2023-11-21

An external validation of the Kidney Donor Risk Index in the UK transplant population in the presence of semi-competing events

Riley, S

10.1186/s41512-023-00159-9
Diagnostic and Prognostic Research
BioMed Central

An external validation of the Kidney Donor Risk Index in the UK transplant population in the presence of semi-competing events

Stephanie Riley[1]*, Kimberly Tam[2], Wai-Yee Tse[3], Andrew Connor[3], Yinghui Wei[1]*

[1]Centre for Mathematical Sciences, School of Engineering, Computing and Mathematics, University of Plymouth, Plymouth, UK

[2]School of Engineering, Computing and Mathematics, University of Plymouth, Plymouth, UK

[3] Department of Renal Medicine, South West Transplant Centre, University Hospitals Plymouth NHS Trust, Plymouth, UK

*Correspondence to:

Stephanie Riley stephanie.riley@plymouth.ac.uk

Yinghui Wei yinghui.wei@plymouth.ac.uk

Abstract

Background Transplantation represents the optimal treatment for many patients with end-stage kidney disease. When a kidney donor is available to a waitlisted patient, clinicians responsible for the care of the potential recipient must make the decision to accept or decline the offer based upon complex and variable information about the donor, the recipient and the transplant process. A clinical prediction model may be able to support clinicians in their decision-making. The Kidney Donor Risk Index (KDRI) was developed in the United States to predict graft failure following kidney transplantation. The survival process following transplantation consists of semi-competing events where death precludes graft failure, but not vice-versa.

Methods We externally validated the KDRI in the UK kidney transplant population, and assessed whether validation under a semi-competing risks framework impacted predictive performance. Additionally, we explored whether the KDRI requires updating. We included 20,035 adult recipients of first, deceased donor, single, kidney-only transplants between January 1st 2004 and December 31st 2018 collected by the UK Transplant Registry and held by NHS Blood and Transplant. The outcomes of interest were one- and five-year graft failure following transplantation. In light of the semi-competing events, recipient death was handled in two ways: censoring patients at the time of death, and modelling death as a competing event. Cox proportional hazard models were used to validate the KDRI when censoring graft failure by death, and cause-specific Cox models were used to account for death as a competing event.

Results The KDRI underestimated event probabilities for those at higher risk of graft failure. For five-year graft failure discrimination was poorer in the semi-competing risks model (0.625, 95% CI: 0.611 to 0.640;0.607, 95% CI: 0.589 to 0.625), but predictions were more accurate (Brier score 0.117, 95% CI:0.112 to 0.121; 0.114, 95% CI:0.109 to 0.118). Calibration plots were similar regardless of whether death was modelled as a competing event or not. Updating the KDRI worsened calibration, but marginally improved discrimination.

Conclusions Predictive performance for one-year graft failure was similar between death-censored and competing event graft failure, but differences appeared when predicting five-year graft failure. The updated index did not have superior performance and we conclude that updating the KDRI in the present form is not required.

Keywords: survival analysis; time-to-event model; competing events; risk prediction; external validation; kidney transplantation

Introduction

For many patients with end-stage kidney disease, transplantation represents the optimal treatment. The demand for deceased donor kidneys in the United Kingdom (UK) greatly outweighs availability (1). It is therefore essential to maximise the number of successful transplants in order to reduce the number of recipients returning to the transplant waiting list or dialysis. A prediction model may provide support to clinicians charged with deciding whether to accept the offer of a donor kidney for an individual patient. Such models can incorporate a large number of donor, recipient, and transplant related factors to produce personalised risk predictions.

In the United States (US) the Kidney Donor Risk Index (KDRI), proposed by Rao et al. (2), is used as part of the allocation process for deceased donor kidneys to those awaiting a kidney transplant. It was originally developed to predict graft failure in first-time, kidney-only, adult transplants with the intention of being used as a decision-making tool at the time of a donor kidney offer. The risk index uses 13 donor-related parameters that would be known by the clinician at the time of the offer including age, height, weight, and history of hypertension and diabetes.

The scientific and clinical practices underpinning the delivery of transplantation services have evolved over time. Further variation exists between different units and countries. As such, prediction models developed in a particular country may not be reliably applicable to populations in other countries in the future. It is therefore essential to externally validate proposed prediction models when considering their use in different populations and to revisit these validations over time (3-5).

We sought to validate the predictive performance of the KDRI in the UK kidney transplantation population. In our systematic review (6), we found that the KDRI has been validated in different populations across the globe (7-15). In the UK, Watson et al. (14) assessed the performance in transplants performed between 2000 and 2007. The KDRI showed moderate discrimination

in predicting the earliest of graft failure and death (C-index 0.63). The calibration has not previously been assessed in the UK kidney transplant population.

The survival process following transplantation consists of semi-competing events, where a terminal event precludes the observation of a non-terminal event, but not vice-versa. Specifically, in the context of kidney transplant survival outcomes, once a patient has died, we can no longer observe whether they experience graft failure. However, if a patient suffered graft failure then we could still observe their death. In the existing literature on prediction models for graft failure, death is often not treated as a competing event, rather graft failure is censored by death or they are combined to predict a composite event.

The original KDRI defined graft failure as the earliest of graft failure or death. Predicting a composite outcome assumes that predictors have the same effect on both outcomes of interest (16) and, in doing so, researchers shift the attention from the primary clinical endpoint of the proposed prediction model to one that may not be of clinical interest. Censoring the primary event of interest by the competing event violates the assumption of non-informative censoring typically used in standard time-to-event methods and can lead to bias in the cumulative incidence estimator, such that the sum of the individual event estimators exceeds the estimator of the composite event (17, 18). Recent work has also noted the importance of accounting for competing events in external validation studies as well as in model development (19, 20).

The aim of this study was to externally validate the KDRI in the UK kidney transplantation adult population. Additionally, we aimed to explore whether modelling death as a competing event, rather than censoring recipients at the time of death, influences the predictive performance of the KDRI. Furthermore, we assessed whether updating the KDRI was required to improve predictions for graft failure.

Methods

This study was reported in accordance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement (21).

Source of data

This was a cohort study based on registry database collected by the UK Transplant Registry (UKTR) and held by NHS Blood and Transplant (NHSBT). Recipients were transplanted in the United Kingdom between January 1st 2004 and December 31st 2018. Recipients were followed-up until March 31st 2021.

Participants

Adult recipients (aged 18 years and above) of a first, deceased donor, single, kidney-only transplant were included. Where recipients have had multiple transplants within the study period, only their first one was used for analysis. Recipients of en-bloc, or multiple organ transplants (such as combined kidney and pancreas transplants) were not included.

