

Covariates adjustment questioned conclusions of predictive analyses: an illustration with the Kidney Donor Risk Index

Etienne Dantan, Florent Le Borgne, Magali Giral, Angelina Dion, Anne-Hélène Querard, Yohann Foucher

▶ **To cite this version:**

Etienne Dantan, Florent Le Borgne, Magali Giral, Angelina Dion, Anne-Hélène Querard, et al.. Covariates adjustment questioned conclusions of predictive analyses: an illustration with the Kidney Donor Risk Index. Journal of Clinical Epidemiology, 2021, 135, pp.103-114. 10.1016/j.jclinepi.2021.02.007 . hal-03237474

HAL Id: hal-03237474

https://nantes-universite.hal.science/hal-03237474

Submitted on 24 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Covariates adjustment questioned conclusions of predictive analyses: an illustration with the Kidney Donor Risk Index

E. Dantan[1], F. Le Borgne[1,2], M. Giral[3,4], A. Dion[1], AH Querard[1,5], Y. Foucher[1,6]

1. INSERM UMR 1246 - SPHERE, Nantes University, Tours University, Nantes, France.
2. IDBC/A2com, Pacé, France.
3. CRTI UMR 1064, Inserm, Université de Nantes, ITUN, CHU Nantes, RTRS Centaure, Nantes, France.
4. Centre d'Investigation Clinique en Biothérapie, Nantes, France.
5. Department of Nephrology, Dialysis and Transplantation, Departmental Hospital of Vendée, La Roche -sur- Yon, France.
6. Centre Hospitalier Universitaire de Nantes, Nantes, France.

Corresponding author: Dr. Etienne DANTAN. INSERM UMR 1246 - SPHERE, Institut de Recherche en Santé 2, 22 boulevard Bénoni Goullin, 44200 Nantes, France. Phone: +33 (0)2 53 00 91 23. Email: Etienne.Dantan@univ-nantes.fr

Running title: Covariates in predictive studies: an illustration in kidney transplantation

Abbreviations: 95%CI: 95% Confidence Interval; AUC: Area Under the Curve; BMI: Body Mass Index; CMV: Cytomegalovirus; CVA: Cerebro-Vascular Accident; DIVAT: Données Informatisées et VAlidées en Transplantation; DSA: Donor Specific Antibody; HCV: Hepatitis C Virus; HLA: Human Leukocyte Antigen; HR: Hazard Ratio; KDRI: Kidney Donor Risk Index; IPW: Inverse Probability Weighting; NA: Not Available (missing data); SD: Standard Deviation.

Abstract

Objective. We aimed to illustrate that considering covariates can lead to meaningful interpretation of the discriminative capacities of a prognostic marker. For this, we evaluated the ability of the Kidney Donor Risk Index (KDRI) to discriminate kidney graft failure risk.

Study design and Setting. From 4114 French patients, we estimated the adjusted area under the time-dependent ROC curve by standardizing the marker and weighting the observations. By weighting the contributions, we also studied the impact of KDRI-based transplantations on the patient and graft survival.

Results. The covariate-adjusted AUC varied from 55% [95%CI:51%–60%] for a prognostic up to 1 year post-transplantation to 56% [95%CI: 52%–59%] up to 7 years. The Restricted Mean Survival Time (RMST) was 6.44 years for high-quality graft recipients [95%CI:6.30–6.56] and would have been 6.31 years [95%CI: 6.13–6.46] if they had medium-quality transplants. The RMST was 5.10 years for low-quality graft recipients [95%CI: 4.90–5.31] and would have been 5.52 years [95%CI: 5.17–5.83] if they had medium-quality transplants.

Conclusions. We demonstrated that the KDRI discriminative capacities were mainly explained by the recipient characteristics. We also showed that counterfactual estimations, often used in causal studies, are also interesting in predictive studies, especially regarding the new available methods.

Keywords: Discrimination, Prognostic, Counterfactual prediction, Kidney transplantation, Kidney Donor Risk Index.

What is new ?

- The standardized and weighted time-dependent ROC curve allows to estimate covariate-adjusted discriminative capacities.
- We report for the first time that the capacities of the KDRI to discriminate recipients according to their risk of graft failure were mainly the consequence of the transplant allocation policy in place over the study period.
- We reported close counterfactual post-transplantation outcomes of the recipients regardless of the quality of the graft they received.
- Considering covariates in predictive studies may lead to a meaningful interpretation of the intrinsic predictive performance of a marker or a scoring system.
- Novel methodologies exist to address covariates in predictive studies; they deserve a higher visibility and need to be disseminated.

1. Introduction

Many methods are available to adjust for covariates in causal studies. In contrast, adjustment covariates are quasi-systematically ignored in predictive studies. Discriminative capacities are often reported using the unadjusted AUC of the ROC curve, reflecting the observed discriminative capacities of a marker in a population. While this quantity is important, a complementary covariate-adjusted estimation may be of interest. As demonstrated by Janes and Pepe [1], the variables associated with a marker can constitute nuisances in generalizing the results. For instance, in a multicentric study, the center can explain by itself a part of the discriminative capacities. The ROC curve estimation in a counterfactual population where the prognostic marker distribution is independent of covariates would estimate its intrinsic discriminative performance. Few methods exist to estimate the intrinsic discriminative capacities of a prognostic marker in the presence of covariates [2–4].

In kidney transplantation, Rao et al. [5] proposed the Kidney Donor Risk Index (KDRI), a score of the donor marginality aiming to discriminate grafts according to their risk of failure. The KDRI is calculated using the following donor characteristics: age, ethnicity, height, weight, histories of hypertension and diabetes, serum creatinine, hepatitis C virus (HCV) serology, Cerebro-Vascular Accident (CVA) as the cause of death, and donation after cardiac death. Rao et al. [5] used a multivariable Cox model to construct the KDRI. It was adjusted for the recipient characteristics and the year of the transplant. However, the authors did not consider these covariates when assessing its' discriminative capacities.

Whilst the KDRI was developed in the United States and has been used in its kidney allocation system since 2014 [6,7], only a few external validation studies exist. Three studies performed in the Netherlands, Germany and United-Kingdom concluded to its low discriminative capacities [8–10]. Collectively, these studies suffered from the same limitation as the first internal validation conducted by Rao et al. [5]. They investigated the raw (unadjusted) discriminative capacities. In addition to the unadjusted discriminative capacities, i.e. conditional on the observed validation cohort, considering that elderly patients with more frequent comorbidities were more likely to receive grafts from older donors with several graft failure risk factors can be meaningful. Indeed, the predictive capacities of the KDRI may be related in part to the recipient's characteristics. Recently, Clayton et al. confirmed the low discriminative capacity of the KDRI from an external validation in the Australian and New Zealand kidney recipient population [11]. Interestingly, they also reported that the discriminative capacities increased using a survival model adjusted for KDRI, transplant, and recipient characteristics. Whilst this assessed the discriminative performance of a more global score incorporating the KDRI and the adjustment variables, it did not allow to consider the intrinsic discriminative performance of the KDRI by isolating its discriminative effect independently of the adjustment covariates. Complementary to unadjusted predictive performances, the concept of covariate-adjusted predictive capacities is of primary importance to evaluate the intrinsic discriminative performances of the KDRI, i.e. independently of the recipient characteristics. Indeed, the KDRI is mainly used before graft allocation, for instance to identify marginal transplants dedicated to the older recipients or to systematically transport these organs using machine perfusion.

