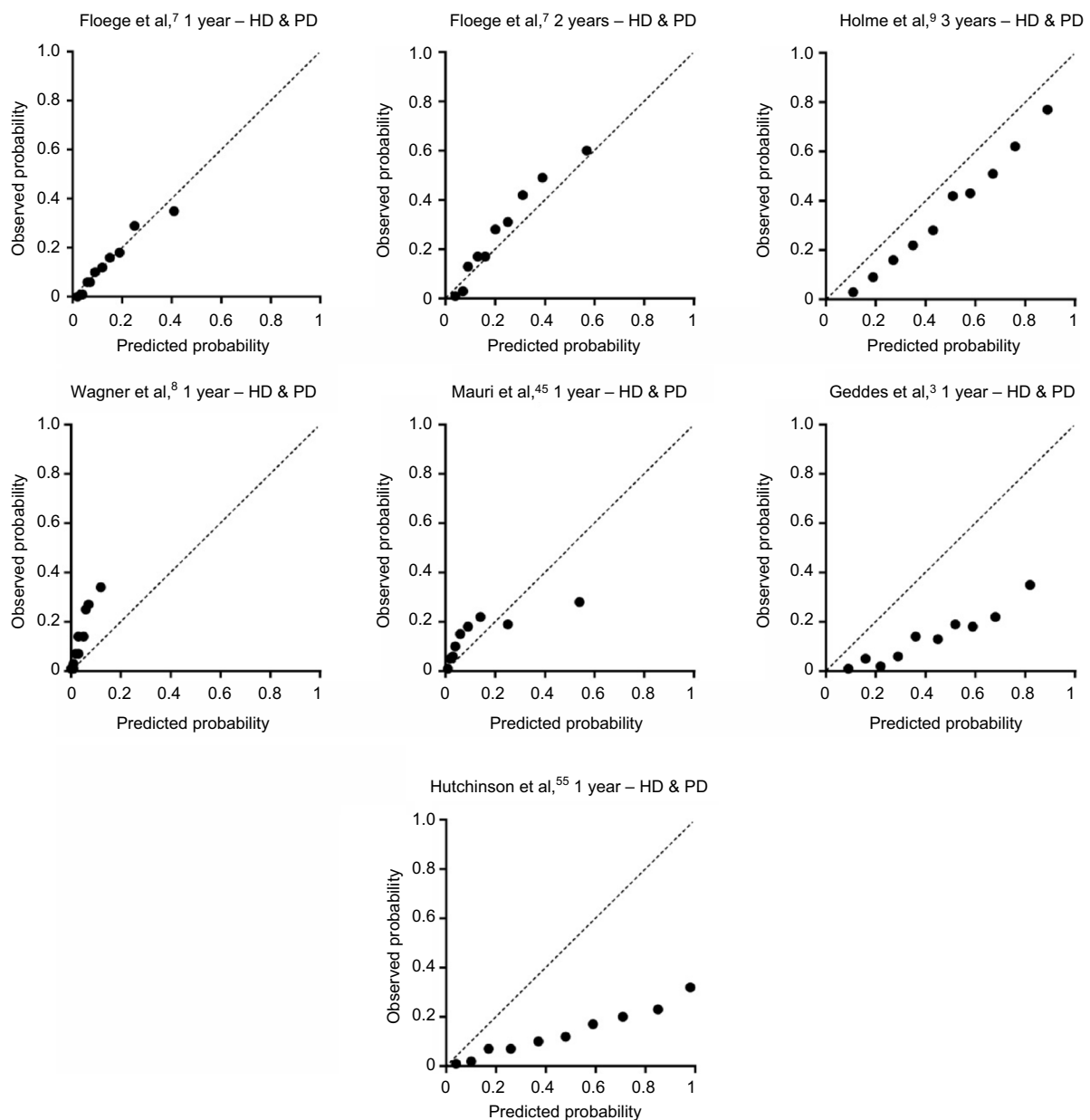


Figure 4 Calibration plots per model.