Outcomes

We assessed the performance of the KDRI for predicting graft failure one year and five years following kidney transplantation. Graft failure was defined as the time from transplantation until either return to dialysis or re-transplantation. In light of the semi-competing events, recipient death was handled in two ways: censoring patients at the time of death, and modelling death as a competing event. The original KDRI was intended to predict graft failure, however, the model was developed to predict a composite outcome of time to the earliest of death or graft failure.

Missing data

Recipients with both event time and indicator missing were excluded from analysis. Missing values for donor height, weight, ethnicity, history of hypertension, history of diabetes, cause of death (cerebrovascular accident or not), creatinine value, and Hepatitis C virus status were

imputed using multiple imputation with chained equations (22), assuming that the data were missing at random. None of the donors were missing age or type (deceased cardiac or deceased brain donor). Hence, these variables were not imputed but were included in the imputation model, along with the Aalen-Johansen estimates for the cumulative hazard. Continuous variables were imputed using predictive mean matching to ensure that implausible values were not imputed, such as negative values for height and weight.

12.78% of the patients had incomplete information for calculating the KDRI, therefore we determined at least thirteen imputed data sets were required (23). Fifteen imputed data sets were generated. For continuous variables, imputations were checked by comparing the distributions between imputed data sets. For binary and categorical variables, we checked whether the counts were similar between imputations (see Supplementary Material).

Parameter estimates and model performance measures, along with the associated standard errors, were pooled across the imputed data sets according to Rubin's rules (24). These pooled estimates and standard errors were used to construct 95% confidence intervals using the 97.5^{th} quantile of the t-distribution.

Sample size

The suitability of the sample size was determined according to the methods of Riley et al. (25). While the sample size for this study was fixed (20,035 recipients), we also explored the mean standard error of the calibration slope for a range of sample sizes. These, along with further details on the sample size calculation, can be found in the Supplementary Material.

In the development article (2) the KDRI was split into quintiles, and Kaplan-Meier curves of the probability of graft survival were reported for each. We read the survival probabilities for the minimum and maximum quantiles and explored the sample size required for survival probabilities within that range. For one-year graft failure, we considered survival probabilities 0.875, 0.901, 0.927, and 0.953, and for five-year graft failure 0.635, 0.697, 0.760, and 0.822.

For one- and five-year graft failure, the mean standard error for the calibration slope varied between 0.053 and 0.092, and 0.036 and 0.051, respectively for the survival probabilities under consideration (Table 1). We deemed these to be acceptable.

Summary statistics

Median time to graft failure was calculated using the Kaplan-Meier method, whereby the median time is given by the time at which the probability of survival is 0.5. Median follow-up times were calculated using the reverse Kaplan-Meier method. This is similar to the Kaplan-Meier method except the censoring indicator is treated as an event indicator.

Model performance

Discrimination

Discrimination measures the rank separation between those who experience the outcome of interest and those who do not. For example, a model that discriminates well will predict a higher risk for a recipient that experiences graft failure than one who does not.

The discrimination was assessed using the time-dependent area under receiver operating characteristic curve (AUC) (27). Values typically range between 0.5 and 1, where 1 indicates perfect discrimination and 0.5 shows that predictions are as accurate as flipping a coin.

Calibration

Calibration is used to measure the agreement between observed and predicted risks. Here we used the observed event proportion as a proxy for observed risk. As some patients were censored prior to the event time horizon, it was not possible to calculate the observed event proportion. To overcome this we used a jack-knife approach to calculate pseudo-observations, which were then used as proxy measures of event indicators for censored patients (28).

Calibration plots using pseudo-observations with local weighted regression smoothing were assessed in each imputed dataset. For models that are well calibrated the smoothed curve lies on the diagonal line that runs through the origin.

Further, we calculated the calibration slope, where a value equal to one indicates perfect calibration. A calibration slope less than 1 suggests that predictions are too high for recipients with high event probabilities and too low for those with low probabilities. Conversely, a calibration slope greater than 1 suggests that recipients with high observed risk are under-estimated, and those with low observed risk are over-estimated. The calibration slope was calculated using a generalised linear model with pseudo-observations as the outcome and the complementary log-log transformed predicted risks as both an offset and covariate . The coefficient of the transformed risks indicates how far the calibration slope differs from 1, thus the calibration slope is given by summing these two values.

Given that the baseline survival value for one- and five-years following transplantation were not reported in the original article, it was calculated within the UKTR data. Consequently, the calibration may appear more optimistic since the baseline cumulative incidence, which is required to calculate the absolute risk of graft failure, was estimated in the same cohort.

Overall prediction accuracy

The Brier score measures prediction error by estimating the squared difference between the event indicator and estimated risk (30). Values closer to zero indicate a more accurate prediction model.

Validation of the KDRI

The KDRI (2) can be calculated using

$$\begin{aligned}\text{KDRI} = \exp\{&-0.0194\ \text{I}[\text{age} < 18\text{yr}](\text{age} - 18\text{yr}) + \ 0.0128(\text{age} - 40) \\ &+ 0.0107\ \text{I}[\text{age} > 50](\text{age} - 50) - \frac{0.0464(\text{height} - 170)}{10} \\ &- \frac{0.0199\ \text{I}[\text{weight} < 80](\text{weight} - 80)}{5} \\ &+ 0.1790\ \text{I}[\text{ethnicity African American}] \\ &+ 0.1260\ \text{I}[\text{history of hypertension}] + 0.1300\ \text{I}[\text{history of diabetes}] \\ &+ 0.0881\ \text{I}[\text{cause of death cerebrovascular accident (CVA)}] \\ &+ 0.2200(\text{creatinine} - 1) - 0.2090\ \text{I}[\text{creatinine} > 1.5](\text{creatinine} - 1.5) \\ &+ 0.2400\ \text{I}[\text{Hepatitis C virus (HCV) positive}] \\ &+ 0.1330\ \text{I}[\text{deceased cardiac donor (DCD)}]\},\end{aligned}$$

where $\text{I}[.]$ is an indicator function which is equal to 1 if the criteria in [.] are satisfied and 0 otherwise. The KDRI originally derived by Rao et al. also considered transplant related factors, such as cold ischaemic time, human leukocyte antigen mismatch, and whether it was an en-bloc or double kidney transplant. In practice, only the donor related factors are used to calculate the KDRI (31). With this in mind, we validated the donor-only KDRI.

For each recipient, we calculated the linear predictor of the KDRI, by applying the natural logarithm to the index. Cox proportional hazards models (32) were used to assess the performance of the KDRI for predicting death-censored graft failure. In the presence of competing events, the Cox model can lead to biased risk estimation. As alternatives, researchers typically use either the cause-specific Cox (33) or the Fine-Gray model (34). The Fine-Gray model is often preferred when the goal is prediction rather than association (35). However, in some instances, it is possible for the sum of patient-specific event probabilities, which should be constrained between zero and one, to exceed one (36). Therefore, we used the cause-specific Cox models to validate the KDRI when accounting for death as a competing event.

Updating the KDRI

To assess whether the KDRI required updating, we re-estimated the coefficients used in the original index. For the KDRI to be applicable in the UK cohort we substituted African American ethnicity for Black ethnic origin. Variables were centred in the same way that they were in the original KDRI, and no further variable selection was undertaken. We re-estimated the coefficients by censoring graft failure at the time of death using a Cox proportional hazards model, and accounting for death as a competing event using a cause-specific Cox model.

The coefficients were re-estimated in each of the 15 imputed datasets, and the performance of those updated models were individually assessed. The re-estimated coefficients and performance measures were then pooled according to Rubin's rules.

When updating the KDRI we assessed the predictive performance in the same group of recipients that were used to update the KDRI. This will naturally produce optimistic results. To account for optimism in the numerical summaries of predictive performance we used Harrell's bias correction method (37), with 100 bootstrap samples. Calibration plots have not been adjusted for optimism and thus represent the apparent calibration.

Software

Multiple imputation was performed using Stata/MP 16.1 (38). All other analyses were conducted in R 4.1.2 (39).

Results

Summary statistics

In total 20,134 deceased donor single kidney-only recipients who received a transplant between January 1st 2004 and December 31st 2018 in the UK were eligible for inclusion (Figure 1). Eleven of the recipients had missing time-to-event information for both graft failure and death, and 88 missing for graft failure only. Therefore 20,035 transplants were included in our analysis.

The end of the follow-up period was March 31st 2021. The median follow-up time was 5.96 years and maximum follow-up time 17.05 years. The minimum probability of survival for graft failure and for death was both above 0.5 , hence we did not observe the median survival time for either outcome in this study. At the end of the follow-up period 13,724 (68.50%) recipients were alive with a functioning graft. 2,675 (13.35%) recipients experienced graft failure only, and 904 (4.51%) died following graft failure. 2,732 (13.64%) recipients died with a functioning graft (Figure 1).

By the end of the first year following transplantation 1,050 (5.24%) recipients had experienced graft failure only, and 186 (0.93%) died following graft failure. A total of 497 (2.48%) recipients had died with a functioning graft. By five years, 1,936 (9.66%) recipients experienced graft failure alone, and 456 (2.28%) died following graft failure. 1,509 (7.53%) transplant recipients died with a functioning graft.

A summary of the donor characteristics used to calculate the KDRI, including the number of missing values, is presented in Table 2, and a summary of recipient characteristics can be found in the Supplementary Material. Donors were aged between one and 85 years old, where 626 (3.13%) were younger than 18 years of age and 10,925 (54.53%) were older than 50. 10,646 (53.14%) donors weighed less than 80kg. Creatinine was greater than 1.5mg/dl for 1,720 (8.59%) donors.

No values were missing for donor age and type of donor. 1,688 (8.43%) donors had missing values for creatinine, the most of any variables required for calculating the KDRI. 2,560 (12.78%) were missing at least one value required to calculate the KDRI.

The distribution of the KDRI in the original article by Rao et al. (2) was similar in shape to that of the transplants included in this analysis (Figure 2). However, median KDRI values were higher in the UK cohort; 1.32 compared with 1.05 in the US cohort used to develop the index.

External validation of the KDRI

One-year graft failure

The KDRI discriminated moderately well for predicting one-year graft failure. The time-dependent AUC was 0.607 (95% CI: 0.589 to 0.625) and 0.610 (95% CI: 0.592 to 0.628) with and without accounting for competing events, respectively (*Table 3).*

Calibration plots and slopes were similar when modelling graft failure while censoring for death and death as a competing event (Figure 3a, Figure 3b). The KDRI was well calibrated for recipients with predicted risks less than 10%, but calibration was worse for those with predicted risk above this value. Only 1,100 (5.5%) were at a higher risk than 10%, and for those the KDRI underestimated the risk of graft failure. Calibration slopes were, respectively, 1.074 (95% CI: 0.878 to 1.271) and 1.075 (95% CI: 0.877 to 1.272) for predicting death-censored and competing event graft failure.

Predictive accuracy was the same regardless of whether death was handled as a competing event or not with reported Brier scores equal to 0.058 for both types of outcomes.

Five-year graft failure

Five years following kidney transplantation the time-dependent AUC was slightly lower when predicting graft failure with death as a competing event (0.611, 95% CI: 0.597 to 0.625) as opposed to censoring at the time of death (0.625, 95% CI: 0.611 to 0.640).

Using calibration plots, predicted risks using the KDRI were generally similar to the observed proportion of recipients who experienced graft failure (Figure 4a, Figure 4b). Calibration was poorest for those at higher risk of graft failure. The risk of graft failure was underestimated for recipients at a higher risk. The calibration slopes were 0.964 (95% CI: 0.827 to 1.100) when censoring recipients at the time of death, and 0.979 (95% CI: 0.835 to 1.123) when modelling death as a competing event.

Predictions were less accurate for five-year compared with one-year graft failure. Brier scores differed for death-censored and competing event graft failure (0.117,95% CI: 0.112 to 0.121; and 0.114, 95% CI: 0.109 to 0.118 respectively).

Updating the KDRI

To update the KDRI, we re-estimated the coefficients used in the original index in the UK kidney transplant population. No additional predictors were considered. In the updated models, the estimates were similar regardless of whether death was modelled as a competing event or not (Table 4). Confidence intervals for the effect of ethnicity were much wider than in the original index, likely because only 1.11% of donors were of Black ethnic origin.

The effect of age for those under 18 and over 50 years, and height were not found to be associated with graft failure. Additionally, donor ethnicity, and donor HCV status were not significantly associated with graft survival.

One-year graft failure

Discrimination was similar for predicting death-censored graft failure (time-dependent AUC 0.614) and graft failure with death as a competing event (time-dependent AUC 0.608) (Table 5).

Calibration slopes were 1.096 when censoring at the time of death, and 1.068when modelling death as a competing event. Calibration plots were similar for both types of graft failure and clearly showed that low risks were over-estimated and high risks were under-estimated (Figure 3c, Figure 3d).

There was no difference in prediction accuracy whether accounting for death as a competing event or not, with Brier scores equal to 0.058 for both cases.

Five-year graft failure

Discrimination was lower when modelling graft failure with death as a competing event (time-dependent AUC 0.629 and 0.614, respectively) for the updated index (Table 5).

We found calibration slopes were 1.016 for death-censored graft failure and 1.002 when modelling death as a competing event. The calibration plots (Figure 4c, Figure 4d) showed miscalibration for recipients at the highest and lowest predicted risks, and 95% confidence intervals were much wider at the tails of the curve.

Prediction accuracy was slightly improved when modelling death as a competing event compared to censoring at the time of death with Brier scores equal to 0.114 and 0.117, respectively.

Discussion

Principal findings

In external validation the KDRI had moderate discrimination and was generally well calibrated for predicting graft failure one year and five years following kidney transplantation. For predicting one-year graft failure discrimination, calibration and predictive accuracy did not differ depending on how death prior to graft failure was handled. Discrimination was higher for predicting five-year graft failure. Predictions were more accurate for early graft failure compared to those at five years following transplantation.

Calibration slopes indicated miscalibration in the KDRI, however, the corresponding 95% confidence intervals were wide. In calibration plots for both outcomes, miscalibration was mainly driven by recipients at higher risk, where the event probabilities were generally underestimated. However, it should be noted that the baseline survival was not reported in the original article and as such has been estimated within the same cohort as is being validated. Therefore, the calibration of the KDRI in the UK kidney transplant population may be more optimistic.

Updating the KDRI in the UK kidney transplant population yielded similar coefficients, but some prognostic factors were no longer associated with graft failure. Given that the coefficient

estimates did not differ between the Cox and the cause-specific Cox models, it is unsurprising that there was little difference between the predictive performance of those models.

Strengths and limitations

To our knowledge, this is the first study to assess the performance of the KDRI under a semi-competing risks framework for first, deceased donor, single, kidney-only, and adult transplants. Our work included all eligible kidney transplants that occurred in the UK during the study period, with a long follow-up period. The KDRI was previously validated in the UK by Watson et al. (14) using information on kidney transplants that occurred between 2000 and 2007. Therefore, our work serves to assess whether the KDRI is still relevant in the UK kidney transplant population.

The baseline survivor function was recalibrated for our cohort; therefore, calibration may seem more optimistic in this external validation of the KDRI. From the current analyses we cannot comment on the clinical utility of the index in the UK kidney transplant population. Further work is required to determine whether the KDRI is clinically relevant in practice.

Few recipients experienced a competing event (died with a functioning graft), which may explain why little difference was found in predictive performance when considering death as a competing event and censoring graft failure at the time of death. There is a lack of guidance concerning under what situation ignoring the competing risk elements can impact the performance of the prediction models. Externally validating and updating a model under the competing risk framework serves as a sensitivity analysis to evaluate the developed models which ignore the competing risk elements. Future work could explore to what extent the proportion of non-terminal events censored by the terminal events impacts the predictive performance. This can potentially lead to recommendations for practice for when it is necessary to account for competing events, and when traditional methods, such as the Cox proportional hazards model, might suffice.

Results in context

The KDRI only considers donor-related variables to predict graft failure in the recipient of the kidney transplant. Additional donor variables may improve predictive performance. The Maryland Aggregate Pathology Index (MAPI) (40), for example, utilises information gathered from biopsies of donor kidneys, and has shown higher discrimination in internal and external validation (9, 41). Such additional information may be able to improve performance. However, it may not be practical in a decision-making tool since, in the UK, this information may not be known at the time of the offer of a donor kidney. Additionally, utilising information about the recipient and the transplant process, or other existing indices which incorporate these variables, could also improve predictions.

A validation study in the US (9) evaluated the predictive performance of the KDRI two years following transplantation and showed poor discrimination with time-dependent AUC equal to 0.45. External validation in Australia and New Zealand (15) reported C-index 0.63 (95% CI: 0.60 to 0.65) for predicting death-censored graft failure. In Canada the KDRI showed moderate discrimination with C-index equal to 0.59 (13). The KDRI was previously validated using data from kidney transplants performed in the UK between 2000 and 2007 (14), and reported a C-index of 0.63.

Zhong et al. (42) also assessed whether the original KDRI required updating using information on kidney transplants performed between 2000 and 2016 in the US. Their updated index showed marginally higher discrimination than the original KDRI (original KDRI C-index 0.651; updated KDRI C-index 0.652), however the calibration was not assessed. This study also determined that there is little to be gained in updating the KDRI.

Conclusions

The Kidney Donor Risk Index, originally developed in the US population, showed moderate predictive performance overall in our external validation in the UK kidney transplant population. The use of a semi-competing risks framework made a slight difference when

predicting five-year graft failure compared to censoring for death. The updated index had slightly improved discrimination but was poorly calibrated for those with the highest and lowest risk of graft failure. Therefore, we conclude that updating the KDRI in the present form is not required.

Declarations

Ethics approval and consent to participate

This study is approved by the University of Plymouth Faculty Research Ethics committee (reference number 2631), and by the NHS Health Research Authority (REC reference: 21/HRA/2130, IRAS ID: 296108). Consent was not required as the data are non-identifiable and pre-existing from usual care, held by NHSBT. NHSBT are reliant on the General Data Protection Regulation Article 6(1)(e) – Performance of a public task. Under Article 9(2)(h), (i) and (j), NHSBT is permitted to use patient identifiable information for service evaluation and safety monitoring without the consent of patients. This was a retrospective study and no participants were recruited. We comply with NHSBT data access policy, Good Clinical Practice and the NHS Research Governance Framework for Health and Social Care Research (2017).

Consent for publication

Not applicable.

Availability of data and materials

The data underlying this article can be requested from the data controller, NHS Blood and Transplant. The code used to conduct these analyses are available from GitHub: https://github.com/Yinghui-Wei-team/kdrivalidation

Competing interests

The authors declare that they have no competing interests.

Funding

This study is funded by UKRI EPSRC DTP (EP/T518153/1:243290) and University Hospitals Plymouth NHS Trust Renal Research Trust Funds (1374). YW received funding from UKRI EPSRC Impact Acceleration Account (EP/X525789/1).

Authors' contributions

SR: study design, interpretation of data, writing – initial draft, writing – reviewing and editing; KT: writing – reviewing and editing; WT: clinical insights, writing – reviewing and editing; AC: supervision, study design, clinical insights, writing – reviewing and editing; YW: funding acquisition, supervision, study design, interpretation of data, writing – reviewing and editing.

Acknowledgements

The authors gratefully acknowledge NHS Blood and Transplant for access to the UK Transplant Registry data used in this project. We are grateful to all the transplant centres in the UK who contributed data on which this project is based. We thank the two reviewers who provided valuable comments and suggestions.

References

1. NHS. Waiting List: Kidney Transplant 2018 [Available from: https://www.nhs.uk/conditions/kidney-transplant/waiting-list/.
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.
3. Jenkins DA, Martin GP, Sperrin M, Riley RD, Debray TPA, Collins GS, et al. Continual updating and monitoring of clinical prediction models: time for dynamic prediction systems? Diagnostic and Prognostic Research. 2021;5(1):1.
4. Davis SE, Greevy RA, Jr, Fonnesbeck C, Lasko TA, Walsh CG, Matheny ME. A nonparametric updating method to correct clinical prediction model drift. Journal of the American Medical Informatics Association. 2019;26(12):1448-57.
5. Davis SE, Lasko TA, Chen G, Siew ED, Matheny ME. Calibration drift in regression and machine learning models for acute kidney injury. Journal of the American Medical Informatics Association. 2017;24(6):1052-61.
6. Riley S, Zhang Q, Tse W, Connor A, Wei Y. Using information available at the time of donor offer to predict kidney transplant survival outcomes: a systematic review of prediction models. Transplant International. 2022.
7. Calvillo-Arbizu J, Perez-Valdivia MA, Gentil-Govantes MA, Castro-de-la-Nuez P, Mazuecos-Blanca A, Rodriguez-Benot A, et al. Does the Kidney Donor Profile Index (KDPI) predict graft and patient survival in a Spanish population? Nefrologia. 2018;38(6):587-95.
8. Coca A, Arias-Cabrales C, Valencia AL, Burballa C, Bustamante-Munguira J, Redondo-Pachon D, et al. Validation of a survival benefit estimator tool in a cohort of European kidney transplant recipients. Scientific Reports. 2020;10(1):17109.
9. 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. Clinical Transplantation. 2020;34(3):e13802.
10. Massie A, Leanza J, Fahmy L, Chow E, Luo X, Segev D. A risk index for living donor kidney transplantation. American Journal of Transplantation. 2016;16:31.
11. Rehse G, Halleck F, Khadzhynov D, Lehner LJ, Kleinsteuber A, Staeck A, et al. Validation of the Living Kidney Donor Profile Index in a European cohort and comparison of long-term outcomes with US results. Nephrology Dialysis Transplantation. 2019;34(6):1063-70.
12. Rose C, Sun Y, Ferre E, Landsberg D, Gill J. An Examination of the Application of the Kidney Donor Risk Index in British Columbia. Canadian Journal of Kidney Health and Disease. 2018;5.
13. 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? Canadian Journal of Kidney Health & Disease. 2018;5:2054358118791148.
14. Watson CJE, Johnson RJ, Birch R, Collett D, Andrew Bradley J. A simplified donor risk index for predicting outcome after deceased donor kidney transplantation. Transplantation. 2011;15:314-8.
15. 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. Nephrology Dialysis Transplantation. 2019;34(12):2127-31.
16. Gómez G, Lagakos SW. Statistical considerations when using a composite endpoint for comparing treatment groups. Statistics in Medicine. 2013;32(5):719-38.
17. Coemans M, Verbeke G, Döhler B, Süsal C, Naesens M. Bias by censoring for competing events in survival analysis. BMJ. 2022;378:e071349.
18. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. Circulation. 2016;133(6):601-9.

19. Ramspek CL, Teece L, Snell KIE, Evans M, Riley RD, van Smeden M, et al. Lessons learnt when accounting for competing events in the external validation of time-to-event prognostic models. International Journal of Epidemiology. 2021;51(2):615-25.
20. van Geloven N, Giardiello D, Bonneville EF, Teece L, Ramspek CL, van Smeden M, et al. Validation of prediction models in the presence of competing risks: a guide through modern methods. BMJ. 2022;377:e069249.
21. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1-W73.
22. 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.
23. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine. 2011;30(4):377-99.
24. Rubin DB. Multiple imputation for survey nonresponse. New York: Wiley; 1987.
25. Riley RD, Collins GS, Ensor J, Archer L, Booth S, Mozumder SI, et al. Minimum sample size calculations for external validation of a clinical prediction model with a time-to-event outcome. Statistics in Medicine. 2022;41(7):1280-95.
26. 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. Statistics in Medicine. 2019;38(7):1276-96.
27. 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. Statistics in Medicine. 2013;32(30):5381-97.
28. Andersen PK, Pohar Perme M. Pseudo-observations in survival analysis. Statistical Methods in Medical Research. 2010;19(1):71-99.
29. Royston P. Tools for Checking Calibration of a Cox Model in External Validation: Prediction of Population-Averaged Survival Curves Based on Risk Groups. The Stata Journal. 2015;15(1):275-91.
30. Gerds TA, Schumacher M. Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. Biometrical Journal. 2006;48(6):1029-40.
31. Network OPaT. Organ Procurement and Transplantation Network Policies 2023. 143 p.
32. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society: Series B (Methodological). 1972;34(2):187-202.
33. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Statistics in Medicine. 2007;26(11):2389-430.
34. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. Journal of the American Statistical Association. 1999;94(446):496-509.
35. Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic Models With Competing Risks: Methods and Application to Coronary Risk Prediction. Epidemiology. 2009;20(4):555-61.
36. Austin PC, Steyerberg EW, Putter H. Fine-Gray subdistribution hazard models to simultaneously estimate the absolute risk of different event types: Cumulative total failure probability may exceed 1. Statistics in Medicine. 2021;40(19):4200-12.
37. Harrell Jr. FE, Lee KL, Mark DB. Multivariable Prognostic Models: Issues In Developing Models, Evaluating Assumptions And Adequacy, And Measuring And Reducing Errors. Statistics in Medicine. 1996;15(4):361-87.
38. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC; 2019.
39. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
40. 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. American Journal of Transplantation. 2008;8(11):2316-24.

41. 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. Clinical Transplantation. 2014;28(8):897-905.
42. 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.

Tables

Table 1: Mean standard error of calibration slope from simulation study of size 500, assuming a sample size of 2,000. The linear predictor was assumed to follow Log-Norm(log(1.05), 0. 42487).

Survival probability	Event time distribution rate	Mean calibration slope standard error
1-year graft failure		
0.875	0.118	0.053
0.901	0.092	0.061
0.927	0.067	0.074
0.953	0.042	0.092
5-year graft failure		
0.635	0.083	0.036
0.697	0.065	0.039
0.760	0.049	0.044
0.822	0.035	0.051

Table 2: Characteristics of donor patients in the UK kidney transplant population between January 1st 2004 and December 31st 2018. Numerical summaries of variables used to calculate the Kidney Donor Risk Index, including number (and percentage) of missing values.

Variable	Mean [SD] or N (%)	Missing (%)
Age, years	49.91 [15.44]	0 (0)
Height, cm	170.29 [10.80]	269 (1.34)
Weight, kg	77.96 [17.77]	130 (0.65)
Ethnicity		63 (0.31)
Asian	372 (1.86)	
Black	222 (1.11)	
Chinese/Oriental	56 (0.28)	
Mixed	148 (0.74)	
Other	160 (0.80)	
White	19,014 (94.90)	
History of hypertension		716 (3.57)
Yes	5,296 (26.43)	
No	14,023 (69.99)	
History of diabetes		527 (2.63)
Yes	1,275 (6.36)	
No	18,233 (91.01)	
Cause of death		161 (0.80)
CVA	741 (3.70)	
Not CVA	19,133 (95.50)	
Creatinine, mg/dl	0.95 [0.60]	1,688 (8.43)
HCV test result		63 (0.32)
Positive	21 (0.10)	
Negative	19,951 (99.58)	
Donor type		0 (0)
DCD	7,517 (37.52)	
DBD	12,518 (62.48)	

CVA: cerebrovascular accident; HCV: hepatitis C virus; DCD: deceased cardiac donor; DBD: deceased brain donor.

Table 3: Numerical summary of performance of the original Kidney Donor Risk Index for predicting graft failure one year and five years following transplantation while censoring for death and modelling death as a competing event. T-D AUC: time-dependent area under receiver operating curve; CI: confidence interval.

	Censoring at the time of death	Accounting for death as competing event
One-year graft failure		
T-D AUC	0.610	0.607
(95% CI)	(0.592, 0.628)	(0.589, 0.625)
Calibration slope	1.074	1.074
(95% CI)	(0.878, 1.271)	(0.877, 1.272)
Brier Score	0.058	0.058
(95% CI)	(0.055, 0.062)	(0.054, 0.061)
Five-year graft failure		
T-D AUC	0.625	0.611
(95% CI)	(0.611, 0.640)	(0.597, 0.625)
Calibration slope	0.964	0.979
(95% CI)	(0.827, 1.100)	(0.835, 1.123)

Brier Score	0.117	0.114
(95% CI)	(0.112, 0.121)	(0.109, 0.118)

Table 4: Coefficients of the variables used to calculate the Kidney Donor Risk Index from the original development, the updated Cox proportional hazards model and the updated cause-specific Cox model. Variables are centred as they were in the original publication.

Variable	Original	Updated (Cox model)	Updated (Cause-specific Cox model)
Age-18; for donors under	-0.019	-0.032	-0.031
18 years	(-0.031, -0.010)	(-0.096, 0.033)	(-0.095, 0.033)
Age-40, years	0.013	0.016	0.016
	(0.011, 0.015)	(0.009, 0.024)	(0.009, 0.024)
Age-50; for donors over	0.011	0.003	0.003
50 years	(0.005, 0.016)	(-0.010, 0.016)	(-0.010, 0.016)
Height per 10cm increase	-0.046	-0.034	-0.034
	(-0.062, -0.031)	(-0.093, 0.026)	(-0.093, 0.026)
Weight per 5kg increase;	-0.020	-0.035	-0.035
for donors below 80kg	(-0.031, -0.010)	(-0.068, -0.002)	(-0.068, -0.002)
Ethnicity			
Not Black ethnic origin		Reference	
Black ethnic origin	0.179	0.409	0.405
	(0.122, 0.239)	(-0.017, 0.835)	(-0.021, 0.832)
History of hypertension			
No		Reference	

Yes	0.126	0.256	0.255
	(0.077, 0.174)	(0.135, 0.378)	(0.133, 0.376)
History of diabetes			
No		Reference	
Yes	0.130	0.154	0.153
	(0.039, 0.215)	(-0.050, 0.358)	(-0.051, 0.357)
Cause of death CVA			
No		Reference	
Yes	0.088	-0.035	-0.036
	(0.039, 0.131)	(-0.308, 0.238)	(-0.309, 0.237)
Creatinine-1, mg/dl	0.220	0.395	0.395
	(0.157, 0.285)	(0.205, 0.586)	(0.204, 0.586)
Creatinine-1; for donors with creatinine > 1.5, mg/dl	-0.209	-0.500	-0.500
	(-0.301, -0.117)	(-0.819, -0.182)	(-0.818, -0.181)
HCV			
Negative		Reference	
Positive	0.240	0.233	0.234
	(0.122, 0.358)	(-1.286, 1.751)	(-1.284, 1.753)
Donor type			
DBD		Reference	

DCD	0.133	0.113	0.112
	(0.020, 0.247)	(0.005, 0.221)	(0.004, 0.220)

CVA: cerebrovascular accident; DBD: deceased brain donor; DCD: deceased cardiac donor.

Table 5: Numerical summary of performance of the updated Kidney Donor Risk Index for predicting graft failure one year and five years following transplantation while censoring for death and modelling death as a competing event. T-D AUC: time-dependent area under receiver operating curve.

	Censoring at the time of death	Accounting for death as competing event
One-year graft failure		
T-D AUC	0.614	0.608
(Optimism)	(-0.00015)	(0.00252)
Calibration slope	1.096	1.068
(Optimism)	(-0.00049)	(0.02812)
Brier Score	0.058	0.058
(Optimism)	(-0.00004)	(-0.00007)
Five-year graft failure		
T-D AUC	0.629	0.612
(Optimism)	(-0.00019)	(0.00254)
Calibration slope	1.016	1.002
(Optimism)	(-0.00133)	(0.02911)
Brier Score	0.117	0.114
(Optimism)	(0.00006)	(-0.00010)

