

ORIGINAL ARTICLE

Predicting first-year kidney graft loss at the time of donor offer using machine learning and competing-risk analysis in a Latin American cohort

Andrea Garcia-Lopez^{1,2}, Juliana Cuervo-Rojas³, Juan Garcia-Lopez⁴ and Fernando Giron-Luque^{2,5}

¹PhD Program in Clinical Epidemiology, Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine, Pontificia Universidad Javeriana, Bogotá, Colombia, ²Department of Transplant Research, Colombiana de Trasplantes, Bogotá, Colombia, ³Department of Clinical Epidemiology and Biostatistics, Faculty of Medicine, Pontificia Universidad Javeriana, Bogotá, Colombia, ⁴Department of Technology and Informatics, Colombiana de Trasplantes, Bogotá, Colombia and ⁵Department of Transplant Surgery, Colombiana de Trasplantes, Bogotá, Colombia

Correspondence to: Andrea Garcia-Lopez; E-mail: aegarcia@colombianadetrasplantes.com

ABSTRACT

Background. Kidney transplantation faces organ shortages, underscoring the need for early risk stratification of graft loss. We developed and validated a 12-month prediction model that treats death as a competing event.

Methods. We conducted a retrospective cohort study of 2030 adult kidney transplant recipients (2008–2023) from Colombia's largest transplant network. Models included a random survival forest for competing risks (RSF-CR) and Fine–Gray (FG) regression. Internal validation used stratified cross-validation. Model performance was evaluated via discrimination (C-index), calibration and clinical utility (decision curve analysis).

Results. Key predictors included donor type, stroke cause of death (deceased donor), recipient age, donor creatinine, panel reactive antibodies (PRA) class I >20, expanded criteria donor, donor age, years on dialysis, PRA class II >20, donor hypertension, donor–recipient compatibility and retransplantation. The RSF-CR model outperformed the FG, achieving a C-index of 0.87 (versus 0.72) and high sensitivity (88%). It accurately identified low-risk candidates (negative predictive value 98%) and showed a positive net benefit.

Conclusion. We developed and validated a predictive model for first-year graft loss in kidney transplant recipients using a machine learning for competing risks model. The model showed strong discriminative ability and moderate calibration. Further temporal validation in our population and external validation in other clinical contexts is required to ensure its applicability.

Keywords: competing risk, graft failure, kidney transplantation, machine learning, prediction model

Received: 22.12.2025; accepted: 13.3.2026

© The Author(s) 2026. Published by Oxford University Press on behalf of the ERA. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com

KEY LEARNING POINTS

What was known:

- Although first-year graft loss is relatively infrequent, its consequences can be catastrophic for patients, caregivers and transplant teams, and despite generally favourable long-term graft survival and careful preventive strategies, a subset of recipients still experience graft failure within the first year.
- Several donor and recipient risk scores exist, but most were developed in high-income settings, focus on pre- or perioperative variables and seldom account for death as a competing event for graft loss.
- The role of modern machine learning methods for competing risks, and their added value over traditional regression-based models, has been underexplored in kidney transplantation, particularly in Latin American populations.

This study adds:

- We developed and internally validated a 12-month graft loss prediction model using a random survival forest for competing risks built exclusively on information available at the time of donor offer.
- Compared with a Fine–Gray regression model, the machine learning model showed substantially higher discrimination (C-index 0.87 versus 0.72), better identification of low-risk patients (negative predictive value 98%) and a positive net benefit on decision curve analysis.
- Our study provides one of the first large, real-world competing risks machine learning models for kidney graft loss in Latin America, with transparent reporting of discrimination, calibration and clinical utility.

Potential impact:

- This model could support transplant teams in real-time donor acceptance decisions by objectively identifying candidates at risk of first-year graft loss using routinely collected pretransplant variables.
- By separating donor–recipient combinations into low- and higher-risk strata, the tool may improve organ allocation, optimize use of expanded-criteria donors and facilitate risk communication with patients.
- If future temporal and external validations confirm these results, the model could be implemented as a bedside or web-based calculator to enhance individualized prognostication in kidney transplantation programs, including in middle-income countries.

INTRODUCTION

The persistent shortage of donor organs poses a major challenge in kidney transplantation, with thousands of patients remaining on waiting lists annually, many of whom die before receiving a transplant [1, 2]. This scarcity highlights the need for strategies that maximize graft longevity, ensuring each transplanted kidney provides lasting benefits. Precise risk stratification at the time of donor offer is essential to optimize outcomes, allowing clinicians to tailor donor–recipient matching in a way that minimizes both early graft failure and mortality with a functioning graft [3].

Clinical decision-making in transplantation is inherently complex, requiring careful appraisal of donor suitability and recipient risk. A robust pretransplant predictive tool could enhance early identification of high-risk candidates, optimize organ allocation and improve patient and graft outcomes. Although most prognostic models in kidney transplantation focus on medium- to long-term outcomes (5–10 years post-transplant) [4–7], comparatively few address early graft loss—a less frequent but clinically and economically devastating event.

Although prognostic models have become increasingly valuable in clinical practice, their performance in settings with competing risks remains underexplored [3, 8]. Current approaches predominantly rely on Cox regression or survival machine learning techniques [3, 5, 7–9], and few adequately account for competing risks such as when death precedes graft failure [10, 11], which makes the prediction of failure less accurate. Notably, no study to date has applied random survival forest (RSF) methodology with competing risk analysis in kidney transplantation. RSF accommodates non-linear relationships, high-dimensional data and complex interactions without restrictive parametric

assumptions, potentially outperforming traditional predictive methods.

In this study we developed and validated a risk prediction model for first-year graft loss using RSF including death with a functional graft as a competing risk in a large national Colombian cohort. Our objectives were to compare machine learning-based competing risk models with traditional Fine–Gray (FG) regression, evaluate the impact of competing risk adjustment on prediction accuracy and create a practical clinical calculator for risk stratification at the time of donor offer by incorporating time-to-event analysis and proper handling of competing risks. The target population includes adult kidney transplant recipients in Colombia and the model is intended to support clinical decision-making at the time of donor organ offer. It is intended for healthcare professionals involved in transplant evaluation and allocation, such as transplant nephrologists, surgeons and coordinators, with the broader purpose of improving short-term graft survival and resource utilization in transplant programs.

MATERIALS AND METHODS

Study design and population

We conducted a retrospective cohort study of 2030 adult kidney transplant recipients (≥ 18 years of age) who underwent surgery between July 2008 and July 2023 in the Colombiana de Trasplantes, the largest transplant network in Colombia, encompassing four centres nationwide. We excluded patients who developed arterial or venous thrombosis after transplantation because early vascular thrombosis is a surgical/technical complication that typically results in immediate graft loss or

nephrectomy and follows a causal pathway distinct from the immunologic and clinical mechanisms under investigation. The inclusion of these cases would artificially inflate early event rates and bias risk estimates.

Organ allocation in Colombia is regulated by the government through its National Institute of Health (Instituto Nacional de Salud). Deceased-donor kidneys are distributed using geographic criteria (local, regional and national), prioritizing local allocation within each region, and organs are offered nationally when no suitable regional match is identified. Allocation priority is primarily based on donor–recipient compatibility [ABO blood group, human leucocyte antigen (HLA) and age] and recipient-related factors such as time on the waiting list and prior history of living kidney donation. Organ allocation and routine decision-making in Colombia do not incorporate the Kidney Donor Profile Index (KDPI).

After transplantation, recipients are routinely followed by nephrology monthly as part of standard care. The program does not perform protocol biopsies and donor-specific antibody (DSA) monitoring is not routinely available, therefore immunological risk assessment relies on clinical history and baseline immunological information. In our clinical practice, recipients are considered at higher immunological risk if they are retransplant candidates or have evidence of sensitization. Immunosuppression follows a steroid-free maintenance protocol as standard practice. Cardiovascular risk is considered higher when any of the following are present: age >45 years, >4 years on dialysis, diabetes mellitus or established coronary artery disease.

Living donors are typically selected among individuals without major comorbidities. Deceased-donor kidneys are accepted under standard criteria and expanded criteria. However, donation after circulatory death (DCD) is not performed in this setting. During the study period, organ allocation and donor characterization did not routinely incorporate KDPI, and not all KDPI components were consistently registered during the period of study. Therefore, donor risk was described using routinely recorded clinical donor variables and expanded-criteria donor classification.

Data source and collection

The study utilized data from Colombiana de Trasplantes' comprehensive transplant registry. Data collection included baseline donor and recipient characteristics and immunological and clinical parameters. Data access for research was granted on 22 August 2024, after institutional review board approval. To safeguard confidentiality, only the principal investigator (A.G.-L.) had access to identifiable patient information.

Outcomes

The event variable had three possible outcomes: end of follow-up with a functional graft (right censoring), graft loss and death with a functional graft. Our primary outcome was graft loss within the first year post-transplantation, defined as permanent return to dialysis. Death from any cause was considered as a competing event. The outcome was determined based on whichever of the two events—graft loss or death—occurred first. Patients with functioning grafts were right censored either on 31 July 2024 or at the end of the 12-month follow-up period. A total of 205 (10.1%) patients were lost to follow-up and were censored at their last recorded follow-up date.

Independent predictors

The initial selection of independent predictors was informed by a thorough literature review of existing models for kidney transplant outcomes, complemented by clinical expert consultation to ensure the inclusion of clinically relevant variables. We incorporated donor-related characteristics, sociodemographic factors, clinical parameters and recipient medical history.

Donor variables included age, body mass index (BMI), donor type (living versus deceased), hypertension, diabetes, serum creatinine, cause of death and expanded criteria (for deceased donors). Recipient characteristics included sociodemographic factors (age, sex, BMI, adherence, social support network), clinical variables (underlying kidney disease, diabetes, hypertension, peripheral vascular disease, cardiovascular disease, smoking, dialysis type, time on dialysis) and immunological risk factors, including previous transplant, HLA compatibility, panel reactive antibodies (PRA) class I and II and history of blood transfusions (Supplementary File S1, Table A1).

Variable selection was subsequently performed to refine the predictor set to optimize model performance and clinical relevance (Supplementary File S1, Table A4).

Sample size

The sample size was calculated following Steyerberg's recommendations [12] of having at least 10 events per parameter included in the model and a minimum of 100 events in total for validation studies. We expected to observe between 160 and 200 events considering that cumulative graft loss within the first year post-transplant has been reported for 8–10% of cases and we had available data for ≈2000 transplanted patients, which would allow us to include 16–20 parameters in the final model.

Model development, validation and implementation

Model development was a sequential process with the following six steps: data preparation, variable selection, definition of training and validation datasets, model training, model evaluation and model presentation (Fig. 1).

Two types of models were developed: a machine learning model—random survival forest for competing risks (RSF-CR)—and a traditional statistical model for competing risks—FG.

This study adheres to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis using Artificial Intelligence (TRIPOD-AI) guidelines [13].

Step 1: Data preparation

Missing data. We implemented multiple imputation by chained equations (MICE) through the MICE library, performing five imputations with five iterations each. We selected the random forest algorithm as our imputation method due to its demonstrated effectiveness in handling both continuous and categorical variables while capturing complex feature interactions [14]. The imputation process encompassed 12 variables with varying missingness rates: adherence and support network showed identical missingness rates of 10.7%, while HLA compatibility markers (HLA-A, HLA-B and HLA-DR) each exhibited 0.8% missing data. For immunological profiling, we imputed PRA (0.9% for PRA-I and 1.1% for PRA-II), along with key donor characteristics including BMI (2.8% missing), hypertension (2.0%), diabetes (2.2%), cause of death (0.2%) and serum creatinine levels (2.5%).

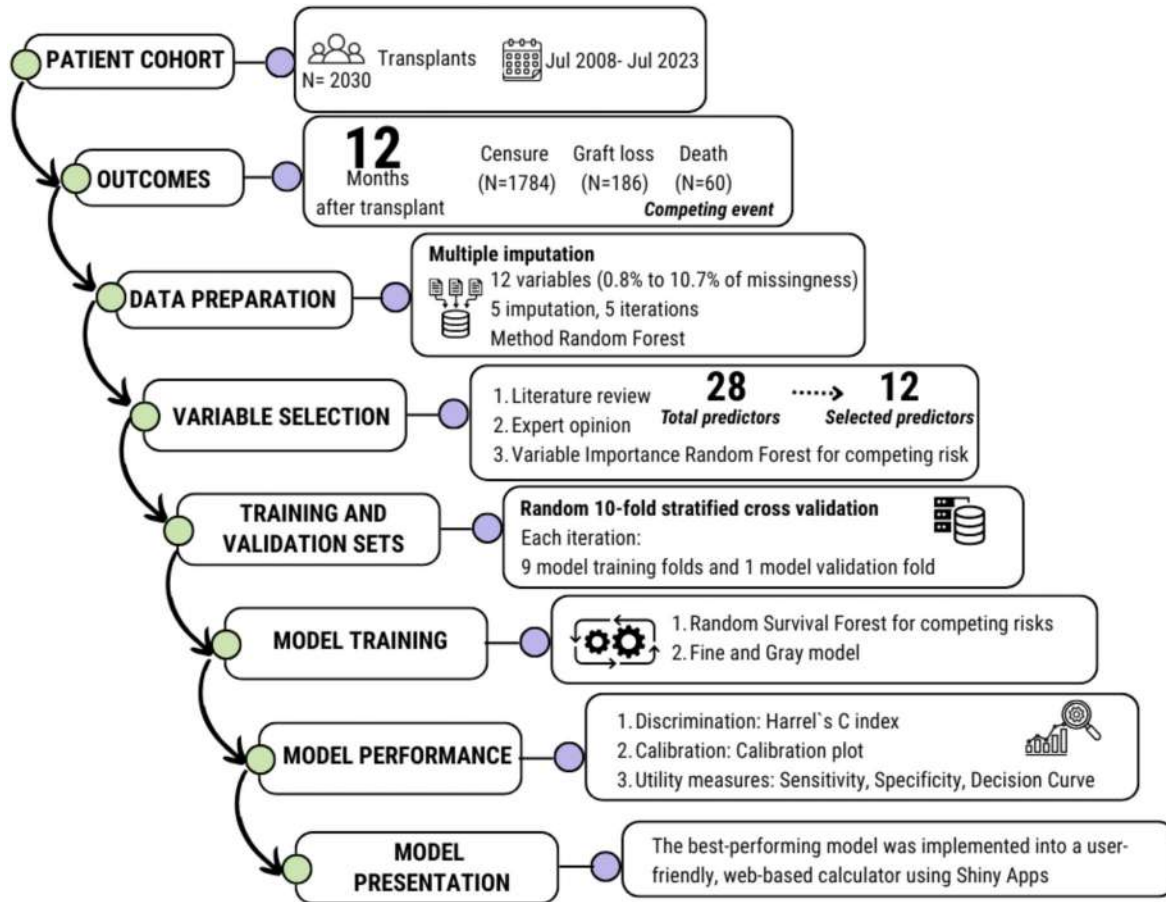


Figure 1: Overview of the study workflow for developing and validating a predictive model for post-transplant outcomes.

Step 2: Variable selection

Relevant variables for the models were identified through a three-step approach:

1. Literature review: We conducted a comprehensive review of prior studies to identify variables consistently associated with the outcome and relevant to our study context.
2. Expert consultation: Clinical and methodological experts were consulted to assess the known and potential clinical significance of candidate variables.
3. RSF-CR feature selection: The RSF-CR method was applied to integrate feature selection directly into the learning process.

Variables were selected based on their variable importance measure (VIMP). Only variables with a VIMP >0 were retained for model training. The final subset of variables was used to train both models (RSF-CR and FG) in subsequent analyses (Supplementary File S1, Table A4).

Step 3: Training and validation data

We implemented a 10-fold stratified cross-validation approach for internal validation. This method was chosen to maintain the proportional distribution of minority classes (graft loss and death) across all folds, given their relatively low incidence in the cohort. Stratification preserved the original event rates in both the training and validation sets, reducing bias and improving generalizability.

The dataset was randomly partitioned into 10 equally sized folds, with each fold containing approximately the same proportion of events (graft loss and death) as the original dataset. In each iteration, nine folds were used for model training and the remaining fold was used for validation. This process was repeated 10 times.

By using stratified cross-validation, we mitigated the risk of unbalanced representation of rare outcomes in individual folds. This approach also helped evaluate model stability across different data subsets, reducing overfitting and improving validity.

Step 4: Model training and validation

RSF-CR model. The RSF-CR model, introduced by Ishwaran et al. [15, 16], extends the standard random forest algorithm originally developed by Breiman [17]. As our primary goal was risk prediction (rather than treatment effect evaluation), we used the event-specific cumulative incidence function (CIF) as the model output instead of the hazard function.

The RSF-CR model was constructed through the following steps:

1. Bootstrap sampling: Draw k bootstrap samples from the training dataset to construct k survival trees.
2. Tree growth for each bootstrap sample: At each node, randomly select candidate variables and split the node to maximize CIF separation using longrankCR from the randomForestSRC package.

3. Stopping rule: Tree growth terminates when terminal nodes reach the minimum size threshold (nodesize parameter).
4. CIF estimation: Compute the cause-specific CIF for each tree.
5. Aggregation: Average CIF estimators across all k trees to obtain the final prediction.

A detailed mathematical description of the RSF-CR methodology is available in the original work by Ishwaran *et al.* [15].

FG model. The FG modelling approach was implemented as follows: the subdistribution hazard function was specified treating competing events through weighted risk sets. Next, previously selected covariates based on the VIMP were incorporated. The model was then fitted using maximum likelihood estimation with inverse probability weighting to handle censoring and competing risks. Subsequently, the CIF was derived from the estimated subdistribution hazards to provide clinically interpretable risk predictions. Finally, model performance was assessed. A detailed mathematical description of the FG methodology is available in the original work by Fine *et al.* [18].

Class imbalance considerations. The primary outcome (graft loss within 1 year) exhibited class imbalance (9.1% event rate), consistent with real-world kidney transplant populations. We intentionally avoided sample balancing techniques (e.g. oversampling or undersampling) to preserve the natural data distribution and ensure a clinically realistic model validation. While class imbalance can affect traditional accuracy metrics, our chosen algorithms—RSF-CR and FG—are inherently robust to imbalance through ensemble learning and maximum likelihood estimation, respectively. Rather than artificially optimizing performance, we prioritized discrimination (C-index) and potential clinical utility [net benefit, decision curve analysis (DCA)], ensuring that model performance reflects meaningful trade-offs in real-world decision-making. This approach aligns with current recommendations favouring external validity over artificial performance enhancement in clinical prediction models [19, 20].

Model output. Both the RSF-CR and FG models generate individualized predictions of the CIF for graft loss, accounting for death as a competing event. For each patient, the model estimates the probability of graft loss at 12 months post-transplant, while appropriately considering the competing risk of death. Although each model generates continuous probabilities, they were translated into risk categories using a threshold to identify high risk of loss. This threshold was selected based on the observed incidence of the outcome, clinical relevance and risk stratification goals.

Step 5: Model performance evaluation

We evaluated model performance following methodological standards for clinical prediction models as outlined by Steyerberg *et al.* [21]. Our comprehensive assessment examined both statistical performance—including discrimination and calibration—and potential clinical utility. Both the RSF-CR and FG models were evaluated for discrimination and calibration. The best-performing model was subsequently evaluated for potential clinical usefulness through DCA (Supplementary File S1, Table A2). This approach aims at selecting a model that provides both statistically robust predictions and potential meaningful support for clinical decision-making [21–23].

Model discrimination. We assessed discrimination with Harrell's C-index and selected the model that achieved superior discrim-

ination (higher C-index) while preserving calibration. Interpretation followed standard benchmarks: 0.5 indicates no discrimination, 0.7 denotes acceptable clinical performance and ≥ 0.8 reflects strong discriminative ability.

Model calibration. We evaluated calibration using calibration curves, which visually compare the predicted and observed probabilities of the primary outcome. The curves indicate the agreement between model predictions and realized event rates across risk groups.

Model potential utility. The potential clinical utility of the final model was assessed using decision analytic measures, including sensitivity, specificity and net benefit, complemented by DCA [21]. Because discrimination and calibration alone do not capture practical value, we quantified the trade-off between benefits (true positives) and harms (false positives) [21]. The DCA was constructed according to the chosen threshold that defined patients as at expected risk (below the threshold) and at high risk of graft loss (above the threshold).

Sensitivity and specificity were reported for a predefined threshold (0.09). Net benefit incorporates the relative clinical consequences of misclassification by weighting false positives according to their harm. We also conducted a threshold-agnostic evaluation by plotting net benefit across a clinically relevant range and comparing the model against two default strategies: 'treat all', where all patients are treated as high risk and receive very intensive follow-up, and 'treat none', where all patients receive usual care (no high-risk management). The model demonstrated superior clinical utility when its net benefit exceeded that of these strategies, indicating its potential to optimize decision-making and clinical management of transplant recipients. All analyses were performed using the *rmda* package in R version 4.5.2 (R foundation for Statistical Computing, Vienna, Austria), with results presented as decision curves for intuitive interpretation.

Step 6: Model presentation

A web-based risk prediction tool. After evaluating both models in terms of statistical performance (discrimination and calibration) and potential clinical utility (DCA), the best-performing model was implemented into a user-friendly, web-based calculator using Shiny apps. This tool provides individualized risk predictions for graft loss while accounting for death as a competing event, facilitating pre-transplant assessments for further prospective validation.

OPEN SCIENCE

Data and code sharing

The code used to develop and evaluate all models is provided in the [Supplementary File S2](#). Access to the dataset may be granted upon reasonable request to the corresponding author, subject to approval by Colombiana de Trasplantes. If approved, data will be shared in anonymized form, in compliance with institutional policies, applicable ethical guidelines and legal regulations.

Research ethics considerations

This study was approved by the Research and Ethics Committee of the Faculty of Medicine of Pontificia Universidad Javeriana (REF 266/2021), in accordance with national and local ethical regulations, as well as the principles outlined in the Declaration of

Helsinki. Given the retrospective design and nature of the study, the Committee waived the requirement for individual informed consent.

RESULTS

Outcome frequency and recipient characteristics

The study had 2030 kidney transplant patients. A total of 186 patients experienced graft loss (9.1%) and 60 patients died with a functional graft (2.96%) within 12 months after transplantation. The mean age of patients who experienced graft loss and death was significantly higher (45.9 and 53.2 years, respectively) compared with the no-event group (43.1 years). Sex distribution was similar across groups, with males representing \approx 59% of the total population. Social support network adequacy and adherence rates were high across all groups, with no significant differences. Notably, patients who lost the graft had a higher prevalence of diabetes (19.4%) compared with the no-event group (16.0%), and their mean BMI was slightly higher (24.1) compared with the no-event group (23.7).

Diabetes prevalence was significantly higher among patients who died (41.7%) compared with the no-event group (16.0%). Peripheral vascular disease was also more common among deceased patients (5.0%) compared with the no-event group (0.7%). BMI was slightly higher in the death group (mean 25.1) compared with the no-event group (mean 23.7). The type of underlying kidney disease varied significantly, with diabetes being more prevalent among deceased patients (35.0%) compared with the no-event group (11.5%). Time on dialysis was longer for patients who died (mean 64.3 months) compared with the no-event group (mean 48.6 months). Compatibility of HLA markers showed no significant differences across groups. The characteristics of recipients are presented in Table 1.

Donor characteristics

Table 2 presents a comparative analysis of donor characteristics across three groups. Deceased donors were predominant, especially in the graft loss (88.2%) and death (71.7%) groups, compared with 64.9% in the no-event group. Living donors accounted for 35.2% of the no-event group but were less represented in those with adverse outcomes. The mean donor age increased progressively from the censored group (40.9 years) to the graft loss (44.6 years) and death (50.9 years) groups, with a similar trend observed in BMI. Comorbidities such as diabetes and hypertension were more prevalent in the graft loss and death groups. Notably, stroke as a cause of death (deceased donors) was reported in half of the donors in the graft loss (50%) and death (50.0%) groups. Serum creatinine levels were slightly elevated in the graft loss group. Additionally, the proportion of expanded criteria donors was significantly higher in the graft loss (39.2%) and death (48.3%) groups compared with the no-event group (21.6%).

Cumulative incidence of graft loss and death

Figure 2 shows the cumulative incidence of graft loss and patient death over the 12-month follow-up period. Graft loss showed a progressive increase from 0.2% at 1 month to 9.1% at 12 months, with the most notable increase occurring within the first 6 months post-transplant (6.4%). In contrast, mortality remained relatively low throughout the period, starting at 0.1% in

the first month and reaching 3.0% by month 12. At 12 months, overall patient and graft survival was 87.8%.