The PROgnosis RESearch Strategy (PROGRESS) group reported that most publications in predictive medicine present model developments, a small number report external validations, and only very few aim to demonstrate their clinical utility [12]. In this study, we propose for the first time an external validation of the KDRI by considering the recipients covariates related to organ allocation policies. We also aimed to assess the impact of the KDRI-based transplantations by predicting the counterfactual post-transplantation patient outcomes according to the quality of the graft they received. Besides the relevance of our study to kidney transplantation, we aimed to illustrate the importance of considering covariates in predictive studies more generally.

2. Materials and methods

2.1. Study Population

Data were extracted from the French multicentric DIVAT cohort (Données Informatisées et VAlidées en Transplantation; www.divat.fr, CNIL decision DR-2025-087 N°914184, 15 February 2015). All participants gave informed consent. A total of 4716 patients met the following inclusion criteria: adult recipients of a single kidney transplanted for the first time between January 2006 and December 2018 from a deceased ABO-compatible donor. We did not include 602 patients due to missing data for at least one variable involved in the KDRI computation. We used the KDRI formula based on ten donor characteristics as proposed by Rao et al. [5]. We assumed all patients were white because the ethnicity was not recorded (according to French law). The final study population included 4114 patients.

2.2. Outcomes

Similar to the Rao et al. study [5], the outcome was the time from the transplantation (baseline) to patient and graft failure, defined as the first event between a return-to-dialysis, a pre-emptive re-transplantation or death with a functioning graft.

2.3. Analysis of the discriminative capacities

We assessed the raw (unadjusted) discriminative capacities using the area under the time-dependent ROC curve (AUC) for prognostic times from one to seven years' post-transplantation [13]. For one center, only 15 patients were still at risk at seven years post-transplantation (Table S1 in Web supplementary materials). The discriminative capacities cannot reasonably be investigated beyond this time.

The prognostic marker's intrinsic discriminative capacities were estimated using the area under the standardized and weighted time-dependent ROC curves, as proposed by Le Borgne et al. [4]. Such ROC curve is defined as the graphical representation of the sensitivity versus (1-specificity), for all the possible threshold of the standardized continuous prognostic marker, both quantities being estimated in a counterfactual population where the baseline covariates have similar distribution between cases (the event occurs before the prognostic time) and controls (event-free at the prognostic time). The mathematical definitions are proposed in Appendix 1 in Web supplementary materials. For each prognostic time, two steps composed this approach. Firstly, we standardized the score according to the covariates among the controls using a multiple linear regression of the log-transformed KDRI as the outcome for the subset of recipients with a participating time greater than the prognostic

time. The multiple linear regression must be estimated for each prognostic time since the controls change given the participating time. We were particularly vigilant to the respect of the linear regression assumptions at each prognostic time from 1 to 7 years post-transplantation: the KDRI was log-transformed to respect the Gaussian assumption, and we checked the homoscedasticity graphically. From this model, we computed the standardized KDRI values for the whole sample, i.e. the standardized residuals. Secondly, we weighted the contribution of each patient by their probability of experiencing the event before the prognostic time. The objective was to consider the discrimination performance of the prognostic score on a counterfactual population where baseline variables have similar distribution between cases and controls. To obtain these weights, we fitted a multivariable Cox model and estimated the cumulative baseline hazard function by using the Breslow estimator. We assumed the censoring process as independent. We checked the log-linearity for quantitative variables by analyzing the Martingale residuals [14]. Proportional hazards assumption was evaluated by plotting log-minus-log survival curves, and by analyzing the scaled Schoenfeld residuals [15]. We stratified the baseline hazard function for non-proportional variables.

As recommended [4], we considered the causes of the KDRI distribution or patient and graft failure risk as adjustment covariates. We did not include donor features or variables on the causal pathway. As suggested by VanderWeele and Shpitser [16], the covariates were selected blindly by two expert nephrologists (MG, AHQ). Any discrepancies were resolved by a literature review and discussion with a third independent nephrologist. The resulting set of covariates was then used in the subsequent two steps, i.e. standardization and weighting.

2.4. Patient-graft survival probabilities given KDRI-based stratification

We considered two KDRI thresholds: the 20th percentile, as used in the US kidney allocation system [6], and the 80th percentile, since recipients of high-level KDRI kidneys may have a long-term benefit despite being marginal transplants [17]. From such thresholds, we thus studied three types of grafts in terms of quality: high, medium and low. We were interested in the possible consequences, in terms of survival probabilities, of a stratified medical decision that would be based on the KDRI. For the construction of the counterfactual worlds and to respect the positivity assumption, we considered that recipients of a high-quality graft could not receive a low-quality one, and vice-versa. More precisely, we defined two scenarios. First, we wondered whether recipients of a high-quality graft would have had a worse prognosis if they had received a medium-quality organ, excluding the possibility that they received a low-quality graft (Scenario #1). Second, we wondered whether recipients of a low-quality graft would have had a better prognosis if they had received a medium-quality organ, excluding the possibility that they received a high-quality graft (Scenario #2).

For the two scenarios, we obtained the raw survival curves using the Kaplan-Meier (KM) survival estimator [18]. We also estimated counterfactual predictions. The covariate-adjusted survival curves by using the Inverse Probability Weighting (IPW) [19]. The corresponding restricted mean survival times were estimated as the area under the covariate-adjusted survival curves until 7 years post-transplantation. We used the weighted Cox model to obtain the corresponding hazard ratio (HR) [20]. More precisely, we estimated the average treatment effect on the treated (ATT) effect [21], taking as the reference the extreme group, i.e. high-quality grafts in Scenario #1 and low-quality grafts in Scenario #2. For each scenario, the covariates previously identified as the cause of the KDRI distribution

or patient and graft failure risk by the two expert nephrologists were included in the logistic regression to compute the propensity scores [22].

2.5. Assumptions related to causal reasoning

Dickerman and Hernán recalled that estimating counterfactual predictions required the same assumptions as those of causal inference [23–25]. To assess if the initial covariates' imbalance between the KDRI strata were straightened after weighting, we estimated the standardized differences in the pseudo-population (Tables 2 and 3) [26]. The positivity assumption required that the conditional probability of receiving every KDRI value is greater than zero. We graphically observed an overlap of the propensity score distributions (Figure S1 in Web supplementary materials). As previously stated, we evaluated the assumptions related to the regressions we used to reduce the risk of model misspecification. The exchangeability assumption implies that we did not omit important covariates. Because we collected all the characteristics related to both the recipient and the donor involved in the graft allocation in France, one can reasonably assume the respect of the exchangeability assumption.

2.6. Software

All analyses were performed using R software version 3.5.1. [27]. The standardized and weighted time-dependent ROC curve analyses and IPW-based analyses were performed using the 0.9 version of the *RISCA* package.

3. Results

3.1. Description of baseline characteristics

The demographic characteristics of donors and recipients are listed in Table 1. We noted that the 4114 included patients were comparable to the 602 patients excluded due to missing data for the KDRI computation (Table S2 in Web supplementary materials). The median KDRI was 1.47 (range from 0.64 to 3.87). The 20^{th} and 80^{th} percentiles were 1.07 and 2.13, respectively. Several recipient characteristics appeared to be differentially distributed between the KDRI-based groups. The recipient age and Body Mass Index (BMI) increased with the KDRI ($p<0.0001$). Relapsing initial diseases were less frequent in the group with a high-level KDRI ($p<0.0001$). The histories of diabetes ($p<0.0001$), hypertension ($p<0.0001$), cardiovascular diseases ($p<0.0001$), dyslipidemia ($p<0.0001$) and neoplasia ($p<0.0001$) were more frequent in the group with a high-level KDRI.