Abbreviations: HD, hemodialysis; PD, peritoneal dialysis.

for better discrimination, which is a population characteristic and has little to do with model fit.⁵⁰ Due to such differences in case-mix and mortality rates, this validation study can be seen as a true test of the transportability of the models to patients from different source populations, in which population differences on a geographical and temporal level can be a great strength.²³ If a model performs well across a spectrum of different populations, a universally applicable model may be within reach.

Another factor that complicates comparison is the different prediction time frames used by the studies. The larger the time frame, the more difficult it is to predict a patient's death. For a better comparison and validation, we urge studies

to include the baseline hazard function of a Cox prediction model or at least the baseline hazard value for clinically relevant time points, as specified in the TRIPOD guidelines.

To conclude, we warn against drawing general conclusions on which validated model is “the best”, as the different models were tested on transportability to a different degree and reproducibility was not always apparent. Clinical expertise remains leading when interpreting the differences between derivation and validation cohorts and differences between model performance.

This study contains a number of great strengths. To our knowledge, this is the first study to conduct a systematic review of existing models for predicting mortality in dialysis

Table 5 Calibration results of external validation for HD and PD patients

Reference	Time frame	Original population	Calibration: HD and PD	
			Calibration slope (SE)	Calibration-in-the-large*
Floege et al ⁷	1 year	HD	0.71 (0.04)	14.0% vs 13.5%
Floege et al ⁷	2 years	HD	0.83 (0.04)	22.1% vs 26.9%
Holme et al ⁹	3 years	HD	0.96 (0.05)	47.6% vs 37.2%
Wagner et al ⁸	1 year	HD and PD	0.82 (0.04)	4.0% vs 13.5%
	2 years		0.82 (0.04)	8.9% vs 26.9%
	3 years		0.82 (0.04)	14.2% vs 37.2%
Mauri et al ⁴⁵	1 year	HD	0.42 (0.04)	12.0% vs 12.6%
Geddes et al ³	1 year	HD and PD	0.82 (0.05)	41.7% vs 13.5%
	2 years		0.82 (0.05)	59.3% vs 26.9%
	3 years		0.82 (0.05)	69.7% vs 37.2%
	5 years		0.82 (0.05)	80.9% vs 58.2%
Hutchinson et al ⁵⁵	1 year	HD	0.50 (0.03)	45.3% vs 13.5%
	2 years		0.50 (0.03)	51.4% vs 26.9%
	3 years		0.50 (0.03)	60.3% vs 37.2%
	5 years		0.50 (0.03)	70.8% vs 58.2%

Note: *Predicted vs observed risk of death.

Abbreviations: HD, hemodialysis; PD, peritoneal dialysis; SE, standard error.

Table 6 Calibration results of external validation stratified for dialysis modality

Reference	Time frame	Original population	Calibration: HD		Calibration: PD	
			Calibration slope (SE)	Calibration-in-the-large*	Calibration slope (SE)	Calibration-in-the-large*
Floege et al ⁷	1 year	HD	0.67 (0.04)	16.7% vs 16.9%	0.76 (0.08)	9.1% vs 7.1%
Floege et al ⁷	2 years	HD	0.79 (0.05)	26.2% vs 30.9%	0.91 (0.08)	14.4% vs 19.5%
Holme et al ⁹	3 years	HD	0.92 (0.06)	52.8% vs 42.2%	1.04 (0.09)	37.8% vs 27.3%
Wagner et al ⁸	1 year	HD and PD	0.85 (0.05)	5.0% vs 16.9%	0.87 (0.08)	2.1% vs 7.1%
	2 years		0.85 (0.05)	11.1% vs 30.9%	0.87 (0.08)	4.7% vs 19.5%
	3 years		0.85 (0.05)	17.6% vs 42.2%	0.87 (0.08)	7.7% vs 27.3%
Mauri et al ⁴⁵	1 year	HD	0.34 (0.05)	16.9% vs 15.7%	0.78 (0.15)	2.9% vs 6.7%
Geddes et al ³	1 year	HD and PD	0.75 (0.06)	47.6% vs 16.9%	1.02 (0.10)	30.7% vs 7.1%
	2 years		0.75 (0.06)	66.0% vs 30.9%	1.02 (0.10)	46.8% vs 19.5%
	3 years		0.75 (0.06)	76.1% vs 42.2%	1.02 (0.10)	57.7% vs 27.3%
	5 years		0.75 (0.06)	86.1% vs 64.2%	1.02 (0.10)	71.2% vs 44.3%
Hutchinson et al ⁵⁵	1 year	HD	0.46 (0.03)	51.6% vs 16.9%	0.56 (0.05)	33.6% vs 7.1%
	2 years		0.46 (0.03)	58.1% vs 30.9%	0.56 (0.05)	38.8% vs 19.5%
	3 years		0.46 (0.03)	67.3% vs 42.2%	0.56 (0.05)	47.2% vs 27.3%
	5 years		0.46 (0.03)	77.5% vs 64.2%	0.56 (0.05)	58.4% vs 44.3%

