# Predicting outcomes in patients with kidney disease: methodology and clinical applications

Ramspek, C.L.


# Predicting Outcomes in Patients with Advanced Chronic Kidney Disease

# 6

# Towards the best kidney failure prediction tool: a systematic review and selection aid

Chava L. Ramspek, Ype de Jong, Friedo W. Dekker, Merel van Diepen

*Nephrol Dial Transplant. 2020 Sep 1;35(9):1527-1538.*

## Abstract

**Background.** Prediction tools that identify CKD (Chronic Kidney Disease) patients at a high risk of developing kidney failure have potential for large clinical value, but limited uptake. The aim of the current study is to systematically review all available models predicting kidney failure in CKD patients, organize empirical evidence on their validity, and ultimately provide guidance in the interpretation and uptake of these tools.

**Methods.** PubMed and Embase were searched for relevant articles. Titles, abstracts and full-text articles were sequentially screened for inclusion by two independent researchers. Data on study design, model development and performance were extracted. The risk of bias and clinical usefulness were assessed and combined in order to provide recommendations on which models to use.

**Results.** Out of 2183 screened studies, a total of 42 studies were included in the current review. Most studies showed high discriminatory capacity and the included predictors had large overlap. Overall, the risk of bias was high. Just under half the studies (48%) presented enough detail for the use of their prediction tool in practice and few models were externally validated.

**Conclusions.** The current systematic review may be used as a tool to select the most appropriate and robust prognostic model for various settings. Though some models showed large potential, many lacked clinical relevance due to being developed in a prevalent patient population with a wide range of disease severity. Future research efforts should focus on external validation and impact assessment in clinically relevant patient populations.

## Background

Chronic kidney disease (CKD) may lead to kidney failure, though rates of progression vary substantially between individuals.[1] Prediction tools that can identify patients at high risk of developing kidney failure could have a large clinical value. They could be used to inform individualized decision making, be employed in determining the appropriate time for referral to nephrologists, and be used in the planning and preparation of renal replacement therapy (RRT). Prediction tools might also offer opportunities for risk stratification in research and improvement of health policies.[2]

Multiple prediction models have been developed to identify individuals at high risk of kidney failure, and have been previously described in two systematic reviews.[3 4] Many of these models showed good predictive abilities in development. However, despite nephrologists and patients acknowledging a lack of prognosis discussions in practice, clinical uptake of these tools is still limited.[5] Policy makers also seem hesitant in endorsing prediction tools. The most recent KDIGO guideline recommends the use of prediction models for timely referral for planning RRT.[6] The guideline, however, fails to provide guidance on which risk prediction tool should be used to do so.

The lack of uptake by clinicians and policymakers has been partly attributed to substandard methodology, lack of external validation and shortage of easy calculation options.[7] The last two published reviews in 2012 and 2013 both included 8 studies on prediction of kidney failure in CKD patients.[3 4] Since then the number of available models has increased exceedingly. A new systematic review of the available models is the first step towards the use and recommendation of robust prognostic tools. The aim of the current study is therefore to systematically review all available models predicting kidney failure in CKD patients, organize empirical evidence on their validity, and ultimately provide guidance in the selection of the best prediction tool for various settings.



## Methods

### Data sources and searches

The current review was framed by the search for prognostic prediction models for CKD patients, predicting the future event of kidney failure. To ensure transparent reporting and accurate study appraisal, the PRISMA, TRIPOD and CHARMS guidelines were followed where applicable.[8-10] We searched PubMed and Embase databases on the 31st of December, 2017 for English language studies regarding risk prediction in CKD patients. The search strategies were designed to include relevant development, validation and implementation studies, and are provided in the supplement.

## Study selection

Titles, abstracts and full-text articles were sequentially screened for inclusion by two independent researchers (CLR and YdJ). Discrepancies on inclusion of full-text articles were solved by consulting a third co-author (MvD). Articles were included if they met the following pre-defined selection criteria: 1) The study must develop, validate, update or implement a multivariate prognostic prediction model, with a prediction research question as aim, as opposed to an etiological or methodological goal. 2) The study must present at least one measure to assess model performance. 3) The study population must consist of adult CKD patients. 4) The study outcome must include kidney failure or end-stage renal disease. The references of included studies and related reviews were manually screened in order to identify additional relevant studies.

## Data extraction and quality assessment

Following selection, two reviewers (CLR and YdJ) independently conducted the data extraction and quality assessment. Discrepancies were discussed with input from an additional co-author (MvD) where necessary. Conform CHARMS recommendations, information on the source of data, population, outcome, sample size, missing data, model development, and model performance were extracted and summarized. Additionally, data on external validations of models were extracted. Furthermore, the risk of bias and clinical usefulness were judged by both reviewers independently. In order to facilitate further comparison, studies were grouped by study population which ranged from very broad (general CKD) to specific CKD subgroups such as IgA-nephropathy or diabetic nephropathy. Quality and risk of bias were assessed in both development and validation studies by making use of a novel tool, the Prediction study Risk Of Bias Assessment Tool (PROBAST). Though this tool has yet to be published in its complete form, there is no other formal risk of bias assessment available that is applicable to prediction studies. The PROBAST is specifically designed for systematic reviews of prediction studies and uses a domain-based approach with 23 signalling questions that categorize the risk of bias into high, low or unclear for 5 separate domains: participant selection, predictors, outcome, sample size and missing data, and analysis. It also assesses usability of a model. It has been used in multiple reviews in the past year and was presented in part at the 2016 Cochrane Colloquia. [11] The final test version of PROBAST was obtained through personal email contact with the first author dr. R. Wolff.

## Data synthesis

Given the multitude of different models and heterogeneity in study characteristics, we opted for a narrative synthesis of results supported by extensive tables and figures with study characteristics listed per article. Model performance was evaluated by examining the discrimination and calibration of included prediction tools. Discrimination is most

often described by the C-statistic and indicates how well the model discriminates between patients with and without the event of interest. It lies between 0.5 and 1, where 0.5 is similar to tossing a coin and 1 indicates perfect discrimination. [12] Important to take into account, is that the C-statistic of the same model can vary highly, dependent on the population on which the model is tested. When a population is heterogeneous in the predictors that make up the prediction tool, the C-statistic may increase substantially. [13] Calibration on the other hand, describes the agreement between the absolute number of predicted events and observed events population wide. It is best represented in a plot, wherein the predicted probability of kidney failure is plotted against the observed rate of kidney failure. [12]To evaluate the sample size and risk of overfitting in development studies, the events per candidate predictor (EPV) were extracted. A minimum of 10 events per candidate predictor has been suggested as rule of thumb for an acceptable sample size in model development studies.[14] For external validation studies it has been recommended to include a minimum of 100 events in total to obtain a precise estimate of performance. [15]

## Results

### Study selection

The study selection process is described in a flowchart (Figure 1). Overall, 2183 titles were identified, of which 431 abstracts were assessed, and 90 full-text publications were evaluated in-depth. From these articles, a final 42 studies met all inclusion criteria and were included in the current review. Most full-text exclusions were due to the predicted outcome not including kidney failure or the lack of a multivariate model. Though prediction research has seen a great surge in nephrology the last few years, the first included predictive model was already published in 1986 for IgA-nephropathy patients. Since the beginning of the 2000's a substantial increase of published models is apparent, as can be seen in Figure 2. Though the number of developed models has increased almost every year, the number of validation studies has remained small. Out of the 42 included studies, 7 studies exclusively externally validated already existing models.[16-22] Besides development, 10 studies also externally validated their own or previously published models. Disconcertingly, no study assessing the impact of using such a prediction tool was found, which is ultimately the only way of assessing whether the model can improve patient care.