3.2. Follow-up description

Among the 4114 patients, 530 patients returned to dialysis, five were preemptively re-transplanted, and 416 died with a functioning graft. The cumulative follow-up covered 16906 patient-years. The median follow-up time was 4.77 years (reverse Kaplan-Meier estimator). For the entire cohort, the patient-graft survival probabilities at 2-, 5- and 7-years post-transplantation were 89.0% (95%CI from 88.0% to 90.0%), 78.6% (95%CI from 77.1% to 80.1%) and 69.0% (95%CI from 67.1% to 71.0%), respectively. Figure 1 presents the patient-graft survival curves according to the KDRI-based strata and illustrates a more deteriorated survival for the high KDRI values (LogRank test, $p<0.05$).

3.3. Discriminative capacities of the KDRI

Figure 2 presents the raw discriminative capacities, with AUCs of the ROC curves ranging from 65% (95%CI from 62% to 68%) for a prognostic of patient and graft failure up to 1-year post-transplantation to 68% (95%CI from 65% to 70%) for a prognostic up to 7-years.

The set of adjustment covariates predefined by two expert nephrologists included risk factors of patient and graft failure (recipient age, recipient BMI, cold ischemia time, time spent on dialysis, relapsing initial nephropathy, histories of diabetes, hypertension, cardiovascular disease, neoplasia, HLA A-B-DR incompatibilities, pre-transplantation anti-HLA immunization of class I and II and transplantation center) and characteristics associated with the KDRI distribution (recipient age, recipient BMI, cold ischemia time and transplantation center). History of diabetes and transplantation center were considered as stratification variables due to non-proportionality. These recipient and transplantation characteristics constituted the set of covariates used to estimate the standardized and weighted time-dependent ROC curves. The corresponding covariate-adjusted AUC varied from 55% (95%CI from 51% to 60%) for a prognostic of patient and graft failure up to 1-year post-transplantation to 56% (95%CI from 52% to 59%) up to 7- years post-transplantation (Figure 2).

3.4. Patient-graft survival probabilities given KDRI-based stratification

Figure 1 presents the raw patient and graft survival curves. The survival rate at 7-years post-transplantation was 81.3% (95%CI from 77.7% to 85.0%) for high-quality transplants, 71.8% (95%CI from 69.3% to 74.3%) for the recipients of medium-quality transplants, and 46.8%

(95%CI from 41.9% to 52.2%) for low-quality transplants. For a cohort followed-up to seven years, the unadjusted mean patient survival times with a functioning graft were 6.37 years (95%CI from 6.23 years to 6.49 years) for the recipients of high-quality grafts, 5.96 years (95%CI from 5.91 years to 6.08 years) for recipients of medium-quality grafts and 4.99 years (95%CI from 4.79 years to 5.18 years) for the recipients of low-quality grafts. The mean survival time of recipients of high-quality grafts was significantly 4.83 months longer (95%CI from 2.45 to 6.24 months) compared to recipients of medium-quality grafts. The mean survival time of recipients of low-quality grafts was significantly 11.72 months shorter (95%CI from 9.53 to 14.65 months) compared to recipients of medium-quality grafts.

For the two counterfactual prediction scenarios, the set of adjustment covariates associated with the patient and kidney graft failure were considered. Table 2 presents the characteristics of the pseudo-population obtained in Scenario #1: the imbalance of covariates was straightened. We estimated a non-significant difference in the risk of patient and graft failure between recipients of a high-quality graft and the same patients if they had received a medium-quality graft (HR = 0.80, 95%CI from 0.58 to 1.09). The adjusted patient-graft survival curves were close (Figure 3). For a cohort followed-up to 7 years, the mean patient survival time with a functioning graft was 6.44 years for the recipients of high-quality graft (95%CI from 6.30 years to 6.56 years) versus 6.31 years (95%CI from 6.13 years to 6.46 years) if they had received a medium-quality organ. This corresponded to a non-significant increase of 6 weeks (95%CI from -4 weeks to 17 weeks) attributable to the increase in the graft quality.

Table 3 presents the distribution of the characteristics of the pseudo-population obtained for Scenario #2 with well-balanced covariates. The patient and graft failure risk among low-

quality graft recipients tended to be higher than that of the same patients if they had received a medium-quality organ (HR = 1.34, 95%CI from 1.03 to 1.76). The corresponding adjusted patient-graft survival curves are presented in Figure 3. For a cohort followed-up to seven years, the mean patient survival time with a functioning graft was 5.10 years for recipients of low-quality grafts (95%CI from 4.90 years to 5.31 years) and would have been 5.52 years (95%CI from 5.17 years to 5.83 years) if they had received a medium-quality graft. This corresponded to a significant decrease of 5 months due to the lower transplant quality (95%CI from 1 months to 38 months).

4. Discussion

In this study, we illustrate that considering covariates could be meaningful in the evaluation of discriminative capacities. The unadjusted AUC we reported varied between 0.65 and 0.68 depending on the prognostic time. These results are comparable to those initially proposed by Rao et al. [5] (AUC = 0.62) and are also concordant with the subsequent external validations (AUC = 0.62 in Peters-Sengers et al. [8], AUC = 0.66 in Rehse et al. [9] and AUC=0.63 in Clayton et al. [11]). These values correspond to the observed probabilities that for a randomly selected pair of patients with and without patient and graft failure at the prognostic time, the patient with the failure has a greater KDRI than the other patient. In addition, we estimated the covariate-adjusted AUC ranged from 0.54 to 0.57 depending on the prognostic time. In a counterfactual population where the donor and recipients would have been independent, these values corresponded to the expected probabilities that for a randomly selected pair of patients with and without patient and graft failure at the

prognostic time, the patient with the failure has a greater KDRI than the other patient. These new results question the KDRI as an indicator to evaluate the intrinsic quality of a graft from a deceased donor. The differences between unadjusted and covariate-adjusted AUCs show that an important part of the discriminative capacities of the KDRI observed in the cohort is based on the graft allocation policy and not on the quality of the graft itself.

Furthermore, we investigated the causal effect on the patient-graft survival of the transplant quality evaluated by the KDRI. We proposed two counterfactual worlds where recipients of high-quality graft would have received medium-quality grafts, and recipients of low-quality grafts would have received medium-quality grafts. In each scenario, we estimated low differences, suggesting that the KDRI-based graft quality has no relevant impact on patient-graft survival in France. As previously stated for the Expanded Criteria Donor classification of kidney transplants [28], these results also suggests a possible expansion of the donor pool may be possible, and is especially pertinent given the ongoing increase in the worldwide donor organ shortfall Notably, Aubert et al. demonstrated that considering low-quality grafts could reduce the discard rate of deceased donor kidneys [29]. The decision to discard a donor organ is usually made at the time of retrieval and is primarily based on the intrinsic quality of the graft, while the potential recipients have not yet been identified. Our covariate-adjusted results performed among kidney transplant recipients challenges the KDRI as a relevant tool to evaluate such an intrinsic quality. Besides, this quality must be relatively evaluated regarding the impact of delaying transplantation of candidates who would have received a marginal transplant. But such comparison between recipients of a marginal graft versus comparable patients not still transplanted is methodologically challenging [30].

We used the standardized and weighted time-dependent ROC curve to assess the discriminative capacities of a marker in the presence of adjustment covariates and censored times-to-event. Other methods are possible. One can stratify the analyses according to covariate levels and compute the weighted mean of the estimands (AUC, survival rates, etc.). However, this approach can be applied when the number of covariates is low, and continuous covariates must be categorized. To evaluate the discriminative capacities when there is no censoring, one can alternatively use the theory of placement values [31,32]. With censored time-to-event data, several authors have proposed covariate-specific estimation of ROC curves: semi-parametric estimator, non-parametric kernel-based estimator [33], and predictive value curves [34]. These methods result in a subject-specific estimation for each value of the covariates, while the marginal approach we used allows a global evaluation, comparable to the methods based on the placement values [4].

