

Personalized Prognostication of Kidney Transplant Candidates on the Waiting List: A Machine Learning Approach

Satoru Kawakita (✉ skawakita@terasaki.org)

Terasaki Institute for Biomedical Innovation <https://orcid.org/0000-0002-3781-5786>

Omesh Ranasinghe

Terasaki Research Institute

Emily Wood

Terasaki Research Institute

Jennifer Beaumont

Terasaki Research Institute

Research article

Keywords: Organ transplantation, machine learning, personalized medicine, predictive model

Posted Date: November 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-108492/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Kidney transplantation (KT) remains the best available treatment for end-stage kidney disease. Currently, there is an unmet need for personalized clinical decision aids that help clinicians and KT candidates on the waitlist (WL) make informed decisions. To this end, we took a machine learning (ML)-based approach by using random survival forest with competing risks (RSF-CR) to develop a personalized prognostication system to predict the patient outcomes.

Methods

We obtained data from the Scientific Registry of Transplant Recipients (n=65,583) and developed RSF-CR models to predict the cumulative incidence functions (CIFs) of receipt of KT (CIF_{KT}) and WL removal due to death or deteriorating condition (CIF_{death}) at one year post-listing. Hyperparameter tuning was performed with the integrated Brier score of out-of-bag predicted CIFs as the performance metrics. Feature selection was performed based on variable importance computed by the optimized model. Final model performance was evaluated on a hold-out validation set via receiver operating characteristics (ROC) analysis, decision curve analysis, and additional performance measures (e.g. recall, precision, and F1).

Results

The selected model revealed 16 and 18 prognostic factors for the prediction of CIF_{KT} and CIF_{death} respectively. The model also demonstrated good prognostic ability with an area under the ROC curve and an integrated Brier score of 73.4 [72.7, 74.0] and 16.1 [0.0, 40.5] for CIF_{KT} and 75.0 [73.7, 76.4] and 4.6 [0.0, 11.4] for CIF_{death} on a validation dataset. Dichotomization at selected threshold probabilities yielded a precision, recall, and F1 of 0.927, 0.824, and 0.873 for CIF_{KT} and 0.975, 0.785, and 0.869 for CIF_{death} .

Conclusions

RSF-CR revealed new predictors of the potential outcomes at one year post-listing for KT candidates. While the class imbalance problem was observed, adjustment of the threshold probability allowed the model to become a binary classifier with outstanding performance.

This work demonstrated the value of using RSF-CR in the identification of prognostic factors and personalized prediction of the outcomes for KT candidates on WL. More research is warranted to fully unlock the potential of ML in providing personalized medicine.

Introduction

Kidney transplantation (KT) remains the best available treatment for end-stage renal disease (ESRD); however, the need for kidneys continues to far exceed the number of organs available. At the end of 2018,

there were approximately 92,000 KT candidates on the waitlist (WL), and during the same year, 14,784 and 6,120 received deceased (DDKT) and living donor KT (LDKT) respectively, 4,193 died, and 4,240 became too sick for transplant while waiting for a kidney.¹

Keeping patients well-informed of likely outcomes and treatment options is an essential component of patient-centered care and informed decision-making. It is thus critical to develop and adopt clinical decision aids that healthcare services may use to provide patients with accurate and complete information². KT candidates often wonder and debate between multiple treatment options (e.g. whether to keep receiving statin or go on dialysis while waiting for a kidney)³. Consequently, clinicians are frequently asked by KT candidates about the likelihood of receiving a kidney during a clinical consultation. In the United States, the answer to the question primarily depends on the median waiting time – the time to when 50% of candidates received KT in the past, which is calculated based on a Kaplan-Meier survival curve⁴. While providing this statistic does give patients an answer, this value fails to fully account for several important factors such as the presence of competing risks and inter-individual differences that make each case unique severely limiting the accuracy of the information provided. Therefore, addressing these shortcomings of the current practice is imperative for clinicians to provide more personalized care and improve informed decision-making for KT candidates on WL.

Competing risks are factors that preclude the main event of interest being observed, and therefore should be considered accordingly. The Kaplan-Meier survival estimate treats competing risks merely as censored events resulting in the overestimation of the survival probability⁵. Particularly for receipt of KT, WL removal due to death or deteriorating health condition may be considered a competing risk and should be treated as such. Previously, Cox regression models with competing risks were developed using the Fine-Gray method with a priori variables to predict the probabilities of death, KT, and remaining on WL at the time of listing⁴. However, while the use of a set of preselected variables helps ensure a model is clinically relevant, there is a potential risk of missing important factors that are indeed predictive of the outcome. Additionally, in December 2014, the new Kidney Allocation System (KAS) went into effect in the United States, introducing several major allocation policy changes for deceased KT donors⁶. As a result, there has been a considerable impact on the rates of deceased donor KT especially for highly-sensitized patients with high calculated panel reactive antibody (cPRA)^{7,8} further limiting the validity of the previously developed model in the post-KAS era. Finally, more studies are needed to elucidate the advantages and disadvantages of the random forests-based algorithm proposed recently by Ishwaran et al.⁹ over the Fine-Gray method. Random survival forests with competing risks (RSF-CR) is a non-parametric machine learning (ML) algorithm that does not make any assumptions about the underlying distributions in the training data and automatically accounts for high-level interactions and higher-order terms in the features being analyzed. This allows the algorithm to learn complex patterns effectively from time-to-event data and achieve the calculation of variable importance and outcome prediction without preselection of a priori variables.

In this study, we employed the RSF-CR method to those who were listed after the introduction of the KAS to construct prognostic models that compute the one year post-listing probabilities of two potential outcomes: receipt of KT and WL removal due to death or deteriorating condition.

Materials And Methods

Study design

This study used data from the Scientific Registry of Transplant Recipients (SRTR). The SRTR data system includes data on all donor, wait-listed candidates, and transplant recipients in the US, submitted by the members of the Organ Procurement and Transplantation Network (OPTN). The Health Resources and Services Administration (HRSA), U.S. Department of Health and Human Services provides oversight to the activities of the OPTN and SRTR contractors. The data on adult kidney transplant candidates listed between 12/04/2014 and 12/04/2016 was used to develop ML models to predict two potential outcomes: receipt of KT (primary outcome) and WL removal due to death or deteriorating condition (secondary outcome) (Figure 1). WL removal for other reasons was censored. The dataset was split on a WL date of 12/05/2015 into development and validation sets.

Model training (development)

Grid search was performed with nine different model configurations to tune three model parameters: the number of trees (ntree), node size (nodesize), and the number of variables selected for each tree (mtry). BS at one year post-listing with standard deviation and IBS were used as the performance metrics to select the final model for validation. The algorithm was implemented using the randomForestSRC package in R¹⁰.

Feature selection

With regards to feature selection, RSF-CR is considered an embedded method in which feature selection is integrated as part of the learning¹¹. As such, variables that are predictive of the outcome were determined using the best-performing model based on VIS. VIS of less than 0 indicates a non-informative factor and thus was removed. VIS was calculated by taking the difference between prediction error under the predictor perturbed via random node assignment and the original predictor for each tree and averaged over x number of trees. The accuracy generally is higher for a lower number of trees used for aggregation, which ranges between one and the number of trees used to build the forest⁹. We averaged the values over 10 trees for the current study to achieve good accuracy while keeping the computation time and required memory manageable.

RSF-CR

RSF-CR as proposed by Ishwaran et al. is an extension of the vanilla random forest algorithm developed by Brieman. As our main aim is to assist decision-making rather than investigate treatment options with

the models, we used the event-specific CIF rather than the hazard function as the model output. The CIF is defined as the probability P of experiencing an event of type j given a set of covariates x by time t :

$$F_j(t|x) = P\{T^0 \leq t, \delta^0 = j|x\},$$

where T^0 is the event time observed and δ^0 is the event type of interest. For performance evaluation, out-of-bag (OOB) CIF was used, which is derived from an OOB ensemble constructed with data that is excluded from each bootstrap sample.

The following steps describe how RFS-CR is constructed:

1. Take k bootstrap samples from training data for the k trees to be included in the forest.
2. Grow a risk tree for each bootstrap sample. At each node, randomly select l candidate variables. The node is split on a candidate variable such that it maximizes a splitting rule (i.e. modified Gray's test).
3. Grow the tree until a terminal node has no less than m unique cases.
4. Compute the cause-specific CIF, cause-specific hazard function, cause-specific mortality, and forest event-free survival for each tree.
5. Aggregate the trees by taking the average of each estimator over k trees.

Thorough mathematical explanations are provided in detail by Ishwaran et al. in their original article⁹.

Model validation and evaluation

Model validation was conducted on the validation set. Overall predictive performance and discrimination were evaluated using time-dependent IBS, and ROC-AUC. IBS is an overall performance measure calculated by integrating BS over all available times. ROC-AUC is another indicator of performance based on the area under the true positive rate (sensitivity) versus false positive rate (1-specificity) curve. Error analysis was performed by bootstrapping with replacement to estimate the 95% confidence intervals. As probability-based measures suffer from the presence of an imbalance in the dataset, we also evaluated precision (positive predictive value), recall (sensitivity), and F1 of the classifiers by dichotomizing the predicted probabilities at an optimal threshold value determined by decision curve analysis. F1 is a harmonic mean of precision and recall¹² defined as:

$$F_1 = 2 * \frac{\text{precision} + \text{recall}}{\text{precision} + \text{recall}}$$

Model calibration was assessed via calibration plot by plotting the observed prevalence against the mean predicted probability for the deciles of the predicted probability¹³. Additional analyses were carried out in patient subgroups by cPRA category, OPTN region, listing year, and race where we compared the mean predicted probability with observed prevalence in each of the subgroups. Additionally, decision curve analysis was performed to assess the clinical utility of the final models. To develop decision curves, net

benefits (NBs) were plotted against threshold probabilities of zero through one with an increment of 0.01 for three different scenarios: all patients are treated, no patients are treated, and only selected patients are treated using the prognostic systems. The net benefit was calculated by using the following formula:

$$NB = \frac{TP}{n} - \frac{FP}{n} \left(\frac{p_t}{1-p_t} \right),$$

where TP = true positive count, n = sample size, FP = false positive count, and p_t = threshold probability¹⁴. The calculation of NB has been adjusted from its original form such that it accounts for censored events and competing risks accordingly¹⁵. Reduction in avoidable treatment per 100 patients (N_{100}) was then computed as follows:

$$N_{100} = (NB_m - NB_{treat\ all}) \left(\frac{p_t}{1-p_t} \right) * 100$$

where NB_m = net benefit of the model and $NB_{treat\ all}$ = net benefit of treat all.

