

## Original Article

# Risk prediction models for graft failure in kidney transplantation: a systematic review

Rémi Kaboré<sup>1,2</sup>, Maria C. Haller<sup>3,4,5</sup>, Jérôme Harambat<sup>1,2,6,7</sup>, Georg Heinze<sup>3</sup> and Karen Leffondré<sup>1,2,6</sup>

<sup>1</sup>University of Bordeaux, ISPED, Centre INSERM U1219-Bordeaux Population Health Research, Bordeaux, France, <sup>2</sup>INSERM, ISPED, Centre INSERM U1219-Bordeaux Population Health Research, Bordeaux, France, <sup>3</sup>Center for Medical Statistics, Informatics and Intelligent Systems (CeMSIIS), Section for Clinical Biometrics, Medical University of Vienna, Vienna, Austria, <sup>4</sup>Department of Nephrology and Hypertension Diseases, Transplantation Medicine and Rheumatology, Krankenhaus Elisabethinen, Linz, Austria, <sup>5</sup>Methods Support Team European Renal Best Practice, Ghent University Hospital, Ghent, Belgium, <sup>6</sup>INSERM, Clinical Investigation Center-Clinical Epidemiology-CIC-1401, Bordeaux, France and <sup>7</sup>Pediatric Nephrology Unit, Pellegrin-Enfants Hospital, Bordeaux University Hospital, Bordeaux, France

Correspondence and offprint requests to: Karen Leffondré; E-mail: karen.leffondre@u-bordeaux.fr

### ABSTRACT

Risk prediction models are useful for identifying kidney recipients at high risk of graft failure, thus optimizing clinical care. Our objective was to systematically review the models that have been recently developed and validated to predict graft failure in kidney transplantation recipients. We used PubMed and Scopus to search for English, German and French language articles published in 2005–15. We selected studies that developed and validated a new risk prediction model for graft failure after kidney transplantation, or validated an existing model with or without updating the model. Data on recipient characteristics and predictors, as well as modelling and validation methods were extracted. In total, 39 articles met the inclusion criteria. Of these, 34 developed and validated a new risk prediction model and 5 validated an existing one with or without updating the model. The most frequently predicted outcome was graft failure, defined as dialysis, re-transplantation or death with functioning graft. Most studies used the Cox model. There was substantial variability in predictors used. In total, 25 studies used predictors measured at transplantation only, and 14 studies used predictors also measured after transplantation. Discrimination performance was reported in 87% of studies, while calibration was reported in 56%. Performance indicators were estimated using both internal and external validation in 13 studies, and using external validation only in 6 studies. Several prediction models for kidney graft failure in adults have been published. Our study highlights the need to better account for competing risks when

applicable in such studies, and to adequately account for post-transplant measures of predictors in studies aiming at improving monitoring of kidney transplant recipients.

**Keywords:** kidney graft loss, prediction model, prognosis, systematic review, transplantation

### INTRODUCTION

Predicting outcomes to guide clinical care, decision-making and resource allocation is a challenging issue in kidney transplantation. Although the survival of kidney transplant recipients has improved during the past 20 years, their life expectancy remains far below that of the general population [1–5]. Anticipating therapeutic procedures in patients at high risk of losing their kidney graft is a key to improve graft survival. Kidney biopsy is the ‘gold standard’ diagnostic tool to assess the graft rejection process and thus identify patients at risk. Unfortunately, this procedure remains invasive and expensive. Therefore, graft rejection initiation is often discovered too late to allow adjustment of care in time. Tools for predicting the long-term risk of graft loss appear to be suitable and cheap alternatives to addressing this difficulty. Several studies were conducted to propose clinical prediction tools for end-stage renal disease, and few systematic reviews of these articles have been realized [6, 7]. However, studies on graft failure prediction seem to be less frequent and, to the best of our knowledge, no systematic review of these tools and related methodological issues has been conducted. Yet,

kidney graft failure prediction raises several methodological issues including the presence of competing risks [8]. The use of available information on biomarkers after transplantation to improve graft failure prediction also raises some methodological challenges. The method used to validate the developed prediction tool is also of importance since its results indicate the relevance and suitability of the prediction tools in clinical practice.

Our objective was therefore to review prognostic studies published in the past 10 years that developed and validated, or only validated, a risk prediction model for graft failure after transplantation in kidney recipients, and to discuss methodological approaches.

## MATERIALS AND METHODS

### Search strategy

A search syntax with Boolean combinations of terms for kidney transplantation and prognostic models was constructed and used to identify risk prediction models for graft failure in kidney transplant recipients in Scopus and PubMed (Table 1) [9]. Detailed search strategies are provided in [Supplementary Appendix 1](#). The results of these searches were limited to studies published in English, German and French between 1 January 2005 and 31 December 2015. Reference lists of included articles were also reviewed for relevant citations.

### Study selection

R.K. screened retrieved articles by title and abstract and two reviewers (R.K. and J.H.) assessed the full text of each potentially relevant study to determine eligibility for inclusion using pre-defined criteria. We included studies that developed and validated a risk prediction model to predict graft failure in kidney transplant recipients as well as studies that validated an existing model with or without model updating. We excluded diagnostic models, narrative reviews, commentaries, case reports and editorials that contained no original data. Prediction models that were not designed to predict graft failure following kidney transplantation or did not perform any validation were also excluded. In case of discordance between the two reviewers, the final decision about inclusion/exclusion of the study was based on a discussion with a third reviewer (K.L.).

### Data extraction

We extracted the following data from each of the included prediction modelling studies adapted from the CHeck list for critical Appraisal and data extraction for systematic reviews of prediction Modelling Studies (CHARMS) checklist: (i) authors name and publication year; (ii) population characteristics including donor type (living/deceased), recipients age range and transplantation period; (iii) definition of graft failure (i.e. return to or initiation of dialysis, re-transplantation and/or death), and if applicable, any mention of competing risks in the paper (i.e. competing risk by death with functioning graft for death-censored graft failure, and competing risk by dialysis/re-transplantation for death with functioning graft), and time

horizon of prediction; (iv) the data and the methods used to develop the model including centres, years of transplantation, sample size, predictors and their time of measurement with respect to transplantation, and the type of statistical model (e.g. Cox regression model, logistic regression model); and (v) the data and the methods used to validate the model including external and internal validation methods, and the prediction performance criteria including indicators of overall performance, discrimination, calibration and reclassification (all explained below). We also retrieved the number of citations of the article in Google Scholar, as well as whether the predictive performance of the proposed tool was compared to other existing tools. R.K. performed the data extraction, the result of which was evaluated and discussed with K.L. until a consensus was reached.

### Overview of methodological approaches

Validation methods of prognosis models have been extensively discussed [10–12]. The overall performance of a model reflects the distance between predicted and observed outcome [13, 14]. Discrimination reflects the ability of the model to distinguish patients who will experience the event from patients who will not. It is usually measured using the concordance (C) statistics, i.e. the area under the receiver operating characteristic (ROC) curve (AUC) [10, 11], which should account for censoring [15, 16] and competing risks [17, 18] if applicable. Calibration reflects the capacity of the model to correctly estimate the probability of the event at an individual level and can be assessed graphically [19]. Reclassification quantifies the improvement of a new risk prediction model compared with an existing one, in terms of classification of patients into those who will experience the event and those who will not [20, 21].

All performance indicators can be calculated on the same data set as the one used to develop the model (internal validation), and/or a separate data set (external validation). Internal validation includes single random split-sampling (e.g. 50% development sample, 50% validation sample) and resampling methods (cross-validation or bootstrapping) based on many repeated splits of the data set [22]. External validation includes temporal (using patients transplanted in different years), spatial (using patients transplanted in different centres) and fully external validation (using patients from different countries) [23].

## RESULTS

The search yielded 2173 citations (Figure 1). A total of 134 citations were retrieved for full text assessment, and 39 of these met our inclusion criteria. Full data extraction table is in [Supplementary Appendix 2](#). Among the 39 studies, 34 developed and validated a new risk prediction model [24–57] and 5 validated an existing one (2 without any model updating [58, 59] and 3 with updating [60–62]). Among the 39 studies, 19 were published after 2010, with important variability between the numbers of citations in Google Scholar (Table 2).

### Studied populations

Most studies ( $n = 31$ ) targeted adult recipients [24–26, 28–55], including one devoted to elderly recipients [35]. No study

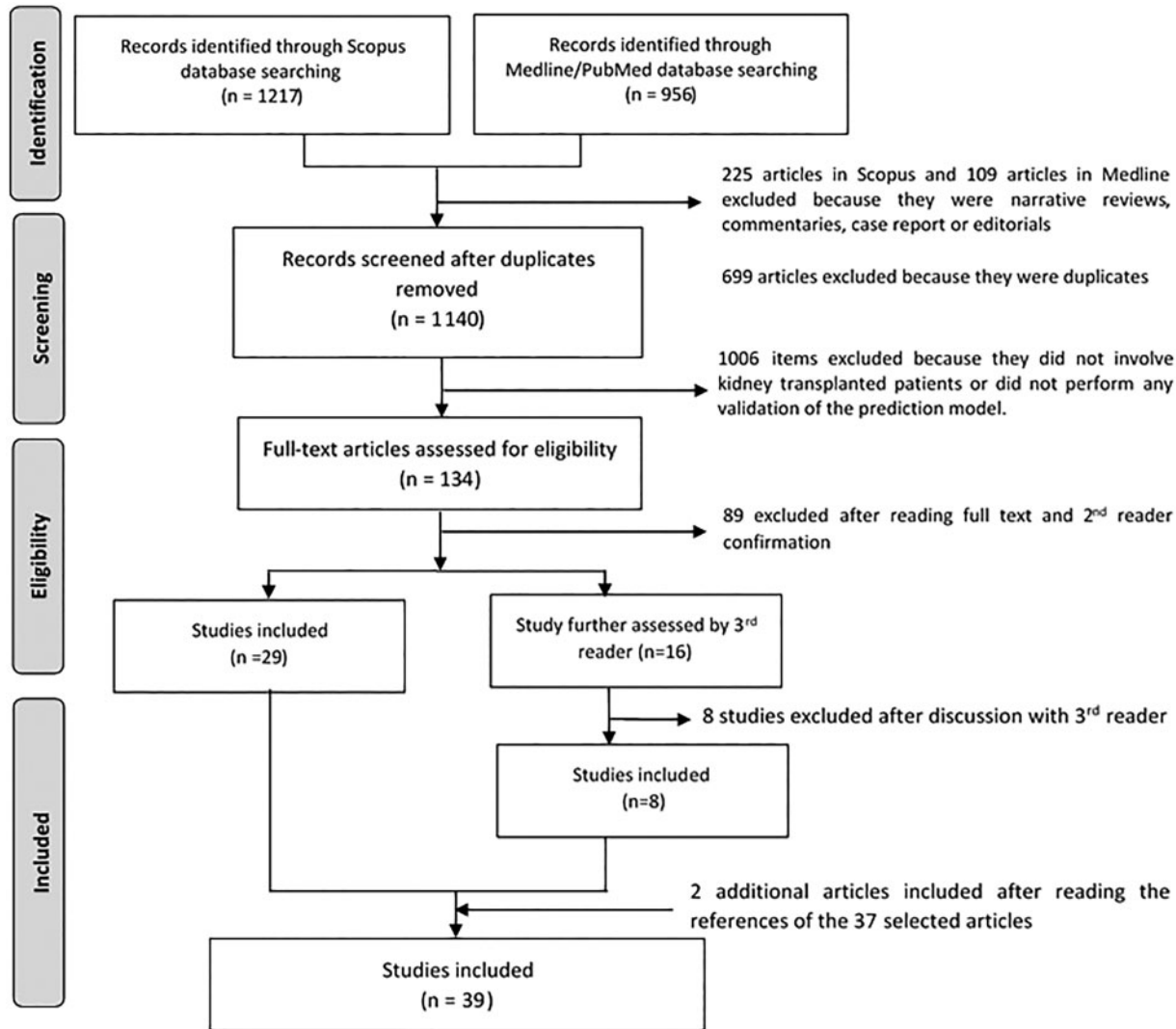


FIGURE 1: Flow chart of systematic review of studies kidney graft loss prediction.

Table 1. Database search strategy

Topic	Specific terms used	Position in article
Kidney transplanted population	'transplant OR transplantation OR graft OR grafting' AND 'kidney OR renal'	TITLE-ABSTRACT TITLE-ABSTRACT
Prediction	'prediction OR predict OR predictive OR probability OR prognosis OR prognostic OR prognostication OR score OR scores' AND	TITLE-ABSTRACT
Modelling	'model OR models OR regression OR equation OR equations OR modeling OR modelling' AND	TITLE-ABSTRACT
Outcome	'failure OR loss OR Death OR mortality OR survival' AND	TITLE-ABSTRACT
Prediction study design	'prediction OR predict OR predictive OR predicting OR validation OR validity OR validated OR cross-validation OR selection OR calibration OR discrimination OR discriminates OR ROC'	TITLE-ABSTRACT-KEYWORDS <sup>a</sup>

<sup>a</sup>Further search in keywords in Scopus database only.

Table 2. Publication and population characteristics

Characteristics	n	%
<b>Type of studies</b>		
Development and validation of a new model	34	87
Validating an existing model		
Without any model updating	2	5
With model updating	3	8
<b>Publication years</b>		
2005–07	6	15
2008–10	14	36
2011–13	14	36
2014–15	5	13
<b>Number of citations in Google Scholar</b>		
<6	9	23
6–15	9	23
16–29	12	31
≥30	9	23
<b>Recipient age</b>		
All ages	1	3
≥16 years	3	8
<b>Adults</b>		
≥18 years	15	38
≥65 years	1	3
Age range unspecified	15	38
Not reported	4	10
<b>Recipient kidney transplant type</b>		
Deceased or living donors	9	23
Living donors only	4	10
Deceased donor only	14	36
Unspecified	12	31
<b>Data used to develop the model</b>		
National registry	17	43
Single centre	14	36
Multicentre	8	21

specifically targeted paediatric recipients, although four studies included all ages or recipients older than 16 years. In total, 4 studies (10%) did not indicate age range of recipients, 14 (36%) studies developed/validated models for recipients of deceased donors only and 12 (31%) did not report the donor type. In total, 17 (43%) studies used data from national registries to develop their model, 14 (36%) used data from a single centre and 8 (21%) from several centres (Table 2).

### Predicted outcomes

Nine articles investigated several types of events including different definitions of graft failure (Table 3). Among the 30 studies predicting one event type only, dialysis or re-transplantation (whichever comes first) was investigated in 10 papers, death was included in the graft failure definition in 11 papers and 9 studies predicted only death. Among these nine studies, one indicated that it was death with functioning graft [61] and four that it was death before or after dialysis/re-transplantation [27, 35, 37, 62], so with or without functioning graft. Four papers just mentioned ‘any cause of death’ [29, 38, 39, 58], which did not indicate whether it was death with functioning graft only, or death with or without functioning graft. Outcomes were mostly predicted at long-term time horizon (i.e. at least 5 years post-transplant) (Table 3), and no study clearly

Table 3. Predicted events, timing of prediction and predictors, and statistical modelling

Study characteristics	n	%
<b>Predicted events</b>		
Graft failure (dialysis/re-transplantation/death with functioning graft)	11	28
Death censored graft failure (dialysis/re-transplantation)	10	25
Death with or without functioning graft	4	10
Death with functioning graft	1	3
Death of ‘any cause’ (without specifying with or without functioning graft)	4	10
Several predicted events	9	23
<b>Predicted time horizon</b>		
Short term (1–4 years) only	7	18
Long term only		
5–10 years	15	38
Years unspecified	3	8
Both short-term and long-term time horizon	11	28
Not reported	3	8
<b>Timing of predictors measurement</b>		
At transplantation only	25	64
After transplantation	14	36
<b>Statistical modelling method</b>		
Cox regression only	23	59
Logistic regression only	5	12
BBNs only	1	3
Based tree model only	3	8
Linear regression model only	1	3
Several types of statistical model used	5	12
Calculation of score index by formula	1	3
<b>Final form of the model used for validation and future use</b>		
Original model	17	44
Score	19	49
Nomogram	1	2
Several forms	2	5

reported if and how competing risks were accounted for when applicable (Supplementary Appendix 2).

### Predictors used

Predictors used to build the models (Figure 2) were measured at time of kidney transplantation only in 25 studies, and at later times during follow-up in 14 studies. The predictors, the values of which were measured at transplantation, only included donor and recipient age, donor type, cause of kidney failure, cold ischemia time, dialysis duration prior to kidney transplantation and body mass index of the donor and the recipient (Supplementary Appendix 2). Other predictors were measured shortly after transplantation [e.g. creatinine in the first week after transplantation or last estimated glomerular filtration rate (eGFR) available [38, 44]] or further from transplantation (e.g. serum creatinine at 6 months [40, 44] or at 1 year [50], blood pressure, and proteinuria at 1 year [33]). Other predictors less commonly used in routine care were also included in some models, such as immunological markers [26] or carotid-femoral pulse wave velocity at 8 weeks after transplantation [29].



**Table 4. Data and methods used to validate the model**

	<i>n</i>	%
<b>Method used to test the performance of the model</b>		
External validation only ( <i>n</i> = 6)		15
Spatial <sup>a</sup>	2	
Temporal <sup>a</sup>	2	
Fully external <sup>a</sup>	2	
Internal validation only ( <i>n</i> = 20)		51
Split-sample only		
50% development, 50% validation	5	
60% development, 40% validation	1	
66% development, 33% validation	2	
Cross-validation only		
5-fold cross-validation	1	
10-fold cross-validation	4	
Unspecified	3	
Bootstrapping only	1	
Several internal validation methods used	1	
Internal validation unspecified	2	
Both internal and external validation ( <i>n</i> = 13)		34
<b>Method used to quantify the performance of the model</b>		
Overall performance ( <i>n</i> = 2)		5
$R^2$ (Coefficient of determination)	2	
Discrimination ( <i>n</i> = 34)		87
AUC (C statistic) only	26	
Classification (sensitivity and specificity) only	3	
Both criteria	5	
Calibration ( <i>n</i> = 22)		56
Hosmer–Lemeshow test only	5	
Calibration slope only	2	
Agreement between predicted and observed graft failures only	12	
Several calibration methods	3	
Reclassification		8
NRI	3	
Other criteria ( <i>n</i> = 1)		3
IPEC score	1	

<sup>a</sup>‘Spatial external validation’ means validated in other centre(s) in the same source population; ‘temporal external validation’ means validated in recent years in the same source population; ‘fully external validation’ means validated in another source population.

them, eight used single random split-sampling only [27, 34, 37–39, 43, 49, 52] and nine used resampling techniques only [26, 28, 29, 32, 36, 45, 48, 55, 62]. External validation was only performed in six studies (two studies used spatial validation [53, 60], two temporal validation [30, 42] and two fully external validation [58, 59]). Both internal and external validation was performed in 13 studies [24, 31, 33, 35, 41, 44, 46, 47, 50, 51, 57, 61, 63].

Among the 31 studies that reported AUC (Table 4), the highest AUC value (0.94) was obtained for an artificial neural networks model predicting dialysis/re-transplantation/death at 5 years in patients transplanted from living donors and who survived beyond 3 months post-transplant [24] (Supplementary Appendix 2). The result of the Hosmer–Lemeshow test was reported in seven papers [24, 32, 35, 44, 53, 54, 58]. The agreement between predicted and observed outcomes was evaluated in 12 studies, but calibration plots were shown in 5 studies only [28, 43, 49–51]. The calibration slope was reported in two studies [29, 41]. Reclassification performance using the Net Reclassification Improvement (NRI) index was reported in three studies [44, 46, 53]. Overall performance using the

coefficient of determination was reported in two studies [29, 51]. The integrated prediction error curve (IPEC) score, which evaluates the predictive performance in the context of high-dimensional survival data [66, 67], was used in one study [47].

## DISCUSSION

A relatively large number of prognostic models have been proposed and validated in the last 10 years to predict graft failure in adult kidney transplant recipients. However, we found substantial variability in data collected and methods used for model development and validation. Notably, the definition of graft failure was particularly variable and included death in about half of the studies. A large number of different predictors have been used in the various models, including a number of predictors measured after transplantation such as serum creatinine [33, 34, 38, 40], eGFR [41, 44, 50], proteinuria [33], acute rejection [33, 39], acute tubular necrosis [38], carotid-femoral pulse wave velocity [29] and use of immunosuppressant [38, 39]. About half of the studies validated their model without external data, using various internal validation techniques including mostly single-split sampling. Most studies reported the discrimination performance of their model using AUC only.

The large variability in the quality of the methods used to develop and validate the risk prediction model, as well as in the quality of reporting, has been observed in many other systematic reviews of prognosis studies conducted in other contexts [68–70]. Such observation led to the development of clear recommendations for assessing the performance of prediction models [10] as well as for reporting the results [71, 72]. Recommendations indicate for example that both discrimination and calibration should be reported, internal validation should be corrected for optimism using resampling techniques, external validation could be useful to assess the generalizability to other similar populations, and performance indicators for survival outcomes should account for censoring and competing risks [10]. While internal validation correcting for over-optimism is absolutely necessary as a first step to evaluate the performance of the model in the population used to develop the model, external validation is essential for subsequent use in clinical practice. Indeed, the overall population of kidney recipients is large and diverse, and clinicians may not be confident enough in a tool that has not been validated in different kidney recipient populations. Our results indeed show that among studies that performed both internal and external validation, the AUC tended to be substantially lower when based on external validation. The AUC also varied a lot depending on the time horizon of prediction, even for a given prediction model. Some studies did not report this time horizon, which makes it impossible to compare the reported AUC to other published AUC.

Methodological issues need to be considered when deciding to use or not a proposed predictive tool in clinical practice. For example, censoring and competing risks are of major importance in our setting since graft failure is a time-to-event (survival) outcome that is systematically subject to censoring because not all patients are usually followed-up over the entire period of prediction. Graft failure may also be subject to

competing risks depending on its exact definition. Indeed, death with functioning graft is a competing risk for death-censored graft failure, and dialysis or re-transplantation are competing risks for death with functioning graft [8]. It is also important to distinguish prediction of death with functioning graft from prediction of death after dialysis or re-transplantation, since both dialysis and re-transplantation modify the risk of death. Dialysis and re-transplantation should thus systematically be considered as competing events of death. Yet four studies mixed both death before and after dialysis or re-transplantation, and four just specified ‘any cause of death’ without clarifying whether it was death with functioning graft only, or whether it included death after dialysis/re-transplantation. While the Cox model censoring at competing events can be used directly to estimate hazard ratios [8, 73], its use to derive predicted probabilities of events and corresponding model performance indicators in the presence of competing risks requires specific methods and software [17, 74–76]. Yet, none of the studies that we reviewed clearly reported if and how competing risks were accounted for when applicable. Thus, it is difficult to assess the accuracy of the reported values of the performance indicators such as AUC, which may be biased if these issues were not correctly accounted for. The bias may be negligible if only few patients experienced the competing event in the population, but potentially important in populations where a non-negligible proportion of patients experience the competing event. A predicting tool that does not correctly account for important competing risks in the population of interest should thus be used with caution in clinical practice, since the individual probabilities derived from this tool may be biased.

For the practical use of predictive tools, it is also important to distinguish the interest of clinicians in predicting the risk of graft failure from transplantation for a new patient given its baseline characteristics, or in predicting the risk of graft failure from a given follow-up visit after transplantation, given the evolution of the patient’s condition since transplantation. Our review found 14 studies that used predictors measured during follow-up after transplantation such as serum creatinine. However, some of these studies did not clearly explain how post-transplant predictors were accounted for, although this is not so straightforward and thus prone to methodological pitfalls. The most popular approach to account for such time-dependent predictors in risk prediction models is the landmark approach [77]. This consists in (i) choosing a landmark time (i.e. a time origin), for example, 1 year after transplantation if one wants to incorporate information on predictors within the first year, (ii) developing the model using all patients that are still followed-up and at risk of graft failure at the landmark time and (iii) predicting the probability for each patient to lose the graft and/or die between the landmark time and a given time horizon. The risk prediction model developed this way can incorporate any measure of the predictors that have been assessed between transplantation and the landmark time. The advantage of this approach is that it allows the use of standard regression models for survival or competing risk data. The limitation is that, if several measurements of the predictors are available between transplantation and the landmark time, one should generally use summary statistics of these repeated post-transplant

measures in the model, such as the last measured value or the percentage decrease of eGFR between transplantation and the landmark time [44]. This was probably the approach used in most of the 14 papers using post-transplant predictors, although only a few of them have clearly reported that only patients who had not experienced the event before the landmark time were used to develop the prediction model [33, 40, 41, 44, 50, 53]. An alternative approach, which would not require summarizing repeated values of quantitative post-transplant predictors (e.g. serum creatinine), would be to use a joint modelling approach, which consists of simultaneously modelling the whole trajectory of the post-transplant predictor and the risk of graft failure [78, 79]. Joint models have recently been used to illustrate their advantages for investigating the association between eGFR trajectories and initiation of renal replacement therapy [80] or death in end-stage renal disease patients [81]. They also have been used to assess the association between serum creatinine trajectories and risk of kidney failure [82]. However, to the best of our knowledge, in spite of their potential for providing accurate dynamically updated predictions, joint models have still not been used to predict graft failure in kidney transplant recipients.

The potential limitations of our study were the restriction to English, German and French languages and exclusion of grey literature (i.e. materials and research produced by organizations outside of the traditional academic publishing and distribution channels). However, we believe that this did not impact our main findings regarding the heterogeneity of methods used.

## CONCLUSION

This review demonstrates that several prediction models for kidney graft failure have been developed and validated in the past 10 years, with substantial variability in the definition of graft failure, included predictors, and methods used for model development and validation. This makes it difficult to specifically recommend the use of a particular proposed predictive model in clinical practice, although some models have shown good predictive performance and could easily be used routinely, e.g. the KTFS for patients with functioning graft at 1 year [33]. However, we can make some strong recommendations for the development and validation of future risk prediction models in kidney transplantation, or for validation of existing ones. First, we specifically recommend clearly defining the types of events included in graft failure definition. If the definition implies competing risks, we strongly recommend adequately accounting for them in model development and in the estimation of performance indicators, as well as to systematically report the method used to account for them. We also recommend clearly stating the time origin (e.g. transplantation or 1 year post-transplantation) and the time horizon (e.g. 5 or 10 years) at which prediction performances are evaluated. If post-transplant predictors are used, repeated measurements of post-transplant predictors should be adequately accounted for in model development and validation. We indeed believe that dynamic predictions are of great potential interest for improving the monitoring of patients, but the statistical model used to compute such dynamic predictions needs to be adequately

developed and validated. We also recommend validating the predictive tool using resampling methods, and reporting prediction performance for different time horizons. External validation should also be strongly encouraged to assess the generalizability in a larger population of kidney recipients. Finally, predictive models that are developed for routine clinical practice should be implemented via a tool that is easy to use in this setting.

## SUPPLEMENTARY DATA

Supplementary data are available online at <http://ndt.oxfordjournals.org>.

## Funding

RK was supported by a PhD research grant from the French ministry of Higher Education and Research.

## CONFLICT OF INTEREST STATEMENT

None declared.

## REFERENCES

1. Lassalle M, Ayav C, Frimat L *et al*. The essential of 2012 results from the French Renal Epidemiology and Information Network (REIN) ESRD registry. *Nephrol Ther* 2015; 11: 78–87
2. Gondos A, Döhler B, Brenner H *et al*. Kidney graft survival in Europe and the United States: strikingly different long-term outcomes. *Transplantation* 2013; 95: 267–274
3. Hariharan S, Johnson CP, Bresnahan BA *et al*. Improved graft survival after renal transplantation in the United States, 1988 to 1996. *N Engl J Med* 2000; 342: 605–612
4. Meier-Kriesche H-U, Schold JD, Kaplan B. Long-term renal allograft survival: have we made significant progress or is it time to rethink our analytic and therapeutic strategies? *Am J Transplant* 2004; 4: 1289–1295
5. Lamb KE, Lodhi S, Meier-Kriesche H-U. Long-term renal allograft survival in the United States: a critical reappraisal. *Am J Transplant* 2011; 11: 450–462
6. Tangri N, Kitsios GD, Inker LA *et al*. Risk prediction models for patients with chronic kidney disease: a systematic review. *Ann Int Med* 2013; 158: 596–603
7. Al-Aly Z. Prediction of renal end points in chronic kidney disease. *Kidney Int* 2013; 83: 189–191
8. Noordzij M, Leffondré K, van Stralen KJ *et al*. When do we need competing risks methods for survival analysis in nephrology? *Nephrol Dial Transplant* 2013; 28: 2670–2677
9. Falagas ME, Pitsouni EI, Malietzis GA *et al*. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* 2008; 22: 338–342
10. Steyerberg EW, Vickers AJ, Cook NR *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; 21: 128–138
11. Tripepi G, Jager KJ, Dekker FW *et al*. Statistical methods for the assessment of prognostic biomarkers (part I): discrimination. *Nephrol Dial Transplant* 2010; 25: 1399–1401
12. Tripepi G, Jager KJ, Dekker FW *et al*. Statistical methods for the assessment of prognostic biomarkers (Part II): calibration and re-classification. *Nephrol Dial Transplant* 2010; 25: 1402–1405

13. Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika* 1991; 78: 691–692
14. Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biom J* 2008; 50: 457–479
15. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; 56: 337–344
16. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med* 2006; 25: 3474–3486
17. Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 2013; 32: 5381–5397
18. Wolbers M, Blanche P, Koller MT *et al*. Concordance for prognostic models with competing risks. *Biostatistics* 2014; 15: 526–539
19. Grønnesby JK, Borgan O. A method for checking regression models in survival analysis based on the risk score. *Lifetime Data Anal* 1996; 2: 315–328
20. Pencina MJ, D'Agostino RB, D'Agostino RB *et al*. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; 27: 157–172
21. Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011; 30: 11–21
22. Steyerberg EW, Eijkemans MJ, Harrell FE *et al*. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making* 2001; 21: 45–56
23. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999; 130: 515–524
24. Akl A, Ismail AM, Ghoneim M. Prediction of graft survival of living-donor kidney transplantation: nomograms or artificial neural networks? *Transplantation* 2008; 86: 1401–1406
25. Baskin-Bey ES, Kremers W, Nyberg SL. A recipient risk score for deceased donor renal allocation. *Am J Kidney Dis* 2007; 49: 284–293
26. Bodonyi-Kovacs G, Putheti P, Marino M *et al*. Gene expression profiling of the donor kidney at the time of transplantation predicts clinical outcomes 2 years after transplantation. *Hum Immunol* 2010; 71: 451–455
27. van Walraven C, Austin PC, Knoll G. Predicting potential survival benefit of renal transplantation in patients with chronic kidney disease. *Can Med Assoc J* 2010; 182: 666–672
28. Tiong HY, Goldfarb DA, Kattan MW *et al*. Nomograms for predicting graft function and survival in living donor kidney transplantation based on the UNOS Registry. *J Urol* 2009; 181: 1248–1255
29. Dahle DO, Eide IA, Åsberg A *et al*. Aortic stiffness in a mortality risk calculator for kidney transplant recipients. *Transplantation* 2015; 99: 1730–1737
30. De Vusser K, Lerut E, Kuypers D *et al*. The predictive value of kidney allograft baseline biopsies for long-term graft survival. *J Am Soc Nephrol* 2013; 24: 1913–1923
31. Einecke G, Reeve J, Sis B *et al*. A molecular classifier for predicting future graft loss in late kidney transplant biopsies. *J Clin Invest* 2010; 120: 1862–1872
32. Lin RS, Horn SD, Hurdle JF *et al*. Single and multiple time-point prediction models in kidney transplant outcomes. *J Biomed Inform* 2008; 41: 944–952
33. Foucher Y, Daguin P, Akl A *et al*. A clinical scoring system highly predictive of long-term kidney graft survival. *Kidney Int* 2010; 78: 1288–1294
34. Fritsche L, Hoerstrup J, Budde K *et al*. Accurate prediction of kidney allograft outcome based on creatinine course in the first 6 months posttransplant. *Transplant Proc* 2005; 37: 731–733
35. Grams ME, Kucirka LM, Hanrahan CF *et al*. Candidacy for kidney transplantation of older adults. *J Am Geriatr Soc* 2012; 60: 1–7
36. Greco R, Papalia T, Lofaro D *et al*. Decisional trees in renal transplant follow-up. *Transplant Proc* 2010; 42: 1134–1136
37. Hemke AC, Heemskerk MBA, van Diepen M *et al*. Survival prognosis after the start of a renal replacement therapy in the Netherlands: a retrospective cohort study. *BMC Nephrol* 2013; 14: 258
38. Hernández D, Rufino M, Bartolomei S *et al*. A novel prognostic index for mortality in renal transplant recipients after hospitalization. *Transplantation* 2005; 79: 337–343
39. Hernández D, Sánchez-Fructuoso A, González-Posada JM *et al*. A novel risk score for mortality in renal transplant recipients beyond the first post-transplant year. *Transplantation* 2009; 88: 803–809



40. Ho J, Wiebe C, Rush DN *et al.* Increased urinary CCL2: Cr ratio at 6 months is associated with late renal allograft loss. *Transplantation* 2013; 95: 595–602
41. Kasiske BL, Israni AK, Snyder JJ *et al.* A simple tool to predict outcomes after kidney transplant. *Am J Kidney Dis* 2010; 56: 947–960
42. Kikić Z, Herkner H, Sengölge G *et al.* Pretransplant risk stratification for early survival of renal allograft recipients. *Eur J Clin Invest* 2014; 44: 168–175
43. Krikov S, Khan A, Baird BC *et al.* Predicting kidney transplant survival using tree-based modeling. *Am Soc Artif Intern Org* 2007; 53: 592–600
44. Moore J, He X, Shabir S *et al.* Development and evaluation of a composite risk score to predict kidney transplant failure. *Am J Kidney Dis* 2011; 57: 744–751
45. Rao PS, Schaubel DE, Guidinger MK *et al.* A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. *Transplantation* 2009; 88: 231–236
46. Loupy A, Lefaucheur C, Vernerey D *et al.* Molecular microscope strategy to improve risk stratification in early antibody-mediated kidney allograft rejection. *J Am Soc Nephrol* 2014; 25: 2267–2277
47. Lowsky DJ, Ding Y, Lee DKK *et al.* A K-nearest neighbors survival probability prediction method. *Stat Med* 2013; 32: 2062–2069
48. Machnicki G, Pinsky B, Takemoto S *et al.* Predictive ability of pretransplant comorbidities to predict long-term graft loss and death. *Am J Transplant* 2009; 9: 494–505
49. Munivenkatappa RB, Schweitzer EJ, Papadimitriou JC *et al.* The Maryland aggregate pathology index: a deceased donor kidney biopsy scoring system for predicting graft failure. *Am J Transplant* 2008; 8: 2316–2324
50. Schnitzler MA, Lentine KL, Axelrod D *et al.* Use of 12-month renal function and baseline clinical factors to predict long-term graft survival: application to BENEFIT and BENEFIT-EXT trials. *Transplantation* 2012; 93: 172–181
51. Schold JD, Kaplan B, Baliga RS *et al.* The broad spectrum of quality in deceased donor kidneys. *Am J Transplant* 2005; 5: 757–765
52. Watson CJE, Johnson RJ, Birch R *et al.* A simplified donor risk index for predicting outcome after deceased donor kidney transplantation. *Transplantation* 2012; 93: 314–318
53. Shabir S, Halimi J-M, Cherukuri A *et al.* Predicting 5-year risk of kidney transplant failure: a prediction instrument using data available at 1 year posttransplantation. *Am J Kidney Dis* 2014; 63: 643–651
54. Gusukuma LW, Junior S, Tedesco H *et al.* Risk assessment score in pre-kidney transplantation: methodology and the socioeconomic characteristics importance. *J Bras Nefrol* 2014; 36: 339–351
55. Tang H, Poynton MR, Hurdle JF *et al.* Predicting three-year kidney graft survival in recipients with systemic lupus erythematosus. *Am Soc Artif Intern Org* 2011; 57: 300–309
56. Tang H, Goldfarb-Rumyantzev AS, Hunter C *et al.* Validating prediction models of kidney transplant outcome using local data. *AMIA Annu Symp Proc* 2007; 11: 1128
57. Brown TS, Elster EA, Stevens K *et al.* Bayesian modeling of pretransplant variables accurately predicts kidney graft survival. *Am J Nephrol* 2012; 36: 561–569
58. Szabó RP, Asztalos L, Szabó L *et al.* Validation of a prognostic function for renal transplant recipients in Hungary. *J Nephrol* 2011; 24: 619–624
59. Gourishankar S, Grebe SO, Mueller TF. Prediction of kidney graft failure using clinical scoring tools. *Clin Transplant* 2013; 27: 517–522
60. Bang K, Lee HK, Huh W *et al.* Assessment of deceased donor kidneys using a donor scoring system. *Yonsei Med J* 2010; 51: 870
61. Moore J, He X, Liu X *et al.* Mortality prediction after kidney transplantation: comparative clinical use of 7 comorbidity indices. *Exp Clin Transplant* 2011; 9: 32–41
62. Jassal SV, Schaubel DE, Fenton SSA. Baseline comorbidity in kidney transplant recipients: a comparison of comorbidity indices. *Am J Kidney Dis* 2005; 46: 136–142
63. Tang H, Hurdle JF, Poynton M *et al.* Validating prediction models of kidney transplant outcome using single center data. *Am Soc Artif Intern Org* 2011; 57: 206–212
64. Foucher Y, Meurette A, Daguin P *et al.* A personalized follow-up of kidney transplant recipients using video conferencing based on a 1-year scoring system predictive of long term graft failure (TELEGRAFT study): protocol for a randomized controlled trial. *BMC Nephrol* 2015; 16: 6
65. Lorent M, Giral M, Pascual M *et al.* Mortality prediction after the first year of kidney transplantation: an observational study on two European cohorts. *PLoS One* 2016; 11: e0155278
66. Porzelius C, Schumacher M, Binder H. A general, prediction error-based criterion for selecting model complexity for high-dimensional survival models. *Stat Med* 2010; 29: 830–838
67. Porzelius C, Schumacher M, Binder H. The benefit of data-based model complexity selection via prediction error curves in time-to-event data. *Comput Stat* 2011; 26: 293–302
68. Collins GS, Mallett S, Omar O *et al.* Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med* 2011; 9: 103
69. 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. *J Clin Epidemiol* 2013; 66: 268–277
70. Mallett S, Royston P, Waters R *et al.* Reporting performance of prognostic models in cancer: a review. *BMC Med* 2010; 8: 21
71. Moons KGM, Altman DG, Reitsma JB *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Int Med* 2015; 162: W1–W73
72. Collins GS, Reitsma JB, Altman DG *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Int Med* 2015; 162: 55–63
73. Andersen PK, Geskus RB, de Witte T *et al.* Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol* 2012; 41: 861–870
74. Wolbers M, Koller MT, Witteman JCM *et al.* Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology* 2009; 20: 555–561
75. Gerds TA, Scheike TH, Andersen PK. Absolute risk regression for competing risks: interpretation, link functions, and prediction. *Stat Med* 2012; 31: 3921–3930
76. Rosthøj S, Andersen PK, Abildstrom SZ. SAS macros for estimation of the cumulative incidence functions based on a Cox regression model for competing risks survival data. *Comp Methods Programs Biomed* 2004; 74: 69–75
77. Van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scand J Stat* 2007; 34: 70–85
78. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; 67: 819–829
79. Proust-Lima C, Taylor JMG. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* 2009; 10: 535–549
80. Asar Ö, Ritchie J, Kalra PA *et al.* Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *Int J Epidemiol* 2015; 44: 334–344
81. McCrink LM, Marshall AH, Cairns KJ. Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients. *Int Stat Rev* 2013; 81: 249–269
82. Fournier M-C, Foucher Y, Blanche P *et al.* A joint model for longitudinal and time-to-event data to better assess the specific role of donor and recipient factors on long-term kidney transplantation outcomes. *Eur J Epidemiol* 2016; 31: 469–479

Received for publication: 1.6.2016; Accepted in revised form: 3.10.2016