Note: *Predicted vs observed risk of death.

Abbreviations: HD, hemodialysis; PD, peritoneal dialysis; SE, standard error.

patients and subsequently validate and compare the existing models. To guarantee study reproducibility and transparency to the greatest extent possible, the PRISMA and TRIPOD guidelines were followed. Furthermore, missing data were accounted for by performing multiple imputations. This allowed us to test each model on the same patient population, and decreased the chance of selection bias.²⁸ Additionally, NECOSAD is a highly representative large cohort of incident dialysis patients, and the large number of events within NECOSAD provide us with unbiased and precise estimations of calibration and discrimination performance measures.^{51,52}

We have chosen to perform very conservative model updating as a sensitivity analysis only, and recommend that

our main (unupdated) results be used when interpreting the performance of the validated models. Since validation means assessing the performance of a predefined model with new data, one could wonder whether model updating leads to a new model, which would then need external validation itself.^{34,53} If the validation results demonstrate that a model is not transportable to a specific (sub)population, this could give incentive for model updating in the context of an impact study or during clinical implementation.²³

It is clear that before clinical implementation of prognostic models is warranted, extensive validation is necessary. The aim of clinical implementation is for such models to function as a useful tool for nephrologists, using evidence-based

medicine that combines individual clinical expertise, patients' choices, and the best external evidence.^{6,54}

For future research, we suggest that more effort is put into validating, building upon, and implementing existing models in different populations. In the emerging field of prediction modeling, it is essential to provide summaries and evaluations of previously conducted studies through external validation, in order to move forward and prevent research waste. This study brings us a step closer to the use of a model with which personalized information on prognosis can be given to dialysis patients, aiding patient-centered decision making.

Acknowledgments

We would like to thank the Dutch Kidney Foundation and the Dutch Kidney Patient Association for their enthusiasm and support. The nursing staff of the 38 different dialysis units, who collected most of the data, are gratefully acknowledged for their assistance. Moreover, we thank the staff of the NECOSAD trial office for assistance in the logistics of this study.

The NECOSAD study group consisted of AJ Apperloo, JA Bijlsma, M Boekhout, WH Boer, PJM van der Boog, HR Büller, M van Buren, FTH de Charro, CJ Doorenbos, MA van den Dorpel, A van Es, WJ Fagel, GW Feith, CWH de Fijter, LAM Frenken, JACA van Geelen, PGG Gerlag, W Grave, JPMC Gorgels, RM Huisman, KJ Jager, K Jie, WAH Koning-Mulder, MI Koolen, TK Kremer Hovinga, ATJ Lavrijssen, AJ Luik, J van der Meulen, KJ Parlevliet, MHM Raasveld, FM van der Sande, MJM Schonck, MMJ Schuurmans, CEH Siegert, CA Stegeman, P Stevens, JGP Thijssen, RM Valentijn, GH Vastenburg, CA Verburgh, HH Vincent, and PF Vos.