Results

Predictive modeling process

To develop reliable and robust RSF-CR models, we followed the procedure as described in Figure 1. First, data were pre-processed as outlined and split into development (training) and validation (test) sets on a listing date of December 4, 2015. The development set was then used to train RSF-CR models. Grid search with nine different model configurations was performed for hyperparameter tuning. Variable importance score (VIS) was obtained from the best performing model to assess the contribution of each predictor to the outcome prediction. Finally, the selected model was validated on the validation set using a series of statistical techniques commonly used for clinical predictive modeling: receiver-operating curve (ROC) analysis, calibration, and decision curve analysis.

Patient characteristics: development and validation cohorts

Table 1 and Supplementary Table 1 show the patient characteristics for the development (n=27,451) and validation sets (n=29,778). Upon comparisons, the two cohorts had similar patient profiles with minor differences. At one year post-listing, the cumulative incidence functions (CIF_{1yr}) were 0.2303 and 0.2449 for KT and 0.042 and 0.040 for WL removal due to death and/or deteriorating health conditions in the development and validation sets respectively (Figure 2). The development cohort had a median follow-up time of 2.16 years whereas the validation cohort had that of 1.61 years.

Model development

Grid search for hyperparameter tuning was performed on the development set to test nine different model configurations with the Brier score (BS) at one year post-listing and integrated Brier score (IBS) for KT (BS_{KT} , IBS_{KT}) and WL removal due to death or deteriorating condition (BS_{death} , IBS_{death}) as the performance metrics. Model 9 exhibited the best predictive performance with an IBS_{KT} of 0.1362, one year post-listing BS_{KT} of 0.021 ± 0.009 , IBS_{death} of 0.0207, and one year post-listing BS_{death} of 0.005 ± 0.003 (Table 2). The selected model had n_{tree} of 1000, m_{try} of 6, and $n_{nodesize}$ of 1. n_{tree} did not have noticeable effect on model performance based on the comparisons of Models 1, 5, and 6 while $n_{nodesize}$ appeared to be the most influential factor as evident from the comparisons of Models 1, 2, 3, 4, and 9.

Identification of prognostic factors

RSF-CR computes a VIS for each variable included in model training thereby providing a deeper understanding of the contribution of each variable to the estimation of CIF_{1yr} . The VIS for predictors of $CIF_{1yr,KT}$ and $CIF_{1yr,death}$ in the selected model are shown in Figure 3 and Supplementary Figure 1 respectively. There were a total of 16 predictors of $CIF_{1yr,KT}$ with the most potent factor being current functional status (VIS = 0.00062) followed by levels of total albumin (VIS = 0.00027). For $CIF_{1yr,death}$, we identified 18 predictors, in which functional status was most predictive of $CIF_{1yr,death}$ (VIS = 0.00067) followed by diabetes (VIS = 0.0001). Interestingly, cPRA category was not selected as a predictor of KT although it is one of the major changes introduced by the new KAS system.

Model validation

The selected model was further validated on the validation set via receiver-operating characteristics (ROC) analysis, calibration plot, patient subgroup analysis, and decision curve analysis. The predicted $CIF_{1yr,KT}$ demonstrated discriminative ability with an area under the ROC curve (AUC-ROC) of 73.4 [72.7, 74.0] and an IBS of 16.1 [0.0, 40.5], and good calibration as observed in the calibration plot (Figure 4). Similarly, $CIF_{1yr,death}$ showed an AUC-ROC of 75.0 [73.7, 76.4], IBS of 4.6 [0.0, 11.4] and calibration comparable to that of $CIF_{1yr,KT}$ (Supplementary Figure 2). To evaluate model calibration in subsets of patients, we further stratified the validation cohort by cPRA bracket, race, region, and wait-listing year (Supplementary Table 2). Although minor discrepancies were found for some groups such as 0.99 cPRA ($\Delta = -5.69$), 0.9-0.97 cPRA ($\Delta = -5.38$), and Region 8 ($\Delta = -4.12$), there was generally good agreement between the mean estimated $CIF_{1yr,KT}$ and observed $CIF_{1yr,KT}$. Next, the clinical utility of the models was assessed via decision curve analysis (Figure 5). Compared with treating all patients and treating none, treating patients based on predicted $CIF_{1yr,KT}$ provided greater net benefit (NB) at certain threshold probabilities. Treatment here refers to any clinical intervention that may be suggested when a patient is predicted to undergo KT at one year post-listing such as discontinuation of statin therapy. NB peaked at a threshold probability of 0.28. At this threshold probability, the NB was 0.217, which translates to a reduction by 10.15 patients in avoidable treatment per 100 patients compared to the treat all strategy. As our dataset contained data imbalance where a disproportional set of patients was censored compared to

those who experienced one of the events of interest, we took a moving threshold approach to address the issue.¹⁶ That is, the optimal threshold probability from the decision curve analysis was used to categorize the patients into two groups: 1 (predicted to receive KT at one year post-listing) versus 0 (predicted not to undergo KT by one year post-listing). When using this cutoff, the model yielded excellent prognostic performance as shown by a confusion matrix and a series of performance measures in Figure 6. There were a total of 6,630 true positive cases, which corresponded to a detection prevalence rate of 0.24 when the observed prevalence rate was 0.27. We placed greater emphasis on the precision (positive predictive value), recall (sensitivity) and F_1 score as there was class imbalance in the dataset with substantially more patients not receiving KT. The precision and recall were 0.927 and 0.824 with the F_1 being 0.873 suggesting that the model is highly capable of accurately differentiating one group from another. Likewise, predicted $CIF_{1yr, death}$ demonstrated a degree of clinical utility (Supplementary Figure 3); using the RSF-CR model will assist clinicians in making clinical decisions about the administration of treatment to prevent death. Moreover, the predicted $CIF_{1yr, death}$ exhibited high predictive performance when dichotomized at a probability of 0.11 at which the NB represents a reduction of avoidable treatment by 1.46 cases per 100 patients (Supplementary Figure 4). The number of true positive cases and detection prevalence reached 1,136 and 0.038 (observed prevalence = 0.049) respectively. The model achieved a precision of 0.975, recall of 0.785, and an F_1 score of 0.869.

Discussion

A recent report from 2018 indicates that there were nearly 90,000 KT candidates listed waiting to receive a kidney at the end of the year¹. The introduction of the new KAS has significantly changed the mechanisms by which donor kidneys are allocated to patients to increase KT access to those who were previously at disadvantage such as HS patients⁷. This paradigm shift in donor allocation has necessitated a new prognostic system that accurately estimates the likelihood of individual patients to receive KT in the post-KAS era. While the Fine and Gray method has conventionally been used to model survival data with competing risks, RSF-CR⁹ is another attractive technique to develop prediction models with such data. Therefore, we employed RSF-CR coupled with the post-KAS data from a national-scale registry database to identify predictors of the outcomes and build robust prognostic models. The final models achieved good discrimination and calibration and exhibited good predictive performance for both KT and WL removal due to death or deteriorating condition at one year post-listing.

We have identified 16 predictors for $CIF_{1yr, KT}$ and 18 predictors for $CIF_{1yr, death}$ based on the VIS computed by the best-performing model. There is a body of evidence that supports the prognostic ability of each of these factors. Poor functional status was shown to be associated with both lower transplant rates¹⁷ and a higher risk of WL mortality¹⁸. Lower albumin level has been described as a risk factor of decline in kidney function and renal diseases such as diabetes nephropathy¹⁹. Diabetes poses a great risk of increased mortality not just among KT candidates but also in the general population²⁰ rendering it one of the important variables considered by the KAS in donor allocation²¹. Inactive status was reported to be

correlated with lower KT rates in a study where candidates initially having inactive status were more likely to be older, female, and African-American and have higher BMI²². Although one of the goals of the KAS is to reduce geographical inequality in donor allocation, it is still prevalent today²³. This explains Organ Procurement and Transplant Network (OPTN) region being one of the predictors for $CIF_{1yr,KT}$. Racial disparities in access to KT are also one of the active areas of research with recent studies showing the importance of addressing this issue in increasing KT rates²⁴⁻²⁶. Interestingly, cPRA was not one of the significant prognostic factors although it is integral part of the new allocation policy. This observation is likely attributed to the definition of KT used in the study, which includes both LDKT and DDKT as the primary scope of the KAS is to improve the rates of DDKT rather than LDKT.

For the present study, we utilized a random forests-based ML algorithm, RSF-CR⁹ to develop prognostic models with survival data with competing risks. RSF, an extension of the Breiman's random forests²⁷ was originally proposed by Ishwaran et al. in 2008²⁸ upon which they built RSF-CR for competing risk applications. The advantages of this algorithm include the ability to (1) "learn" nonlinear relationships between the input features without stringent assumptions regarding the underlying data distributions (e.g. no model assumptions to be met); (2) quantify the variable importance of each factor for estimation of CIF; (3) potentially reveal new prognostic factors in a data-driven approach; (4) be able to account for the presence of competing risks to avoid overestimation of CIF; and (5) leverage the strength of random forests and RSF, both of which have empirically been shown to outperform traditional algorithms such as logistic regression²⁹, decision trees³⁰, and cox regression.²⁸ The developed prognostic system is superior to the current median waiting time approach for KT candidates, which does not account for inter-individual differences and the presence of competing risks. The ML model provides the ability to estimate the likelihood of receiving KT and WL removal due to death or deteriorating health condition in a personalized manner. Based on the predicted probabilities, clinicians and patients can make more informed decisions about their treatment options.

Despite our results showing the potential of RFS-CR in predicting the patient outcomes, our study has several limitations, one of which includes the imbalanced nature of our dataset with substantially more patients being censored compared to those who experienced any of the events of interest. This resulted in some of the performance measures such as ROC-AUC and IBS to be relative rather than absolute rendering comparisons with other studies difficult. As a way to address this issue, we dichotomized the predicted probabilities at a threshold probability and employed the performance measures (precision, recall, and F_1 scores) known to be more appropriate in case of imbalanced datasets³¹. This moving threshold approach is one of the traditional methods to deal with class imbalance¹⁶. Another limitation of the study comes from the use of a registry database. Databases of large size were reported to have high error rates caused by various factors such as human data entry errors and misinterpretation of original documents³². We dedicated a substantial amount of time to data cleaning to minimize this risk. Also, we only validated the model with the registry data as an external dataset was not available. Thirdly, it is relatively recent that the KAS was put into effect, and therefore it was impractical to predict beyond one year post-listing. Finally, as with the original random forests algorithm, RSF-CR, by default, does not

provide the direction of influence for each predictor in the presence of multiple variables. There are several techniques that may be utilized to understand the directionality of each predictor such as Shapley Additive Explanation and Local Interpretable Model-Agnostic Explanations, to name a few³³. Therefore, as a next step, we will perform additional analyses of the predictors along with external model validation to improve both local and global interpretability as well as the robustness of the model.