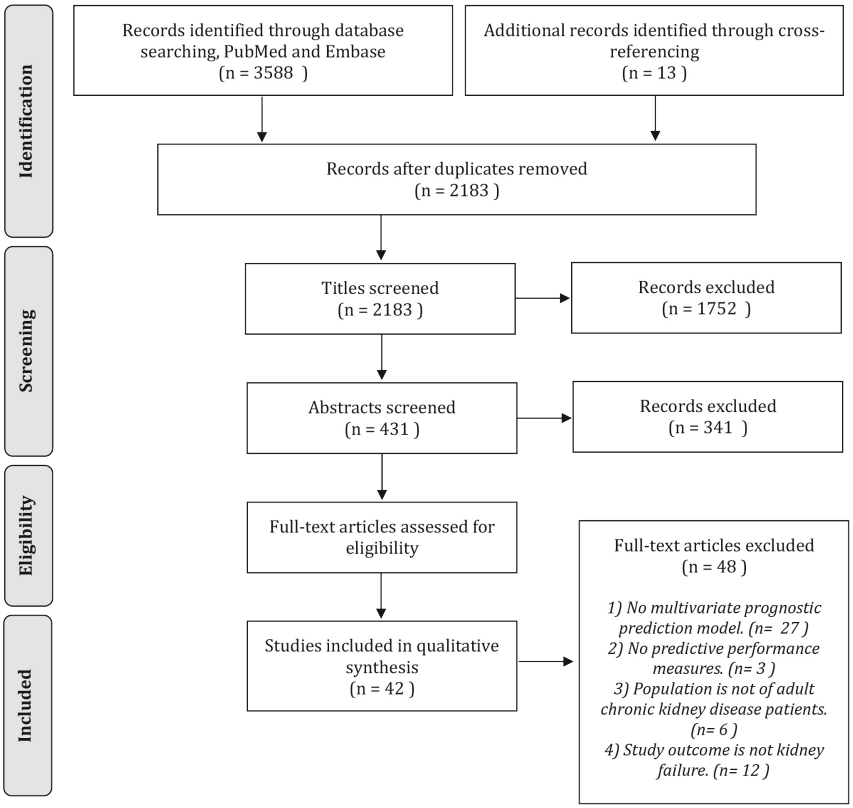



**Figure 1: PRISMA Flow Diagram of study inclusion.**

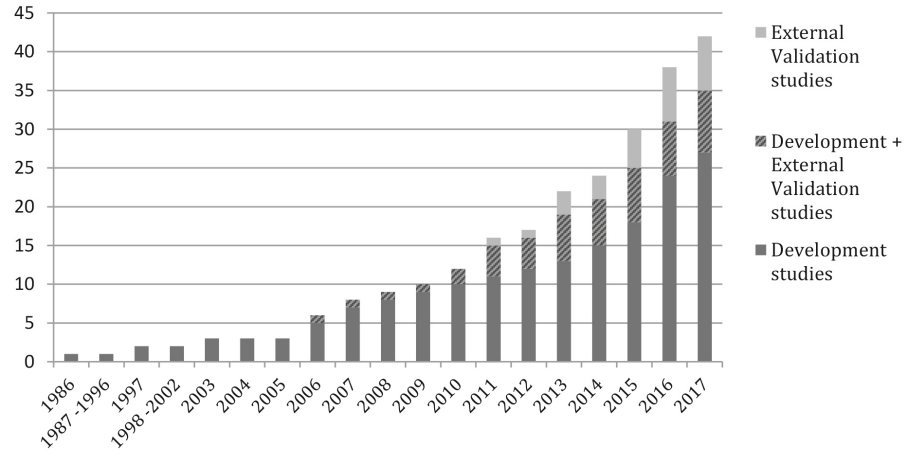
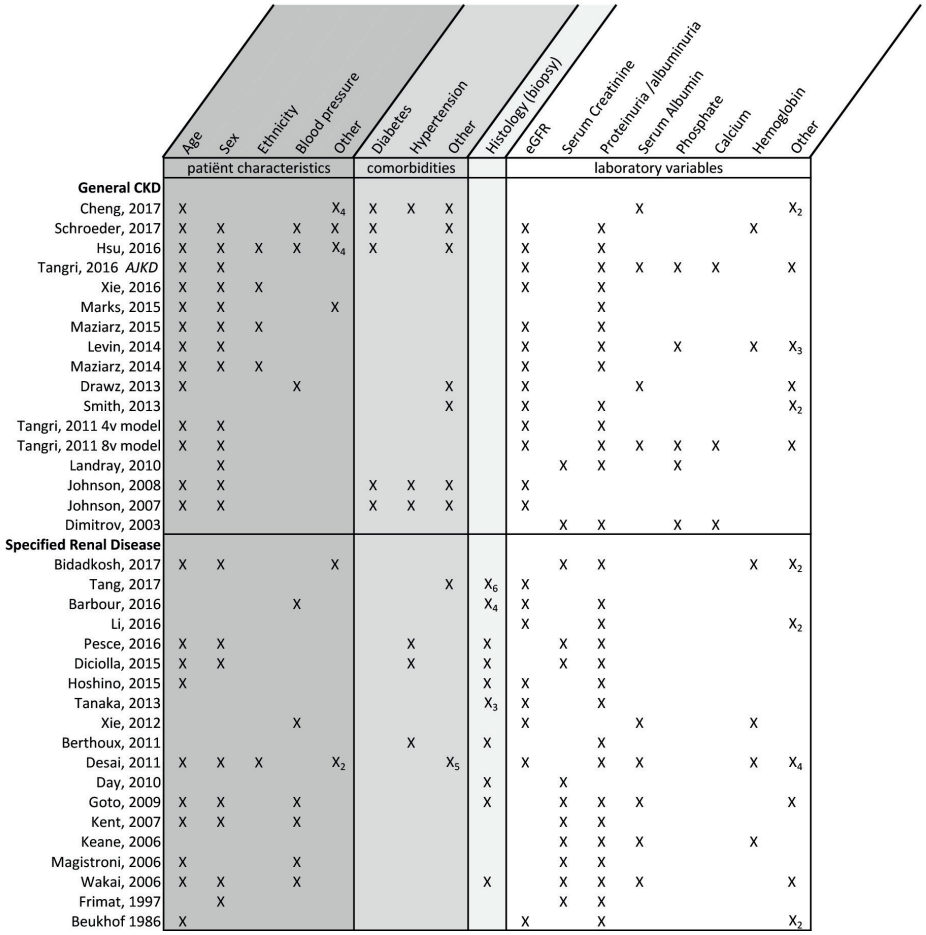



**Figure 2: Cumulative number of published development and validation studies for models that predict kidney failure in CKD patients (N=42).**

## Characteristics of development studies

A total of 35 studies were published on the development of novel tools to predict kidney failure in CKD patients. Generally, a distinction can be made between models developed for a general CKD patient population (n=16), and models developed for a population with a specified primary renal disease (n=19), mainly IgA-nephropathy or diabetic nephropathy. The characteristics of all included development studies are described in Table 1. Since each study developed between 1 and 12 prediction models, the results presented in Table 1 concern the final model(s) as selected by the authors, or the model with the best performance if no final model was suggested. The population size differed greatly between studies and ranged from 75 to 28779 patients. A small sample size was a problem in 17/35 studies, as they had less than 10 events per candidate predictor, thus running the substantial risk of overfitting their model.[14] To assess to what extent these models are overfit, external validation is of key importance. Before validity of these models has been tested they should not be used in practice. For specific renal diseases the baseline was almost always the first biopsy (and disease confirmation), providing a clear moment in time for when to use the prognostic model or score. Models developed in general CKD, however, rarely defined the moment in time when their prediction tool should be used, as most of these studies enrolled prevalent CKD patients with a large range of disease severity. Only two models developed their model on incident patients, who were included at the first referral to a nephrologist.[23 24]There was some variation in outcome definitions, but for most studies renal failure was defined as the need for renal replacement therapy (dialysis start or kidney transplantation). Five studies used eGFR or creatinine as a proxy for kidney failure. Two development studies used RRT start or death as a composite outcome measure. A total of 4 studies did not report their definition of ESRD. The time-frame over which the models predict kidney failure ranged from 6 months to 20 years and 9 studies failed to define a prediction time-frame, presumably using the maximum study follow-up. The specific predictors included per development study are presented in Figure 3. There is a large amount of overlap in final predictors, with almost all studies including age, sex, eGFR (or serum creatinine), proteinuria and histological features for IgA-nephropathy tools.



| | Age | Sex | Ethnicity | Blood pressure | Other | Diabetes | Hypertension | Other | Histology (biopsy) | eGFR | Serum Creatinine | Proteinuria /albuminuria | Serum Albumin | Phosphate | Calcium | Hemoglobin | Other |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | patiënt characteristics | | | | | comorbidities | | | | laboratory variables | | | | | | | |
| **General CKD** | | | | | | | | | | | | | | | | | |
| Cheng, 2017 | X | | | | $X_4$ | X | X | X | | | | X | | | | | $X_2$ |
| Schroeder, 2017 | X | X | | X | X | X | | X | | X | X | | | | | X | |
| Hsu, 2016 | X | X | X | X | $X_4$ | X | | X | | X | X | | | | | | |
| Tangri, 2016 *AJKD* | X | X | | | | | | | | X | | X | X | X | X | | X |
| Xie, 2016 | X | X | X | | | | | | | X | X | | | | | | |
| Marks, 2015 | X | X | | | X | | | | | | X | | | | | | |
| Maziarz, 2015 | X | X | X | | | | | | | X | X | | | | | | |
| Levin, 2014 | X | X | | | | | | | | X | X | | X | | | X | $X_3$ |
| Maziarz, 2014 | X | X | X | | | | | | | X | X | | | | | | |
| Drawz, 2013 | X | | | X | | | | X | | X | | X | | | | | X |
| Smith, 2013 | | | | | | | | X | | X | X | | | | | | $X_2$ |
| Tangri, 2011 4v model | X | X | | | | | | | | X | X | | | | | | |
| Tangri, 2011 8v model | X | X | | | | | | | | X | X | X | X | X | X | | X |
| Landray, 2010 | | X | | | | | | | | | X | X | | X | | | |
| Johnson, 2008 | X | X | | | | X | X | X | | X | | | | | | | |
| Johnson, 2007 | X | X | | | | X | X | X | | X | | | | | | | |
| Dimitrov, 2003 | | | | | | | | | | | X | X | | X | X | | |
| **Specified Renal Disease** | | | | | | | | | | | | | | | | | |
| Bidadkosh, 2017 | X | X | | | X | | | | | | X | X | | | | X | $X_2$ |
| Tang, 2017 | | | | | | | | X | $X_6$ | X | | | | | | | |
| Barbour, 2016 | | | | X | | | | | $X_4$ | X | | X | | | | | |
| Li, 2016 | | | | | | | | | | X | | X | | | | | $X_2$ |
| Pesce, 2016 | X | X | | | | | X | | X | | X | X | | | | | |
| Diciolla, 2015 | X | X | | | | | X | | X | | X | X | | | | | |
| Hoshino, 2015 | X | | | | | | | | X | X | | X | | | | | |
| Tanaka, 2013 | | | | | | | | | $X_3$ | X | | X | | | | | |
| Xie, 2012 | | | | X | | | | | | X | | | X | | | X | |
| Berthoux, 2011 | | | | | | | X | | X | | | X | | | | | |
| Desai, 2011 | X | X | X | | $X_2$ | | | $X_5$ | | X | | X | X | | | X | $X_4$ |
| Day, 2010 | | | | | | | | | X | | X | | | | | | |
| Goto, 2009 | X | X | | X | | | | | X | | X | X | X | | | | X |
| Kent, 2007 | X | X | | X | | | | | | | X | X | | | | | |
| Keane, 2006 | | | | | | | | | | | X | X | X | | | X | |
| Magistroni, 2006 | X | | | X | | | | | | | X | X | | | | | |
| Wakai, 2006 | X | X | | X | | | | | X | | X | X | X | | | | X |
| Frimat, 1997 | | X | | | | | | | | | X | X | | | | | |
| Beukhof 1986 | X | | | | | | | | | X | | X | | | | | $X_2$ |

**Figure 3: predictors included in development studies (N=35).** The inclusion of a predictor is shown as "X". The subscript under X (e.g. "$X_2$") indicates the number of predictors included from that category.

Concerning the reporting of performance measures, discrimination measures were reported far more often than calibration measures. Discrimination in the form of a C-statistic was reported in 28/35 studies. The C-statistic ranged from 0.72 to 0.96 and was generally high, indicating good to excellent discrimination in most studies. Calibration was presented far less frequently, with only 11 studies presenting a calibration plot, bar chart or test.

In order to calculate an individual's risk, the model constant and HRs/regression coefficients per predictor are needed. Many studies only presented HRs per predictor without the constant (intercept or baseline hazard value), and some gave no data on the model equation at all. The full formula for the developed model was presented in only 6/35 studies. Just 3 studies provided a web-calculator for easy use of which two web-calculators

are no longer in working order. A total of 13 studies provided a simplified scoring system. In total 25 final models were validated in some form, either internally and/or externally. Cross-validation, bootstrapping and random split sample were the most used forms of internal validation.

## Characteristics of external validation studies

A total of 17 studies externally validated one or more of the developed prediction tools, the characteristics of these models and validations can be found in Table 2. Most validation studies were performed by the same group of researchers who developed the models, and often presented in the same publication as the development. Compared to the development performance, the C-statistic was lower in 68% of the validations. Two studies updated the validated model by recalibrating the baseline hazard and two studies added predictors to the existing model. In total 5 risk scores predicting prognosis in IgA-nephropathy patients and 7 prognostic tools for general CKD patients were externally validated. Only the Absolute Renal Risk (ARR) score, Goto score and Kidney Failure Risk Equation (KFRE) (3, 4 & 8-variable) were validated multiple times. The largest validation study of the KFRE was performed by Tangri et al. [18], and summarized the validation of the KFRE in more than 30 countries including over half a million patients.

## Risk of bias

Risk of bias was assessed in all 42 included studies, using signalling questions from the PROBAST specified to detecting methodological flaws in both development and validation prediction studies. Overall, the risk of bias was high, as can be seen in Figure 4a and b. Forty-one out of 42 studies received a high bias risk in at least 1 of the 5 domains, the only study with an

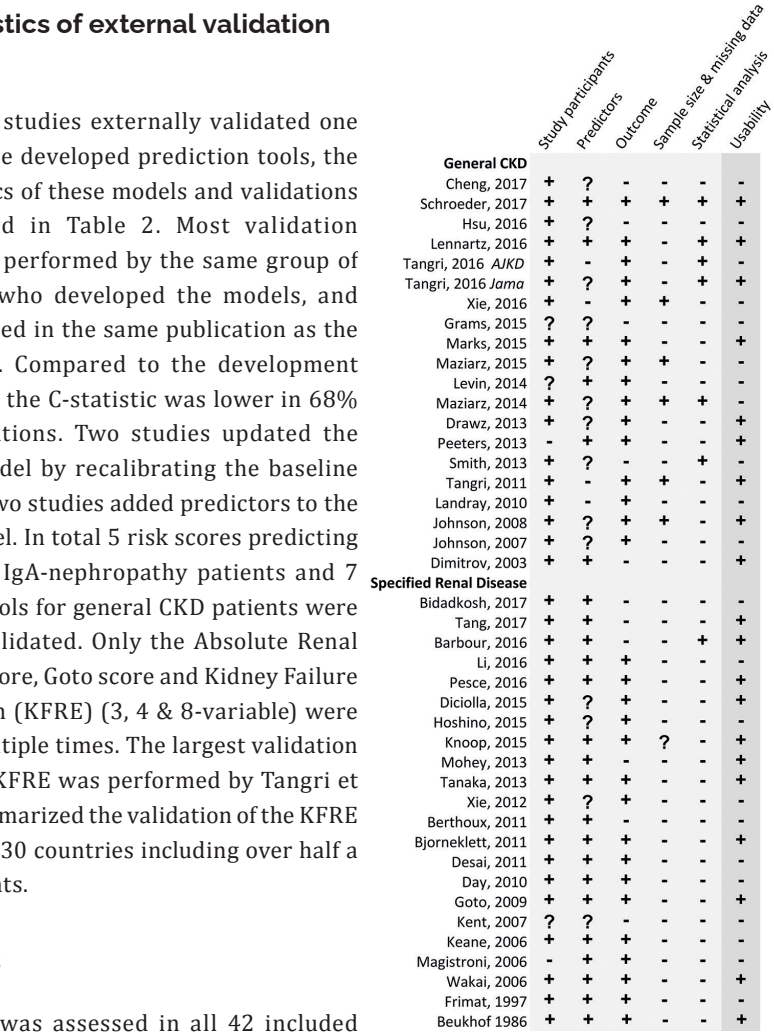



**Figure 4a: Risk of bias and usability of prediction models (N=42).** Assessed using the PROBAST. The five risk of bias domains were evaluated as low risk (+), unclear risk (?), or high risk (-). Usability was evaluated as yes (+) or no (-).

overall low risk of bias was by Schroeder et al.[25] The majority of studies had a high risk of bias in the domain sample size and missing data. This was often due to the use of complete-case analysis, which is generally an inappropriate method of handling missing data. A small sample size was a frequent problem limiting model usage, as a small sample often results in an over-fit model and thereby biased results. In the domain statistical analysis 83% of studies had a high risk of bias. The largest reason was incomplete reporting of performance measures as few studies reported sufficient calibration results. Also, many studies did not correct their model for overfitting through internal validation. The usability of the model was assessed in a separate domain. If the full model formula, a calculator or a risk score with absolute risk table was available the tool was considered usable. Less than half the studies (48%) presented enough detail for the use of their prediction tool in practice. The usable models that specified a prediction time-frame are presented in Figure 5, categorized by type of patient population and outcome. This figure may be employed as selection guide when wanting to calculate an individuals' prognosis, taking into account that many of the models have significant shortcomings and may not be ready for use in clinic.