Disclosure

CL Ramspek received a Kolff student research grant (number 150KK99) from the Dutch Kidney Foundation (De Nierstichting) for her research activities and internship. The other authors report no conflicts of interest in this work.

References

- de Jager DJ, Grootendorst DC, Jager KJ, et al. Cardiovascular and noncardiovascular mortality among patients starting dialysis. *JAMA*. 2009;302(16):1782–1789.
- Collins AJ, Foley RN, Chavers B, et al. United States Renal Data System 2011 annual data report: atlas of chronic kidney disease and end-stage renal disease in the United States. *Am J Kidney Dis*. 2012;59(1 Suppl 1):A7, e1–e420.
- Geddes CC, van Dijk PC, McArthur S, et al. The ERA-EDTA cohort study: comparison of methods to predict survival on renal replacement therapy. *Nephrol Dial Transplant*. 2006;21(4):945–956.
- Kazmi WH, Gilbertson DT, Obrador GT, et al. Effect of comorbidity on the increased mortality associated with early initiation of dialysis. *Am J Kidney Dis*. 2005;46(5):887–896.
- Plantinga LC, Fink NE, Levin NW, et al. Early, intermediate, and long-term risk factors for mortality in incident dialysis patients: the Choices for Healthy Outcomes in Caring for ESRD (CHOICE) study. *Am J Kidney Dis*. 2007;49(6):831–840.
- Chazot C, Jean G. The dynamics of prognostic indicators: toward earlier identification of dialysis patients with a high risk of dying. *Kidney Int*. 2013;84(1):19–21.
- Floege J, Gillespie IA, Kronenberg F, et al. Development and validation of a predictive mortality risk score from a European hemodialysis cohort. *Kidney Int*. 2015;87(5):996–1008.
- Wagner M, Ansell D, Kent DM, et al. Predicting mortality in incident dialysis patients: an analysis of the United Kingdom Renal Registry. *Am J Kidney Dis*. 2011;57(6):894–902.
- Holme I, Fellström BC, Jardín AG, Schmieder RE, Zannad F, Holdaas H. Prognostic model for total mortality in patients with haemodialysis from the Assessments of Survival and Cardiovascular Events (AURORA) study. *J Intern Med*. 2012;271(5):463–471.
- Quinn RR, Laupacis A, Hux JE, Oliver MJ, Austin PC. Predicting the risk of 1-year mortality in incident dialysis patients: accounting for case-mix severity in studies using administrative data. *Medical care*. 2011;49(3):257–266.
- Zhao C, Luo Q, Xia X, et al. Risk score to predict mortality in continuous ambulatory peritoneal dialysis patients. *Eur J Clin Invest*. 2014;44(11):1095–1103.
- Otero-Lopez MS, Martinez-Ocana JC, Betancourt-Castellanos L, Rodriguez-Salazar E, Garcia-Garcia M. Two prognostic scores for early mortality and their clinical applicability in elderly patients on haemodialysis: poor predictive success in individual patients. *Nefrologia*. 2012;32(2):213–220.
- Pollak VE, Pesce A, Kant KS. Continuous quality improvement in chronic disease: a computerized medical record enables description of a severity index to evaluate outcomes in end-stage renal disease. *Am J Kidney Dis*. 1992;19(6):514–522.
- Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
- Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol*. 2008;61(11):1085–1094.
- Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol*. 2008;61(1):76–86.
- Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models – II: external validation, model updating, and impact assessment. *Heart*. 2012;98(9):691–698.
- Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2015;68(1):25–34.
- Bleeker SE, Moll HA, Steyerberg EW, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol*. 2003;56(9):826–832.
- Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55–63.
- Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201–209.
- Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14:40.
- Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279–289.

24. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Heidelberg: Springer; 2015.
25. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ*. 2009;338:b606.
26. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700.
27. Korevaar JC, Jansen MA, Dekker FW, et al. When to initiate dialysis: effect of proposed US guidelines on survival. *Lancet*. 2001;358(9287):1046–1050.
28. de Goeij MC, van Diepen M, Jager KJ, Tripepi G, Zoccali C, Dekker FW. Multiple imputation: dealing with missing data. *Nephrol Dial Transplant*. 2013;28(10):2415–2420.
29. Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Stat Methods Med Res*. 2007;16(3):199–218.
30. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol*. 2009;9:57.
31. Moons KG, Donders RA, Stijnen T, Harrell FE Jr. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59(10):1092–1101.
32. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ*. 2009;338:b604.
33. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models – I: development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683–690.
34. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol*. 2013;13:33.
35. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–138.
36. Janssen KJ, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KG. A simple method to adjust clinical prediction models to local circumstances. *Can J Anaesth*. 2009;56(3):194–201.
37. Chua HR, Lau T, Luo N, et al. Predicting first-year mortality in incident dialysis patients with end-stage renal disease: the UREA5 study. *Blood Purif*. 2014;37(2):85–92.
38. Cohen LM, Ruthazer R, Moss AH, Germain MJ. Predicting six-month mortality for patients who are on maintenance hemodialysis. *Clin J Am Soc Nephrol*. 2010;5(1):72–79.
39. Shiotsu Y, Mori Y, Nishimura M, et al. Prognostic utility of plasma S100A12 levels to establish a novel scoring system for predicting mortality in maintenance hemodialysis patients: a two-year prospective observational study in Japan. *BMC Nephrol*. 2013;14:16.
40. Cao XY, Zhou JH, Cai GY, et al. Predicting one-year mortality in peritoneal dialysis patients: an analysis of the China peritoneal dialysis registry. *Int J Med Sci*. 2015;12(4):354–361.
41. Foley RN, Parfrey PS, Hefferton D, Singh I, Simms A, Barrett BJ. Advance prediction of early death in patients starting maintenance dialysis. *Am J Kidney Dis*. 1994;23(6):836–845.
42. Hemke AC, Heemskerk MB, van Diepen M, Weimar W, Dekker FW, Hoitsma AJ. Survival prognosis after the start of a renal replacement therapy in the Netherlands: a retrospective cohort study. *BMC Nephrol*. 2013;14:258.
43. Hemke AC, Heemskerk MB, van Diepen M, Dekker FW, Hoitsma AJ. Improved mortality prediction in dialysis patients using specific clinical and laboratory data. *Am J Nephrol*. 2015;42(2):158–167.
44. Doi T, Yamamoto S, Morinaga T, Sada KE, Kurita N, Onishi Y. Risk score to predict 1-year mortality after haemodialysis initiation in patients with stage 5 chronic kidney disease under predialysis nephrology care. *PLoS One*. 2015;10(6):e0129180.
45. Mauri JM, Cleries M, Vela E. Design and validation of a model to predict early mortality in haemodialysis patients. *Nephrol Dial Transplant*. 2008;23(5):1690–1696.
46. Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9(5):1–12.
47. Collins GS, Moons KG. Comparing risk prediction models. *BMJ*. 2012;344:e3186.
48. Abbasi A, Peelen LM, Corpeleijn E, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ*. 2012;345:e5900.
49. Steenkamp R, Rao A, Fraser S. UK Renal Registry 18th Annual Report (December 2015) Chapter 5: survival and causes of death in UK adult patients on renal replacement therapy in 2014 – national and centre-specific analyses. *Nephron*. 2016;132 Suppl 1:111–144.
50. 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*. 2010;172(8):971–980.
51. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med*. 2016;35(2):214–226.
52. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005;58(5):475–483.
53. van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19(24):3401–3415.
54. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't – 1996. *Clin Orthop Relat Res*. 2007;455:3–5.
55. Hutchinson TA, Thomas DC, MacGibbon B. Predicting survival in adults with end-stage renal disease: an age equivalence index. *Ann Intern Med*. 1982;96(4):417–423.

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventive initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

systematic reviews, risk and safety of medical interventions, epidemiology and biostatistical methods, and evaluation of guidelines, translational medicine, health policies and economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress