

2022

# Using information available at the time of donor offer to predict kidney transplant survival outcomes: a systematic review of prediction models

Riley, S

<https://pearl.plymouth.ac.uk/handle/10026.1/21692>

---

10.3389/ti.2022.10397

Transplant International

Frontiers

---

*All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.*



# Using Information Available at the Time of Donor Offer to Predict Kidney Transplant Survival Outcomes: A Systematic Review of Prediction Models

Stephanie Riley<sup>1</sup>, Qing Zhang<sup>1</sup>, Wai-Yee Tse<sup>2</sup>, Andrew Connor<sup>2</sup> and Yinghui Wei<sup>1\*</sup>

<sup>1</sup>Centre for Mathematical Sciences, School of Engineering, Computing and Mathematics, University of Plymouth, Plymouth, United Kingdom, <sup>2</sup>Department of Renal Medicine, South West Transplant Centre, University Hospitals Plymouth NHS Trust, Plymouth, United Kingdom

Statistical models that can predict graft and patient survival outcomes following kidney transplantation could be of great clinical utility. We sought to appraise existing clinical prediction models for kidney transplant survival outcomes that could guide kidney donor acceptance decision-making. We searched for clinical prediction models for survival outcomes in adult recipients with single kidney-only transplants. Models that require information anticipated to become available only after the time of transplantation were excluded as, by that time, the kidney donor acceptance decision would have already been made. The outcomes of interest were all-cause and death-censored graft failure, and death. We summarised the methodological characteristics of the prediction models, predictive performance and risk of bias. We retrieved 4,026 citations from which 23 articles describing 74 models met the inclusion criteria. Discrimination was moderate for all-cause graft failure (C-statistic: 0.570–0.652; Harrell's C: 0.580–0.660; AUC: 0.530–0.742), death-censored graft failure (C-statistic: 0.540–0.660; Harrell's C: 0.590–0.700; AUC: 0.450–0.810) and death (C-statistic: 0.637–0.770; Harrell's C: 0.570–0.735). Calibration was seldom reported. Risk of bias was high in 49 of the 74 models, primarily due to methods for handling missing data. The currently available prediction models using pre-transplantation information show moderate discrimination and varied calibration. Further model development is needed to improve predictions for the purpose of clinical decision-making.

## OPEN ACCESS

### \*Correspondence:

Yinghui Wei  
 yinghui.wei@plymouth.ac.uk

Received: 31 January 2022

Accepted: 04 May 2022

Published: 23 June 2022

### Citation:

Riley S, Zhang Q, Tse W-Y, Connor A and Wei Y (2022) Using Information Available at the Time of Donor Offer to Predict Kidney Transplant Survival Outcomes: A Systematic Review of Prediction Models. *Transpl Int* 35:10397. doi: 10.3389/ti.2022.10397

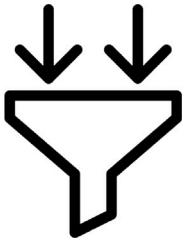
**Systematic Review Registration:** <https://osf.io/c3ehp/l>.

**Keywords:** kidney transplantation, prognosis, systematic review, clinical prediction model, decision-making, kidney transplant outcomes, prediction tool

**Abbreviations:** AUC, area under receiver operating characteristic curve; CHARMS, Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies; DGF, delayed graft function; EPTS, Estimated Post Transplant Survival; ESKD, end-stage kidney disease; KDRI, Kidney Donor Risk Index; LKDPI, Living Kidney Donor Profile Index; MAPI, Maryland Aggregate Pathology Index; NHSBT, National Health Service Blood and Transplant; PITHIA, PreImplantation Trial of Histopathology In renal Allografts; PRISMA, Preferred Reporting Items for Systematic reviews and Meta-Analyses; PROBAST, Prediction Model Risk of Bias Assessment Tool; RoB, risk of bias; UK KDRI, United Kingdom Kidney Donor Risk Index.

## Using information available at the time of donor offer to predict kidney transplant survival outcomes: a systematic review of prediction models

4,026 articles identified from literature databases.



23 articles describing 28 development and validation models, and 46 validation-only models are included.



### Key findings

No existing models accounted for semi-competing risks. Higher values of discrimination in models for patient survival. Most models deemed at a high risk of bias. Moderate discrimination reported. Calibration was seldom assessed.

### Time-to-event outcomes

All-cause graft failure. Death-censored graft failure. Patient survival.



### Discrimination

**All-cause graft failure**  
C-statistic: 0.570-0.652; Harrell's C: 0.580-0.660; AUC: 0.530-0.742.  
**Death-censored graft failure**  
C-statistic: 0.540-0.660; Harrell's C: 0.590-0.700; AUC: 0.450-0.810.  
**Patient survival**  
C-statistic: 0.637-0.770; Harrell's C: 0.570-0.735.

### Risk of bias



Risk of Bias ■ High ■ Unclear ■ Low

### Future development

Semi-competing risks models. Handling of missing data. Assessment of calibration. Sample size calculations.

Stephanie Riley, Qing Zhang, Wai-Yee Tse, Andrew Connor, Yinghui Wei. *Transpl Int* 2022  
doi: 10.3389/ti.2022.10397



GRAPHICAL ABSTRACT |

## INTRODUCTION

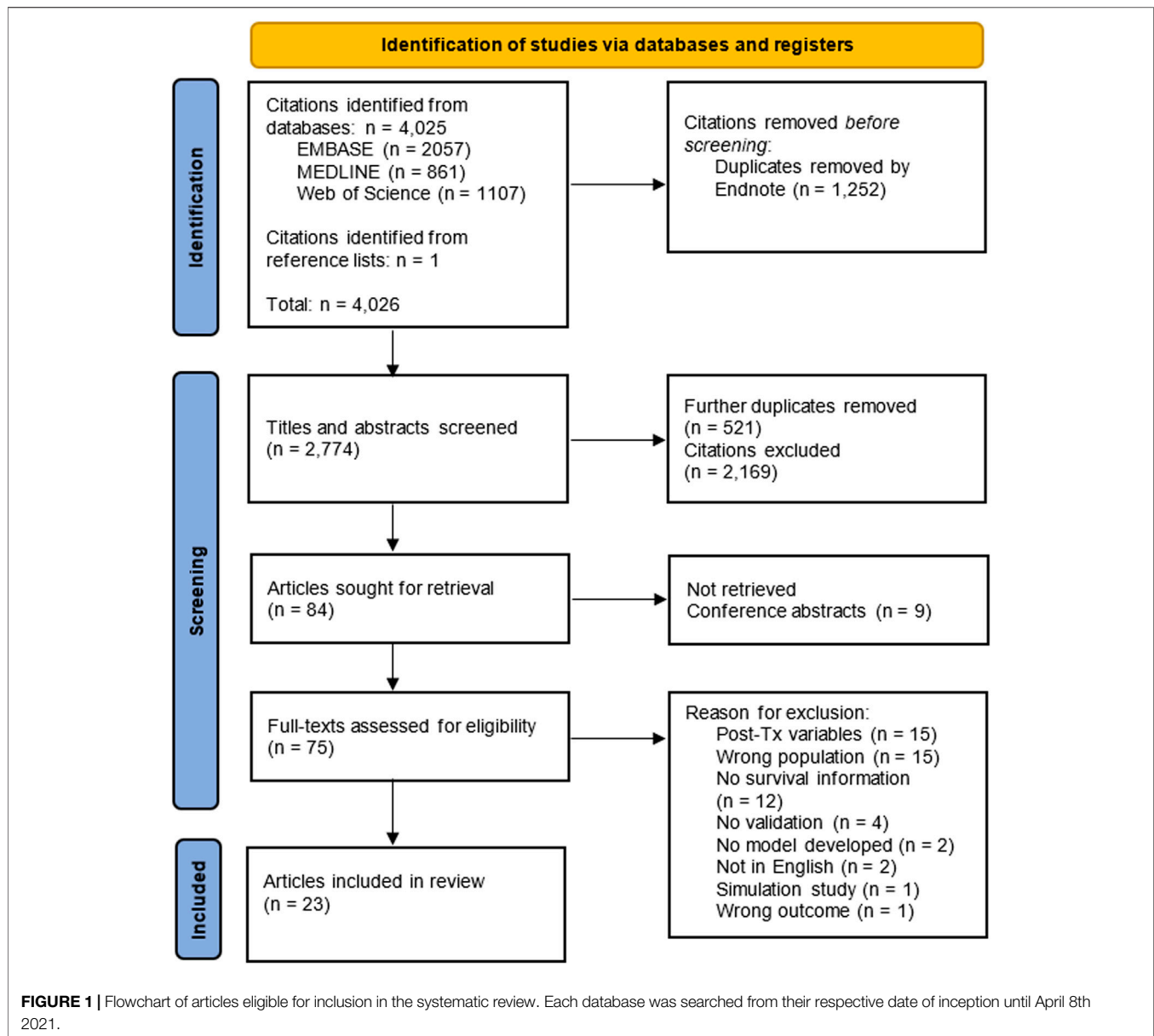
End-stage kidney disease (ESKD) is the most advanced stage of chronic kidney disease. Kidney transplantation is the optimal treatment for many patients with ESKD. In the UK, approximately 3,000 kidney transplants are performed every year, but the number of patients waiting for a transplant is around 5,000 (1). The success, in terms of efficacy and longevity, of an individual transplant will be influenced by a host of factors, some of which can be determined prior to transplantation. A balance must be struck to ensure maximal organ utilisation without compromising transplant outcomes. This is further complicated by the fact that “one size does not fit all”—the definition of a successful transplant will vary depending on the recipient and their clinical scenario. As such, every potential kidney transplant must be carefully considered in the context of the donor and recipient details.

In the UK donor organs are offered through a national donation system, which utilises an algorithm to balance patient priority and the intent to match immunological and additional parameters. The donor offers are reviewed by clinicians acting on behalf of the recipient and a prompt decision must be made to accept or reject each offer. Whether or not to accept a transplant offer remains a challenging clinical decision. Clinical prediction models that utilise information which would commonly be available to the clinician at the time of the donor kidney offer may help to inform the decision-making process.

The anticipated longevity of a kidney transplant is, of course, an important consideration for a clinician faced with the kidney donor acceptance decision. However, given that donor kidneys are a scarce resource and potential recipients must therefore sit on waiting lists, it is often appropriate to balance the anticipated longevity against the alternative of remaining on dialysis. As such, models that can predict graft survival outcomes would be of great clinical utility.

Prediction models have previously been developed for kidney transplant survival outcomes with the aim of advising clinicians at the time of the offer of a donor kidney. The number of articles related to clinical prediction models for kidney transplant survival outcomes is increasing year on year, suggesting a recognition of the clinical interest. The Kidney Donor Risk Index (KDRI) (2), Estimated Post Transplant Score (EPTS), Maryland Aggregate Pathology Index (MAPI) (3) and Living Kidney Donor Profile Index (LKDPI) (4) are commonly reported risk indices. The KDRI and EPTS are part of the kidney allocation system in the US.

The aforementioned risk indices were developed in the US population. A similar index has been produced in the UK (UK KDRI) (5), though is not widely used in practice. In the UK kidney allocation system NHS Blood and Transplant (NHSBT) use their own risk indices for donors and recipients (6). This is to help ensure that the pool of donor kidneys is utilised to best effect. Through this system, for example, younger recipients will typically receive offers of kidneys from younger donors (in order to optimise the chances of transplant longevity) whilst a greater tolerance of less favourable immunological matches will be accepted for older recipients (in order to maximise offers for a



cohort in whom immunological matching is of slightly less importance).

We identified two systematic reviews exploring existing prediction models for kidney transplantation. Kaboré et al. (7) reviewed prediction models for graft outcomes published between 2005 and 2015, while Senanayake et al. (8) reviewed machine learning methods to predict graft failure, delayed graft function (DGF) and acute graft rejection. Since only machine learning models were eligible, their review excluded articles that used the Cox model, which is the model most used for time-to-event analyses.

Both reviews allowed the inclusion of predictors that only become available after transplantation, such as whether patients experienced DGF. To our knowledge this is the first review to

focus only on models that could aid clinical decision-making at the time of the donor offer.

In this systematic review we aim to identify, appraise and summarise existing clinical prediction models for kidney transplant survival outcomes. Only prediction models that use information available at the time of the single kidney-only offer were included, allowing us to focus on models with the most clinical utility.

## METHODS

We prospectively developed a protocol which is publicly available from OSF (9). The findings of this review are reported in

**TABLE 1 |** Summary of sample size used in models for each outcome and model type.

	Number of models	Range	Median	Mean	SD
All-cause graft failure					
Development and validation	11	785–156,069	39,108	41,127	48,719
Validation only	15	416–69,994	5,042	8,641	17,141
Death-censored graft failure					
Development and validation	5	259–10,086	6,662	5,586	4,811
Validation only	19	56–6,405	1,299	3,017	2,909
Patient survival					
Development and validation	11	837–120,818	47,535	41,319	38,270
Validation only	11	935–5,042	4,983	3,323	2,007

Two models with other outcomes which do not fall into the above definitions have sample size of 20,085 and 2,734, respectively.

accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement (10).

### Eligibility Criteria

We included studies with adult recipients (aged 18 years or older) of single, kidney-only transplants. No restrictions were placed on donor type.

No limit was set on publication date. Only full texts published in English were eligible. Conference abstracts without full text were excluded from review.

The outcomes of interest were one or both of the following outcomes, time to graft failure and time to death at any time point following kidney transplantation. Models that did not account for time-to-event information were excluded.

We considered prediction models that make use of information available at the time of a donor kidney offer to inform the acceptance decision. Prediction models developed using predictors that only become available after transplantation were not included, as the decision would have been made by that time.

We included studies which were developed and validated for the outcomes of interest, and validation-only studies which validated existing models developed from independent cohorts. Any measure of predictive performance, such as calibration or discrimination, that was reported alongside a model was considered a form of validation. Validation-only model refers to the case where the current study validates an existing model.

### Information Sources and Search Strategy

Electronic databases Embase, MEDLINE and Web of Science were searched from their respective dates of inception up to April 8th 2021. The search strategy is presented in **Supplementary Table S1**.

All citations from the search results were exported to Endnote, where duplicates were automatically removed from review. Titles and abstracts of all records were independently screened against the above eligibility criteria by two reviewers (SR and QZ) and managed through Rayyan (11). A third reviewer (YW) also independently screened 10% of the titles and abstracts. Two reviewers then independently reviewed full-text reports to assess eligibility (SR and QZ). Any discrepancies were resolved through discussion.

### Data Extraction

Data were extracted from eligible articles according to the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist (12). The full list of data extracted are given in **Supplementary Table S2**. Data were extracted independently by two reviewers (SR and YW) and any discrepancies were resolved through discussion.

### Risk of Bias

We assessed the risk of bias (RoB) in individual models using the Prediction Model Risk of Bias Assessment Tool (PROBAST) (13). Two reviewers (SR and YW) independently determined the RoB of each model and any disagreements were resolved by discussion.

### Outcomes

#### All-Cause Graft Failure

All-cause graft failure, as a composite outcome, is defined as the earliest time to graft failure or death.

#### Death-Censored Graft Failure

Death-censored graft failure considers the time until graft failure, but patients are censored at the time of death. Graft failure and death are semi-competing events (14). Semi-competing events arise when a terminal event precludes a non-terminal event, but not vice-versa (15).

#### Death

This measures time to recipient death, of any cause, as the outcome of interest.

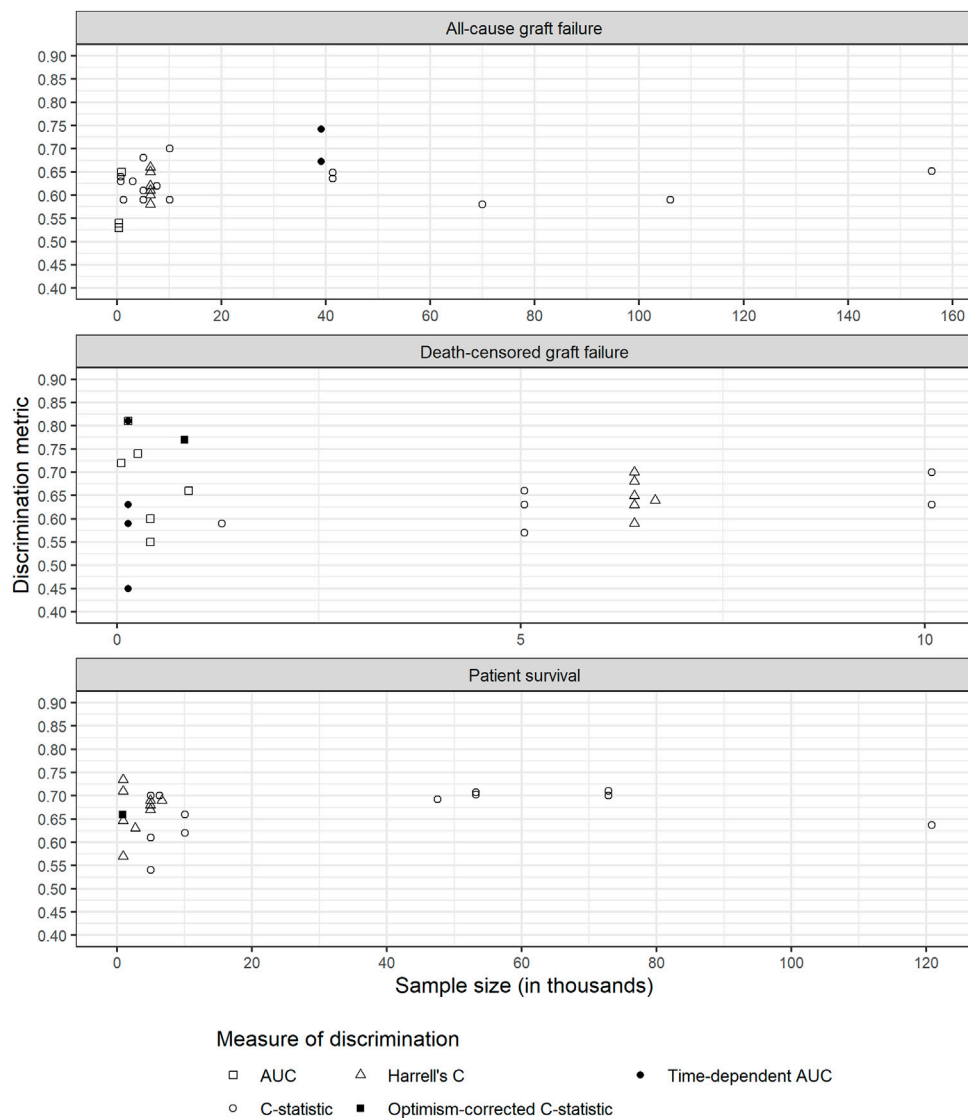
### Analysis

#### Study Characteristics

We summarised the year of publication, geographical location, model type, and model being validated. We explored the discrimination measures by sample size and predictor type (donor, recipient, transplant, or combination of these). For each outcome, we summarised the type of predictors, modelling methods, and methods for handling missing data.

#### Measures of Model Performance

Model performance was evaluated by calibration and discrimination. Calibration assesses the agreement between



**FIGURE 2 |** Discrimination metrics against sample size for each outcome. AUC: area under receiver operating characteristic curve; C-statistic: concordance statistic; Harrell's C: adapts the C-statistic to account for censoring; Optimism-corrected C-statistic: measures the C-statistic while accounting for optimism in model performance; Time-dependent AUC: a measure of the AUC at specified timepoints since time origin.

observed and predicted risk and is often reported through a calibration plot. Discrimination measures a model's ability to separate recipients who will experience the outcome event versus those who will not. It is often measured using Harrell's C statistic, area under receiver operating characteristic curve (AUC) or time-dependent AUC, which account for the censoring of the time-to-event outcome. When a model is developed and internally validated in the same dataset it understandably performs well. Methods to correct for this optimism can be administered using bootstrapping, and resulting measures are referred to as optimism-corrected (16). Where studies did not explicitly state that the C-statistic was adapted for censoring, we elected to report the terminology used in the original articles.

## RESULTS

We retrieved 4,025 citations from three databases through our search and identified one record related to one of the conference abstracts we screened. After the initial screening of titles and abstracts, 75 articles were retrieved for full-text review. Of these records, 23 articles describing 74 models met the inclusion criteria (3–5, 17–36) (Figure 1).

### Characteristics of Included Studies

Of the 74 eligible models, 28 developed and validated a clinical prediction model for our outcomes of interest. The remaining 46 models validated the performance of an existing model in an independent cohort. Articles were published between 2005 and

**TABLE 2** | Summary of commonly reported risk indices for predicting kidney transplant survival outcomes.

Model	Donor	Recipient	Transplant organ/process	Histopathology	Validation studies
EPTS	NA	Age; Diabetes status; Prior solid organ transplants; Time on dialysis	NA	NA	(22) (23) (28)
LKDPI	Age; eGFR; BMI; Ethnicity; History of cigarette use; Systolic blood pressure; Sex; Weight	Sex (compared to donor); Weight (relative to donor weight)	Number of HLA mismatch at HLA-B and HLA-DR; ABO compatibility	NA	(30)
MAPI	NA	NA	NA	Arteriolar hyalinosis; Glomerulosclerosis; Periglomerular fibrosis; Scar Wall-to-lumen ratio interlobular arteries	(25) (29)
UK KDRI	Age; Days in hospital; History of hypertension; Use of adrenaline; Weight	NA	NA	NA	(21)
US KDRI	Age; Cause of death; DCD; Diabetes status; Ethnicity; HCV status; Height History of hypertension Serum creatinine Weight	NA	Cold ischaemic time; Double kidney transplant; En-bloc transplant; Number of HLA mismatch at HLA-B and HLA-DR	NA	(20) (21) (23) (25) (4) (30) (5) (35)

BMI, body mass index; DCD, deceased cardiac donor; eGFR, estimated glomerular filtration rate; HCV, hepatitis C virus; HLA, human leukocyte antigen.

2020; fifteen of the twenty-three articles (65.22%) were published after 2015. Twelve articles used data of recipients from the United States, four from mainland Europe, three from Canada, two from Australia and New Zealand, and one each from the United Kingdom and Thailand. Characteristics of included studies for each model are available in **Supplementary Tables S3–S6**.

In the 28 development and validation models, 27 used the Cox proportional hazards model, while one (17) used a survival random forest. Only eight of the Cox models assessed the proportional hazards assumption.

There was considerable variability in sample sizes used for models (**Table 1**; **Figure 2**). In general, models performing validation alone tended to have smaller sample size. Models with smaller sample sizes did not have noticeably poorer discrimination for any of the outcomes (**Figure 2**).

We considered three types of predictors, donor characteristics, recipient characteristics and transplant process. We found no clear evidence that the type of predictors was associated with better discrimination for any

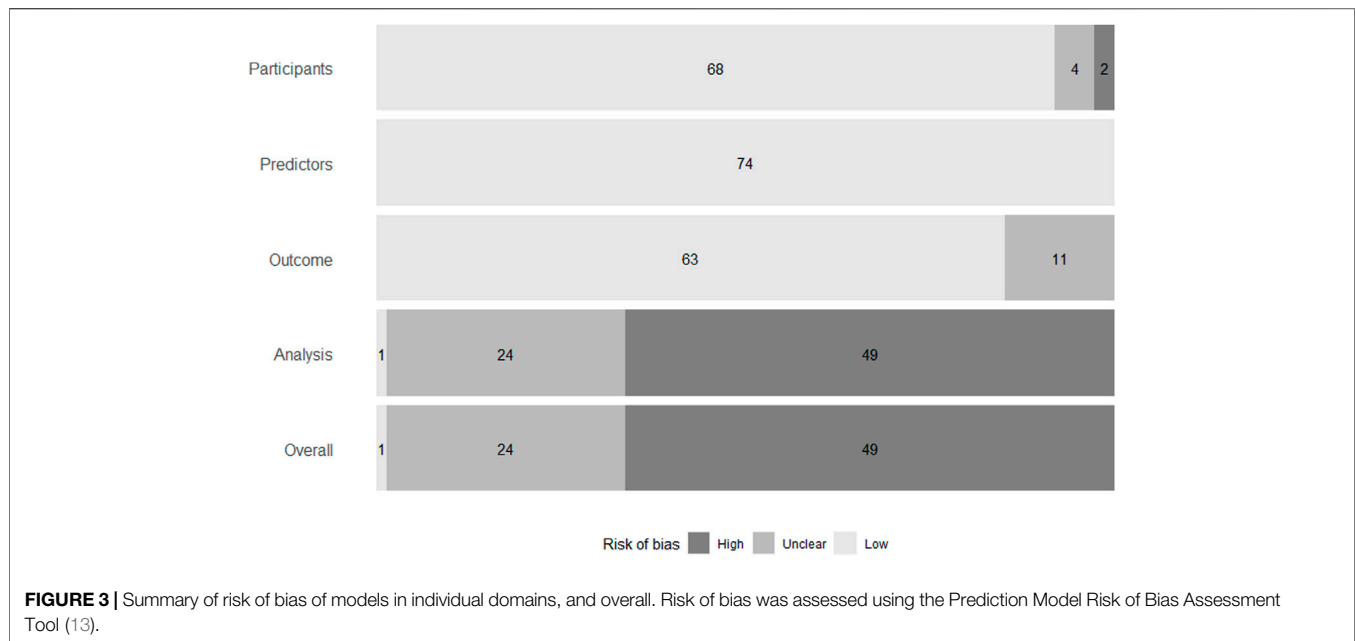
outcome (**Supplementary Figures S1–S3**). Clayton et al. (21) validated the US and UK KDRI, while also adjusting for recipient characteristics and transplant process. Those with higher values of discrimination (models 9 and 12) were adjusted for other donor, recipient and transplant related predictors. This was also observed by Molnar et al. (28). However, this increase could simply be due to having more variables in the model.

Nine of the 28 development and validation models (4, 24, 27, 28) were available in the form of an online tool or calculator. One of the models (32) was presented in the form of a nomogram and another (17) as a contour plot of survival probability.

Commonly validated risk indices, as described in **Table 2**, included the KDRI, EPTS, UK KDRI, LKDPI, and MAPI. Other models validated included those developed by Kasiske et al. (27), Nyberg et al. (37) and Remuzzi et al. (38).

## Risk of Bias

The overall RoB was high in 49 of the 74 models, unclear in 24, and low in only one (**Figure 3**). Of those that were considered a



high RoB overall, all of them were at a high RoB in the analysis domain.

Sample size was reported for all models. However, the number of events were not reported for 37 of the models, therefore it was unclear whether there were a reasonable number of participants with the outcome.

Missing data were not discussed for 12 models. For models that did discuss missing data, 20 performed their analysis based only on those patients that did not have any missing data. This is called a complete-case analysis. All included models reported some measure of discrimination, but calibration was only reported for 13 models. Some models reported the C-statistic but did not discuss whether they had adapted it to account for censoring. This also contributed to a lack of clarity on the suitability of performance measures.

Twenty-two of the 28 development and validation models avoided univariable selection, reducing the possibility of bias in the analysis domain. Sixteen models did not account for overfitting or optimism, rendering them a high RoB.

## Development and Validation Models

### All-Cause Graft Failure

All-cause graft failure was reported in 11 of the 28 development and validation models. Summary data for each model with this outcome are shown in **Supplementary Table S3**.

In eight models, only deceased donor information was used. Three models utilised a combination of living and deceased donors.

Four models performed a complete-case analysis and two models used multiple imputation (39) to handle missing data. Three models imputed values based on mean or median, and two models assigned missing values to a missing category.

All models assessed discrimination. Discrimination measures reported included nine C-statistics (0.59–0.652) and two time-dependent AUC at 20 years (0.673 and 0.752) (**Figure 4**). In four

models that also assessed calibration, two did so using a calibration plot and the remaining two reported the calibration slope (1.04 each).

### Death-Censored Graft Failure

In one of the five models for death-censored graft failure the eligible population was deceased donor kidney recipients whilst in one model it was living donor recipients (**Supplementary Table S4**). Three models utilised a combination of both living and deceased donors.

For death-censored graft failure, four models used multiple imputation and one failed to report any methods for handling of missing data.

All models included at least one measure of discrimination and four evaluated the calibration. Discrimination measures reported included Harrell's C (0.69), AUC (0.74), C-statistic (0.59, 0.63), and optimism-corrected C-statistic (0.66) (**Figure 5**). Three models graphically assessed calibration and one used the Hosmer-Lemeshow test.

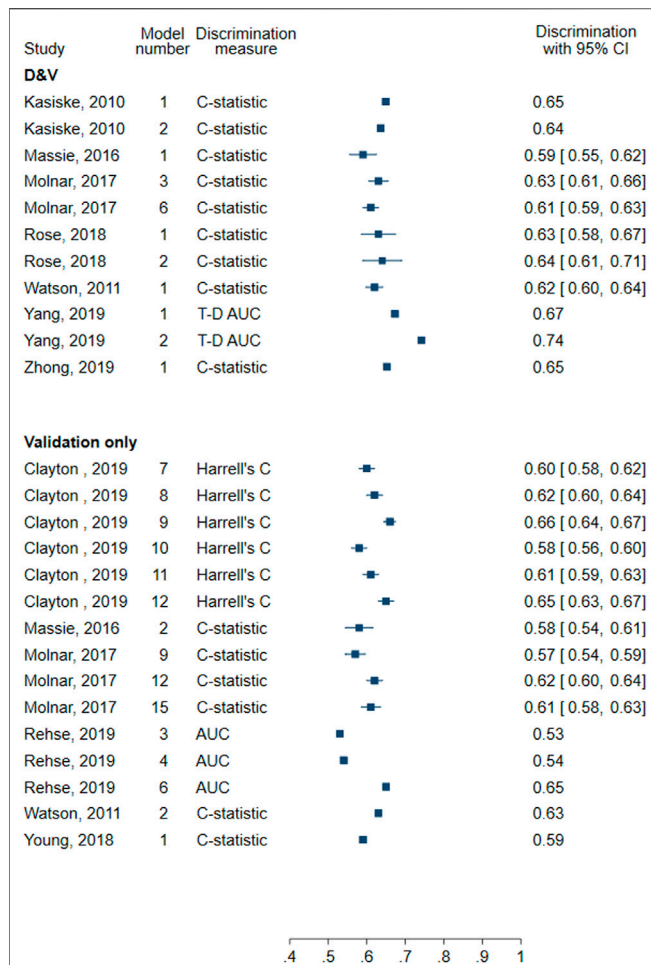
### Patient Survival

Only one of the 11 models utilised living donors, whilst six used deceased donor transplant data and four considered a combination of living and deceased donors (**Supplementary Table S5**).

Eight models handled missing data using multiple imputation and one used single imputation. One model undertook a complete-case analysis and handling of missing data was not reported for one model.

The C-statistic was the most used measure of discrimination (9 models) with reported values between 0.637 and 0.71 (**Figure 6**). Other measures included Harrell's C (0.64) and optimism-corrected C-statistic (0.77). Calibration was also assessed in four models, three of which presented a calibration plot while the other performed the Hosmer-Lemeshow test.





**FIGURE 4 |** Forest plot of discrimination in models for all-cause graft failure. D&V: Development and validation; AUC: area under receiver operating characteristic curve; C-statistic: concordance statistic; Harrell's C: adapts the C-statistic to account for censoring; T-D AUC: Time-dependent AUC, a measure of the AUC at specified timepoints since time origin.

### Validation-Only Models All-Cause Graft Failure

All-cause graft failure was reported in 15 of the 46 validation-only models (**Supplementary Table S3**).

In 10 models, only deceased donor information was used and only living donor data in two models. The remaining three models utilised a combination of living and deceased donors.

Seven conducted a complete-case analysis. Three models used multiple imputation to handle missing data. Two models imputed values based on mean or median. For two models it was unclear how missing data were handled and one model did not discuss missing data.

All models assessed discrimination. Seven assessed discrimination using Harrell's C (0.55–0.66) and six reported the C-statistic (0.57–0.63) (**Figure 4**). Two models used the AUC (0.53–0.65). No models assessed calibration.

### Death-Censored Graft Failure

In 14 of the 19 validation-only models the eligible population was deceased donor kidney recipients whilst in two models it was living donor recipients (**Supplementary Table S4**). Three models utilised a combination of both living and deceased donors.

For death-censored graft failure, six models did a complete-case analysis, three models used multiple imputation and one used median imputation. For three models the methods for handling missing data were unclear, and six did not discuss missing data.

All models evaluated the discrimination, but none assessed the calibration. Four models reported the C-statistic (0.54–0.66) and six reported Harrell's C (0.59–0.70). Five models assessed AUC (0.55–0.81), and four assessed time-dependent AUC evaluated 2 years following transplantation (0.45–0.81) (**Figure 5**).

### Patient Survival

Eight out of 11 models used data from deceased donor transplant recipients and three used a combination of living and deceased donors (**Supplementary Table S5**).

Seven models handled missing data using multiple imputation and one conducted a complete-case analysis. Three models failed to discuss missing data.

Eight models assessed discrimination using Harrell's C (0.57–0.735) and three using the C-statistic (0.66–0.70) (**Figure 6**). Calibration was not assessed in any of the validation-only models for patient survival.

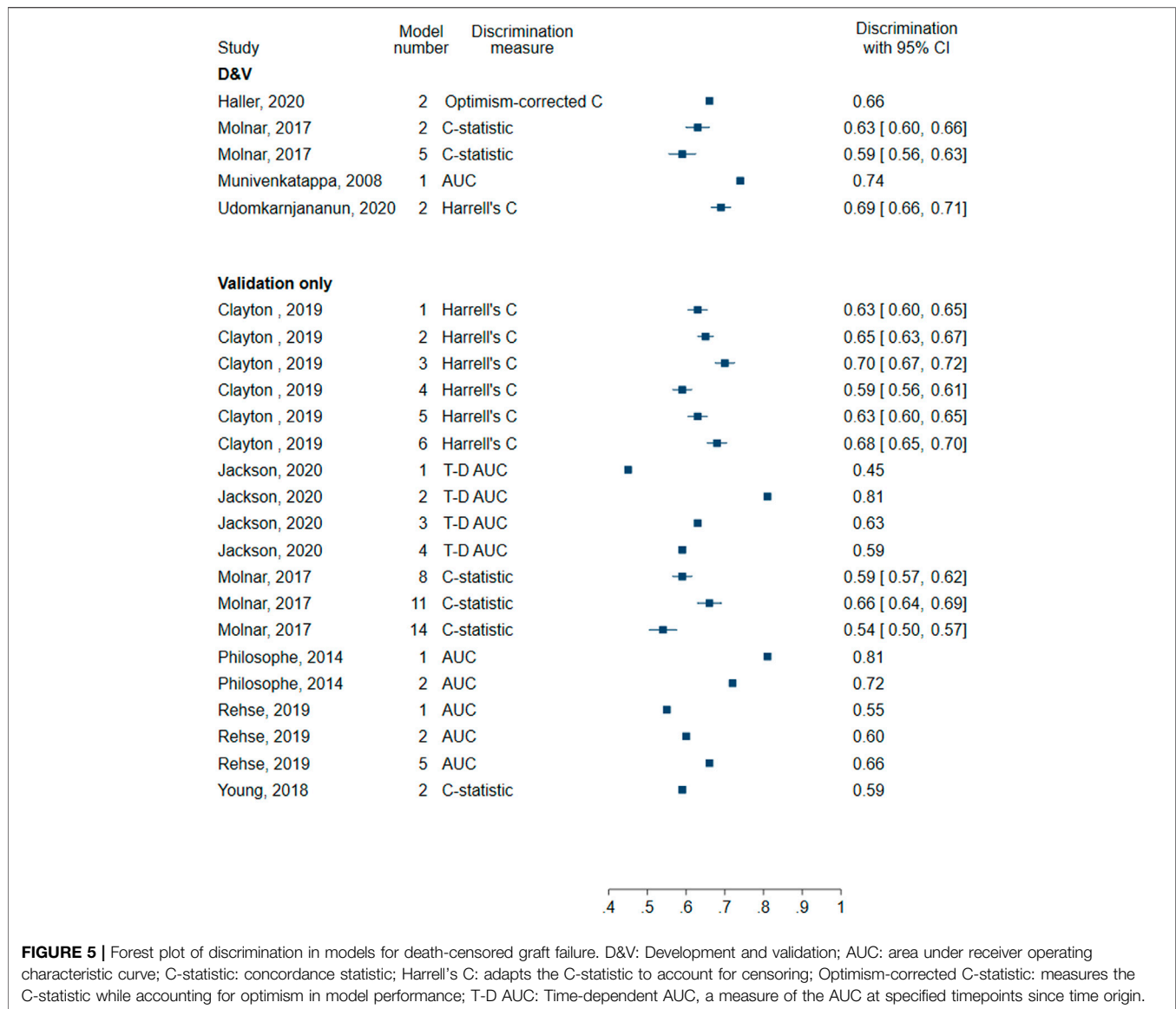
## DISCUSSION

### Principal Findings

Our review focussed on prediction models to inform the kidney donor acceptance decision. Thus, we only included models which used pre-transplantation information. The MAPI (3), for example, utilises histopathological data from pre-transplantation donor kidney biopsies to predict graft survival. However, clinicians in the UK would not typically have access to biopsy results at the time of offer this model has limited utility. The PreImplantation Trial of Histopathology In renal Allografts (PITHIA) (39) is ongoing and assesses whether pre-implantation biopsy analyses improve graft function. As such, there may be scope for the MAPI to be clinically useful.

Discrimination was well reported overall unlike calibration. Existing reviews also observed that calibration is poorly reported (40–43). Without both measures of performance, it is difficult to determine the predictive capability.

Twenty of the 28 development and validation models were developed in the US population, though the discrimination of these models remained similar in external validation in other countries. Overall performance of both development and validation-only models was most determined by measures of discrimination, such as the C-statistic, Harrell's C, or



AUC, which ranged between 0.59 and 0.77 for development and validation models, and 0.45 and 0.81 for validation-only models.

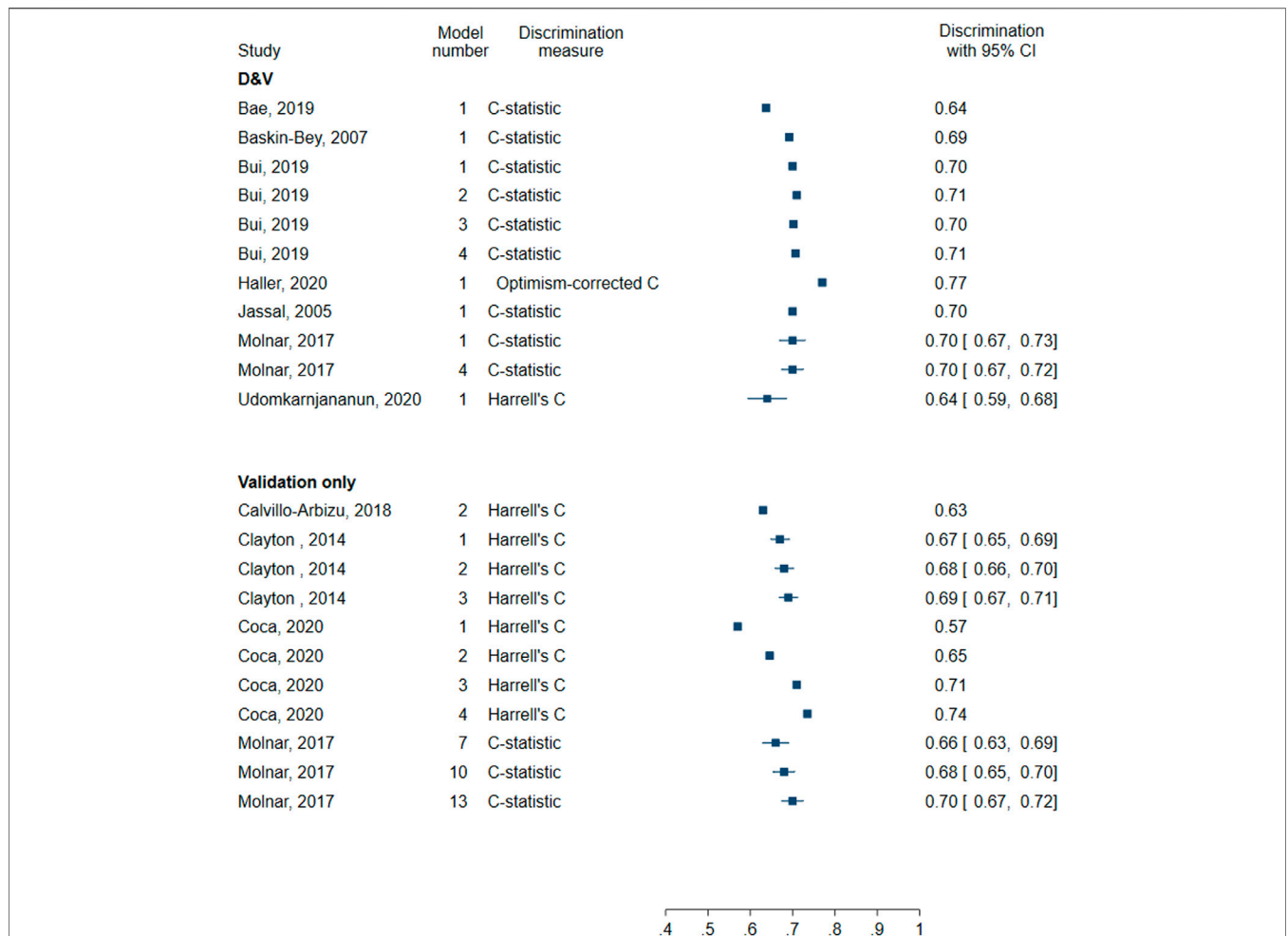
All included models considered the censoring of the time-to-event data using either Cox models or survival random forest. However, models for death-censored graft failure should have ideally considered the semi-competing events graft failure and death. Calvillo-Arbizu et al. (20) noted that death with a functioning graft is a competing event for graft failure but used this as part of the exclusion criteria. Methods such as Fine and Gray (44) and multistate models (45) can be used to account for semi-competing events without discarding the data.

The model by Haller et al. (24), reported optimism-corrected C-statistic 0.77 and showed good calibration. It predicted the survival of recipients of a living donor kidney

using a combination of donor, recipient and transplant factors as predictors. However, it has not been externally validated, so its generalisability to other populations is not known.

The LKDPI by Massie et al. (4) predicted all-cause graft failure with C-statistic 0.59 (95% CI: 0.55–0.62). In external validation studies (30) conducted in Germany, the model continued to show moderate to poor discrimination. The development and validation model by Molnar et al. (28) predicting death-censored graft failure had similarly poor discrimination reporting a C-statistic 0.59 (95% CI: 0.56–0.63) but showed good calibration. No other article externally validated this model.

For deceased donors the model by Yang et al (34) reported time-dependent AUC equal to 0.742 for graft survival, but has not yet been validated externally.



**FIGURE 6** | Forest plot of discrimination in models for patient survival. D&V: Development and validation; C-statistic: concordance statistic; Harrell's C: adapts the C-statistic to account for censoring; Optimism-corrected C-statistic: measures the C-statistic while accounting for optimism in model performance.

The UK KDRI (5) for predicting transplant survival had moderate discrimination with a C-statistic of 0.62. External validation in Australia and New Zealand (21) reported Harrell's C equal to 0.59 (95% CI: 0.56–0.61) and 0.58 (95% CI: 0.56–0.60) for predicting death-censored graft failure and all-cause graft failure, respectively.

Overall RoB was high in 49 out of 74 of the included models, largely due to the analysis methods. One such aspect was the handling of missing data. Twelve models did not discuss missing data at all, and twenty models handled missing data using a complete-case analysis. This analysis approach can lead to biased results (39) due to reduced sample size and increased risk of overfitting. Other methods, such as multiple imputation, are preferred over a complete-case analysis (46).

Overall, there was no clear indication whether the type of predictors affected discrimination (**Supplementary Figures S1–S3**). However, in individual articles we saw that models developed using combinations of type of predictors, rather than donor-only, showed better discrimination. Models with

a small sample size relative to the number of predictors are more susceptible to overfitting (47), which can result in poorer predictive performance.

Sufficient sample size was rarely considered and was one contributing factor to models being deemed at a high RoB. Methods for calculating the effective sample size for the development of a prediction model for time-to-event outcomes have been proposed by Riley et al. (48). A sample size calculation is standard practice in clinical trials, and we believe this practice should cross over into prediction modelling.

### Strengths

To our knowledge this is the first review focusing on prediction models that only use information known prior to transplantation as predictors, and does not restrict to either regression or machine learning methods. Furthermore, we reviewed all articles published from the date of inception of each database, allowing us to maximise the number of articles included.

## Limitations

Our review was restricted to articles published in English. We focussed on models that would be of practical use at the time of an offer of a donor kidney. As such notable models including those by Loupy et al. (49) and Foucher et al. (50), which include post-transplantation information, were not eligible for our review. Based on the existing prediction models, we cannot conclude which methods work better than the other. This opens the opportunity for evaluation, application and testing a range of appropriate methods in future research.

## CONCLUSION

Development of clinical prediction models to inform organ acceptance decision-making should be driven by the clinical utility of such models. The currently available prediction models using pre-transplantation information provide moderate discrimination and varied calibration for patient and graft survival. Sample size calculations, handling of missing data and assessment of calibration are required, alongside better reporting of methods, to increase the quality of the studies. Opportunities to improve predictive performance include the identification of further important predictors and advancement of the development models by acknowledging the complex data such as semi-competing risks between graft failure and death. Until the predictive tools have the desirable performance, they have limited utility in clinical decision-making.

## DATA AVAILABILITY STATEMENT

Data extracted, risk of bias assessment and code are available from Github: <https://github.com/Yinghui-Wei-team/methods-review-kidney-transplant-prediction>.

## REFERENCES

1. Kidney Care UK. *Facts and Stats* (2022). Available from: <https://www.kidneycareuk.org/news-and-campaigns/facts-and-stats/> (Accessed May 19, 2022).
2. Rao PS, Schaubel DE, Guidinger MK, Andreoni KA, Wolfe RA, Merion RM, et al. A Comprehensive Risk Quantification Score for Deceased Donor Kidneys: the Kidney Donor Risk index. *Transplantation* (2009) 88(2): 231–6. doi:10.1097/tp.0b013e3181ac620b
3. Munivenkatappa RB, Schweitzer EJ, Papadimitriou JC, Drachenberg CB, Thom KA, Perencevich EN, et al. The Maryland Aggregate Pathology index: a Deceased Donor Kidney Biopsy Scoring System for Predicting Graft Failure. *Am J Transplant* (2008) 8(11):2316–24. doi:10.1111/j.1600-6143.2008.02370.x
4. Massie AB, Leanza J, Fahmy LM, Chow EKH, Desai NM, Luo X, et al. A Risk index for Living Donor Kidney Transplantation. *Am J Transpl* (2016) 16(7): 2077–84. doi:10.1111/ajt.13709
5. Watson CJ, Johnson RJ, Birch R, Collett D, Bradley JA. A Simplified Donor Risk index for Predicting Outcome after Deceased Donor Kidney Transplantation. *Transplantation* (2011) 93:314–8. doi:10.1097/TP.0b013e31823f14d4

## AUTHOR CONTRIBUTIONS

SR: Design of study, screening of titles, abstracts and full text of all citations, data extraction, risk of bias assessment, analysis of data, writing—original draft, and reviewing and editing. QZ: Screening of titles, abstracts and full text of all citations. W-YT: Clinical insight, writing—reviewing and editing. AC: Supervision, design of study, clinical insight, writing—reviewing and editing. YW: Supervision, design of study, data extraction, risk of bias assessment, analysis of data, writing—reviewing and editing.

## FUNDING

This research is funded by EPSRC DTP (EP/T518153/1:243290) and University Hospitals Plymouth NHS Trust Renal Research Trust Funds (1374).

## CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## ACKNOWLEDGMENTS

We thank Lorna Burns, Information Specialist, for her advice during the development of the search strategy. We thank two reviewers for their helpful comments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontierspartnerships.org/articles/10.3389/ti.2022.10397/full#supplementary-material>

6. NHS Blood and Transplant. *Kidney Transplantation: Deceased Donor Organ Allocation* (2021). Available from: <https://www.odt.nhs.uk/transplantation/tools-policies-and-guidance/policies-and-guidance/>.
7. Kaboré R, Haller MC, Harambat J, Heinze G, Leffondré K. Risk Prediction Models for Graft Failure in Kidney Transplantation: a Systematic Review. *Nephrol Dial Transpl* (2017) 32(Suppl. 1\_2):iii68–ii76. doi:10.1093/ndt/gfw405
8. Senanayake S, White N, Graves N, Healy H, Baboolal K, Kularatna S. Machine Learning in Predicting Graft Failure Following Kidney Transplantation: A Systematic Review of Published Predictive Models. *Int J Med Inform* (2019) 130:103957. doi:10.1016/j.ijmedinf.2019.103957
9. Riley S, Connor A, Wei Y. *Prediction Models for Graft Survival and Patient Survival in Kidney Transplant Recipients: Protocol for a Systematic Review* (2021). Available from: <https://osf.io/c3ehp/> (Accessed May 19, 2022).
10. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 Statement: an Updated Guideline for Reporting Systematic Reviews. *BMJ* (2021) 372:n71. doi:10.1136/bmj.n71
11. Uzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a Web and mobile App for Systematic Reviews. *Syst Rev* (2016) 5(1):210. doi:10.1186/s13643-016-0384-4

12. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *Plos Med* (2014) 11(10):e1001744. doi:10.1371/journal.pmed.1001744
13. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* (2019) 170(1):51–8. doi:10.7326/m18-1376
14. El Ters M, Smith BH, Cosio FG, Kremers WK. Competing Risk Analysis in Renal Allograft Survival: A New Perspective to an Old Problem. *Transplantation* (2021) 105(3):668–76. doi:10.1097/tp.0000000000003285
15. Fine JP, Jiang H, Chappell R. On Semi-competing Risks Data. *Biometrika* (2001) 88(4):907–19. doi:10.1093/biomet/88.4.907
16. Harrell FE, Jr., Lee KL, Mark DB. Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Statist Med* (1996) 15(4):361–87. doi:10.1002/(sici)1097-0258(19960229)15:4<361::aid-sim168>3.0.co;2-4
17. Bae S, Massie AB, Thomas AG, Bahn G, Luo X, Jackson KR, et al. Who Can Tolerate a Marginal Kidney? Predicting Survival after Deceased Donor Kidney Transplant by Donor-Recipient Combination. *Am J Transpl* (2019) 19(2):425–33. doi:10.1111/ajt.14978
18. Baskin-Bey ES, Kremers W, Nyberg SL. A Recipient Risk Score for Deceased Donor Renal Allocation. *Am J Kidney Dis* (2007) 49(2):284–93. doi:10.1053/j.ajkd.2006.10.018
19. Bui K, Kilambi V, Rodrigue JR, Mehrotra S. Patient Functional Status at Transplant and its Impact on Posttransplant Survival of Adult Deceased-Donor Kidney Recipients. *Transplantation* (2019) 103(5):1051–63. doi:10.1097/tp.0000000000002397
20. Calvillo-Arbizu J, Pérez-Valdivia MA, Gentil-Govantes MA, Castro-de-lanuez P, Mazuecos-Blanca A, Rodríguez-Benot A, et al. Does the Kidney Donor Profile Index (KDPI) Predict Graft and Patient Survival in a Spanish Population? *Nefrología (English Edition)* (2018) 38(6):587–95. doi:10.1016/j.nefro.2018.06.014
21. Clayton PA, Dansie K, Sypek MP, White S, Chadban S, Kanellis J, et al. External Validation of the US and UK Kidney Donor Risk Indices for Deceased Donor Kidney Transplant Survival in the Australian and New Zealand Population. *Nephrol Dial Transplant* (2019) 34(12):2127–31. doi:10.1093/ndt/gfz090
22. Clayton PA, McDonald SP, Snyder JJ, Salkowski N, Chadban SJ. External Validation of the Estimated Posttransplant Survival Score for Allocation of Deceased Donor Kidneys in the United States. *Am J Transplant* (2014) 14(8):1922–6. doi:10.1111/ajt.12761
23. Coca A, Arias-Cabrales C, Valencia AL, Burbulla C, Bustamante-Munguira J, Redondo-Pachón D, et al. Validation of a Survival Benefit Estimator Tool in a Cohort of European Kidney Transplant Recipients. *Sci Rep* (2020) 10(1):17109. doi:10.1038/s41598-020-74295-3
24. Haller MC, Wallisch C, Mjøen G, Holdaas H, Dunkler D, Heinze G, et al. Predicting Donor, Recipient and Graft Survival in Living Donor Kidney Transplantation to Inform Pretransplant Counselling: the Donor and Recipient Linked iPREDICTLIVING Tool - a Retrospective Study. *Transpl Int* (2020) 33(7):729–39. doi:10.1111/tri.13580
25. Jackson KR, Munivenkatappa RB, Wesson RN, Garonzik-Wang J, Massie A, Philosophe B. What's the Score? A Comparison of Deceased Donor Kidney Scoring Systems and Correlation with Graft Outcome. *Clin Transpl* (2020) 34(3):e13802. doi:10.1111/ctr.13802
26. Jassal SV, Schaubel DE, Fenton SSA. Baseline Comorbidity in Kidney Transplant Recipients: A Comparison of Comorbidity Indices. *Am J Kidney Dis* (2005) 46(1):136–42. doi:10.1053/j.ajkd.2005.03.006
27. Kasiske BL, Israni AK, Snyder JJ, Skeans MA, Peng Y, Weinhandl ED. A Simple Tool to Predict Outcomes after Kidney Transplant. *Am J Kidney Dis* (2010) 56(5):947–60. doi:10.1053/j.ajkd.2010.06.020
28. Molnar MZ, Nguyen DV, Chen Y, Ravel V, Streja E, Krishnan M, et al. Predictive Score for Posttransplantation Outcomes. *Transplantation* (2017) 101(6):1353–64. doi:10.1097/tp.0000000000001326
29. Philosophe B, Malat GE, Soundararajan S, Barth RN, Manitpisikul W, Wilson NS, et al. Validation of the Maryland Aggregate Pathology Index (MAPI), a Pre-implantation Scoring System that Predicts Graft Outcome. *Clin Transpl* (2014) 28(8):897–905. doi:10.1111/ctr.12400
30. Rehse G, Halleck F, Khadzhyonov D, Lehner LJ, Kleinstuber A, Staack A, et al. Validation of the Living Kidney Donor Profile Index in a European Cohort and Comparison of Long-Term Outcomes with US Results. *Nephrol Dial Transplant* (2019) 34(6):1063–70. doi:10.1093/ndt/gfy118
31. Rose C, Sun Y, Ferre E, Gill J, Landsberg D, Gill J. An Examination of the Application of the Kidney Donor Risk Index in British Columbia. *Can J Kidney Health Dis* (2018) 5:2054358118761052. doi:10.1177/2054358118761052
32. Tiong HY, Goldfarb DA, Kattan MW, Alster JM, Thuita L, Yu C, et al. Nomograms for Predicting Graft Function and Survival in Living Donor Kidney Transplantation Based on the UNOS Registry. *J Urol* (2009) 181(3):1248–55. doi:10.1016/j.juro.2008.10.164
33. Udomkarnjananun S, Townamchai N, Kerr SJ, Tasanarong A, Noppakun K, Lumpaopong A, et al. The First Asian Kidney Transplantation Prediction Models for Long-Term Patient and Allograft Survival. *Transplantation* (2020) 104(5):1048–57. doi:10.1097/tp.0000000000002918
34. Yang J, Sardo Molmenti CL, Cagliani J, Datta H, Grodstein E, Rasul R, et al. Time-Effect of Donor and Recipient Characteristics on Graft Survival after Kidney Transplantation. *Int J Angiol* (2019) 28(4):249–54. doi:10.1055/s-0039-1700500
35. Young A, Knoll GA, McArthur E, Dixon SN, Garg AX, Lok CE, et al. Is the Kidney Donor Risk Index a Useful Tool in Non-US Patients? *Can J Kidney Health Dis* (2018) 5:2054358118791148. doi:10.1177/2054358118791148
36. Zhong Y, Schaubel DE, Kalbfleisch JD, Ashby VB, Rao PS, Sung RS. Reevaluation of the Kidney Donor Risk Index. *Transplantation* (2019) 103(8):1714–21. doi:10.1097/tp.0000000000002498
37. Nyberg SL, Matas AJ, Kremers WK, Thostenson JD, Larson TS, Prieto M, et al. Improved Scoring System to Assess Adult Donors for Cadaver Renal Transplantation. *Am J Transplant* (2003) 3(6):715–21. doi:10.1034/j.1600-6143.2003.00111.x
38. Remuzzi G, Grinyò J, Ruggerenti P, Beatini M, Cole EH, Milford EL, et al. Early Experience with Dual Kidney Transplantation in Adults Using Expanded Donor Criteria. *J Am Soc Nephrol* (1999) 10(12):2591–8. doi:10.1681/asn.v10i122591
39. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls. *BMJ* (2009) 338:b2393. doi:10.1136/bmj.b2393
40. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles Heel of Predictive Analytics. *BMC Med* (2019) 17(1):230. doi:10.1186/s12916-019-1466-7
41. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External Validation of Multivariable Prediction Models: a Systematic Review of Methodological Conduct and Reporting. *BMC Med Res Methodol* (2014) 14(1):40. doi:10.1186/1471-2288-14-40
42. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction Models for Diagnosis and Prognosis of Covid-19: Systematic Review and Critical Appraisal. *BMJ* (2020) 369:m1328. doi:10.1136/bmj.m1328
43. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and Methods in Clinical Prediction Research: A Systematic Review. *Plos Med* (2012) 9(5):e1001221. doi:10.1371/journal.pmed.1001221
44. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc* (1999) 94(446):496–509. doi:10.1080/01621459.1999.10474144
45. Meira-Machado L, de Uña-Álvarez J, Cadarso-Suárez C, Andersen PK. Multistate Models for the Analysis of Time-To-Event Data. *Stat Methods Med Res* (2009) 18(2):195–222. doi:10.1177/0962280208092301
46. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* (2015) 162(1):W1–W73. doi:10.7326/m14-0698
47. Riley RD, Snell KIE, Martin GP, Whittle R, Archer L, Sperrin M, et al. Penalization and Shrinkage Methods Produced Unreliable Clinical Prediction Models Especially when Sample Size Was Small. *J Clin Epidemiol* (2021) 132:88–96. doi:10.1016/j.jclinepi.2020.12.005

48. Riley RD, Snell KI, Ensor J, Burke DL, Harrell Jr FE, Moons KG, et al. Minimum Sample Size for Developing a Multivariable Prediction Model: PART II - Binary and Time-To-Event Outcomes. *Stat Med* (2019) 38(7): 1276–96. doi:10.1002/sim.7992
49. Loupy A, Aubert O, Orandi BJ, Naesens M, Bouatou Y, Raynaud M, et al. Prediction System for Risk of Allograft Loss in Patients Receiving Kidney Transplants: International Derivation and Validation Study. *Bmj* (2019) 366: l4923. doi:10.1136/bmj.l4923
50. Foucher Y, Daguin P, Akl A, Kessler M, Ladrière M, Legendre C, et al. A Clinical Scoring System Highly Predictive of Long-Term Kidney Graft Survival. *Kidney Int* (2010) 78(12):1288–94. doi:10.1038/ki.2010.232

Copyright © 2022 Riley, Zhang, Tse, Connor and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.