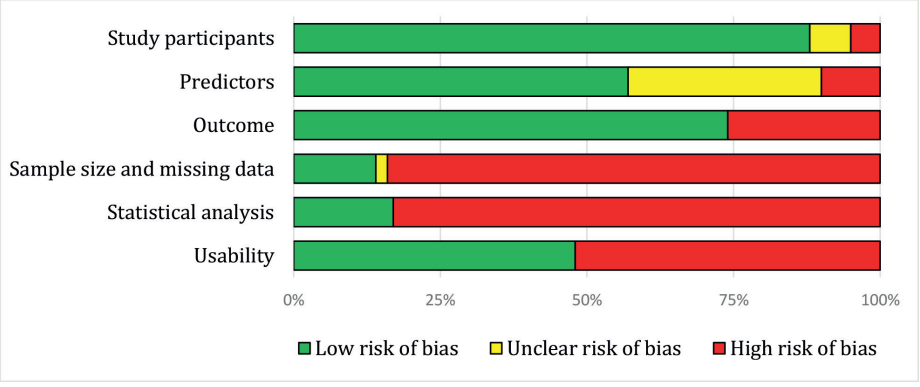



**Figure 4b: PROBAST risk of bias summary for all studies (N=42).**

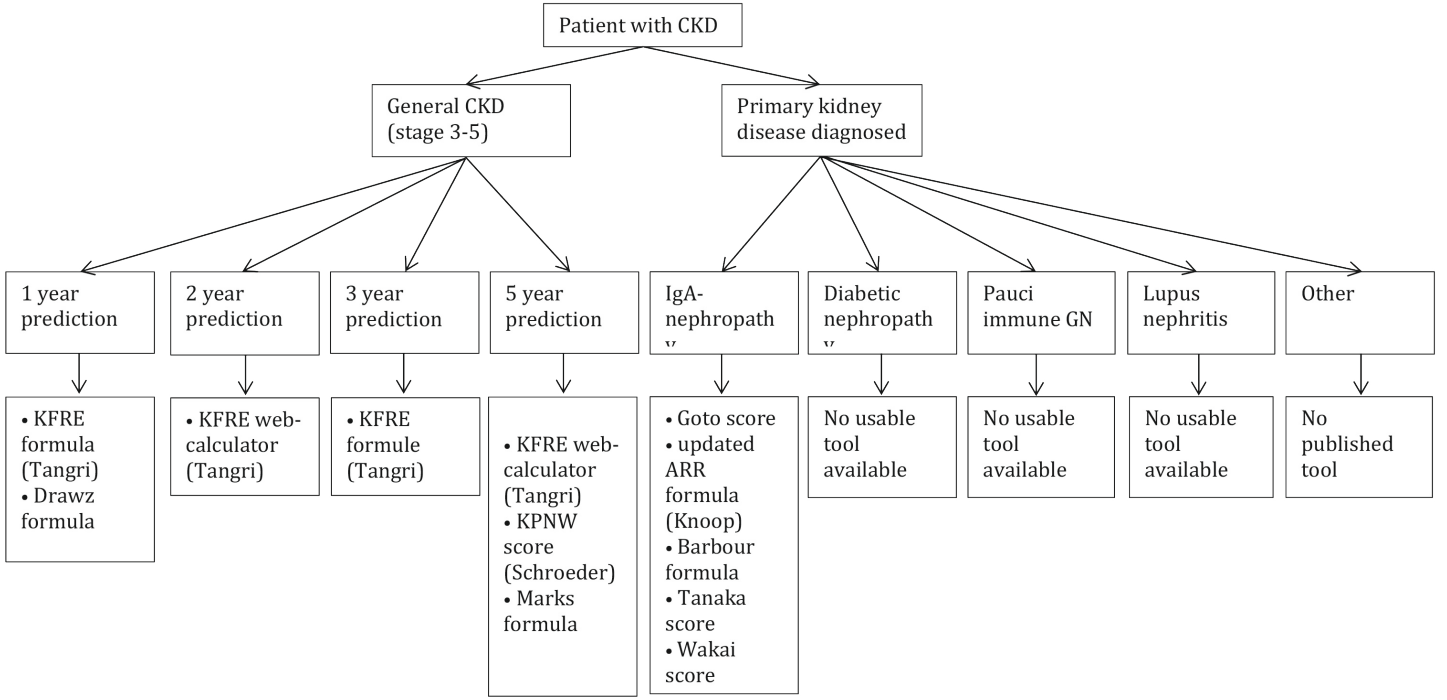



**Figure 5: Model selection guide for CKD patients.** In this graph, only models that allow calculation of an individual's prognosis and are therefore labelled as usable are included. This entails that these models provide either a full formula, score with absolute risk table or (currently working) web-calculator for a specified prediction time-frame. For categories containing multiple models, the risk of bias combined with evidence of external validity was weighed in determining the model order, starting with the most valid and least biased models. Nevertheless, many of the models listed below have significant shortcomings and should be used with caution and understanding.

**Table 1: baseline characteristics of model development studies (N=35).**

| Study | Country | Design | Population | GFR mean | N total, N events | Outcome | Time-frame (years) | EPV | Model type | N pre-dictors | Internal validation | C- sta-tistic | Cali-bration | Pre-sented model |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *General CKD* | | | | | | | | | | | | | |
| Cheng[31] 2017 | Taiwan | Single-centre cohort | General CKD stage 4 | - | 463, 132 | GFR < 15 | 0.5 | 3 | CART | 11 | cross-validation | D: 0.72 | - | Decision rules |
| Schroeder[25] 2017 | USA | Multi-centre cohort | ‡ General CKD stage 3-4 | 47 | 22460, 737 | RRT | 5 | 74 | Cox | 8 | bootstrap *(+external)* | IV: 0.96 | D: plot | Formula[‡‡] & score |
| Hsu[32] 2016 | USA | Cohort | General CKD GFR 20-70 | 44 | 2466, 581 | RRT,50% GFR↓ | - | 36 | Cox | 12 | - | D: 0.89 | - | HRs |
| Tangri[23] 2016 *AJKD* | Canada | Single-centre cohort | General CKD stage 3-5 | 36 | 3004, 344 | RRT | Dynamic | 43 | Cox | 8 | bootstrap, cross-validation | IV: 0.91 | D: plot, test | Formula |
| Xie[33] 2016 | USA | Multi-centre cohort | * General CKD stage 3-5 | 49 | 28779, 1730 | RRT | 1, 3, **5**, 7 | 115 | Cox | 5 | cross-validation | IV: 0.92 | - | HRs |
| Marks[28] 2015 | Schot-land | Multi-centre cohort | General CKD stage 3-5 | 33 | 3396, 142 | RRT | 5 | 24 | Logis-tic | 5 | - *(external)* | D: 0.94 | D: test | Formula |
| Maziarz[34] 2015 | USA | Multi-centre cohort | * General CKD stage 3-5 | - | 28779, 1730 | RRT | 1, 3, **5** | 115 | Cox | 5 | cross-validation | IV: 0.92 | - | HRs |
| Levin[35] 2014 | Canada | Multi-centre cohort | General CKD stage 3-4 | 28 | 2402, 142 | RRT | 1 | 9 | Cox | 7 | bootstrap | D: 0.87 | D: test | HRs |
| Maziarz[36] 2014 | USA | Multi-centre cohort | * General CKD stage 3-5 | - | 16656, 959 | RRT | 1, 3, **5** | 63 | Cox | 5 | cross-validation | IV: 0.90 | - | - |
| Drawz[27] 2013 | USA | Single-centre cohort | † General CKD stage 4-5 | 25 | 1866, 77 | RRT | 1 | 4 | Cox | 6 | bootstrap *(+external)* | IV: 0.86 | - | Formula |

**Table 1:** Continued.

| Study | Country | Design | Population | GFR mean | N total, N events | Outcome | Time-frame (years) | EPV | Model type | N pre-dictors | Internal validation | C- sta-tistic | Cali-bration | Pre-sented model |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Smith[37] 2013 | England | Multi-centre cohort | General CKD stage 3-4 | 32 | 158, 40 | Death, RRT | 2 | 4 | Cox | 10 | - | D: 0.81 | - | HRs |
| Tangri[24] 2011 | Canada | Single-centre cohort | General CKD stage 3-5 | 36 | 3449, 386 | RRT | 1,3,**5** | 16 | Cox | 4, 8 | - *(external)* | 4v: 0.91 8v: 0.92 | - | Formula &web-calculator |
| Landray[38] 2010 | UK | Single-centre cohort | General CKD stage 3-5 | 22 | 382, 190 | RRT | - | 4 | Cox | 4 | - *(external)* | D: 0.87 | D: plot | HRs |
| Johnson[39] 2008 | USA | Multi-centre cohort | ǂ General CKD stage 3-4 | - | 9782, 323 | RRT | 5 | 54 | Cox | 6 | bootstrap | IV: 0.89 | D: plot | Score |
| Johnson[40] 2007 | USA | Multi-centre cohort | ǂ General CKD stage 3-5 | - | 6541, 369 | RRT | 5 | 41 | Cox | 6 | - | D: 0.91 | - | HRs |
| Dimitrov [41] 2003 | Italy | RCT | General CKD GFR 20-70 | 43 | 344, 80 | ESRD | - | 7 | ANN | 4 | - | D: 0.80 | - | decision-tree |
| | *Specified Renal Disease* | | | | | | | | | | | | | |
| Bidadkosh [42], 2017 | Multi-national | RCT | Diabetic nephropathy | 33 | 861, 60 | ESRD, 40% GFR↓ | - | 6 | Cox | 8 | - | D: 0.79 | - | - |
| Tang[43] 2017 | China | Single-centre cohort | Lupus Nephritis | 78 | 599, 145 | RRT, 50% GFR↓, GFR<15 | - | 4 | Cox | 8 | split sample | - | - | HRs & score |
| Barbour[44] 2016 | Multi-national | Multi-centre cohort | IgA-nephropathy | 68 | 901, 162 | GFR<15, 50% GFR↓ | 5 | 21 | Cox | 8 | bootstrap | D: 0.80 | D: plot | Formula |

**Table 1:** Continued.

| Study | Country | Design | Population | GFR mean | N total, N events | Outcome | Time-frame (years) | EPV | Model type | N pre-dictors | Internal validation | C- sta-tistic | Cali-bration | Pre-sented model |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Li[45] 2016 | Taiwan | Single-centre cohort | Diabetic nephropathy | - | 131, 22 | RRT | - | 2 | Cox | 4 | Cross-validation | D: 0.90 | - | HRs & score |
| Pesce[46] 2016 | Multi-national | Multi-centre cohort | IgA-nephropathy | 87 | 1040, 241 | Time to ESRD | 3-8 | 24 | ANN | 6 | split sample + cross-validation | IV: 0.90 | - | Web-calculator (out of service) |
| Diciolla[47] 2015 | Multi-national | Multi-centre cohort | IgA-nephropathy | - | 1040, 241 | RRT | 5 | 40 | ANN | 6 | cross-validation | - | - | Web-calculator (out of service) |
| Hoshino[48] 2015 | Japan | Single-centre cohort | Diabetic nephropathy | 44 | 205, - | RRT | 10 | - | Cox | 4 | cross-validation | IV: 0.93 | - | - |
| Tanaka[49] 2013 | Japan | Multi-centre cohort | IgA-nephropathy | - | 698, 73 | RRT | 5 | 7 | Cox | 5 | - *(external)* | D: 0.87 | D: plot, test | HRs & score |
| Xie[50] 2012 | China | Single-centre cohort | IgA-nephropathy | 88 | 619, 67 | ESRD | 2,5,**10** | 2 | Cox | 4 | - | D: 0.85 | - | HRs |
| Berthoux [30], 2011 | France | Single-centre cohort | IgA-nephropathy | 75 | 332, 45 | Death, RRT | 10, 20 | 8 | Score | 3 | - *(external)* | - | - | HRs & score |
| Desai[51] 2011 | Multi-national | Multi-centre RCT | Diabetic nephropathy | 35 | 995, 222 | RRT | - | 6 | Cox | 19 | bootstrap | D: 0.85 | - | HRs |
| Day[52] 2010 | UK | Single-centre cohort | Pauci immune GN | - | 390, 54 | RRT | 1 | 9 | Cox | 2 | - | D: 0.83 | - | HRs |
| Goto[29] 2009 | Japan | Multi-centre cohort | †† IgA-nephropathy | - | 2283, 252 | RRT | 10 | 18 | Cox | 8 | bootstrap+ split-sample | D: 0.94 IV: 0.94 | - | Score |

**Table 1:** Continued.

| Study | Country | Design | Population | GFR mean | N total, N events | Outcome | Time-frame (years) | EPV | Model type | N pre-dictors | Internal validation | C- sta-tistic | Cali-bration | Pre-sented model |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kent[53] 2007 | Multi-national | Multiple RCT's | Non-diabetic CKD | - | 1860, 311 | RRT/100% s.creat↑ | - | 62 | Cox | 5 | - | D: 0.83 | D: plot, test | HRs |
| Keane[54] 2006 | Multi-national | RCT | diabetic nephropathy | - | 1513, 341 | RRT | - | 12 | Cox | 4 | jackknife | - | D: plot | HRs |
| Magistroni [55] 2006 | Italy | Single-centre cohort | IgA-nephropathy | 83 | 237, 40 | RRT | 10 | 2 | Cox | 4 | - *(external)* | - | - | Score |
| Wakai[56] 2006 | Japan | Multi-centre cohort | †† IgA-nephropathy | - | 2269, 207 | ESRD | 7 | 19 | Cox | 8 | bootstrap + split-sample | D: 0.94 IV: 0.93 | - | Score |
| Frimat[57] 1997 | France | Single-centre cohort | IgA-nephropathy | - | 210, 33 | RRT | 7 | 2 | Cox | 7 | - | - | D: plot | Score |
| Beukhof[58] 1986 | Nether-lands | Single-centre cohort | IgA-nephropathy | 94 | 75, 14 | RRT | 10 | 1 | Cox | 5 | - | - | - | Nomo-gram |

**Notes:** " - " not reported; ‡ Both studies by Johnson overlap in patient population and include the same predictors. The study by Schroeder updates this same model (the KPNW) by including additional predictors, and excluding some original predictors.; ‡‡ The formula as provided in the supplement of Schroeder's article does not provide the knot locations for spline terms, these are available from the authors upon request* The study by Xie and Maziarz 2015 include the same patient population, part of this population is included in Maziarz 2014. All three studies include the same predictors in the same 4 models, but re-estimate beta coefficients for different subsets. The population is of underserved/uninsured patients; † population of veterans 65+ years old; †† overlap in patient population, the study by Goto has an extended follow-up of three years in the same cohort as the study by Wakai.
**Abbreviations:** (e)GFR, (estimated) glomerular filtration rate in ml/min/1.73m$^2$; N, number; EPV, events per variable/candidate predictor; D, development; IV, internal validation; CKD, chronic kidney disease; CART, a classification and regression tree; RRT, renal replacement therapy; HRs, hazard ratios; ESRD, end-stage renal disease; ANN, artificial neural networks; s. creat, serum creatinine.

**Table 2: characteristics of external validation studies and model performance in validations (N=17).**

| Study | Model validated | Inde-pendent | Vali-dation type | Country | Population | GFR mean | N total, N events | Out-come | Time-frame (years) | Model updated | C-statistic | Cali-bra-tion | Updated model presented |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *General CKD* | | | | | | | | | | | | | |
| Schroeder[25] 2017 | KPNW Model (Johnson) | No | External | USA | General CKD stage 3-4 | 48 | 16553, 360 | RRT | 5 | Baseline hazard recalibrated to Colorado | 0.95 | Plot | No |
| Lennartz[19] 2016 | KFRE 4v (Tangri) | No | External | Germany | General CKD stage 2-4 | 46 | 565, 52 | RRT | 3 | Baseline hazard, addition ultrasound parameters | 4v, update: 0.91, 0.91 | Plot | Formula |
| Tangri[18] 2016 *Jama* | KFRE 4v & 8v (Tangri) | No | External | >30 countries | General CKD stage 3-5 | 46 | 721357, 23829 | RRT | 2, **5** | Baseline hazard recalibrated to Europe | 4v, 8v: 0.88, 0.88 | Plot | Formula & Web-calculator |
| Grams[16] 2015 | KFRE 4v (Tangri) | Yes | External | USA | * CKD, GFR 20-65 | - | 1094, - | RRT | 1, **5** | No | 0.83 | - | - |
| Marks[28] 2015 | model 7 (Marks) | No | Temporal | Scotland | General CKD stage 3-5 | 47 | 18687, 222 | RRT | 5 | No | 0.96 | Plot, test | - |
| | KFRE 3v & 4v (Tangri) | Yes | External | Scotland | General CKD stage 3-5 | 47 | 18687, 222 | RRT | 5 | No | 3v, 4v: 0,94, 0.95 | Plot | - |
| Levin[35] 2014 | KFRE 8v (Tangri) | No | External | Canada | General CKD stage 3b-4 | 28 | 2402, 142 | RRT | 1 | Coefficients re-estimated, biomarkers added | 8v, update: 0.86, 0.87 | - | HRs |

**Table 2:** Continued.

| Study | Model validated | Inde-pendent | Validation type | Country | Population | GFR mean | N total, N events | Out-come | Time-frame (years) | Model updated | C-statistic | Cali-bration | Updated model presented |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Drawz[27] 2013 | VA risk score (Drawz) | No | Geographic | USA | ** General CKD stage 4-5 | 25 | 819, 33 | GFR<15, RRT | 1 | No | 0.82 | - | - |
| | KFRE 8v (Tangri) | Yes | External | USA | ** General CKD stage 4-5 | 25 | 2684, 110 | GFR<15, RRT | 1 | No | 0.78 | - | - |
| Peeters[17] 2013 | KFRE 3v, 4v & 8v (Tangri) | Yes | External | The Neth-erlands | General CKD stage 3-5 | 33 | 595, 114 | RRT | 5 | No | 3v, 4v, 8v: 0.88,0.88, 0.89 | Plot, test | - |
| Tangri[24] 2011 | KFRE 3v,4v & 8v (Tangri) | No | External | Canada | General CKD stage 3-5 | 31 | 4942, 1177 | RRT | 1, 3, **5** | No | 3v, 4v, 8v: 0.79,0.83, 0.84 | Plot, test | - |
| Landray[38] 2010 | CRIB score (Landray) | No | External | UK | General CKD stage 3-5 | 22 | 213, 66 | RRT | - | No | 0.91 | Plot | - |
| ***Specified Renal Disease*** | | | | | | | | | | | | | |
| Knoop[21] 2015 | ARR Score (Berthoux) | Yes | External | Norway | IgA-nephropathy | - | 1134, 320 | Death, RRT | 5, **10**, 15 | Coefficients re-estimated, age & GFR added | 0.79 update: 0.89 | - | Formula |
| Mohey[22] 2013 | ARR Score (Berthoux) | No | External | France | Secondary IgA -nephropathy | 82 | 74, 19 | GFR<15, death | 10, 20 | No | - | - | - |
| Tanaka[49] 2013 | Tanaka score | No | External | Japan | IgA-nephropathy | - | 702, 85 | RRT | 5 | No | 0.89 | plot, test | - |

**Table 2:** Continued.

| Study | Model validated | Inde-pendent | Validation type | Country | Population | GFR mean | N total, N events | Out-come | Time-frame (years) | Model updated | C-statistic | Cali-bration | Updated model presented |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Xie[50] 2012 | Goto score | Yes | External | China | IgA-nephropathy | 88 | 619, 67 | ESRD | 2,5,**10** | No | 0.82 | - | - |
| | RENAAL score (Keane) | Yes | External | China | IgA-nephropathy | 88 | 619, 67 | ESRD | 2,5,**10** | No | 0.79 | - | - |
| | ARR Score (Berthoux) | Yes | External | China | IgA-nephropathy | 88 | 619, 67 | ESRD | 2,5,**10** | No | 0.73 | - | - |
| Berthoux[30], 2011 | ARR Score (Berthoux) | No | Temporal | France | IgA-nephropathy | - | 250, 38 | Death, RRT | 10, 20 | No | - | - | - |
| Bjorneklett[20] 2011 | Goto score | Yes | External | Norway | IgA-nephropathy | 67 | 633, 146 | RRT | 10, 20 | Coefficients re-estimated, classification simplified | - | - | No |
| Magistroni[55] 2006 | Magistroni score | No | External | Italy | IgA-nephropathy | - | 73, 8 | RRT | 10 | No | - | - | - |

**Notes:** " - " not reported; * hypertensive CKD population; ** population of veterans 65+ years old; (e)GFR, (estimated) glomerular filtration rate in ml/min/1.73m$^2$; N, number; EPV, events per variable/candidate predictor; CKD, chronic kidney disease; RRT, renal replacement therapy; HRs, hazard ratios; ESRD, end-stage renal disease.

## Discussion

This systematic review provides an overview of all development and validation studies of predictive models for progression of CKD to kidney failure. Since the last reviews on this topic, the number of publications has more than doubled.[3] Most included studies report high model performance measures, implying that calculating an individual's risk of renal failure with high accuracy is attainable. This is further emphasized by the similar predictors included in various models. There were, however, substantial shortcomings in many publications. As in many medical prediction studies, etiological and prediction goals were often confused, limiting interpretability and applicability.[7 26] Firstly, more than half the tools provided insufficient details to calculate an individual's prognosis of kidney failure, rendering it useless to its intended purpose. Secondly, the clinical relevance of many models is limited due to the selection of derivation population. Thirdly, a high risk of bias was observed across studies, mainly due to high risk of overfitting, inadequate handling of missing data and incomplete reporting of performance measures. Fourthly, sufficient validation was largely lacking, increasing research waste and limiting reliability of models. And finally, not a single impact study on the effect of clinical uptake has been performed. It is, therefore, not surprising that clinical uptake of models remains sporadic and guidelines on which model to use are lacking.

Providing absolute evidence for the single 'best' prognostic tool to use is complicated by differences between studies, mainly concerning varying study populations, use of different prediction baselines, use of varying time-frames and multiple outcome definitions. A selection guide including all usable models is presented, that may assist clinicians and patients in choosing the tool appropriate to their setting (Figure 5). There are many factors to take into account when selecting the most appropriate model, depending on the user's wishes and specific clinical setting. Users should be wary of overfitting in models developed on small sample size studies and we would advise against use of these models, unless validated in a sufficiently large sample. Based on our results we would advise the use of a tool with an overall low risk of bias, which has shown good performance in external validation in a similar population to the population in which use is intended, and ideally has been assessed in an impact study.

For kidney failure prediction in a general CKD cohort with stage 3-5 patients, we would recommend the 4- or 8-variable KFRE, as it has been externally validated extensively for a time-frame of 2 and 5 years. Though the development study potentially introduced bias by selecting predictors that were recorded up to 365 days after prediction baseline and by using univariate analysis to select predictors, the model has shown consistently good performance in CKD stage 3-5 patients from less-biased external validation studies.[18 24] Alternatively, for 5 year predictions the KPNW model as updated and externally validated by Schroeder also has great potential, mainly due to its methodological rigor and low risk of bias, though it is less easy to use than the KFRE.[25] Various other general CKD models showed promising results in development, but should be further externally validated to ensure consistency of performance before clinical use. [23 27 28] For prediction of disease

progression in IgA-nephropathy patients, a large number of models are available. These models, however, were generally developed on a small sample size and often had a high risk of bias. The most evidence on validity was found for the risk scores developed by Goto et al and the ARR (by Berthoux).[29 30] The Goto score does contain some risk of bias due to a complete-case analysis and univariate selection of predictors, but was developed on a relatively large sample size and has been externally validated twice. Though the ARR score was developed using questionable model building methods and with incomplete reporting of performance, this score has been externally validated the most times and a recently updated version presented by Knoop et al. shows great potential.[21]

Clinical relevance proved to be largely lacking for many of the included models in the current review. Specifically models for general CKD patients were often developed on prevalent patients with a large range of disease severity, and did not specify a specific time-point when the model should be used. Prediction of renal failure can be extremely accurate when using a population with GFR's ranging from 10 to 60 ml/min/1.73m/1.73m$^2$. However, in practice, such tools would probably be employed for a more homogeneous group of patients in which it is clinically relevant to discuss prognosis. The predictive capacities of the model would be lower in such a population. This is exemplified in the KFRE validation performed by Peeters et al., where the AUC of the 4-variable KFRE dramatically decreased from 0.88 in the whole population (CKD stage 3-5) to 0.71 in the more relevant population of CKD stage 4 patients.[17] Another factor limiting usability and interpretability is that a number of studies didn't define a prediction time-frame. Finally, the definition of outcome differs between studies. The use of composite endpoints is particularly problematic, as it limits the value of the model for clinicians, as each separate endpoint requires different interventions. In conclusion, an ideal model is developed for one clearly defined clinically meaningful and objective endpoint in a population for which prediction is clinically relevant. Few models included in this review met these recommendations and this lack of clinical relevance could be a large contributor to the slow uptake seen in practice.

Despite the limited uptake and discussed shortcomings of existing tools, risk prediction models for kidney failure have a large potential for improving patients' decision making, treatment and overall health. In future studies, there is need for improvement of quality of reporting and used methodology. As the majority of models included had a high risk of bias, these models should not be implemented unless their validity is proven in unbiased external validation studies. Hopefully, efforts such as the TRIPOD guidelines will improve these inadequacies and result in more robust, usable and unbiased prognostic tools.[9] To limit research waste and improve clinical uptake, it is firstly of crucial importance that development studies provide enough model information (formula/score with absolute risk table) to enable use. For specific renal diseases and homogenous patient populations, there certainly appears to be space for improvement in model development. For populations in which multiple models are available, we advise that future research focusses on the updating, validation and implementation of these existing prognostic tools. Previous studies have shown that the combination of well-established clinical risk factors and kidney disease markers can accurately predict renal failure in a general CKD

population. Therefore, one might advise to focus resources on updating models for more clinically relevant populations in an unbiased fashion. To do so, comprehensive validation of multiple models in different settings is key. Additionally, translation of mathematical model formulas to simple tools such as web-calculators, and enabling automated uptake is of great importance for integration into daily clinical routine. Ultimately, impact studies will be necessary to determine whether the implementation of such tools truly improve patient outcomes. Ideally such impact studies would be randomized controlled trials and would assess the effect of implementing a prediction model in clinical practice. Different outcomes might be considered as end-points in such studies, partly dependent on the time of prediction. Relevant outcomes might be timely referral to nephrologists, timely placement of vascular access, better informed patients, improved quality of life and possibly even improved survival.

The current review has a number of strengths. First of all, we expect to have included a complete overview of existing models. Furthermore, this is the first study on kidney failure models to perform a formal risk of bias assessment aimed specifically at prediction research. The study is limited by the inclusion of only English language articles. Also, the differences in case-mix and characteristics of included studies makes it difficult to directly compare their performances. Herein we are limited by the lack of validation studies that compare multiple models in the same cohort. Finally, we limited the scope of this review to models predicting kidney failure, though other outcomes such as death or cardiovascular events may also have significant clinical value.

In conclusion, this study provides a systematic overview of existing models for predicting progression to kidney failure in CKD patients. The results may be used as a tool to select the most appropriate and robust prognostic model for various settings. Finally, we hope the current review motivates researchers in this field to decrease the generation of new models and combine efforts to explore, analyse and update existing models in clinically relevant settings, in order to ultimately stimulate clinical uptake and improve patient outcomes.