With the advent of advances in computing power, ML algorithms, and data generating capability, ML has generated an unprecedented volume of opportunities for various sectors including healthcare. While ML has gained increasing attention in the healthcare industry, it has seen a unique set of challenges due to the concomitant bioethical concerns surrounding the use of complex ML algorithms with poor interpretability^{34,35}. This has led to the preferred use of algorithms such as decision trees and logistic regression that have high model transparency. However, more convoluted algorithms such as ensemble learning and deep learning models have been demonstrated to predict clinical outcomes with higher accuracy while being more efficient at handling complex data such as images and electric medical records³⁶. Hence more research is needed to elucidate AI's capability in medicine while advancing ML algorithms and interpretability. In addition to predicting patient prognosis, ML models have potential in other aspects of patient care, such as diagnosis, treatment, and clinical workflow to increase efficiency and augment the work of clinicians. This synergy between clinicians and AI suggests that ML-based systems may make patient care accessible to a larger patient population while ensuring high-quality of service. This also highlights the importance of a concept known as "Augmented Intelligence", in which AI works to aid human physicians to maximize their performance. In addition, AI's potential has started to be realized by the pharmaceutical industry and clinical researchers. Among a number of ways AI may help the industry, one of the key applications of ML is to allow more systematic and accurate risk quantification, which may lead to higher success rates of clinical trials. As clinical studies for drug development require substantial amounts of resources, strategies to reduce trial failure are imperative. ML algorithms may be part of such strategies as a tool that aids in patient stratification, treatment response identification, and/or subgroup identification³⁷.

Conclusions

In all, here we demonstrated the potential of RSF-CR in both identification of prognostic factors and personalized prediction of the outcomes for KT candidates on WL. This strategy using the ML techniques is more ideal than the current median time-based approach as it enables a personalized prognostication of KT candidates helping patients make informed decisions. As ML progressively finds its place in medicine, more research is required to ensure successful clinical implementation and fully unlock the potential of AI in personalized medicine.

Abbreviations

AI: Artificial Intelligence

BMI: Body mass index

BS: Brier score

CIF: Cumulative incidence

CPRA: Calculated panel reactive antibody

DDKT: Deceased donor kidney transplant

ESRD: End stage renal disease

IBS: Integrated brier score

KAS: Kidney allocation system

KT: Kidney transplant

LDKT: Living donor kidney transplant

ML: Machine learning

NB: Net benefit

OPTN: Organ Procurement and Transplant Network

ROC: Receiver operating characteristics

RSF-CR: Random survival forest with competing risks

SRTR: Scientific Registry of Transplant Recipients

VIS: Variable importance score

WL: Waitlist

Declarations

Ethics approval and consent to participate:

The patient level data was used under Data Use Agreement (Project #9604) for the current study and made available by SRTR Contracting Officer's Representative.

Consent for publication:

Not applicable

Availability of data and materials:

The authors do not own the data, which was used under license for the current study. All relevant data are available from the SRTR. Interested researchers may request access to the data from <https://www.srtr.org/requesting-srtr-data/data-requests/>.

Competing interests:

The author(s) declare no competing interests.

Funding:

The data reported here have been supplied by the Hennepin Healthcare Research Institute (HHRI) as the contractor for the Scientific Registry of Transplant Recipients (SRTR). The interpretation and reporting of these data are the responsibility of the author(s) and in no way should be seen as an official policy of or interpretation by the SRTR or the U.S. Government.

Acknowledgment:

Data was provided by the Scientific Registry of Transplant Recipients. We would like to thank Dr. Matthew Everly for his support and guidance towards the current study.

Author contributions:

SK conceived, designed, and performed predictive modeling and analysis, and wrote the manuscript. OR critically reviewed the manuscript and contributed to manuscript development. EW critically reviewed the manuscript and contributed to manuscript development. JB contributed to the analysis design and manuscript development.

References

1. Hart, A. *et al.* OPTN/SRTR 2018 Annual Data Report: Kidney. *Am. J. Transplant.***20**, 20–130 (2020).
2. Van Der Veer, S. N., Bekker, H. L. & Van Biesen, W. How to enhance patient experiences of decision-making about kidney transplantation? *Nephrology Dialysis Transplantation* vol. 29 1991–1993 (2014).
3. Morton, R. Do dialysis decision aids improve treatment decision-making? *Peritoneal Dialysis International* vol. 36 359–361 (2016).
4. Hart, A., Salkowski, N., Snyder, J. J., Israni, A. K. & Kasiske, B. L. Beyond median waiting time: Development and validation of a competing risk model to predict outcomes on the kidney transplant waiting list. *Transplantation***100**, 1564–1570 (2016).
5. Andersen, K., Geskus, R. B., Witte, T. De & Putter, H. Competing risks in epidemiology: possibilities and pitfalls. **41**, 861–870 (2012).

6. Stewart, D. E. & Klassen, D. K. Early experience with the new kidney allocation system: A perspective from UNOS. *Clin. J. Am. Soc. Nephrol.***12**, 2063–2065 (2017).
7. Stewart, D. E. *et al.* Changes in Deceased Donor Kidney Transplantation One Year After KAS Implementation. *Am. J. Transplant.***16**, 1834–1847 (2016).
8. Jackson, K. R. *et al.* The national landscape of deceased donor kidney transplantation for the highly sensitized: Transplant rates, waitlist mortality, and posttransplant survival under KAS. *Am. J. Transplant.***19**, 1129–1138 (2019).
9. Ishwaran, H. *et al.* Random survival forests for competing risks. *Biostatistics***15**, 757–773 (2014).
10. Ishwaran H, K. U. Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC). R package version 2.9.3, <https://cran.r-project.org/package=randomForestSRC>. (2020).
11. Jović, A., Brkić, K. & Bogunović, N. A review of feature selection methods with applications. in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings* 1200–1205 (Institute of Electrical and Electronics Engineers Inc., 2015). doi:10.1109/MIPRO.2015.7160458.
12. Lipton, Z. C., Elkan, C. & Naryanaswamy, B. Optimal thresholding of classifiers to maximize F1 measure. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 8725 LNAI 225–239 (Springer Verlag, 2014).
13. Steyerberg, E. W. *et al.* Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* vol. 21 128–138 (2010).
14. Vickers, A. J. & Elkin, E. B. Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Making***26**, 565–74 (2006).
15. ddsjoberg/dca: Decision Curve Analysis. <https://github.com/ddsjoberg/dca>.
16. Collell, G., Prelec, D. & Patil, K. Reviving Threshold-Moving: a Simple Plug-in Bagging Ensemble for Binary and Multiclass Imbalanced Data. (2016).
17. Reese, P. P. *et al.* Functional Status, Time to Transplantation, and Survival Benefit of Kidney Transplantation Among Wait-Listed Candidates. *Am. J. Kidney Dis.***66**, 837–845 (2015).
18. Perito, E. R., Bucuvalas, J. & Lai, J. C. Functional status at listing predicts waitlist and posttransplant mortality in pediatric liver transplant candidates. *Am. J. Transplant.***19**, 1388–1396 (2019).
19. Zhang, J. *et al.* The Level of Serum Albumin Is Associated with Renal Prognosis in Patients with Diabetic Nephropathy. *J. Diabetes Res.***2019**, (2019).
20. Tancredi, M. *et al.* Excess Mortality among Persons with Type 2 Diabetes. *N. Engl. J. Med.***373**, 1720–1732 (2015).
21. Kidney Allocation System - Professional Education - OPTN. <https://optn.transplant.hrsa.gov/learn/professional-education/kidney-allocation-system/>.
22. Grams, M. E., Massie, A. B., Schold, J. D., Chen, B. P. & Segev, D. L. Trends in the inactive kidney transplant waitlist and implications for candidate survival. *Am. J. Transplant.***13**, 1012–1018 (2013).

23. Zhou, S. *et al.* Geographic disparity in kidney transplantation under KAS. *Am. J. Transplant.***18**, 1415–1423 (2018).
24. Waterman, A. D., Rodrigue, J. R., Purnell, T. S., Ladin, K. & Boulware, L. E. Addressing Racial and Ethnic Disparities in Live Donor Kidney Transplantation: Priorities for Research and Intervention. *Semin. Nephrol.***30**, 90–98 (2010).
25. Zhang, X. *et al.* Racial/ethnic disparities in waitlisting for deceased donor kidney transplantation 1 year after implementation of the new national kidney allocation system. *Am. J. Transplant.***18**, 1936–1946 (2018).
26. Waterman, A. D. *et al.* Modifiable patient characteristics and racial disparities in evaluation completion and living donor transplant. *Clin. J. Am. Soc. Nephrol.***8**, 995–1002 (2013).
27. Breiman, L. Random forests. *Mach. Learn.***45**, 5–32 (2001).
28. Ishwaran, H., Kogalur, U. B., Blackstone, E. H. & Lauer, M. S. Random survival forests. *Ann. Appl. Stat.***2**, 841–860 (2008).
29. Couronné, R., Probst, P. & Boulesteix, A. L. Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics***19**, 270 (2018).
30. Esmaily, H. *et al.* A Comparison between Decision Tree and Random Forest in Determining the Risk Factors Associated with Type 2 Diabetes. *J. Res. Health Sci.***18**, (2018).
31. Jeni, L. A., Cohn, J. F. & De La Torre, F. Facing imbalanced data - Recommendations for the use of performance metrics. in *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013* vol. 2013 245–251 (NIH Public Access, 2013).
32. Goldberg, S. I., Niemierko, A. & Turchin, A. Analysis of data errors in clinical research databases. *AMIA Annu. Symp. Proc.***2008**, 242–246 (2008).
33. Elshawi, R., Al-Mallah, M. H. & Sakr, S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Med. Inform. Decis. Mak.***19**, (2019).
34. Vayena, E., Blasimme, A. & Cohen, I. G. Machine learning in medicine: Addressing ethical challenges. *PLoS Med.***15**, e1002689 (2018).
35. Char, D. S., Shah, N. H. & Magnus, D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. *N. Engl. J. Med.***378**, 981–983 (2018).
36. Rajkomar, A., Dean, J. & Kohane, I. Machine learning in medicine. *New England Journal of Medicine* vol. 380 1347–1358 (2019).
37. Vamathevan, J. *et al.* Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery* vol. 18 463–477 (2019).

Tables

Table 1. Patient demographics.

	Development	Validation
n	27,451	29,778
Age at listing, mean (SD)	52.04 (12.98)	52.32 (13.03)
Male, n (%)	17153 (62.5)	18469 (62.0)
Race, n (%)		
Asian	1970 (7.2)	2161 (7.3)
Black	8167 (29.8)	8595 (28.9)
Multiple	175 (0.6)	219 (0.7)
Native	227 (0.8)	271 (0.9)
Pacific	154 (0.6)	143 (0.5)
White	16758 (61.0)	18389 (61.8)
Blood type, n (%)		
A	9046 (33.0)	9665 (32.5)
AB	1102 (4.0)	1092 (3.7)
B	4025 (14.7)	4447 (14.9)
O	13278 (48.4)	14574 (48.9)
BMI, mean (SD)	28.88 (5.61)	30.48 (220.38)
Previous transplant, n (%)	3680 (13.4)	3932 (13.2)
CPRA category, n (%)		
<0.8	24759 (90.2)	27086 (91.0)
0.8-0.89	558 (2.0)	614 (2.1)
0.9-0.97	780 (2.8)	787 (2.6)
0.98	200 (0.7)	191 (0.6)
0.99	363 (1.3)	343 (1.2)
1	791 (2.9)	757 (2.5)
Dialysis, n (%)	20237 (73.7)	21525 (72.3)
OPTN region, n (%)		
1	1075 (3.9)	1206 (4.0)
2	3907 (14.2)	4145 (13.9)

3	3751 (13.7)	3808 (12.8)
5	2781 (10.1)	3193 (10.7)
6	4616 (16.8)	5122 (17.2)
7	917 (3.3)	819 (2.8)
8	2361 (8.6)	2507 (8.4)
9	1624 (5.9)	1681 (5.6)
0	1820 (6.6)	2167 (7.3)
10	2306 (8.4)	2362 (7.9)
11	2293 (8.4)	2768 (9.3)
Follow-up time, median (IQR)	2.16 [0.75, 2.86]	1.61 [0.70, 1.99]

BMI = body mass index; CPRA = calculated panel reactive antibody; IQR = interquartile range; OPTN = Organ Procurement and Transplantation Network; SD = standard deviation;

Table 2. Model training results.

Model	ntree (1000*)	nodesize (15*)	mtry (6*)	1 year BS ± SE (KT)	IBS (KT)	1 year BS ± SE (death)	IBS (death)
1	1000	5	6	0.07±0.033	0.1363	0.025±0.015	0.0208
2	1000	15	6	0.11±0.052	0.1365	0.032±0.02	0.0208
3	1000	45	6	0.138±0.065	0.1367	0.036±0.022	0.0208
4	1000	135	6	0.154±0.073	0.1368	0.038±0.023	0.0208
5	2000	15	6	0.11±0.052	0.1365	0.032±0.02	0.0208
6	4000	15	6	0.11±0.052	0.1365	0.032±0.02	0.0208
7	1000	15	12	0.104±0.048	0.1365	0.032±0.02	0.0208
8	1000	15	3	0.119±0.056	0.1366	0.033±0.021	0.0208
9	1000	1	6	0.021±0.009	0.1362	0.005±0.003	0.0207

*Default value. BS = Brier score; IBS = integrated Brier score; KT = kidney transplant; SE = standard error.

Figures

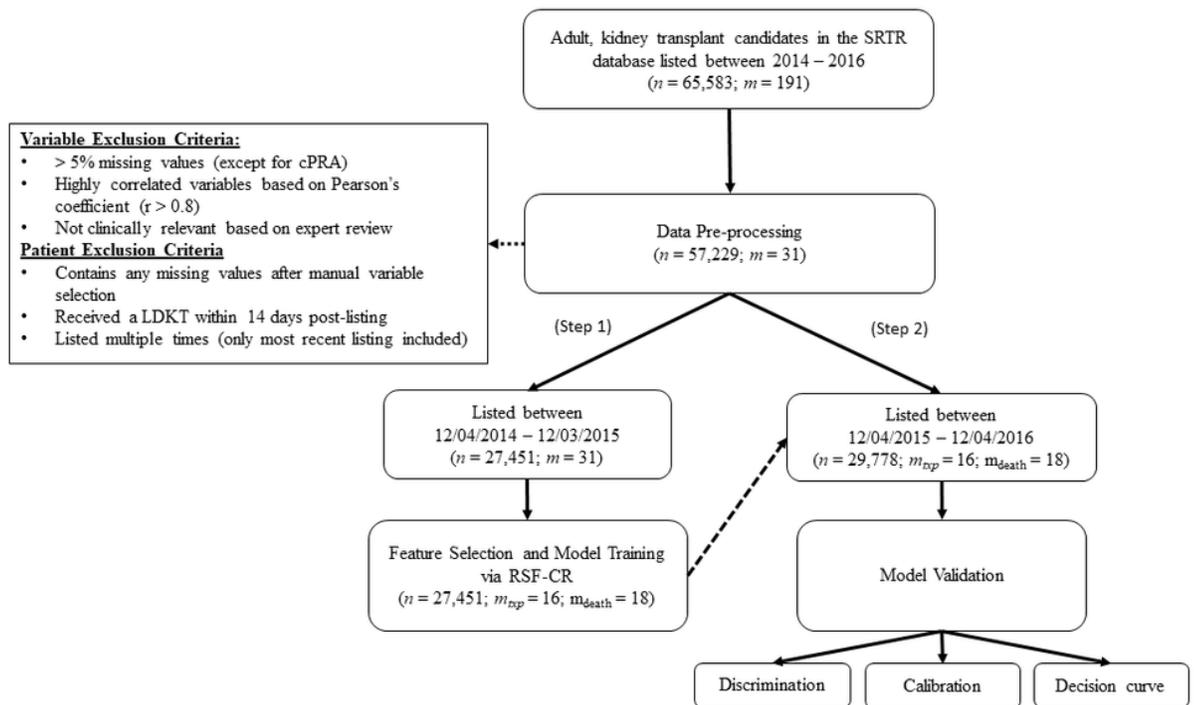


Figure 1

Study design. Data were obtained from the Scientific Registry of Transplant Recipients on adult kidney transplant candidates and were pre-processed as described. The cleaned dataset was split into a development set and validation set. Nine RSF-CR models were trained with the development to predict the CIF for KT and WL removal due to death or deteriorating condition using one year post-listing BS and IBS as the performance metrics. The best model was then used to unveil prognostic factors for each prediction task. Finally, the predictive performance was tested in the validation cohort for discrimination, calibration and net benefit. The letters n and m represent the numbers of patients and variables respectively.

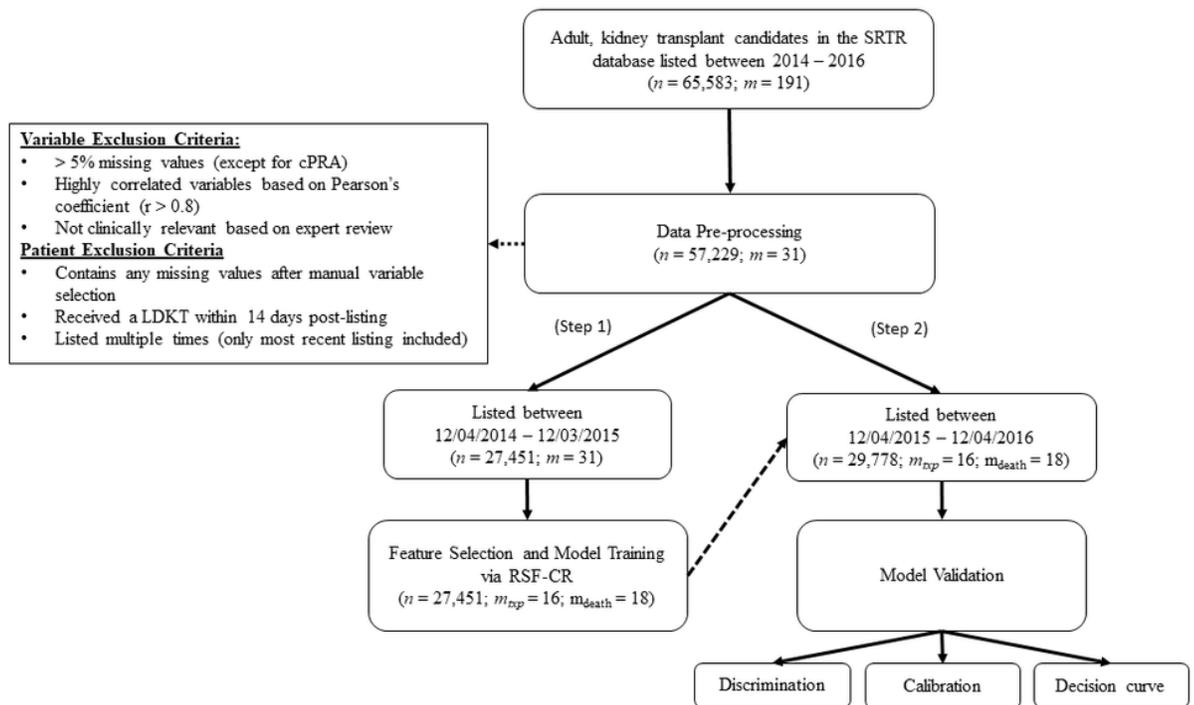


Figure 1

Study design. Data were obtained from the Scientific Registry of Transplant Recipients on adult kidney transplant candidates and were pre-processed as described. The cleaned dataset was split into a development set and validation set. Nine RSF-CR models were trained with the development to predict the CIF for KT and WL removal due to death or deteriorating condition using one year post-listing BS and IBS as the performance metrics. The best model was then used to unveil prognostic factors for each prediction task. Finally, the predictive performance was tested in the validation cohort for discrimination, calibration and net benefit. The letters n and m represent the numbers of patients and variables respectively.

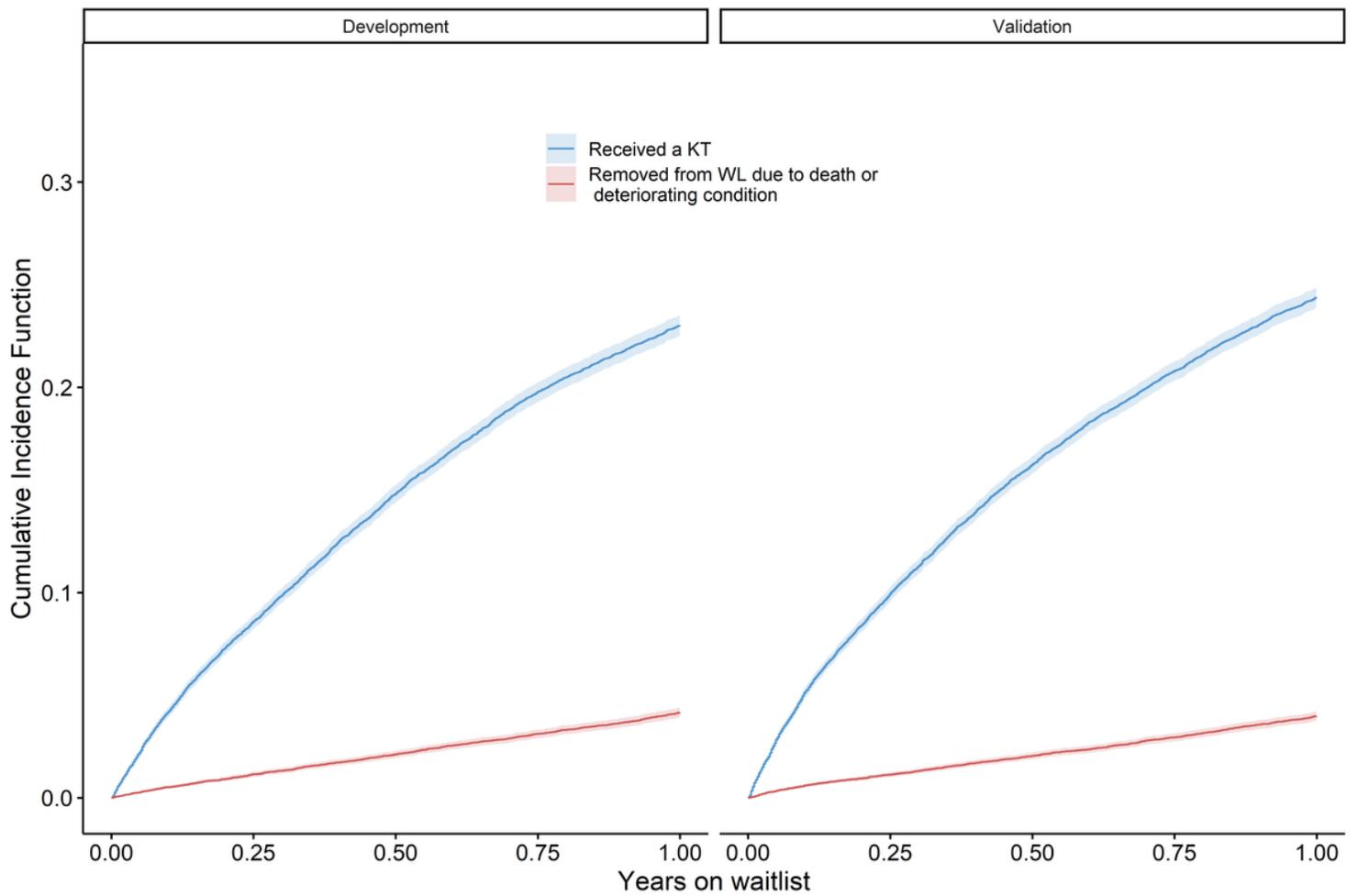


Figure 2

Cumulative incidence curves for KT and WL removal due to death or deteriorating condition. The curves represent the CIFs for the two events of interest during the first year of listing. At one year post-listing, the CIFs were 0.2303 and 0.2449 for KT and 0.042 and 0.040 for WL removal due to death or deteriorating health condition in the development and validation sets respectively.

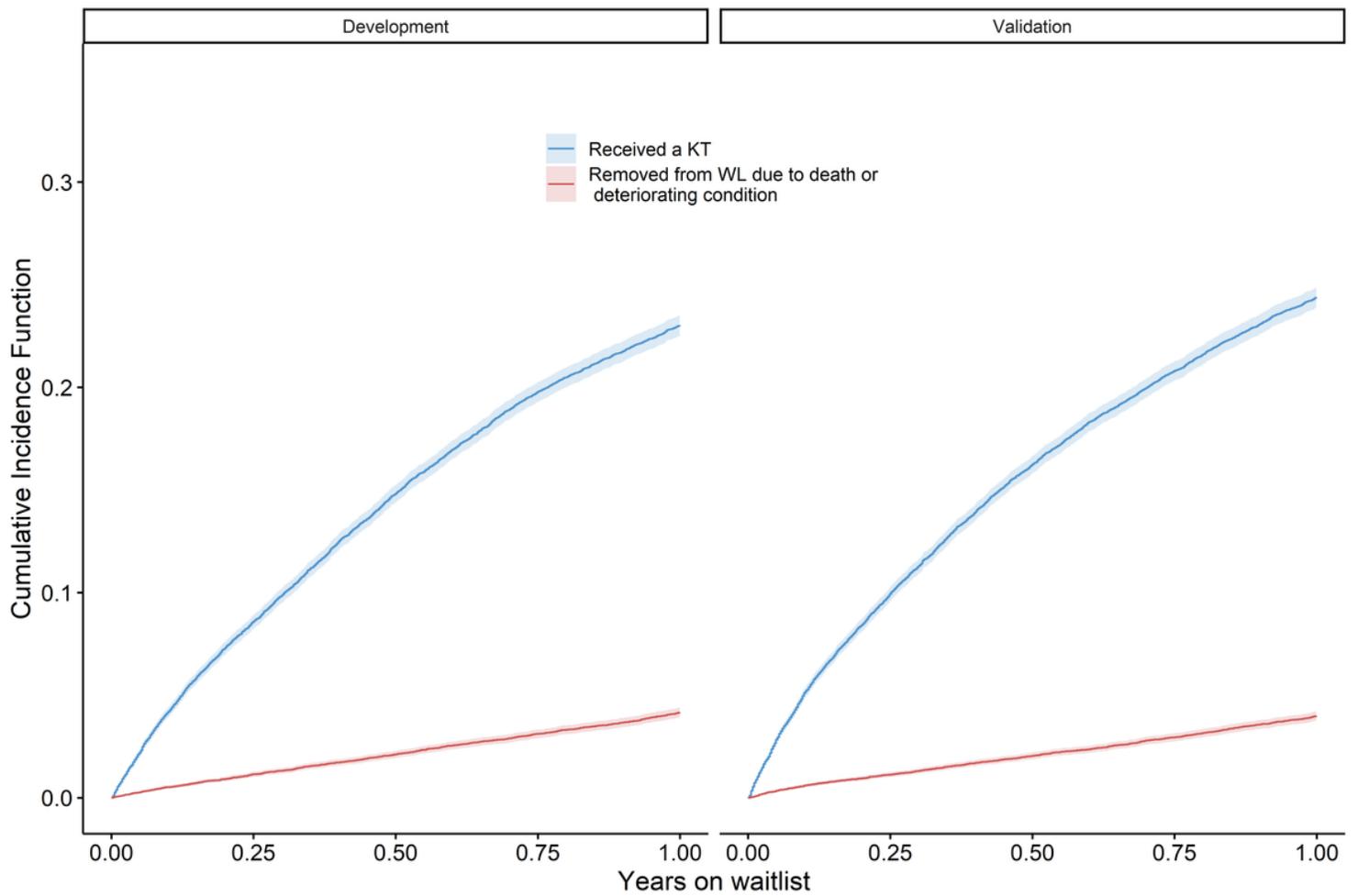


Figure 2

Cumulative incidence curves for KT and WL removal due to death or deteriorating condition. The curves represent the CIFs for the two events of interest during the first year of listing. At one year post-listing, the CIFs were 0.2303 and 0.2449 for KT and 0.042 and 0.040 for WL removal due to death or deteriorating health condition in the development and validation sets respectively.