Figures

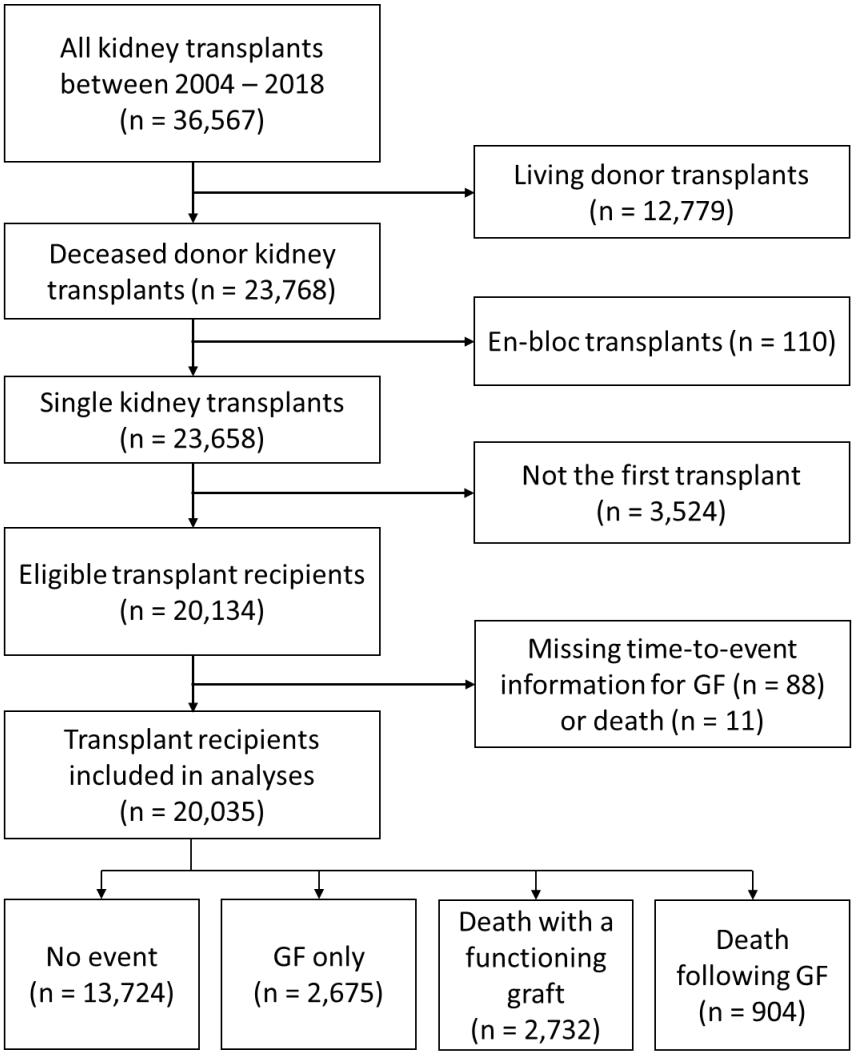

Figure 1: Flowchart of eligible transplant recipients for inclusion in analyses. GF: graft failure.

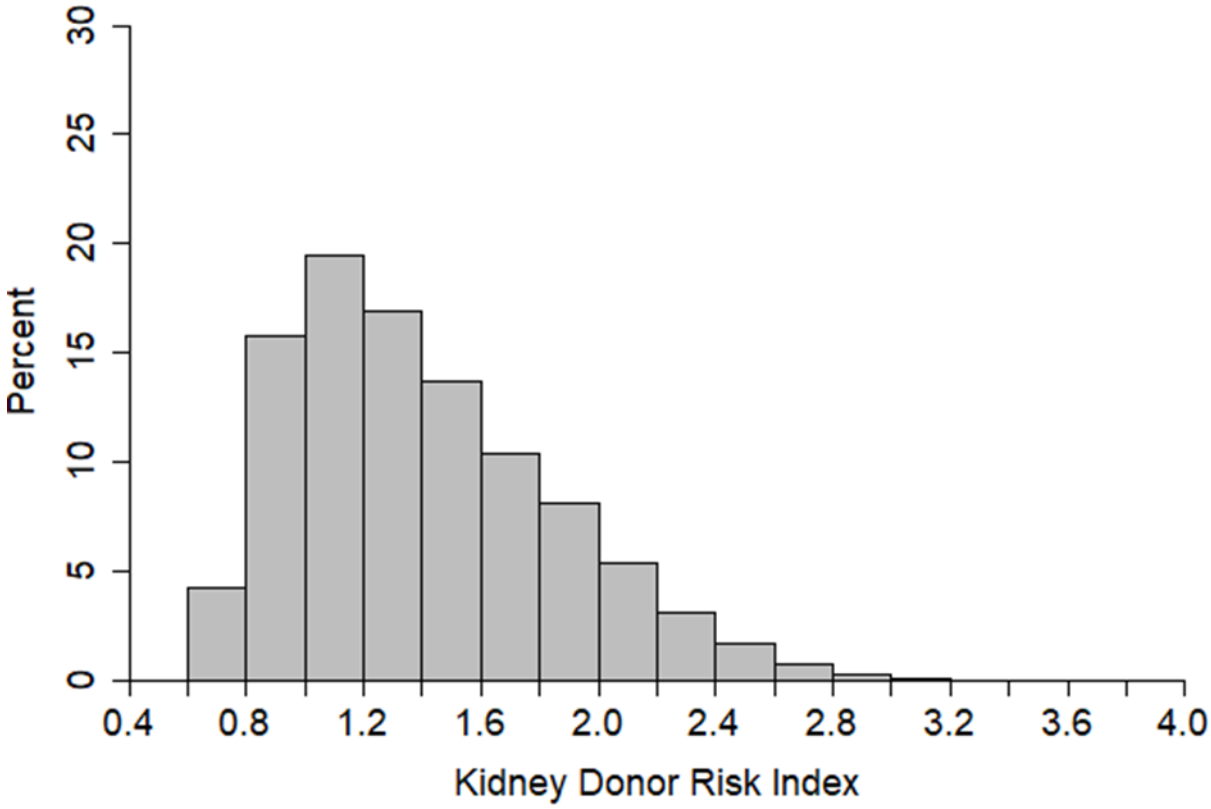

Figure 2: Distribution of the original Kidney Donor Risk Index in UK kidney transplantation population.