## References


1. Chapter 2: Definition, identification, and prediction of CKD progression. *Kidney international supplements* 2013;3(1):63-72. doi: https://doi.org/10.1038/kisup.2012.65
2. Kadatz MJ, Lee ES, Levin A. Predicting Progression in CKD: Perspectives and Precautions. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2016;67(5):779-86. doi: 10.1053/j.ajkd.2015.11.007 [published Online First: 2016/01/05]
3. Tangri N, Kitsios GD, Inker LA, et al. Risk prediction models for patients with chronic kidney disease: a systematic review. *Annals of internal medicine* 2013;158(8):596-603. doi: 10.7326/0003-4819-158-8-201304160-00004 [published Online First: 2013/04/17]
4. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS medicine* 2012;9(11):e1001344. doi: 10.1371/journal.pmed.1001344 [published Online First: 2012/11/28]
5. Schell JO, Patel UD, Steinhauser KE, et al. Discussions of the kidney disease trajectory by elderly patients and nephrologists: a qualitative study. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2012;59(4):495-503. doi: 10.1053/j.ajkd.2011.11.023 [published Online First: 2012/01/10]
6. Chapter 5: Referral to specialists and models of care. *Kidney international supplements* 2013;3(1):112-19. doi: https://doi.org/10.1038/kisup.2012.68
7. Collins GS, Omar O, Shanyinde M, et al. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. *Journal of clinical epidemiology* 2013;66(3):268-77. doi: 10.1016/j.jclinepi.2012.06.020 [published Online First: 2012/11/03]
8. 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 (Clinical research ed)* 2009;339:b2700. doi: 10.1136/bmj.b2700 [published Online First: 2009/07/23]
9. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ (Clinical research ed)* 2015;350:g7594. doi: 10.1136/bmj.g7594 [published Online First: 2015/01/09]
10. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS medicine* 2014;11(10):e1001744. doi: 10.1371/journal.pmed.1001744 [published Online First: 2014/10/15]
11. PROBAST: a risk of bias tool for prediction modelling studies. In: Challenges to evidence-based health care and Cochrane. . Abstracts of the 24th Cochrane Colloquium; 2016 2016 23-27 Oct; Seoul, Korea. John Wiley & Sons.
12. Royston P, Moons KG, Altman DG, et al. Prognosis and prognostic research: Developing a prognostic model. *BMJ (Clinical research ed)* 2009;338:b604. doi: 10.1136/bmj.b604 [published Online First: 2009/04/02]
13. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *American journal of epidemiology* 2010;172(8):971-80. doi: 10.1093/aje/kwq223 [published Online First: 2010/09/03]
14. Wynants L, Bouwmeester W, Moons KG, et al. A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data. *Journal of clinical epidemiology* 2015;68(12):1406-14. doi: 10.1016/j.jclinepi.2015.02.002 [published Online First: 2015/03/31]
15. 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):214-26. doi: 10.1002/sim.6787 [published Online First: 2015/11/11]
16. Grams ME, Li L, Greene TH, et al. Estimating time to ESRD using kidney failure risk equations: Results from the African American Study of Kidney Disease and Hypertension (AASK). *American Journal of Kidney Diseases* 2015;65(3):394-402. doi: http://dx.doi.org/10.1053/j.ajkd.2014.07.026

17. Peeters MJ, Van Zuilen AD, Van Den Brand JAJG, et al. Validation of the kidney failure risk equation in European CKD patients. *Nephrology Dialysis Transplantation* 2013;28(7):1773-79. doi: http://dx.doi.org/10.1093/ndt/gft063
18. Tangri N, Grams ME, Levey AS, et al. Multinational Assessment of Accuracy of Equations for Predicting Risk of Kidney Failure: A Meta-analysis. *Jama* 2016;315(2):164-74. doi: 10.1001/jama.2015.18202 [published Online First: 2016/01/13]
19. Lennartz CS, Pickering JW, Seiler-Mussler S, et al. External Validation of the Kidney Failure Risk Equation and Re-Calibration with Addition of Ultrasound Parameters. *Clinical journal of the American Society of Nephrology : CJASN* 2016;11(4):609-15. doi: 10.2215/cjn.08110715 [published Online First: 2016/01/21]
20. Bjorneklett R, Vikse BE, Bostad L, et al. Long-term risk of ESRD in IgAN; validation of Japanese prognostic model in a Norwegian cohort. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2012;27(4 ):1485-91. doi: 10.1093/ndt/gfr446 [published Online First: 2011/08/09]
21. Knoop T, Vagane AM, Vikse BE, et al. Addition of eGFR and age improves the prognostic absolute renal risk-model in 1,134 norwegian patients with IgA nephropathy. *American journal of nephrology* 2015;41(3):210-19. doi: http://dx.doi.org/10.1159/000381403
22. Mohey H, Laurent B, Mariat C, et al. Validation of the absolute renal risk of dialysis/death in adults with IgA nephropathy secondary to Henoch-Schonlein purpura: a monocentric cohort study. *BMC nephrology* 2013;14:169. doi: 10.1186/1471-2369-14-169 [published Online First: 2013/08/07]
23. Tangri N, Inker LA, Hiebert B, et al. A Dynamic Predictive Model for Progression of CKD. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2017;69(4 ):514-20. doi: 10.1053/j.ajkd.2016.07.030 [published Online First: 2016/10/04]
24. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *Jama* 2011;305(15 ):1553-9. doi: 10.1001/jama.2011.451 [published Online First: 2011/04/13]
25. Schroeder EB, Yang X, Thorp ML, et al. Predicting 5-Year Risk of RRT in Stage 3 or 4 CKD: Development and External Validation. *Clinical journal of the American Society of Nephrology : CJASN* 2017;12(1 ):87-94. doi: 10.2215/cjn.01290216 [published Online First: 2016/12/29]
26. van Diepen M, Ramspek CL, Jager KJ, et al. Prediction versus aetiology: common pitfalls and how to avoid them. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2017;32(suppl_2):ii1-ii5. doi: 10.1093/ndt/gfw459 [published Online First: 2017/03/25]
27. Drawz PE, Goswami P, Azem R, et al. A simple tool to predict end-stage renal disease within 1 year in elderly adults with advanced chronic kidney disease. *Journal of the American Geriatrics Society* 2013;61(5 ):762-8. doi: 10.1111/jgs.12223 [published Online First: 2013/04/27]
28. Marks A, Fluck N, Prescott GJ, et al. Looking to the future: predicting renal replacement outcomes in a large community cohort with chronic kidney disease. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2015;30(9 ):1507-17. doi: 10.1093/ndt/gfv089 [published Online First: 2015/05/07]
29. Goto M, Wakai K, Kawamura T, et al. A scoring system to predict renal outcome in IgA nephropathy: a nationwide 10-year prospective cohort study. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2009;24(10):3068-74. doi: 10.1093/ndt/gfp273 [published Online First: 2009/06/12]
30. Berthoux F, Mohey H, Laurent B, et al. Predicting the risk for dialysis or death in IgA nephropathy. *Journal of the American Society of Nephrology : JASN* 2011;22(4):752-61. doi: 10.1681/asn.2010040355 [published Online First: 2011/01/25]
31. Cheng LC, Hu YH, Chiou SH. Applying the Temporal Abstraction Technique to the Prediction of Chronic Kidney Disease Progression. *Journal of medical systems* 2017;41(5 ):85. doi: 10.1007/s10916-017-0732-5 [published Online First: 2017/04/13]
32. Hsu CY, Xie D, Waikar SS, et al. Urine biomarkers of tubular injury do not improve on the clinical model predicting chronic kidney disease progression. *Kidney international* 2017;91(1 ):196-203. doi: 10.1016/j.kint.2016.09.003 [published Online First: 2016/12/29]

33. Xie Y, Maziarz M, Tuot DS, et al. Risk prediction to inform surveillance of chronic kidney disease in the US Healthcare Safety Net: a cohort study. *BMC nephrology* 2016;17(1):57. doi: 10.1186/s12882-016-0272-0 [published Online First: 2016/06/10]
34. Maziarz M, Black RA, Fong CT, et al. Evaluating risk of ESRD in the urban poor. *Journal of the American Society of Nephrology* 2015;26(6):1434-42. doi: http://dx.doi.org/10.1681/ASN.2014060546
35. Levin A, Rigatto C, Barrett B, et al. Biomarkers of inflammation, fibrosis, cardiac stretch and injury predict death but not renal replacement therapy at 1 year in a Canadian chronic kidney disease cohort. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2014;29(5 ):1037-47. doi: 10.1093/ndt/gft479 [published Online First: 2013/12/29]
36. Maziarz M, Chertow GM, Himmelfarb J, et al. Homelessness and risk of end-stage renal disease. *Journal of health care for the poor and underserved* 2014;25(3 ):1231-44. doi: 10.1353/hpu.2014.0136 [published Online First: 2014/08/19]
37. Smith ER, Lee D, Cai MM, et al. Urinary neutrophil gelatinase-associated lipocalin may aid prediction of renal decline in patients with non-proteinuric Stages 3 and 4 chronic kidney disease (CKD). *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2013;28(6 ):1569-79. doi: 10.1093/ndt/gfs586 [published Online First: 2013/01/19]
38. Landray MJ, Emberson JR, Blackwell L, et al. Prediction of ESRD and death among people with CKD: the Chronic Renal Impairment in Birmingham (CRIB) prospective cohort study. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2010;56(6):1082-94. doi: 10.1053/j.ajkd.2010.07.016 [published Online First: 2010/11/03]
39. Johnson ES, Thorp ML, Platt RW, et al. Predicting the risk of dialysis and transplant among patients with CKD: a retrospective cohort study. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2008;52(4 ):653-60. doi: 10.1053/j.ajkd.2008.04.026 [published Online First: 2008/07/01]
40. Johnson ES, Thorp ML, Yang X, et al. Predicting renal replacement therapy and mortality in CKD. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2007;50(4):559-65. doi: 10.1053/j.ajkd.2007.07.006 [published Online First: 2007/09/29]
41. Dimitrov BD, Ruggenenti P, Stefanov R, et al. Chronic nephropathies: individual risk for progression to end-stage renal failure as predicted by an integrated probabilistic model. *Nephron Clinical practice* 2003;95(2):c47-59. doi: 10.1159/000073668 [published Online First: 2003/11/12]
42. Bidadkosh A, Lambooy SPH, Heerspink HJ, et al. Predictive Properties of Biomarkers GDF-15, NTproBNP, and hs-TnT for Morbidity and Mortality in Patients With Type 2 Diabetes With Nephropathy. *Diabetes care* 2017;40(6 ):784-92. doi: 10.2337/dc16-2175 [published Online First: 2017/03/28]
43. Tang Y, Qin W, Peng W, et al. Development and validation of a prediction score system in lupus nephritis. *Medicine* 2017;96(37 ):e8024. doi: 10.1097/md.0000000000008024 [published Online First: 2017/09/15]
44. Barbour SJ, Espino-Hernandez G, Reich HN, et al. The MEST score provides earlier risk prediction in lgA nephropathy. *Kidney international* 2016;89(1):167-75. doi: 10.1038/ki.2015.322 [published Online First: 2016/01/14]
45. Li HY, Lin HA, Nien FJ, et al. Serum Vascular Adhesion Protein-1 Predicts End-Stage Renal Disease in Patients with Type 2 Diabetes. *PloS one* 2016;11(2):e0147981. doi: 10.1371/journal.pone.0147981 [published Online First: 2016/02/06]
46. Pesce F, Diciolla M, Binetti G, et al. Clinical decision support system for end-stage kidney disease risk estimation in IgA nephropathy patients. *Nephrology Dialysis Transplantation* 2016;31(1):80-86. doi: http://dx.doi.org/10.1093/ndt/gfv232
47. Diciolla M, Binetti G, Di Noia T, et al. Patient classification and outcome prediction in IgA nephropathy. *Computers in biology and medicine* 2015;66( ):278-86. doi: 10.1016/j.compbiomed.2015.09.003 [published Online First: 2015/10/11]
48. Hoshino J, Mise K, Ueno T, et al. A pathological scoring system to predict renal outcome in diabetic nephropathy. *American journal of nephrology* 2015;41(4-5):337-44. doi: 10.1159/000431333 [published Online First: 2015/06/13]

49. Tanaka S, Ninomiya T, Katafuchi R, et al. Development and validation of a prediction rule using the Oxford classification in IgA nephropathy. *Clinical journal of the American Society of Nephrology : CJASN* 2013;8(12 ):2082-90. doi: 10.2215/cjn.03480413 [published Online First: 2013/11/02]
50. Xie J, Kiryluk K, Wang W, et al. Predicting progression of IgA nephropathy: new clinical progression risk score. *PloS one* 2012;7(6 ):e38904. doi: 10.1371/journal.pone.0038904 [published Online First: 2012/06/22]
51. Desai AS, Toto R, Jarolim P, et al. Association between cardiac biomarkers and the development of ESRD in patients with type 2 diabetes mellitus, anemia, and CKD. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2011;58(5):717-28. doi: 10.1053/j.ajkd.2011.05.020 [published Online First: 2011/08/09]
52. Day CJ, Howie AJ, Nightingale P, et al. Prediction of ESRD in pauci-immune necrotizing glomerulonephritis: quantitative histomorphometric assessment and serum creatinine. *American journal of kidney diseases : the official journal of the National Kidney Foundation* 2010;55(2):250-8. doi: 10.1053/j.ajkd.2009.10.047 [published Online First: 2010/01/05]
53. Kent DM, Jafar TH, Hayward RA, et al. Progression risk, urinary protein excretion, and treatment effects of angiotensin-converting enzyme inhibitors in nondiabetic kidney disease. *Journal of the American Society of Nephrology : JASN* 2007;18(6):1959-65. doi: 10.1681/asn.2006101081 [published Online First: 2007/05/04]
54. Keane WF, Zhang Z, Lyle PA, et al. Risk scores for predicting outcomes in patients with type 2 diabetes and nephropathy: the RENAAL study. *Clinical journal of the American Society of Nephrology : CJASN* 2006;1(4):761-7. doi: 10.2215/cjn.01381005 [published Online First: 2007/08/21]
55. Magistroni R, Furci L, Leonelli M, et al. A validated model of disease progression in IgA nephropathy. *Journal of nephrology* 2006;19(1):32-40. [published Online First: 2006/03/09]
56. Wakai K, Kawamura T, Endoh M, et al. A scoring system to predict renal outcome in IgA nephropathy: from a nationwide prospective study. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 2006;21(10):2800-8. doi: 10.1093/ndt/gfl342 [published Online First: 2006/07/11]
57. Frimat L, Briancon S, Hestin D, et al. IgA nephropathy: prognostic classification of end-stage renal failure. L'Association des Nephrologues de l'Est. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association* 1997;12(12):2569-75. [published Online First: 1998/02/12]
58. Beukhof JR, Kardaun O, Schaafsma W, et al. Toward individual prognosis of IgA nephropathy. *Kidney international* 1986;29(2):549-56. [published Online First: 1986/02/01]

## Supplemental Material for Chapter 6

Search strategies used on December 31st 2017.

PubMed:

("ESRD"[ti] OR "ESKD"[ti] OR ((end stage*[ti] OR endstage*[ti]) AND ("renal"[ti] OR kidney*[ti])) OR "Kidney Failure, Chronic"[majr] OR "Chronic Kidney Failure"[ti] OR "Chronic Renal Failure"[ti] OR "Renal Insufficiency, Chronic"[majr] OR "chronic Renal Insufficiency"[ti] OR "chronic kidney Insufficiency"[ti] OR "CKD"[ti] OR "chronic kidney disease"[ti] OR "chronic kidney diseases"[ti] OR nephropath*[ti]) AND ("predictive model"[ti] OR "predictive models"[ti] OR predictive model*[ti] OR "prediction model"[ti] OR "prediction models"[ti] OR prediction model*[ti] OR "prediction rule"[ti] OR "prediction rules"[ti] OR "predictive rule"[ti] OR "predictive rules"[ti] OR "prognostic model"[ti] OR "prognostic models"[ti] OR prognostic model*[ti] OR "risk score"[ti] OR "risk scores"[ti] OR "score"[ti] OR "scoring"[ti] OR "predictive"[ti] OR "predicting"[ti] OR "predict" [ti] OR "predicts" [ti] OR "prediction"[ti] OR "Risk Assessment"[Majr] OR "risk assessment"[ti] OR "risk assessments"[ti]) AND English[lang]

Embase:

("ESRD".ti. OR "ESKD".ti. OR ((end stage*.ti. OR endstage*.ti.) AND ("renal".ti. OR kidney*.ti.)) OR exp *chronic kidney failure/ OR "Chronic Kidney Failure".ti. OR "Chronic Renal Failure".ti. OR "chronic Renal Insufficiency".ti. OR "chronic kidney Insufficiency".ti. OR "CKD".ti. OR "chronic kidney disease".ti. OR "chronic kidney diseases".ti. OR nephropath*.ti.) AND ("predictive model".ti. OR "predictive models".ti. OR predictive model*.ti. OR "prediction model".ti. OR "prediction models".ti. OR prediction model*.ti. OR "prediction rule".ti. OR "prediction rules".ti. OR "predictive rule".ti. OR "predictive rules".ti. OR "prognostic model".ti. OR "prognostic models".ti. OR prognostic model*.ti. OR "risk score".ti. OR "risk scores".ti. OR "score".ti. OR "scoring".ti. OR "predictive".ti. OR "predicting".ti. OR "predict" .ti. OR "predicts" .ti. OR "prediction".ti. OR exp *"Risk Assessment"/ OR "risk assessment".ti. OR "risk assessments".ti.) AND English.lg. NOT (conference OR conference abstract OR conference paper OR "conference review").pt.