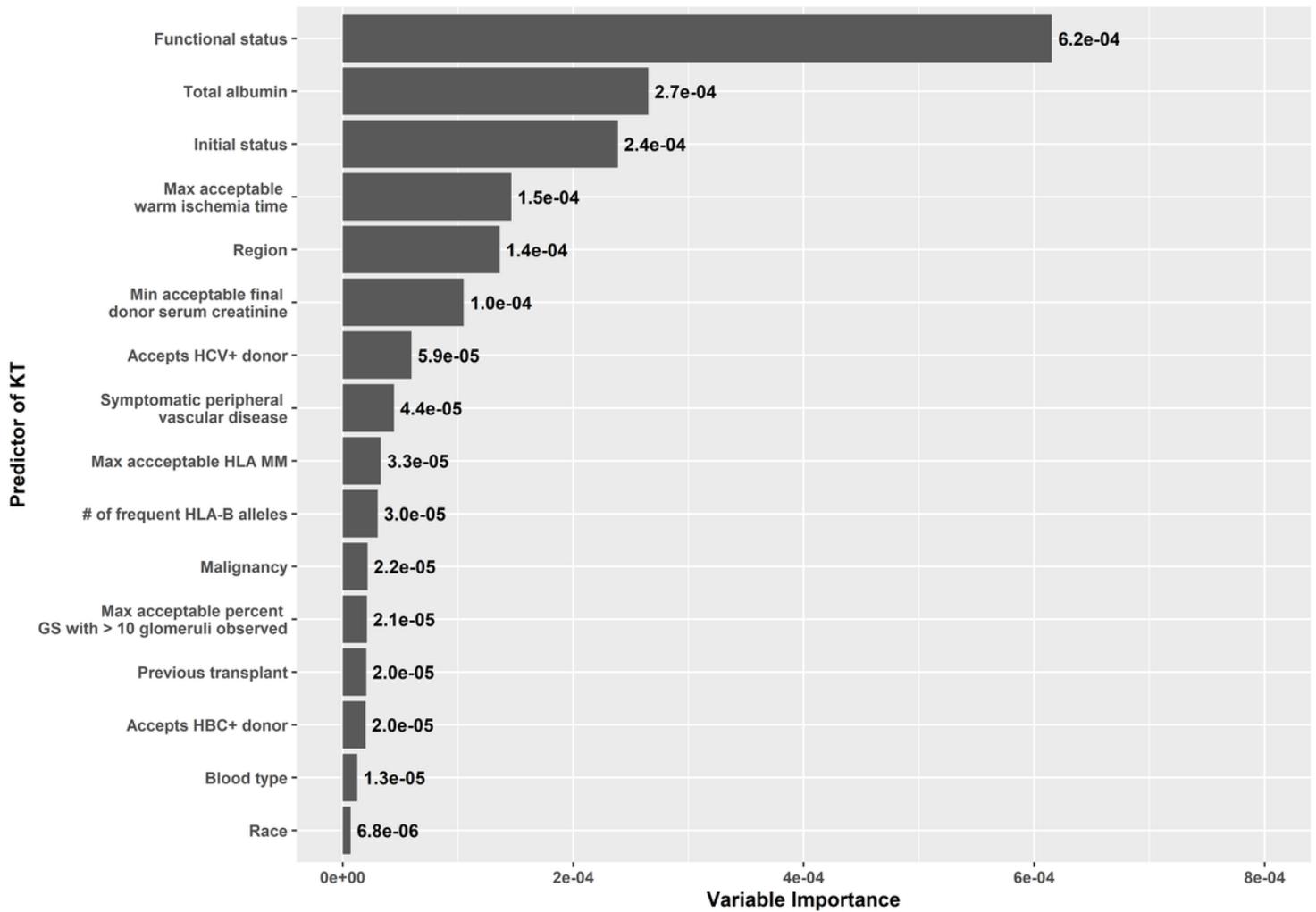


Figure 3

Predictors of one year post-listing CIF for KT. VIS was computed for each predictor unveiled by the best performing model. A total of 16 predictors were identified for prediction of CIF1yr, KT.

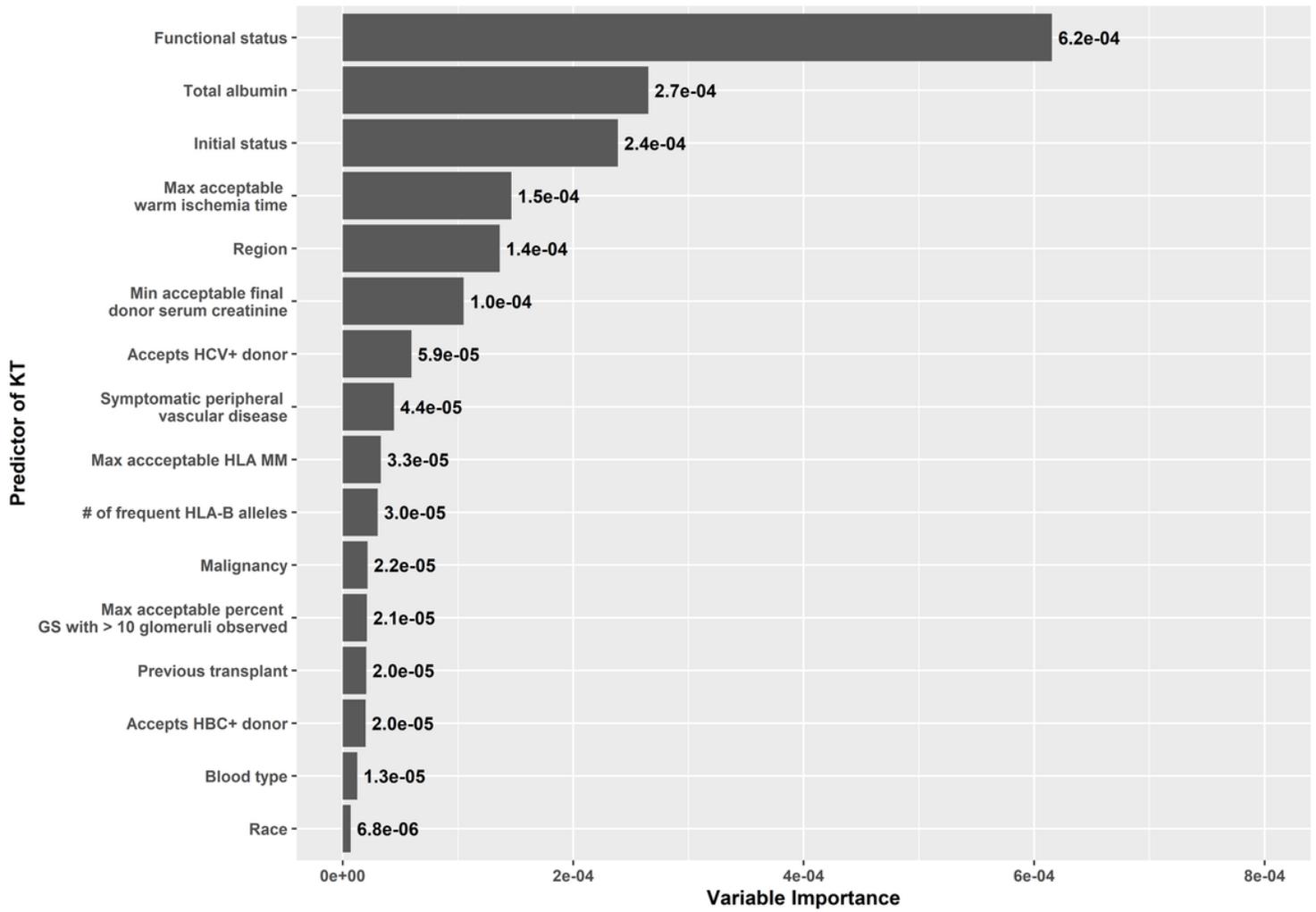


Figure 3

Predictors of one year post-listing CIF for KT. VIS was computed for each predictor unveiled by the best performing model. A total of 16 predictors were identified for prediction of CIF1yr, KT.

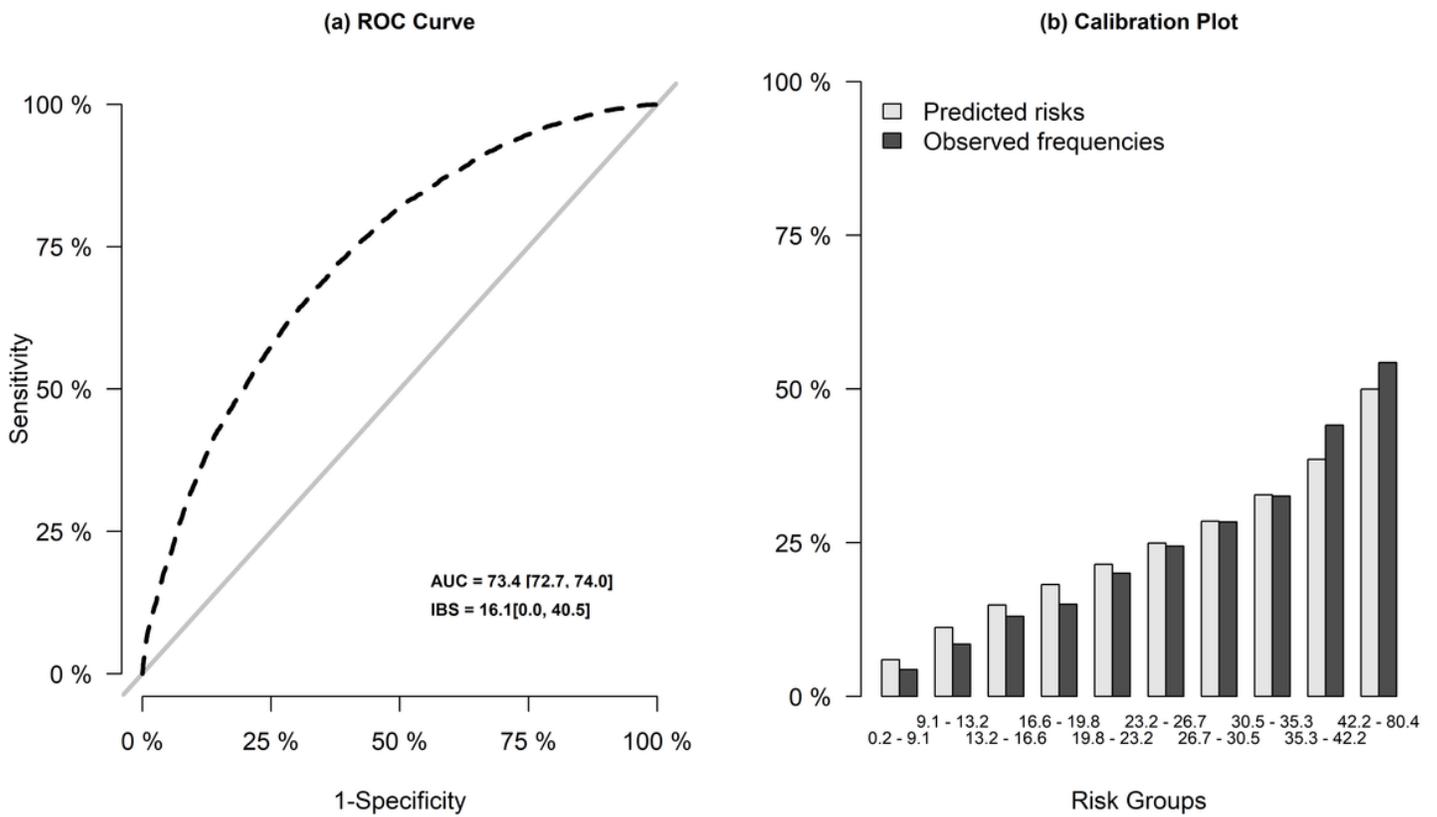


Figure 4

Performance of predicted CIF1yr, KT on a validation set. (a) ROC curve with ROC-AUC and IBS for predicted CIF1yr, KT from the best model. (b) Calibration plot. Predicted CIF1yr, KT values were grouped into deciles and the respective mean and observed prevalence of KT per quantile were plotted against each other.