Our study represents a single demonstrative example of the importance of questioning the adjustment covariates in a predictive study. It does not mean that covariates must be considered in each study aiming to evaluate predictive performance. However, many situations must question analysts about the use of these new methods. In biological studies for instance, patient age or gender often explain both the distribution of a biomarker and the outcome. A part of the apparent discriminative capacities of a biomarker may in fact be due to the patient characteristics [4]. Another example is the possible role of clusters, such as the center or the geographical area. The level of biomarkers may depend on the center-specific conditions of storage or handling and impact the discriminative capacities [35]. Besides, beyond our study focused on the intrinsic discriminative capacities, it may also be of

interest to assess the added prognostic value of a marker over existing scores or over individual predictors [36].

There are several limitations to our study. First, we performed a complete case analysis and we did not include patients with missing data concerning variables used in the KDRI. This represented 602 patients, i.e. 13% of the whole sample, and this probably reduced the study's statistical power. Second, whilst we limited the prognostic up to seven years post-transplantation due to the number of at-risk patients, longer-term results are essential to better appraise the KDRI performance. Third, we performed an external validation of the KDRI by considering the same survival outcome as in the original study of Rao et al. Nevertheless, another study could be to assess the capacities of the KDRI to discriminate the risk of return to dialysis, in competition with the patient death. In such a case, the methods should consider death as a competing risk. But to the best of our knowledge, the extension of covariate-adjusted ROC curve for competing events do not exist. Fourth, we could not exclude the potential bias induced by non-observed covariates. But, the risk factors of patient and kidney graft failure were available in the DIVAT cohort, especially those used for graft allocation.

In conclusion, our study illustrates the possible relevance of considering covariate-adjustement in predictive studies. In kidney transplantation, it allowed us to estimate the low intrinsic discriminative capacities of the KDRI. Novel methodologies exist to consider adjustment covariates in predictive studies, but they still need to be disseminated for a better use in practice.

Funding

This work was supported by the French Biomedicine Agency (reference: AOR Greffe 2016) and the French National Research Agency (ANR) (Research in Informatic and Statistic for Cohort Analyses, www.labcom-risca.com, reference: ANR-16-LCV1-0003-01).

Acknowledgments

We wish to thank Dr. Rémi Lenain for his help in the covariate selection. We also would like to thank members of the DIVAT consortium* and the clinical research assistant team (S. Le Floch, A. Petit, J. Posson, C. Scellier, V. Eschbach, K. Zurbonsen, C. Dagot, F. M'Raiagh, V. Godel, X. Longy, P. Przednowed). We are also grateful to Roche Pharma, Novartis, Sanofi and Astellas laboratories for supporting the DIVAT cohort (www.divat.fr) as the CENTAURE foundation (www.fondation-centaure.org).

*** Données Informatisées et VAlidées en Transplantation, DIVAT Cohort Collaborators (Medical Doctors, Surgeons, HLA Biologists)**

Nantes : *Gilles Blancho, Julien Branchereau, Diego Cantarovich, Agnès Chapelet, Jacques Dantal, Clément Deltombe, Lucile Figueres, Claire Garandeau, Magali Giral, Caroline Gourraud-Vercel, Maryvonne Hourmant, Georges Karam, Clarisse Kerleau, Aurélie Meurette, Simon Ville, Christine Kandell, Anne Moreau, Karine Renaudin, Anne Cesbron, Florent Delbos, Alexandre Walencik, Anne Devis* **; Paris-Necker :** *Lucile Amrouche, Dany Anglicheau, Olivier Aubert, Lynda Bererhi, Christophe Legendre, Alexandre Loupy, Frank Martinez, Rébecca Sberro-Soussan, Anne Scemla, Claire Tinel, Julien Zuber* **; Nancy :** *Pascal Eschwege, Luc*

Frimat, Sophie Girerd, Jacques Hubert, Marc Ladriere, Emmanuelle Laurain, Louis Leblanc, Pierre Lecoanet, Jean-Louis Lemelle **; Lyon E. Hériot :** *Lionel Badet, Maria Brunet, Fanny Buron, Rémi Cahen, Sameh Daoud, Coralie Fournie, Arnaud Grégoire, Alice Koenig, Charlène Lévi, Emmanuel Morelon, Claire Pouteil-Noble, Thomas Rimmelé, Olivier Thaunat* **; Montpellier :** *Sylvie Delmas, Valérie Garrigue, Moglie Le Quintrec, Vincent Pernin, Jean-Emmanuel Serre.*

References

[1] Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. Am J Epidemiol 2008;168:89–97. https://doi.org/10.1093/aje/kwn099.

[2] Janes H, Pepe MS. Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve. Biometrika 2009;96:371–82. https://doi.org/10.1093/biomet/asp002.

[3] Pardo-Fernandez JC. A review on ROC curves in the presence of covariates. Revstat Stat J 2014;12:21–41.

[4] Le Borgne F, Combescure C, Gillaizeau F, Giral M, Chapal M, Giraudeau B, et al. Standardized and weighted time-dependent receiver operating characteristic curves to evaluate the intrinsic prognostic capacities of a marker by taking into account confounding factors. Stat Methods Med Res 2017:0962280217702416. https://doi.org/10.1177/0962280217702416.

[5] Rao PS, Schaubel DE, Guidinger MK, Andreoni KA, Wolfe RA, Merion RM, et al. A comprehensive risk quantification score for deceased donor kidneys: the kidney donor

risk index. Transplantation 2009;88:231–6. https://doi.org/10.1097/TP.0b013e3181ac620b.

[6] Israni AK, Salkowski N, Gustafson S, Snyder JJ, Friedewald JJ, Formica RN, et al. New national allocation policy for deceased donor kidneys in the United States and possible effect on patient outcomes. J Am Soc Nephrol JASN 2014;25:1842–8. https://doi.org/10.1681/ASN.2013070784.

[7] Hart A, Smith JM, Skeans MA, Gustafson SK, Stewart DE, Cherikh WS, et al. OPTN/SRTR 2015 Annual Data Report: Kidney. Am J Transplant 2017;17 Suppl 1:21–116. https://doi.org/10.1111/ajt.14124.

[8] Peters-Sengers H, Heemskerk MBA, Geskus RB, Kers J, Homan van der Heide JJ, Berger SP, et al. Validation of the Prognostic Kidney Donor Risk Index Scoring System of Deceased Donors for Renal Transplantation in the Netherlands. Transplantation 2018;102:162–70. https://doi.org/10.1097/TP.0000000000001889.

[9] Rehse G, Halleck F, Khadzhynov D, Lehner LJ, Kleinsteuber A, Staeck A, et al. Validation of the Living Kidney Donor Profile Index in a European cohort and comparison of long-term outcomes with US results. Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc 2019;34:1063–70. https://doi.org/10.1093/ndt/gfy118.

[10] Watson CJE, Johnson RJ, Birch R, Collett D, Bradley JA. A Simplified Donor Risk Index for Predicting Outcome After Deceased Donor Kidney Transplantation. Transplantation 2012;93:314. https://doi.org/10.1097/TP.0b013e31823f14d4.

[11] Clayton PA, Dansie K, Sypek MP, White S, Chadban S, Kanellis J, et al. External validation of the US and UK kidney donor risk indices for deceased donor kidney transplant survival in the Australian and New Zealand population. Nephrol Dial Transplant Off Publ

Eur Dial Transpl Assoc - Eur Ren Assoc 2019;34:2127–31. https://doi.org/10.1093/ndt/gfz090.

[12] Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Med 2013;10. https://doi.org/10.1371/journal.pmed.1001381.

[13] Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000;56:337–44.

[14] Therneau TM, Grambsch PM, Fleming TR. Martingale-based residuals for survival models. Biometrika 1990;77:147–60. https://doi.org/10.1093/biomet/77.1.147.

[15] Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994;81:515–26. https://doi.org/10.1093/biomet/81.3.515.

[16] VanderWeele TJ, Shpitser I. A new criterion for confounder selection. Biometrics 2011;67:1406–13. https://doi.org/10.1111/j.1541-0420.2011.01619.x.

[17] Massie AB, Luo X, Chow EKH, Alejo JL, Desai NM, Segev DL. Survival benefit of primary deceased donor transplantation with high-KDPI kidneys. Am J Transplant Off J Am Soc Transplant Am Soc Transpl Surg 2014;14:2310–6. https://doi.org/10.1111/ajt.12830.

[18] Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–81.

[19] Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. Comput Methods Programs Biomed 2004;75:45–9. https://doi.org/10.1016/j.cmpb.2003.10.004.

[20] Le Borgne F, Giraudeau B, Querard AH, Giral M, Foucher Y. Comparisons of the performance of different statistical tests for time-to-event analysis with confounding factors: practical illustrations in kidney transplantation. Stat Med 2016;35:1103–16. https://doi.org/10.1002/sim.6777.

[21] Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. Stat Med 2014;33:1242–58. https://doi.org/10.1002/sim.5984.

[22] Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable Selection for Propensity Score Models. Am J Epidemiol 2006;163:1149–56. https://doi.org/10.1093/aje/kwj149.

[23] Dickerman BA, Hernán MA. Counterfactual prediction is not only for causal inference. Eur J Epidemiol 2020;35:615–7. https://doi.org/10.1007/s10654-020-00659-8.

[24] Hernan MA, Robins JM. Causal Inference. Taylor & Francis; 2019.

[25] van Geloven N, Swanson SA, Ramspek CL, Luijken K, van Diepen M, Morris TP, et al. Prediction meets causal inference: the role of treatment in clinical prediction models. Eur J Epidemiol 2020;35:619–30. https://doi.org/10.1007/s10654-020-00636-1.

[26] Ali MS, Groenwold RHH, Belitser SV, Pestman WR, Hoes AW, Roes KCB, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. J Clin Epidemiol 2015;68:112–21. https://doi.org/10.1016/j.jclinepi.2014.08.011.

[27] R Development Core Team . R: A Language and Environment for Statistical Computing. Vienna, Austria: 2010.

[28] Querard AH, Le Borgne F, Dion A, Giral M, Mourad G, Garrigue V, et al. Propensity score-based comparison of the graft failure risk between kidney transplant recipients of standard and expanded criteria donor grafts: Toward increasing the pool of marginal donors. Am J Transplant Off J Am Soc Transplant Am Soc Transpl Surg 2018. https://doi.org/10.1111/ajt.14651.

[29] Aubert O, Reese PP, Audry B, Bouatou Y, Raynaud M, Viglietti D, et al. Disparities in Acceptance of Deceased Donor Kidneys Between the United States and France and Estimated Effects of Increased US Acceptance. JAMA Intern Med 2019. https://doi.org/10.1001/jamainternmed.2019.2322.

[30] Lenain R, Boucquemont J, Leffondré K, Couchoud C, Lassalle M, Hazzan M, et al. Clinical trial emulation by matching time-dependent propensity scores: the example of estimating impact of kidney transplantation. Epidemiol Camb Mass 2020. https://doi.org/10.1097/EDE.0000000000001308.

[31] Pepe MS, Longton G. Standardizing diagnostic markers to evaluate and compare their performance. Epidemiol Camb Mass 2005;16:598–603. https://doi.org/10.1097/01.ede.0000173041.03470.8b.

[32] Pepe MS, Cai T. The analysis of placement values for evaluating discriminatory measures. Biometrics 2004;60:528–35. https://doi.org/10.1111/j.0006-341X.2004.00200.x.

[33] Song X, Ma S, Huang J, Zhou X-H. A semiparametric approach for the nonparametric transformation survival model with multiple covariates. Biostat Oxf Engl 2007;8:197–211. https://doi.org/10.1093/biostatistics/kxl001.

[34] Rodríguez-Álvarez MX, Meira-Machado L, Abu-Assi E, Raposeiras-Roubín S. Nonparametric estimation of time-dependent ROC curves conditional on a continuous covariate. Stat Med 2016;35:1090–102. https://doi.org/10.1002/sim.6769.

[35] Janes H, Longton G, Pepe M. Accommodating Covariates in ROC Analysis. Stata J 2009;9:17–39.

[36] Xanthakis V, Sullivan LM, Vasan RS, Benjamin EJ, Massaro JM, D'Agostino RB, et al. Assessing the incremental predictive performance of novel biomarkers over standard predictors. Stat Med 2014;33:2577–84. https://doi.org/10.1002/sim.6165.

Tables and Figures

Table 1: Description of recipients, donors and transplantation characteristics for the cohort (n=4114) and for 3 types of kidney donor graft defined given KDRI quintiles (High-quality graft: lowest KDRI quintile, Medium-quality graft: three intermediate KDRI quintiles, Low-quality graft: highest quintile)

	Missing data	Global N=4114	High-quality graft (KDRI from 0.64 to 1.07) N=837	Medium-quality graft (KDRI from 1.07 to 2.13) N=2457	Low-quality graft (KDRI from 2.13 to 3.87) N=820	p-value
Quantitative characteristics - mean ± SD						
Recipient age (years)	0	53.07 ± 13.69	37.96 ± 11.71	53.57 ± 10.23	67.01 ± 7.69	<0.0001
Recipient BMI (kg/m²)	28	24.97 ± 4.52	23.37 ± 4.53	25.27 ± 4.54	25.69 ± 4.07	<0.0001
Donor age (years)[a]	0	54.10 ± 16.57	30.84 ± 9.89	55.17 ± 9.43	74.65 ± 5.80	<0.0001
Donor Serum Creatinine (mg/dL)[a]	0	91.43 ± 61.10	91.07 ± 67.15	94.00 ± 64.89	84.10 ± 38.02	0.0003
Donor height (cm)[a]	0	169.66 ± 9.76	174.44 ± 9.17	169.87 ± 9.20	164.15 ± 9.18	<0.0001
Donor weight (kg)[a]	0	74.34 ± 16.11	73.66 ± 14.97	75.58 ± 16.43	71.35 ± 15.85	<0.0001
Cold Ischemia Time (hours)	22	17.69 ± 6.81	17.46 ± 6.82	17.70 ± 6.99	17.89 ± 6.22	0.4450
Time spent on dialysis (years)	18	3.18 ± 3.13	3.15 ± 3.18	3.31 ± 3.20	2.84 ± 2.82	0.0008
Categorical characteristics - N (%)						
Recipient men	0	2553 (62.06)	515 (61.53)	1509 (61.42)	529 (64.51)	0.2689
Pre-transplant dialysis technique	12					0.1031
Preemptive transplantation		435 (10.60)	98 (11.75)	239 (9.76)	98 (11.98)	
Peritoneal dialysis		376 (9.17)	64 (7.67)	241 (9.84)	71 (8.68)	
Hemodialysis		3291 (80.23)	672 (80.58)	1970 (80.41)	649 (79.34)	
Relapsing initial nephropathy	2	972 (23.64)	265 (31.66)	582 (23.71)	125 (15.24)	<0.0001
History of diabetes	0	775 (18.84)	69 (8.24)	475 (19.33)	231 (28.17)	<0.0001
History of hypertension	0	3314 (80.55)	616 (73.60)	2006 (81.64)	692 (84.39)	<0.0001
History of cardiovascular disease	0	1488 (36.16)	187 (22.34)	900 (36.63)	401 (48.90)	<0.0001
History of dyslipidemia	0	1482 (36.01)	156 (18.64)	926 (37.69)	400 (48.78)	<0.0001
History of neoplasia	0	438 (10.65)	39 (4.66)	238 (9.69)	161 (19.63)	<0.0001
History of urological disease	0	444 (10.79)	88 (10.51)	277 (11.27)	79 (9.63)	0.4062
More than 4 HLA A-B-DR incompatibilities	41	677 (16.62)	159 (19.23)	346 (14.20)	172 (21.26)	<0.0001
Pre-transplantation anti-HLA immunization of class I	276	1323 (34.47)	292 (37.24)	763 (33.38)	268 (34.90)	0.1392
Pre-transplantation anti-HLA immunization of class II	309	1175 (30.88)	249 (32.25)	701 (30.94)	225 (29.61)	0.5318
Donor men	1	2406 (58.50)	596 (71.21)	1455 (59.22)	355 (43.35)	<0.0001
Positive donor HCV serology[a]	0	34 (0.83)	2 (0.24)	24 (0.98)	8 (0.98)	0.1094
Donation after Cardiac Death[a]	0	238 (5.79)	37 (4.42)	199 (8.10)	2 (0.24)	<0.0001
Donor history of diabetes[a]	0	312 (7.58)	10 (1.19)	136 (5.54)	166 (20.24)	<0.0001
Donor history of hypertension[a]	0	1273 (30.94)	5 (0.60)	705 (28.69)	563 (68.66)	<0.0001
Donor CMV infection	3	2243 (54.56)	394 (47.07)	1337 (54.46)	512 (62.52)	<0.0001
Recipient CMV infection	46	2568 (63.13)	515 (62.20)	1539 (63.33)	514 (63.46)	0.8231
Donor CVA death[a]	0	2144 (52.11)	126 (15.05)	1339 (54.50)	679 (82.80)	<0.0001
Delayed Graft Function	192	1003 (25.57)	148 (18.39)	631 (26.92)	224 (28.98)	<0.0001
Depleting induction therapy	11	2053 (50.04)	406 (48.68)	1301 (53.04)	346 (42.40)	<0.0001
Expanded Donor Criteria[b]	0	1922 (46.72)	0 (0)	1103 (44.89)	819 (99.88)	<0.0001

[a] Variable in the KDRI definition that cannot be considered as a potential adjustment covariate

[b] Old definition of marginal graft defined from donor characteristics (donor age, donor SCr, donor history of hypertension, Donor CVA death) that cannot be considered as an adjustment covariate

Table 2: Description of recipient transplantation characteristics for the pseudo-population obtained after ATT weighting in Scenario #1 (patients receiving a high-quality graft compared to the same patients receiving a medium-quality graft). Standardized differences were provided for the group comparison in the pseudo-population and for the initial group comparison (87 patients discarded due to missing data on adjustment covariates)

	High-quality graft in the pseudo-population N=750 (50.5%)	**Medium-quality graft in the pseudo-population N=735 (49.5%)**	**Standardized differences in the pseudo-population (%)**	**Standardized differences between the two initial groups (%)**
Quantitative characteristics - mean ± SD				
Recipient age (years)	37.9 ± 11.5	38.2 ± 11.3	3.34	145.00
CIT (hours)	17.4 ± 6.8	17.3 ± 6.9	1.89	2.90
Time spent on dialysis (years)	3.0 ± 2.9	3.4 ± 3.0	11.14	8.62
Categorical characteristics - N (%)				
Recipient BMI (kg/m²)			1.89	25.37
<18	61 (8.1)	58 (7.9)		
[18-30[	618 (82.4)	603 (82.1)		
≥30	71 (9.5)	74 (10.0)		
Recipient men	459 (61.2)	460 (62.6)	2.84	0.25
Dialysis technique			16.09	11.22
Preemptive transplantation	93 (12.4)	57 (7.8)		
Peritoneal dialysis	57 (7.6)	70 (9.5)		
Hemodialysis	600 (80.0)	608 (82.7)		
Relapsing initial disease	237 (31.6)	239 (32.5)	1.90	16.52
History of diabetes	63 (8.4)	55 (7.5)	3.18	31.57
History of hypertension	557 (74.3)	559 (76.0)	4.02	19.46
History of cardiovascular disease	174 (23.2)	172 (23.4)	0.47	30.85
History of dyslipidemia	144 (19.2)	179 (24.3)	12.41	44.08
History of neoplasia	36 (4.8)	41 (5.5)	3.29	19.88
History of urological disease	76 (10.1)	94 (12.8)	8.26	4.40
More than 4 HLA A-B-DR incompatibilities	142 (18.9)	138 (18.8)	0.28	12.26
Pre-transplantation anti-HLA immunization of class I	271 (36.1)	268 (36.5)	0.70	6.46
Pre-transplantation anti-HLA immunization of class II	238 (31.7)	236 (32.1)	0.80	2.07
Recipient CMV infection	457 (61.3)	428 (59.3)	4.21	3.65
Depleting induction therapy	361 (48.1)	399 (54.3)	12.32	10.13

Table 3: Description of recipient transplantation characteristics for the pseudo-population obtained after ATT weighting in Scenario #2 (patients receiving a low-quality graft compared to the same patients receiving a medium-quality graft). Standardized differences were provided for the group comparison in the pseudo-population and for the initial group comparison (92 patients discarded due to missing data on adjustment covariates)

	Medium-quality graft in the pseudo-population N=779 (51.7%)	**Low-quality graft in the pseudo-population N=728 (48.3%)**	**Standardized difference in the pseudo-population (%)**	**Standardized difference between the two initial groups (%)**
Quantitative characteristics - mean ± SD				
Recipient age (years)	67.5 ± 8.1	66.8 ± 7.8	8.94	143.28
CIT (hours)	18.7 ± 6.6	17.9 ± 6.4	12.83	3.75
Time spent on dialysis (years)	2.7 ± 2.7	2.8 ± 2.7	4.53	17.70
Categorical characteristics - N (%)				
Recipient BMI (kg/m²)			4.10	8.69
<18	22 (2.8)	16 (2.2)		
[18-30[	633 (81.3)	594 (81.6)		
≥30	124 (15.9)	118 (16.2)		
Recipient men	494 (63.3)	466 (64.0)	1.42	5.56
Dialysis technique			6.68	10.50
Preemptive transplantation	108 (13.9)	92 (12.6)		
Peritoneal dialysis	53 (6.8)	61 (8.4)		
Hemodialysis	618 (79.3)	575 (79.0)		
Relapsing initial disease	93 (12.0)	111 (15.2)	9.49	22.66
History of diabetes	220 (28.2)	205 (28.2)	0.02	21.35
History of hypertension	690 (88.6)	626 (86.0)	7.72	10.23
History of cardiovascular disease	390 (50.0)	354 (48.6)	2.77	23.25
History of dyslipidemia	416 (53.3)	370 (50.8)	5.03	24.52
History of neoplasia	167 (21.4)	141 (19.4)	5.06	26.79
History of urological disease	96 (12.3)	71 (9.8)	8.16	5.67
More than 4 HLA A-B-DR incompatibilities	173 (22.3)	156 (21.4)	1.99	18.48
Pre-transplantation anti-HLA immunization of class I	276 (35.4)	255 (35.0)	0.82	4.14
Pre-transplantation anti-HLA immunization of class II	260 (33.3)	214 (29.4)	8.53	3.00
Recipient CMV infection	507 (65.4)	462 (64.3)	2.39	2.38
Depleting induction therapy	364 (46.7)	309 (42.5)	8.45	21.52

Figure 1: Patient-graft survival probability given the post-transplantation time from the Kaplan-Meier estimator for the 3 KDRI quintiles based group (High-quality graft: KDRI<1st quintile (1.07), Medium-quality graft: 1st quintile (1.07) ≤ KDRI< 3rd quintile (2.13), Low-quality graft: 3rd quintile (2.13) ≤ KDRI)

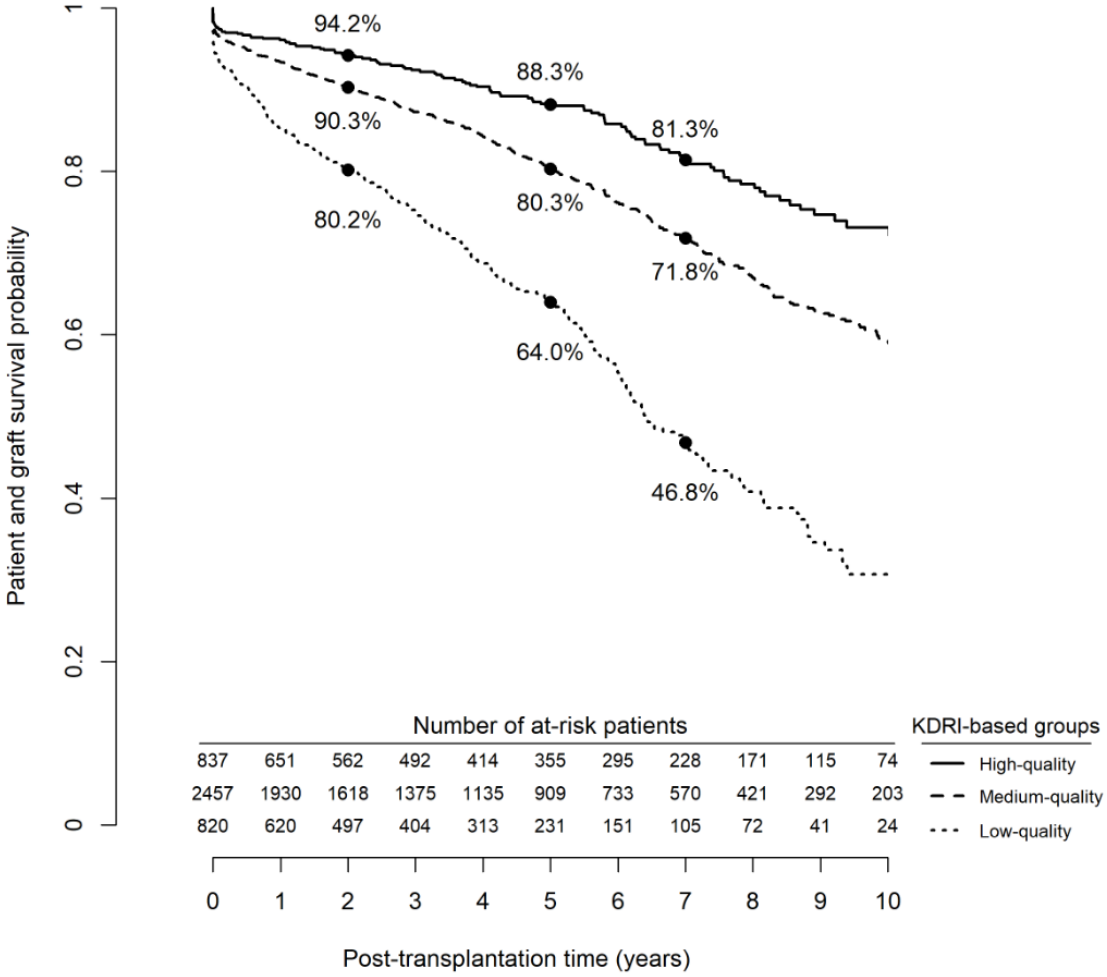


Figure 2: Unadjusted area under the time-dependent ROC curve (AUC) (dot) and AUC of the standardized and weighted time-dependent ROC curves (triangle) given prognostic time horizon, and their corresponding 95% confidence interval

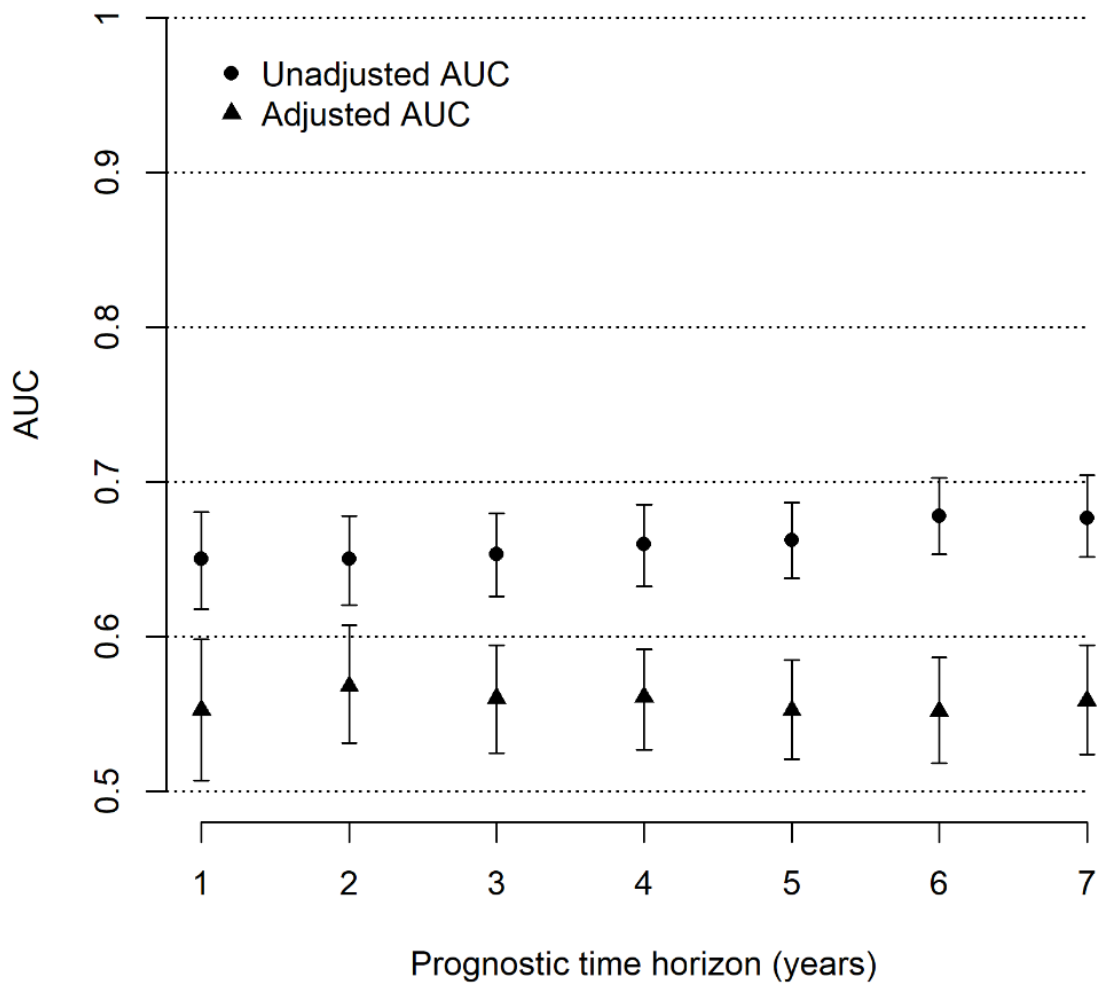
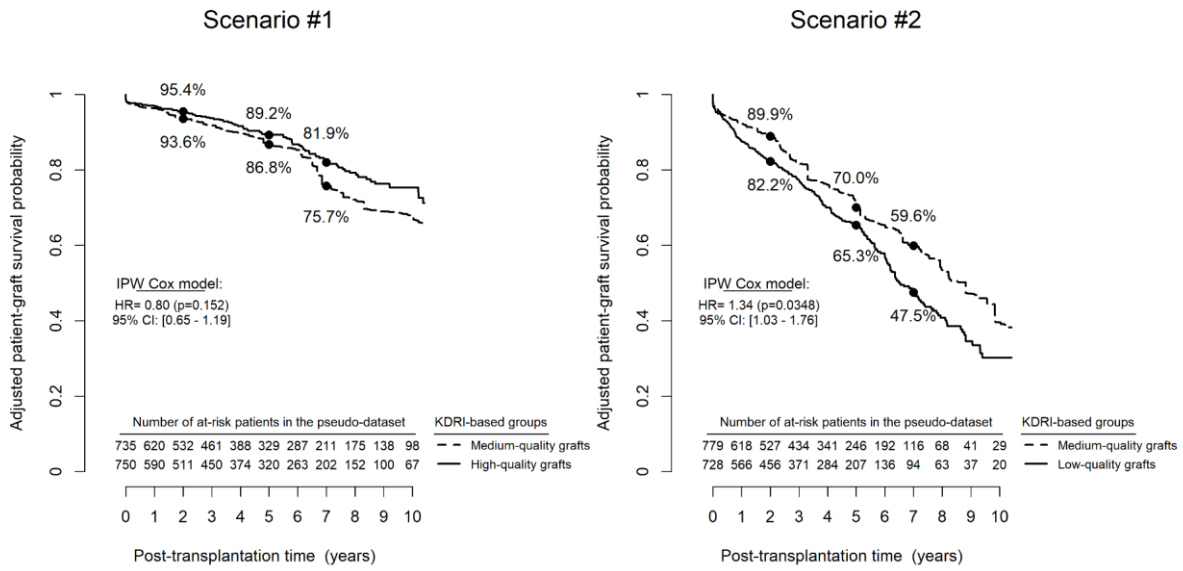