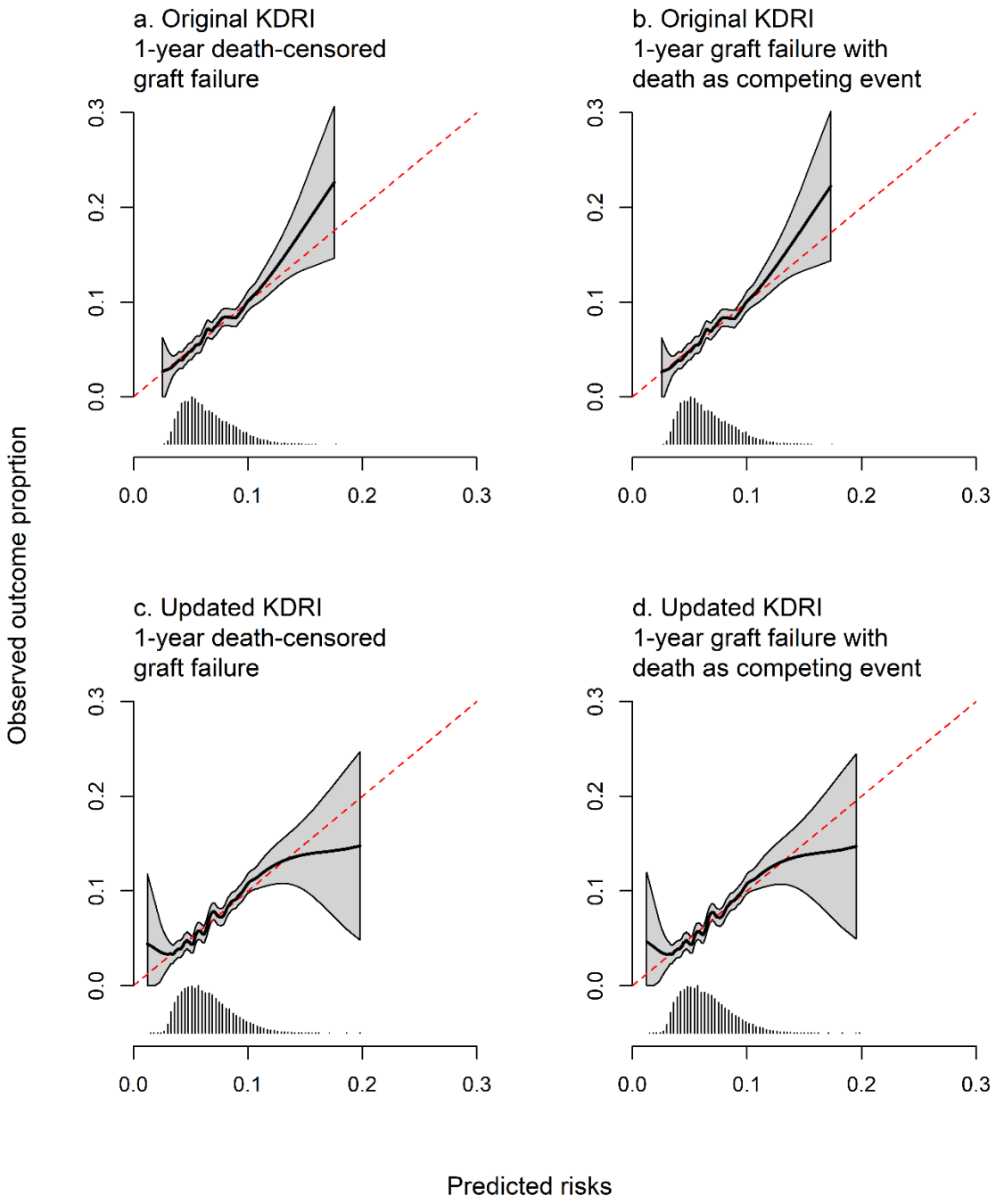

Figure 3: Calibration plots for the original and updated KDRI for predicting 1-year graft failure. The left panels show graft failure censoring at the time of death, and the right panels treat death as a competing event. Below each plot is a histogram of predicted risks. The dashed red line indicates perfect calibration.

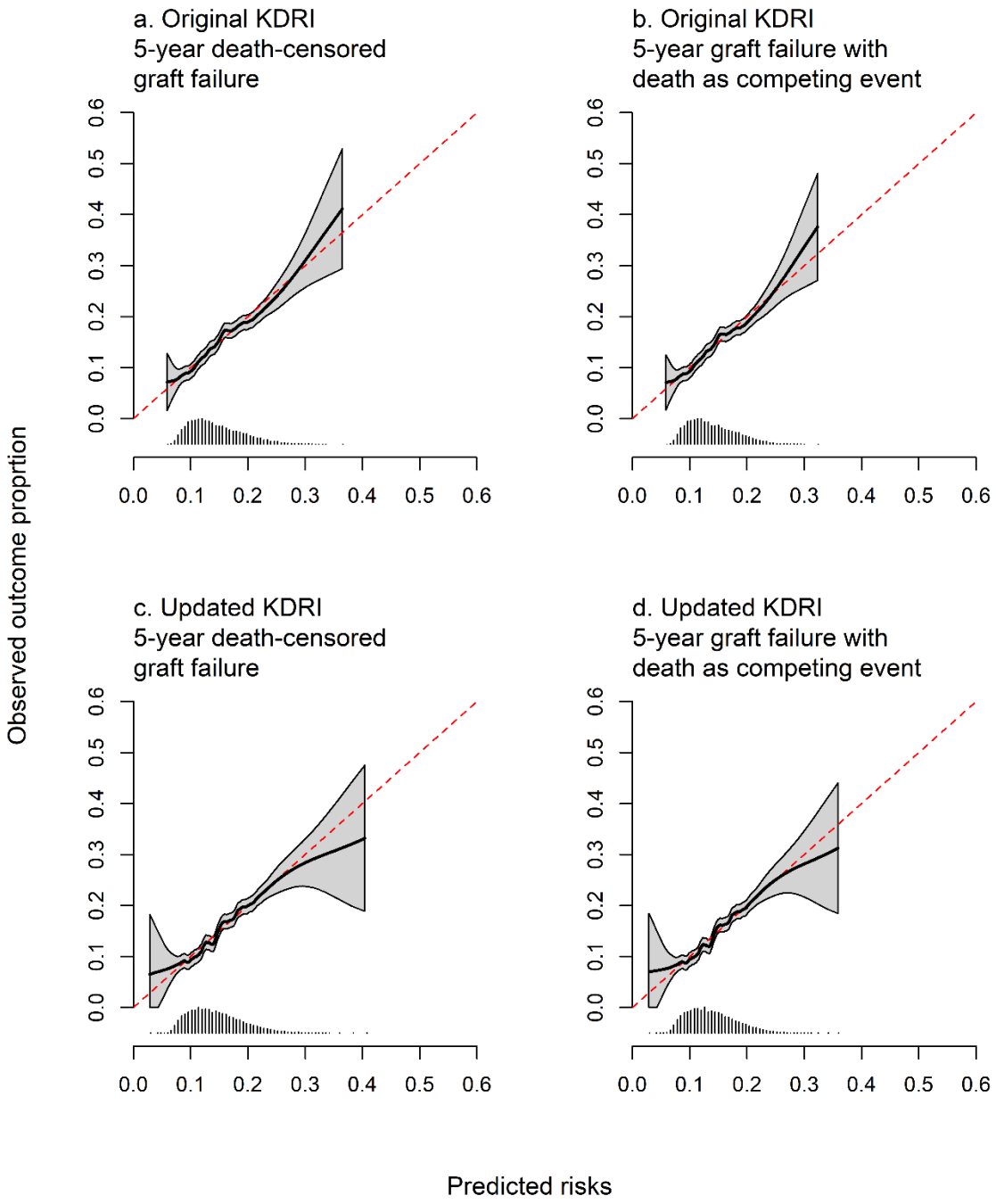

Figure 4: Calibration plots for the original and updated KDRI for predicting 5-year graft failure. The left panels show graft failure censoring at the time of death, and the right panels treat death as a competing event. Below each plot is a histogram of predicted risks. The dashed red line indicates perfect calibration.