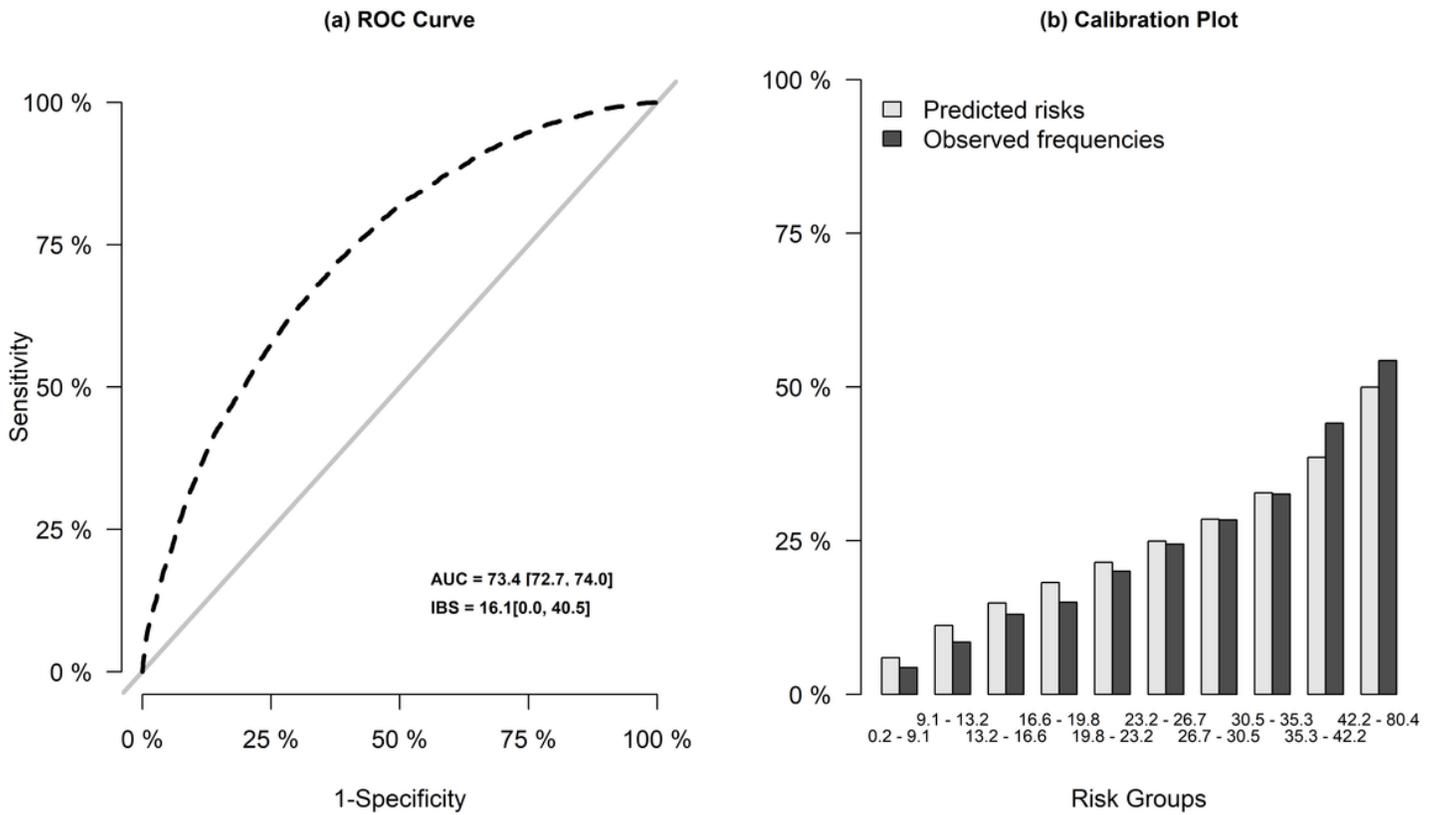


Figure 4

Performance of predicted CIF1yr, KT on a validation set. (a) ROC curve with ROC-AUC and IBS for predicted CIF1yr, KT from the best model. (b) Calibration plot. Predicted CIF1yr, KT values were grouped into deciles and the respective mean and observed prevalence of KT per quantile were plotted against each other.

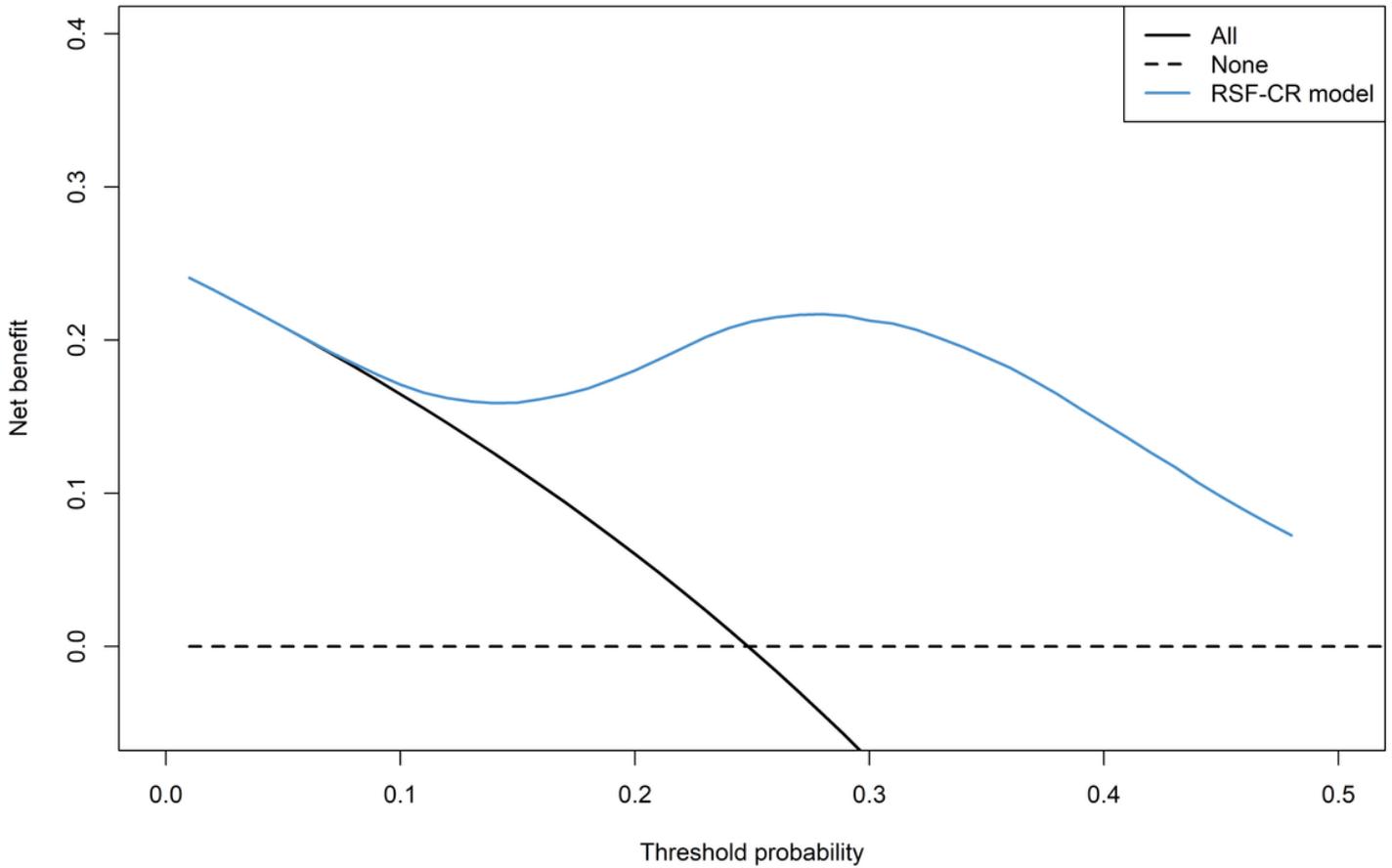


Figure 5

Decision curve analysis of predicted CIF1yr, KT. NB was quantified and plotted across potential threshold probabilities for three treatment strategies: (1) treat none; (2) treat all; and (3) treat based on the RSF-CR model.

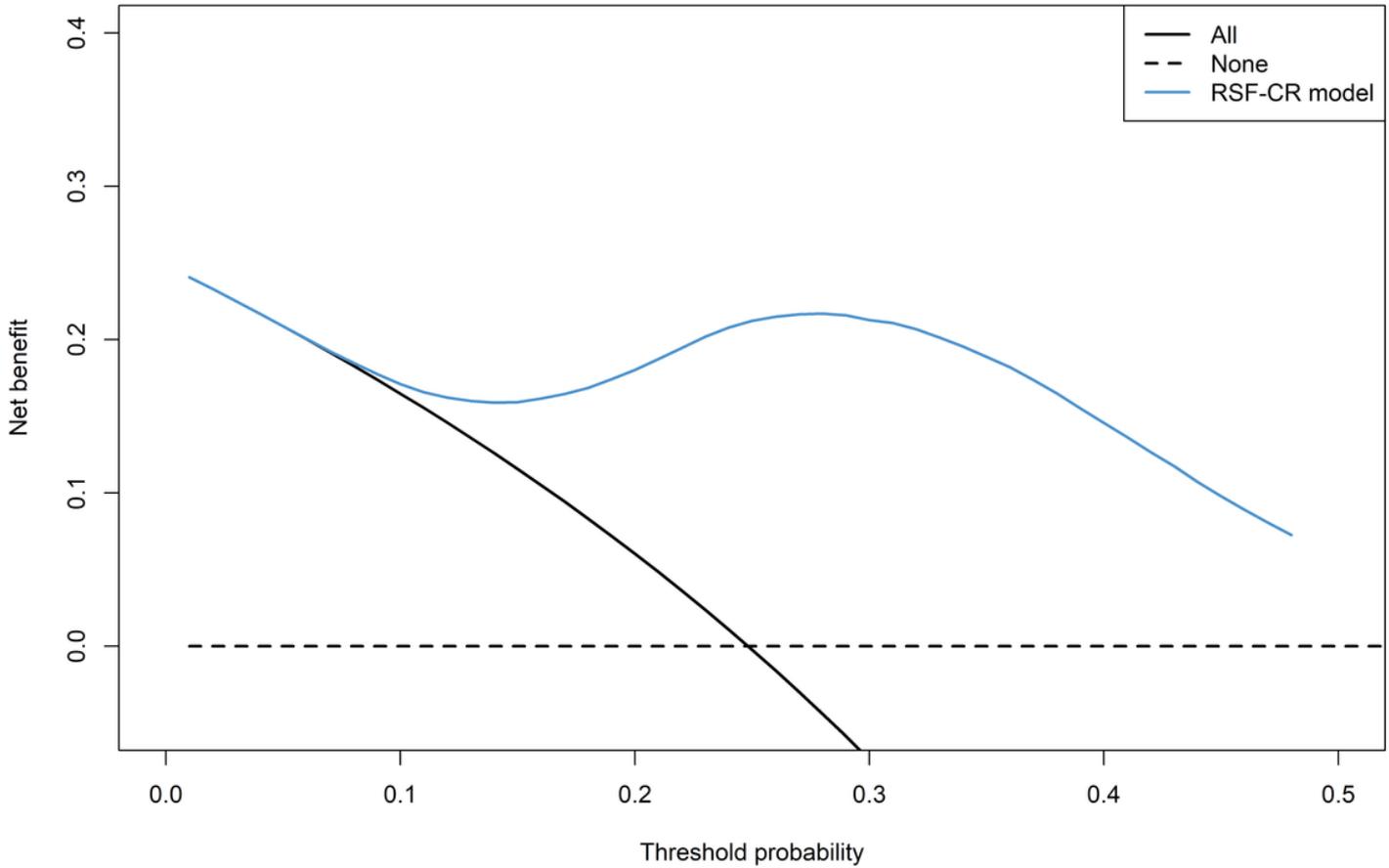


Figure 5

Decision curve analysis of predicted CIF1yr, KT. NB was quantified and plotted across potential threshold probabilities for three treatment strategies: (1) treat none; (2) treat all; and (3) treat based on the RSF-CR model.