Variable selection

A total of 28 variables were initially screened. These variables were incorporated into an RSF-CR model. VIMP was assessed and ultimately 12 variables with a VIMP $>$ 0 were retained. These 12 variables were subsequently used as predictors in both the FG and RSF-CR model. The VIMP for the final model is shown in Fig. 3. The complete variable selection process can be found at [Supplementary File S1, Table A4](#).

Training and validation cohorts

A 10-fold cross-validation was conducted. One-fold was randomly selected to define the training and validation cohorts. The training ($n = 1827$) and validation ($n = 203$) cohorts showed comparable distributions in key variables: donor serum creatinine [mean \pm standard deviation (SD): 0.92 ± 0.37 versus 0.89 ± 0.70], donor age (41.8 ± 14.0 versus 39.4 ± 14.8 years) and donor BMI (25.4 ± 3.75 versus 24.9 ± 3.98). Recipient age (43.8 ± 13.2 versus 42.3 ± 13.2 years) and time on dialysis (50.3 ± 46 versus 47.7 ± 43 months) were closely aligned. Categorical variables—such as donor type (deceased donors 67.8% versus 61.6%), causes of death, expanded criteria donors (24.6% versus 19.2%), donor hypertension (17.1% versus 12.8%) and donor diabetes (2.2% versus 1.0%)—showed broadly similar patterns, with small differences consistent with sampling variability in the smaller validation set ([Supplementary File S1, Table A3](#)).

MODEL TRAINING AND VALIDATION

Random survival forest for competing risks

To optimize the RSF-CR model, we conducted systematic hyperparameter tuning using the randomForestSRC package (version 2025) in R [24]. The number of trees was tested across a range of 500–1500 to ensure robust ensemble learning while maintaining computational efficiency. The mtry parameter, which controls the number of variables randomly sampled at each split, was evaluated between 1 and 9, with lower values enhancing sensitivity. Node size, determining the granularity of terminal nodes, was adjusted between 10 and 20 observations to balance model complexity and generalizability. CIF thresholds ranging from 0.09 to 0.3 were tested to determine the optimal cut-off, with the lower bound (0.09) selected based on the observed event incidence in the study population (9%). This approach ensured that the threshold reflected the underlying epidemiology while aiming to maximize sensitivity and select the model with the highest C-index.

FG model

The FG subdistribution hazard model was trained to evaluate the risk of graft loss in the presence of death as a competing event. The model accounts for the cumulative incidence of graft loss while appropriately weighting competing events to avoid biased predictions. To ensure comparability with the RSF-CR analysis, the same covariates were included. All analyses were implemented in R using the randomForestSRC package (version 2025) [24], cmprsk package (version 2024) [25] and riskRegression package in R version 2025 [26, 27]

Table 1: Characteristics of kidney transplant recipients according to the observed outcome.

Characteristics	No event (n = 1784)	Graft loss (n = 186)	Death (n = 60)	Total (N = 2030)	P-value
Age (years, mean (SD))	43.1 (13.1)	45.9 (13.5)	53.2 (11.8)	43.6 (13.2)	<.001
Sex, n (%)					.303
Female	736 (41.3)	66 (35.5)	20 (33.3)	822 (40.5)	
Male	1048 (58.7)	120 (64.5)	40 (66.7)	1208 (59.5)	
Social support, n (%)					.775
Functional	932 (52.2)	94 (50.5)	35 (58.3)	1061 (52.3)	
Inadequate	852 (47.8)	92 (49.5)	25 (41.7)	969 (47.7)	
Non-adherence, n (%)	146 (8.2)	17 (9.4)	163 (8.3)	146 (8.2)	.229
History of smoking, n (%)	489 (27.4)	58 (31.2)	20 (33.3)	567 (27.9)	.554
History of diabetes, n (%)	286 (16.0)	36 (19.4)	25 (41.7)	347 (17.1)	<.001
History of coronary disease, n (%)	57 (3.2)	8 (4.3)	4 (6.7)	69 (3.4)	.451
Peripheral vascular disease, n (%)	13 (0.7)	2 (1.1)	3 (5.0)	18 (0.9)	.006 94
BMI, mean (SD)	23.7 (3.86)	24.1 (3.89)	25.1 (3.38)	23.8 (3.85)	.0364
Underlying kidney disease, n (%)					.006 98
Congenital	31 (1.7)	1 (0.5)	0 (0)	32 (1.6)	
Unknown	830 (46.5)	78 (41.9)	23 (38.3)	931 (45.9)	
Diabetes	206 (11.5)	30 (16.1)	21 (35.0)	257 (12.7)	
Glomerular	406 (22.8)	42 (22.6)	11 (18.3)	459 (22.6)	
Hypertension	181 (10.1)	23 (12.4)	2 (3.3)	206 (10.1)	
Obstructive	57 (3.2)	5 (2.7)	1 (1.7)	63 (3.1)	
Other	73 (4.1)	7 (3.8)	2 (3.3)	82 (4.0)	
Dialysis type					.287
Haemodialysis	920 (51.6)	114 (61.3)	32 (53.3)	1066 (52.5)	
Peritoneal dialysis	688 (38.6)	61 (32.8)	23 (38.3)	772 (38.0)	
Non-dialysis therapy	176 (9.9)	11 (5.9)	5 (8.3)	192 (9.5)	
Months on dialysis, mean (SD)	48.6 (45.4)	59.2 (46.1)	64.3 (47.5)	50.0 (45.7)	.001 66
Previous transplant, n (%)	56 (3.1)	6 (3.2)	2 (3.3)	64 (3.2)	1
Transplant number, n (%)					.527
1	1728 (96.9)	181 (97.3)	58 (96.7)	1967 (96.9)	
2	52 (2.9)	4 (2.2)	1 (1.7)	57 (2.8)	
3	4 (0.2)	1 (0.5)	1 (1.7)	6 (0.3)	
History of blood transfusions, n (%)	682 (38.2)	83 (44.6)	17 (28.3)	782 (38.5)	.132
PRA class I >20, n (%)	531 (29.8)	71 (38.2)	11 (18.3)	613 (30.2)	.0206
PRA class II >20, n (%)	545 (30.5)	69 (37.1)	12 (20.0)	626 (30.8)	.0789
Compatibility HLA-A, n (%)					.414
0	485 (27.2)	54 (29.0)	24 (40.0)	563 (27.7)	
1	911 (51.1)	96 (51.6)	28 (46.7)	1035 (51.0)	
2	388 (21.7)	36 (19.4)	8 (13.3)	432 (21.3)	
Compatibility HLA-B, n (%)					.193
0	603 (33.8)	72 (38.7)	30 (50.0)	705 (34.7)	
1	874 (49.0)	86 (46.2)	24 (40.0)	984 (48.5)	
2	307 (17.2)	28 (15.1)	6 (10.0)	341 (16.8)	
Compatibility HLA-DR, n (%)					.856
0	291 (16.3)	34 (18.3)	12 (20.0)	337 (16.6)	
1	900 (50.4)	99 (53.2)	31 (51.7)	1030 (50.7)	
2	593 (33.2)	53 (28.5)	17 (28.3)	663 (32.7)	

Variable definitions can be found in [Supplementary File S1](#).

MODEL PERFORMANCE

Model discrimination

Table 3 presents a comparative analysis of performance metrics between the FG and RSF-CR model. The RSF-CR model achieved superior overall discrimination, as evidenced by a higher C-index of 0.87, indicating a strong ability to distinguish between patients at different risks of graft loss with death as a competing event, while the FG model showed acceptable performance with a C-index of 0.72.

Model calibration

Calibration curves for both models are shown in Fig. 4. The x-axis represents the predicted risk and the y-axis the observed risk, both expressed as percentages. The diagonal line corresponds to perfect calibration, where predicted probabilities equal observed outcomes. The stepped black line depicts the actual observed incidence across prediction intervals, reflecting the model's calibration performance. Overall, the RSF-CR model curve tracks the reference line more closely than the FG model, indicating better calibration; however, at higher predicted risks it falls below

Table 2: Characteristics of donors according to the observed outcome.

Characteristics	No event (n = 1784)	Graft loss (n = 186)	Death (n = 60)	Total (N = 2030)	P-value
Donor type, n (%)					<.001
Deceased	1157 (64.9)	164 (88.2)	43 (71.7)	1364 (67.2)	
Living	627 (35.1)	22 (11.8)	17 (28.3)	666 (32.8)	
Age (years), mean (SD)	40.9 (13.9)	44.6 (14.0)	50.9 (14.4)	41.5 (14.1)	<.001
BMI, mean (SD)	25.3 (3.74)	25.7 (3.86)	26.4 (4.44)	25.4 (3.78)	.131
Diabetes, n (%)	35 (2.0)	6 (3.2)	2 (3.3)	43 (2.1)	.628
Hypertension, n (%)	273 (15.3)	49 (26.3)	16 (26.7)	338 (16.7)	<.001
Stroke cause of death (deceased donor), n (%)	604 (33.9)	93 (50.0)	30 (50.0)	727 (35.8)	<.001
Creatinine (mg/dl), mean (SD)	0.912 (0.357)	1.01 (0.511)	0.927 (0.325)	0.921 (0.373)	.0106
Expanded criteria, n (%)	386 (21.6)	73 (39.2)	29 (48.3)	488 (24.0)	<0.001

Variable definitions can be found in [Supplementary File S1](#).

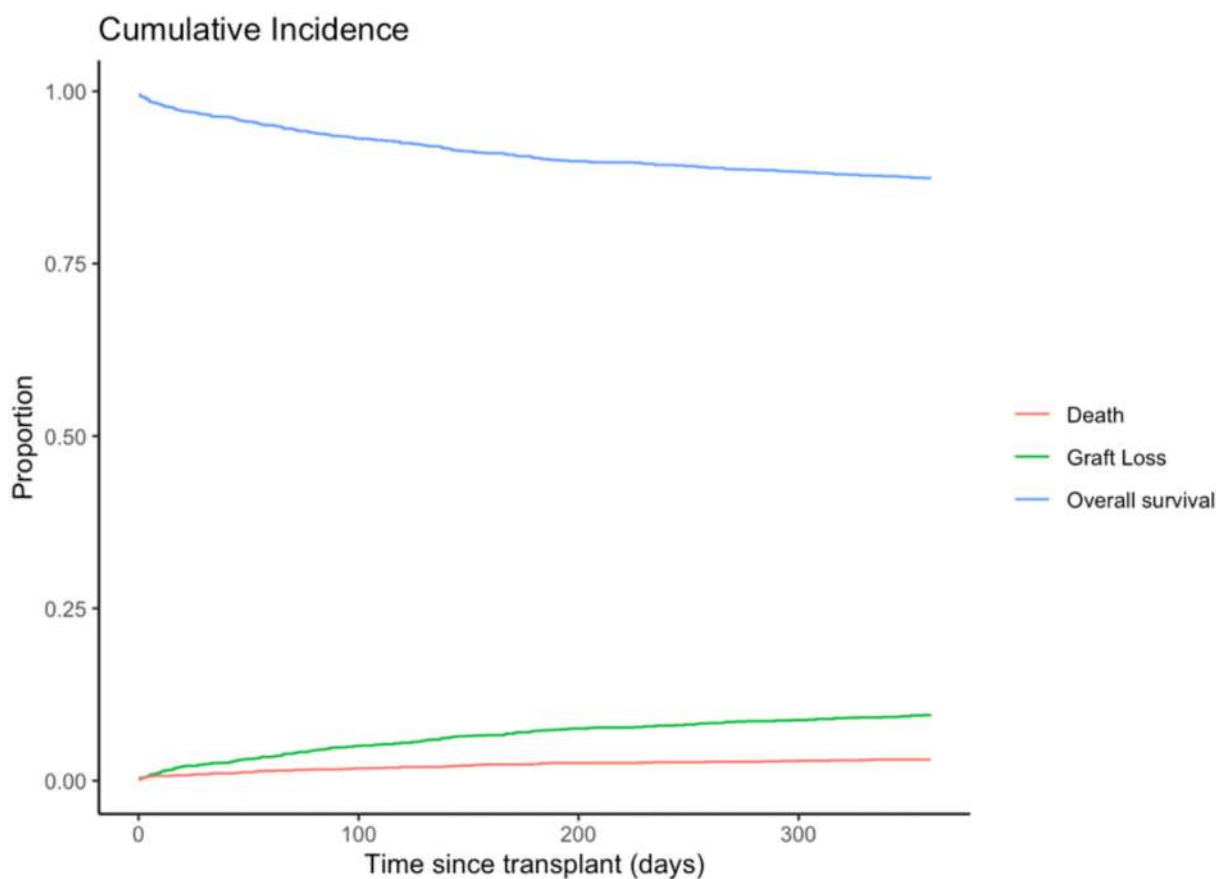


Figure 2: Cumulative incidence curves for post-transplant outcomes during the first year post-transplant. The graph illustrates three key endpoints: death, graft loss (failure of the transplanted organ) and overall survival (composite of death-censored graft survival). Time is displayed in days post-transplant (x-axis), with the y-axis representing the cumulative probability of each event.

the diagonal, suggesting risk overestimation. In contrast, the FG model tends to underestimate at low risks and overestimate at high risks of graft loss.

MODEL UTILITY MEASURES

Sensitivity, specificity and predictive values

Table 3 summarizes model potential utility (sensitivity, specificity and predictive values). At the selected threshold, the RSF-CR model achieved higher sensitivity (88% versus 56%) and a

slightly better F1 score (0.66 versus 0.63), whereas the FG model showed greater specificity (79% versus 55%) and positive predictive value (PPV; 21% versus 15%). The RSF-CR model had a higher negative predictive value (NPV; 95% for FG, 98% for RSF-CR). These profiles translate into different error patterns: the RSF-CR model generates more false positives while the FG model produces more false negatives. After evaluating multiple thresholds, we chose the operating point that maximized sensitivity while retaining a reasonable specificity and the highest concordance index (C-index). This clinically oriented choice prioritizes minimizing false negatives—so most high-risk patients are

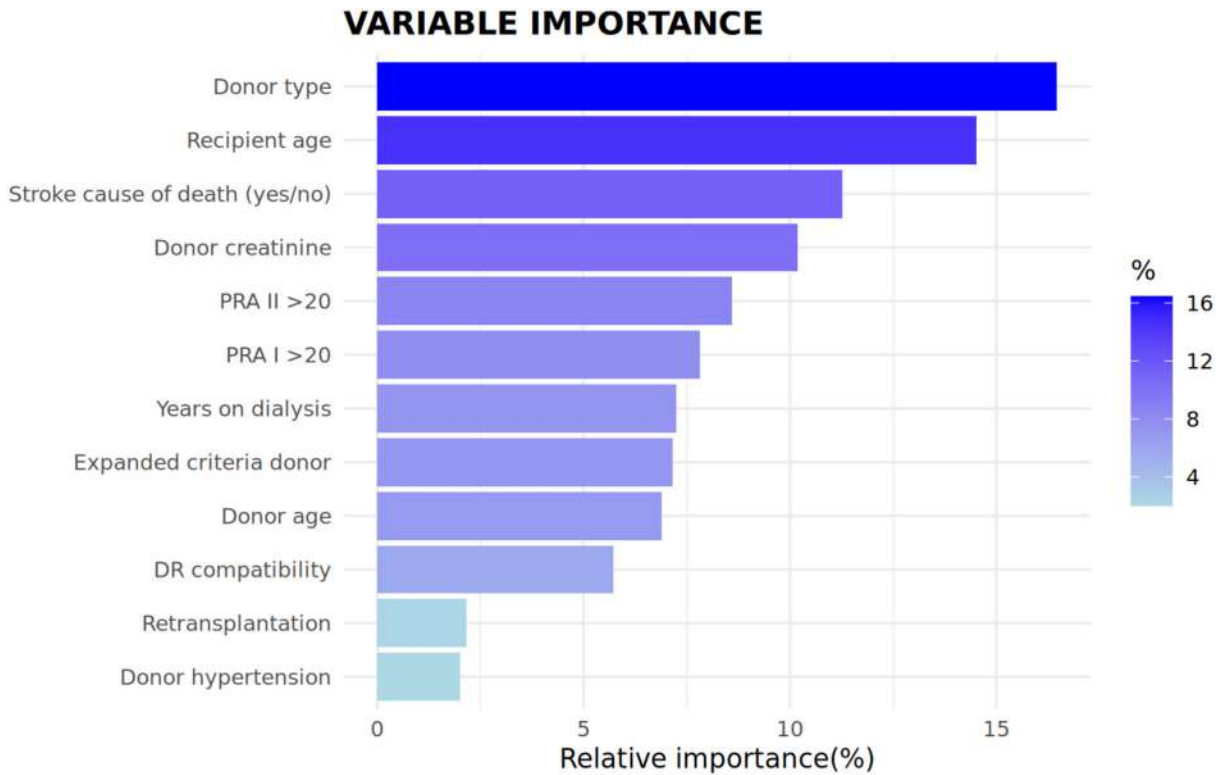


Figure 3: Variable importance plot for the prediction model. Bars display the relative contribution of donor- and recipient-level predictors to the RSF-CR model's performance. 'Donor type' was the most influential variable (highest importance). Importance values are expressed as percentages, scaled to the maximum contribution.

Table 3: Comparison of performance metrics for the prediction of graft loss with death as a competing event using the FG and RSF-CR models.

Metric	FG	RSF-CR
C-index	0.72	0.87
Accuracy	0.75	0.57
F1 score	0.63	0.66
Sensitivity	56%	88%
PPV	21%	15%
Specificity	79%	55%
NPV	95%	98%

These models were developed using a CIF threshold of 0.09, selected to maximize sensitivity and identify the model with the best C-index. The RSF-CR model was trained using the following hyperparameters: ntree = 1000, nodesize = 18, mtry = 3, importance = TRUE, splitrule = 'logrankCR', classwt = class_weights [class_weights <- c(1, 50, 1)], nsplit = 5, seed = 1 and block.size = 1.

identified early for intensified monitoring and implementation of preventive measures—accepting some increase in false positives (i.e. intensified follow-up rather than unnecessary transplant exclusion).

Decision curve for the final model

The DCA in Fig. 5 evaluates the potential clinical utility of the RSF-CR model. The y-axis shows the standardized net benefit—the balance between benefits (true positives) and harms (false positives)—across a range of risk thresholds on the x-axis. Model performance is compared with two default strategies: 'treat all',

where all patients are treated as high risk and receive intensified follow-up, and 'treat none', where all patients receive usual care (no high-risk management).

The RSF-CR model yields a higher net benefit than both defaults across a clinically relevant threshold range (≈0.09–0.25), indicating added value by directing intensive follow-up to patients with higher predicted risk with closer monitoring or preventive strategies while avoiding unnecessary intensification in lower-risk patients. Although the net benefit decreases as thresholds become higher, it consistently remains superior to the default strategies, supporting its potential clinical utility.

Model presentation

Based on the performance and potential clinical utility of the RSF-CR model, we developed an interactive, web-based calculator to facilitate individualized risk assessment for graft loss. This tool provides clinicians with real-time estimates of the 1-year cumulative incidence of graft loss, explicitly accounting for death as a competing event. The calculator is publicly available at <https://kidneytransplantmodels.shinyapps.io/Pretransplant-calculator/>

DISCUSSION

In this study, we developed and internally validated a novel model using RSF-CR to predict first-year graft loss at the time of kidney offer. The model incorporates both donor and recipient characteristics and accounts for death as a competing event. The model showed strong discriminative performance (C-index: 0.87), outperforming the traditional FG model (C-index: 0.72). A

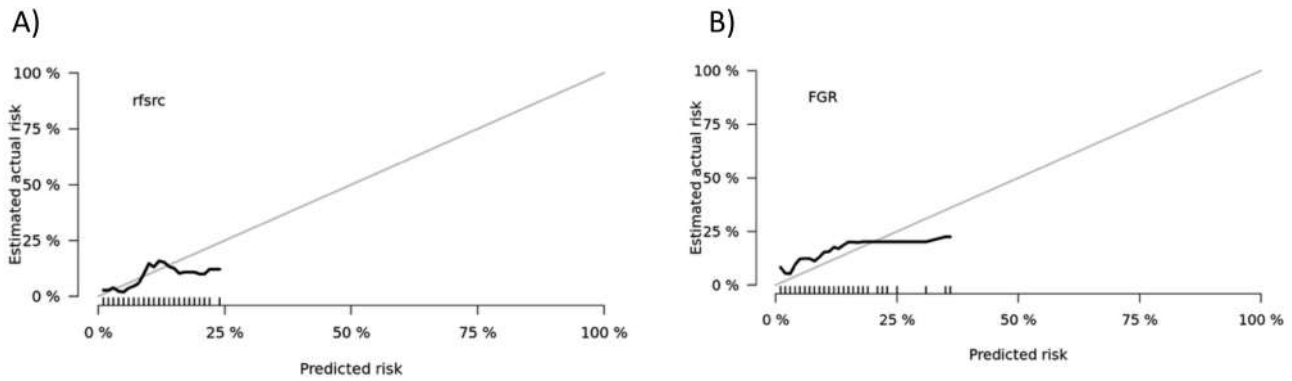


Figure 4: Calibration plots: (A) RSF-CR model and (B) FG model. Calibration plots comparing predicted risk (x-axis) and observed risk (y-axis). The diagonal line represents perfect calibration (predicted = observed).

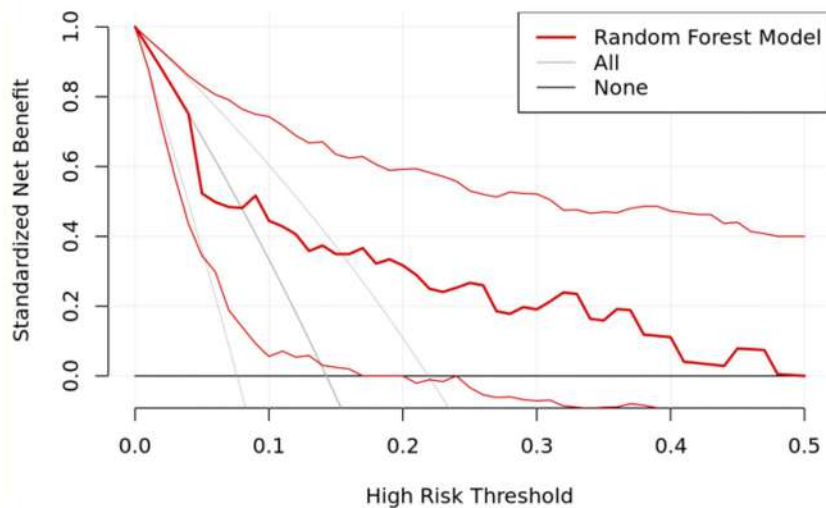


Figure 5: DCA of the RSF-CR model. Standardized net benefit (y-axis) for the model (red) is plotted against two default strategies—'all' (grey: everyone treated as with a high risk of graft loss) and 'none' (black: usual care without high-risk management)—across probability thresholds (x-axis). The model outperforms both defaults for thresholds ≈ 0.09 – 0.25 .

total of 12 clinically relevant predictors were included and the model is presented as an accessible web-based calculator that might be used to support real-time clinical decision-making during donor offer evaluation once further external validation is conducted.

Key findings and clinical implications

The model identified several clinically established predictors of graft loss, including donor-related factors [expanded criteria status, donor type (living versus deceased), age, hypertension history, donor biopsy and cause of death], recipient characteristics (age, time on dialysis, previous transplant, diabetes) and immunological characteristics (PRA class I >20, PRA class II >20, donor-recipient compatibility) [3, 5, 9, 28]. The predominance of donor-related variables in our analysis underscores their critical role in determining early graft outcomes, consistent with current understanding of how donor quality and senescence interact with recipient factors to influence transplant success. Notably, our findings align with known pathophysiological mechanisms, reinforcing the model's biological plausibility and clinical relevance.

Some traditionally recognized risk factors—such as hepatitis C infection status (uncommon in our study population) and primary chronic kidney disease aetiology (often undocumented due to limited access to kidney biopsy)—did not emerge as significant predictors in our cohort. This population-specific variation underscores the importance of developing locally tailored prediction models that accurately reflect the clinical and epidemiological context [21, 23, 28–30].

Comparative analysis showed that the RSF-CR model offered superior discrimination and a sensitivity-first profile (C-index 0.87 versus 0.72, sensitivity 88% versus 56%, F1 0.66 versus 0.63), together with a higher NPV (98% versus 95%). This supports its use as a screening tool to avoid missing truly high-risk cases. In contrast, the FG approach had better specificity (79% versus 55%) and a slightly higher PPV (21% versus 15%), but at the cost of substantially lower sensitivity—implying more false negatives and a greater risk of overlooking high-risk pairs. The RSF-CR's high false positivity and low PPV indicates that some pairs flagged as high risk would ultimately not experience 1-year graft loss. Therefore, a high-risk classification should prompt targeted multidisciplinary reassessment and risk mitigation strategies rather than automatic organ refusal.