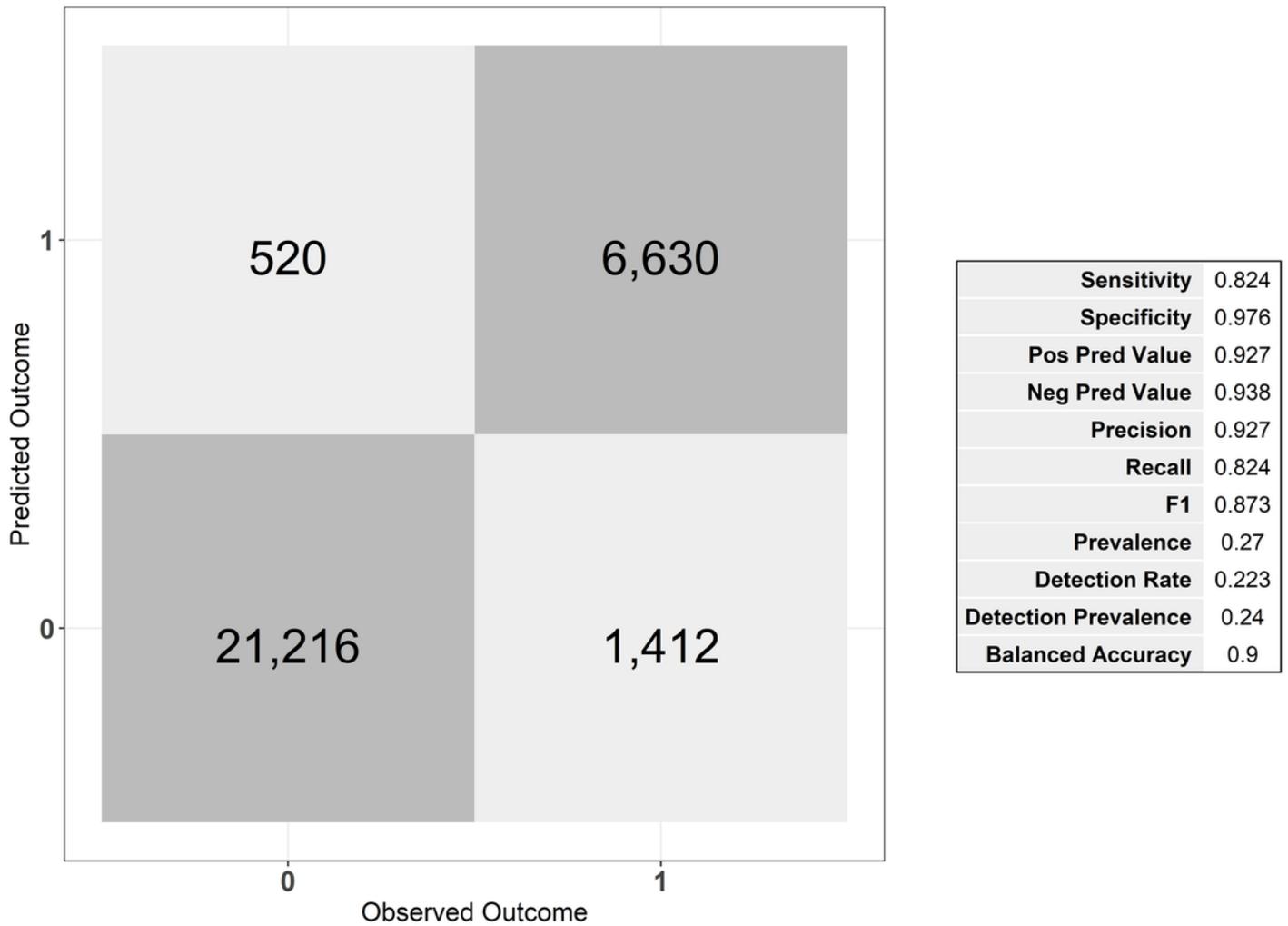


Figure 6

Confusion matrix of the RSF-CR classifier for CIF1yr, KT. Patients in the validation cohort were grouped into four categories: true positive, false positive, true negative, false negative based on the dichotomized version of predicted CIF1yr, KT. Observed outcome of 1 and 0 indicate “observed” and “not observed” while predicted outcome of 1 and 0 indicate “above the threshold” and “below the threshold” respectively.

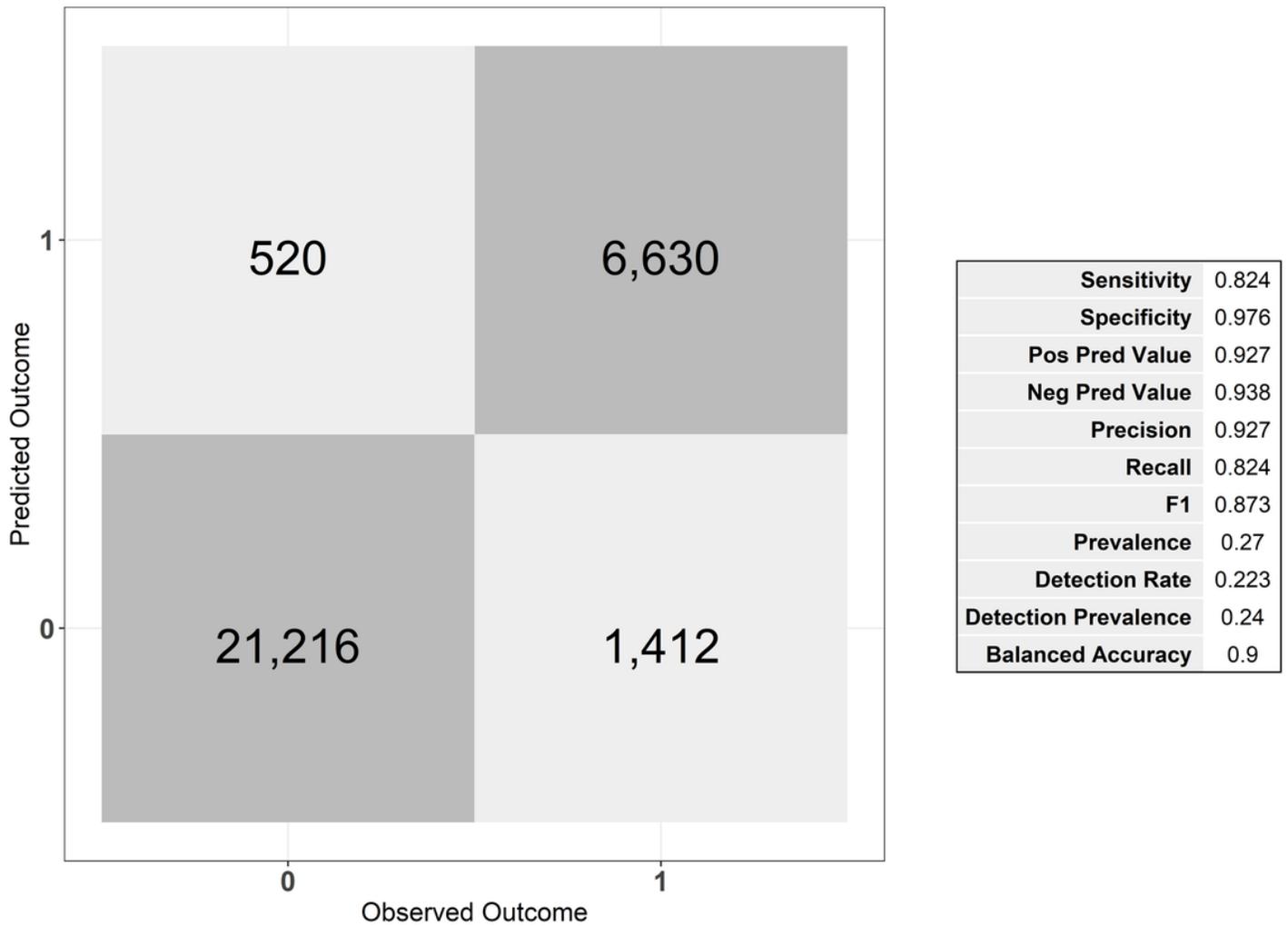


Figure 6

Confusion matrix of the RSF-CR classifier for CIF1yr, KT. Patients in the validation cohort were grouped into four categories: true positive, false positive, true negative, false negative based on the dichotomized version of predicted CIF1yr, KT. Observed outcome of 1 and 0 indicate “observed” and “not observed” while predicted outcome of 1 and 0 indicate “above the threshold” and “below the threshold” respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure1.tiff](#)
- [SupplementaryFigure1.tiff](#)
- [SupplementaryFigure2.tiff](#)
- [SupplementaryFigure2.tiff](#)
- [SupplementaryFigure3.tiff](#)
- [SupplementaryFigure3.tiff](#)

- [SupplementaryFigure4.tiff](#)
- [SupplementaryFigure4.tiff](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable2.xlsx](#)