Importantly, this model has undergone only internal validation, thus its clinical applicability needs further evaluation. External validation across diverse centres and prospective evaluation of real-world impact are still required. In the interim, the model may be used to trigger a targeted, preferably multidisciplinary, clinical reassessment that considers unmodeled factors, leverages complementary diagnostic information and guides perioperative and post-transplant mitigation strategies (e.g. optimizing cold ischaemia, tailoring induction and maintenance immunosuppression and intensifying monitoring). Such a risk-adapted approach supports organ stewardship and patient safety while recognizing the constraints imposed by limited organ availability.

Comparison with other models

Our RSF-CR model builds upon and complements existing prediction tools for kidney transplant outcomes [4–6, 31–38], as identified in the systematic review by Riley *et al.* [3]. While previous models have provided valuable frameworks for risk assessment, our approach introduces several refinements tailored to enhance potential clinical utility and methodological robustness. The systematic review identified 28 models predicting graft loss, most of which were developed in US populations. These models have significantly advanced the field, although their application in other healthcare settings requires further validation. This study reflects real-world kidney transplantation practice in a Latin American middle-income setting, where follow-up is delivered through structured monthly nephrology visits, but certain tools commonly used in high-income registries—such as protocol biopsies, routine DSA surveillance, DCD donation and KDPI-based donor stratification—are not routinely implemented. Our study expands the set of prediction tools by incorporating data from a Colombian cohort, offering insights that may be applicable to similar middle-income clinical settings. On the other hand, the KDPI/Kidney Donor Risk Index (KDRI) were not part of the organ allocation system in our setting and were therefore not routinely calculated or recorded during the study period; consequently, we did not include the KDPI/KDRI as predictors, prioritizing the use of routinely available donor and recipient information to maximize the model's applicability in our clinical context. We acknowledge, however, that this choice may limit direct comparability with widely used international models developed in registries where the KDPI/KDRI are standard.

Existing models predominantly focus on long-term outcomes (5–10 years post-transplant) [4–7], which are essential for understanding overall graft longevity. Our work complements these efforts by targeting medium-term graft loss (1-year post-transplant), a period with distinct clinical implications for patient management and resource allocation.

Notably, some models (e.g. the Maryland Aggregate Pathology Index) [31] leverage histopathological variables, which can offer granular prognostic insights. However, such data are not universally available at the time of donor offer. Our model prioritizes routinely collected clinical variables, ensuring broader applicability across diverse clinical environments.

Methodologically, current models primarily use Cox regression and do not account for competing risks, such as patient death [3]. Our RSF-CR model addresses this by incorporating competing risk analysis, aligning with contemporary recommendations. This approach provides a more accurate prediction of graft failure risk as it accounts for the possibility of death with a functional graft [39].

Calibration, a key aspect of model performance, has been infrequently reported in previous studies [40, 41]. In our validation, we evaluated both discrimination and calibration, applying rigorous internal validation procedures. Although the RSF-CR model showed a slight overestimation at high risks, its overall calibration supports its clinical applicability. Furthermore, while existing tools are often embedded in static scoring systems, we developed an interactive web-based calculator that may be used for real-time risk assessment, facilitating further validation to support clinical decision-making during donor evaluation [21].

Strengths

This study introduces significant methodological innovations in kidney transplant outcome prediction. To our knowledge, this represents the first predictive model that successfully integrates machine learning techniques with competing risk analysis specifically designed for first-year graft outcome assessment. This novel approach overcomes the restrictive parametric assumptions inherent in traditional regression methods while properly accounting for mortality as a competing event, thereby addressing well-documented limitations in existing prognostic tools.

The research derives additional strength from its incorporation of comprehensive clinical data from Colombia's largest transplant network, providing particularly valuable insights for middle-income healthcare systems, where such predictive models are scarce yet critically needed.

Our study followed current methodological standards throughout its design and implementation. This study employs advanced techniques for handling missing data through multiple imputation and utilizes stratified cross-validation to ensure robust internal validation of results. Furthermore, strict adherence to TRIPOD-AI guidelines guarantees transparent reporting of all model development and validation processes. Importantly, the inclusion of DCA suggests the net potential clinical benefit of the model across a range of threshold probabilities, supporting its potential for real-world application in clinical decision-making.

A practical feature of this work is the development of an interactive web-based calculator that translates the model into an accessible tool. Rather than serving as a definitive clinical instrument, it provides a straightforward way to explore individualized risk estimates. Importantly, this also offers a practical platform for conducting future external validations, which are essential before broader adoption in routine clinical practice.

Limitations

While this study has the potential to advance kidney transplant outcome prediction, several limitations must be acknowledged. First, the dataset exhibited significant class imbalance, with censored cases substantially outnumbering graft loss and death events. Although machine learning methods like RSF-CR are relatively robust to imbalance, this skewed distribution may still influence model performance metrics, particularly for minority classes. We mitigated this through stratified sampling, but the imbalance remains an inherent limitation of working with real-world transplant data where unfavourable outcomes are fortunately rare. This imbalance is also a reflection of the actual clinical distribution, and thus some degree of model performance is intentionally sacrificed in favour of clinical utility.

Importantly, although this study draws on the largest transplant network in Colombia, the overall sample—and especially

the number of events—is modest. Accordingly, our findings should be viewed as preliminary, pending confirmation in larger cohorts and external validation.

Second, the model currently lacks external validation beyond our institutional network. While internal validation provides initial evidence of good overall discrimination, calibration and potential clinical utility, the model requires validation with new patients in our network and across other Colombian transplant centres with varying patient characteristics and clinical protocols and in international cohorts, particularly from other Latin American countries with similar healthcare systems.

Future work

These limitations highlight important next steps: prospective validation across diverse Latin American settings, potential model recalibration for specific subpopulations and eventual integration with existing international risk prediction systems. Completing this work will be essential for transforming this model into a widely applicable clinical decision aid.

CONCLUSIONS

We developed and internally validated a 12-month graft loss prediction model for kidney transplant recipients using RSF-CR. The model showed strong discrimination and moderate calibration. DCA suggested a potential net benefit advantage over two default strategies—‘treat all’ (treat everyone as high risk with intensified follow-up) and ‘none’ (usual care, no high-risk management)—across clinically relevant thresholds. Presented as a user-friendly web application, the tool could support real-time discussions during donor offers. However, the model produced a high rate of false positives, which may limit direct clinical adoption without threshold optimization and calibration review. External validation across transplant centres and populations—particularly in other Latin American settings—remains essential to establish generalizability and clinical impact. With such validation, the tool could help optimize donor-recipient matching and improve graft outcomes.

ACKNOWLEDGEMENTS

We would like to thank Colombiana de Trasplantes for their support and for providing the necessary data for this study. The authors would like to thank the many members of the Javeriana Clinical Epidemiology PhD Program for their constructive discussions related to the topic of this article.

AUTHORS' CONTRIBUTIONS

Andrea Garcia-Lopez (Conceptualization, Data curation, Methodology, Supervision, Formal analysis, Validation, Visualization, Writing—original draft, Writing—review & editing), Juliana Cuervo-Rojas (Conceptualization, Methodology, Investigation, Supervision, Validation, Writing—review & editing), Juan Garcia-Lopez (Conceptualization, Data curation, Formal analysis, Writing—review & editing) and Fernando Giron-Luque (Conceptualization, Investigation, Project administration, Supervision, Validation, Writing—review & editing).

SUPPLEMENTARY DATA

Supplementary data are available at [Clinical Kidney Journal](#) online.

CONFLICT OF INTEREST STATEMENT

None declared.

FUNDING

None declared.

DATA AVAILABILITY STATEMENT

The data underlying this article will be shared upon reasonable request to the corresponding author.

REFERENCES

1. Matas AJ, Montgomery RA, Schold JD. The organ shortage continues to be a crisis for patients with end-stage kidney disease. *JAMA Surg* 2023;158:787–8. <https://doi.org/10.1001/jamasurg.2023.0526>
2. Saidi RF, Kenari SKH. Challenges of organ shortage for transplantation: solutions and opportunities. *Int J Organ Transplant Med* 2014;5:87–96.
3. Riley S, Zhang Q, Tse WY et al. Using information available at the time of donor offer to predict kidney transplant survival outcomes: a systematic review of prediction models. *Transpl Int* 2022;35:10397. <https://doi.org/10.1111/tri.13580>
4. Haller MC, Wallisch C, Mjølén 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;33:729–39. <https://doi.org/10.1111/tri.13580>
5. Kasiske BL, Israni AK, Snyder JJ et al. A simple tool to predict outcomes after kidney transplant. *Am J Kidney Dis* 2010;56:947–60. <https://doi.org/10.1053/j.ajkd.2010.06.020>
6. Molnar MZ, Nguyen DV, Chen Y et al. Predictive score for posttransplantation outcomes. *Transplantation* 2017;101:1353–64. <https://doi.org/10.1097/TP.0000000000001326>
7. Senanayake S, Kularatna S, Healy H et al. Development and validation of a risk index to predict kidney graft survival: the kidney transplant risk index. *BMC Med Res Methodol* 2021;21:127. <https://doi.org/10.1186/s12874-021-01319-5>
8. Kaboré R, Haller MC, Harambat J et al. Risk prediction models for graft failure in kidney transplantation: a systematic review. *Nephrol Dial Transplant* 2017;32(Suppl 2):ii68–76.
9. Mark E, Goldsman D, Gurbaxani B et al. Using machine learning and an ensemble of methods to predict kidney transplant survival. *PLoS One* 2019;14:e0209068. <https://doi.org/10.1371/journal.pone.0209068>
10. Truchot A, Raynaud M, Helanterä I et al. Competing and noncompeting risk models for predicting kidney allograft failure. *J Am Soc Nephrol* 2025;36:688–701. https://journals.lww.com/jasn/fulltext/2025/04000/competing_and_noncompeting_risk_models_for.16.aspx
11. Pinto-Ramirez J, Garcia-Lopez A, Salcedo-Herrera S et al. Risk factors for graft loss and death among kidney transplant recipients: a competing risk analysis. *PLoS One* 2022;17:e0269990. <https://doi.org/10.1371/journal.pone.0269990>
12. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. 2nd ed. Cham: Springer International, 2019.
13. Collins GS, Moons KGM, Dhiman P et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods.

- BMJ 2024;**385**:e078378. <https://doi.org/10.1136/bmj-2023-078378>
14. Hong S, Lynn HS. Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. *BMC Med Res Methodol* 2020;**20**:199. <https://doi.org/10.1186/s12874-020-01080-1>
 15. Ishwaran H, Gerds TA, Kogalur UB et al. Random survival forests for competing risks. *Biostatistics* 2014;**15**:757–73. <http://doi.org/10.1093/biostatistics/kxu010>
 16. Ishwaran H, Kogalur UB, Blackstone EH et al. Random survival forests. *Ann Appl Stat* 2008;**2**:841–60. <http://arxiv.org/abs/0811.1645>
 17. Breiman L. Random forests. *Machine Learning* 2001;**45**:5–32. <https://doi.org/10.1023/A:1010933404324>
 18. Fine JP, Gray RJ. A proportional hazards model for the sub-distribution of a competing risk. *J Am Stat Assoc* 1999;**94**:496–509. <https://doi.org/10.1080/01621459.1999.10474144>
 19. Carrier A, Luijken K, de Hond A et al. The harms of class imbalance corrections for machine learning based prediction models: a simulation study. *Stat Med* 2025;**44**:e10320. <https://doi.org/10.1002/sim.10320>
 20. Piccininni M, Wechsung M, Van Calster B et al. Understanding random resampling techniques for class imbalance correction and their consequences on calibration and discrimination of clinical risk prediction models. *J Biomed Inform* 2024;**155**:104666. <https://doi.org/10.1016/j.jbi.2024.104666>
 21. Binuya MAE, Engelhardt EG, Schats W et al. Methodological guidance for the evaluation and updating of clinical prediction models: a systematic review. *BMC Med Res Methodol* 2022;**22**:316. <https://doi.org/10.1186/s12874-022-01801-8>
 22. Cowley LE, Farewell DM, Maguire S et al. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. *Diagn Progn Res* 2019;**3**:16. <https://doi.org/10.1186/s41512-019-0060-y>
 23. Steyerberg EW, Vickers AJ, Cook NR et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;**21**:128–38. https://journals.lww.com/epidem/fulltext/2010/01000/assessing_the_performance_of_prediction_models__a.22.aspx
 24. Ishwaran H, Kogalur UB. randomForestSRC: Fast unified random forests for survival, regression, and classification (RF-SRC). 2025. <https://cran.r-project.org/package=randomForestSRC> (date last accessed).
 25. Gray B. cmprsk: subdistribution analysis of competing risks. 2024. <https://CRAN.R-project.org/package=cmprsk> (9 February 2026, date last accessed)
 26. Gerds TA, Ohlendorff JS, Ozenne B. riskRegression: risk regression models and prediction scores for survival analysis with competing risks. 2025. <https://CRAN.R-project.org/package=riskRegression> (9 February 2026, date last accessed)
 27. Monterrubio-Gómez K, Constantine-Cooke N, Vallejos CA. A review on statistical and machine learning competing risks methods. *Biometrical J* 2024;**66**:e2300060. <https://doi.org/10.1002/bimj.202300060>
 28. Foroutan F, Friesen EL, Clark KE et al. Risk factors for 1-year graft loss after kidney transplantation: systematic review and meta-analysis. *Clin J Am Soc Nephrol* 2019;**14**:1642–50. https://journals.lww.com/cjasn/fulltext/2019/11000/risk_factors_for_1_year_graft_loss_after_kidney.15.aspx
 29. Collins GS, Reitsma JB, Altman DG et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;**13**:1. <https://doi.org/10.1186/s12916-014-0241-z>
 30. Kappen TH, Vergouwe Y, van KWA et al. Adaptation of clinical prediction models for application in local settings. *Med Decis Making* 2012;**32**:E1–10. <https://doi.org/10.1177/0272989X12439755>
 31. Philosophe B, Malat GE, Soundararajan S et al. Validation of the Maryland (MAPI), a pre-implantation scoring system that predicts graft outcome. *Clin Transplant* 2014;**28**:897–905. <https://doi.org/10.1111/ctr.12400>
 32. Rose C, Sun Y, Ferre E et al. An examination of the application of the kidney donor risk index in British Columbia. *Can J Kidney Health Dis* 2018;**5**:2054358118761052. <https://doi.org/10.1177/2054358118761052>
 33. Udomkarnjananun S, Townamchai N, Kerr SJ et al. The first Asian kidney transplantation prediction models for long-term patient and allograft survival. *Transplantation* 2020;**104**:1048–57. <https://doi.org/10.1097/TP.0000000000000918>
 34. Yang J, Sardo Molmenti CL, Cagliani J et al. Time-effect of donor and recipient characteristics on graft survival after kidney transplantation. *Int J Angiol* 2019;**28**:249–54.
 35. Baskin-Bey ES, Kremers W, Nyberg SL. A recipient risk score for deceased donor renal allocation. *Am J Kidney Dis* 2007;**49**:284–93. <https://doi.org/10.1053/j.ajkd.2006.10.018>
 36. Bae S, Massie AB, Thomas AG et al. Who can tolerate a marginal kidney? Predicting survival after deceased donor kidney transplant by donor-recipient combination. *Am J Transplant* 2019;**19**:425–33. <https://doi.org/10.1111/ajt.14978>
 37. Calvillo-Arbizu J, Pérez-Valdivia MA, Gentil-Govantes MA et al. Does the Kidney Donor Profile Index (KDPI) predict graft and patient survival in a Spanish population? *Nefrologia (Engl Ed)* 2018;**38**:587–95.
 38. Clayton PA, Dansie K, Sypek MP 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**:2127–31. <https://doi.org/10.1093/ndt/gfz090>
 39. Noordzij M, Leffondré K, van Stralen KJ et al. When do we need competing risks methods for survival analysis in nephrology? *Nephrol Dial Transplant* 2013;**28**:2670–7. <https://doi.org/10.1093/ndt/gft355>
 40. Van Calster B, McLernon DJ, van Smeden M et al. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;**17**:230. <https://doi.org/10.1186/s12916-019-1466-7>
 41. Collins GS, de Groot JA, Dutton S et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;**14**:40. <https://doi.org/10.1186/1471-2288-14-40>

Received: 22.12.2025; accepted: 13.3.2026

© The Author(s) 2026. Published by Oxford University Press on behalf of the ERA. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site-for further information please contact journals.permissions@oup.com