Figure 3: Adjusted patient-graft survival curves estimated by the IPW Kaplan–Meier estimator for recipients of a high-quality graft (solid line) and the same recipients receiving a medium-quality graft (dashed line) in Scenario #1 (87 patients discarded due to missing data on adjustment covariates) and for recipients of a low-quality graft (solid line) and the same recipients receiving a medium-quality graft (dashed line) in Scenario #2 (92 patients discarded due to missing data on adjustment covariates)



Web supplementary materials

Appendix 1: Estimators of the standardized and weighted sensitivity and specificity as defined in Le Borgne et al. (2017).

We assume a sample of n independent and identically distributed subjects with the observed data of $\{T_i, \delta_i, X_i, Z_i; i = 1, \dots, n\}$. T_i equals $min(T_i^*, C_i)$ where T_i^* being the time to the event and C_i the censoring time for subject i. δ_i is the event indicator, X_i the prognostic marker and Z_i the vector of p covariates.

For a threshold v, a patient i is considered at high-risk of event if $X_i > v$. The time-dependent sensitivity is the probability of being at high-risk among patients experiencing the event before τ:

$$Se(\tau, v) = P(X > v | T \leq \tau)$$

The time-dependent specificity is the probability of being at low-risk among patients surviving until τ:

$$Sp(\tau, v) = P(X \leq v | T > \tau)$$

Le Borgne et al. (2017) recommended to standardize the marker X according to the covariates among the controls (event-free at the prognostic time). Let X' be the standardized prognostic marker. To account for covariates Z, Le Borgne et al. (2017) proposed the standardized and weighted time-dependent sensitivity as $Se(\tau, v) = \mathbb{E}[Se(\tau, v|Z)]$, after standardizing the prognostic marker X. Similarly, the standardized and

weighted time-dependent specificity is defined as $Sp(\tau, v) = \mathbb{E}[Sp(\tau, v|Z)]$. Both quantities can be estimated as follows:

$$\widehat{Se}(\tau, v) = \frac{\sum_{i=1}^{n} \delta_i \mathbb{1}\{T_i \leq \tau, X_i' > v\} / \{\hat{S}_C(T_i) p_i(\tau)\}}{\sum_{i=1}^{n} \delta_i \mathbb{1}\{T_i \leq \tau\} / \{\hat{S}_C(T_i) p_i(\tau)\}}$$

$$\widehat{Sp}(\tau, v) = \frac{\sum_{i=1}^{n} \mathbb{1}\{T_i > \tau, X_i' \leq v\} / \{1 - p_i(\tau)\}}{\sum_{i=1}^{n} \mathbb{1}\{T_i > \tau\} / \{1 - p_i(\tau)\}}$$

To account for censoring, the uncensored subjects are weighted by the censoring survival probability $\hat{S}_C(T_i)$, estimated by the Kaplan-Meier estimator. To account for adjustment covariates, additional weights were proposed to create a counterfactual population in which the distribution of covariates is similar between cases and controls. More precisely, the uncensored subjects are weighted by the probability of experiencing the event before time τ according the covariates, $\hat{p}_i(\tau) = P(T_i^* \leq \tau | Z_i)$, while the censored subjects by the estimated survival probability, $1 - \hat{p}_i(\tau) = P(T_i^* > \tau | Z_i)$.

The intrinsic discriminative capacity of the marker X can be summarized by the area under standardized and weighted time-dependent ROC curve. Such ROC curve is a graphical representation of the weighted sensitivity versus (1-specificity), for all the threshold of the continuous prognostic marker which has previously been standardized according to the covariates among the controls.

Table S1: Number of at-risk patients in each transplantation center according to the prognostic time

Prognostic time (years)	Lyon (n=898)	Montpellier (n=1061)	Nantes (n=1131)	Paris-Necker (n=1024)
1	506	888	984	823
2	319	793	837	728
3	260	689	707	615
4	182	579	592	509
5	91	482	486	436
6	40	385	392	362
7	15	293	307	288
8	6	211	239	208
9	1	141	159	147
10	0	85	111	105

Table S2: Description of the cohort according to the inclusion or non-inclusion due to missing data for KDRI calculation

	Included patients (n=4114)		Non-included patients (n=602)		p-value
	NA	estimations	NA	estimations	
Quantitative characteristics - mean ± SD					
Recipient age (years)	0	53.07 ± 13.69	0	54.63 ± 13.31	0.0075
Recipient BMI (kg/m²)	28	24.97 ± 4.52	20	25.06 ± 4.23	0.6398
Donor age (years)	0	54.10 ± 16.57	33	55.26 ± 16.19	0.1108
Donor Serum Creatinine (mg/dL)	0	91.43 ± 61.10	22	91.82 ± 56.25	0.8755
Donor height (cm)	0	169.66 ± 9.76	229	168.45 ± 9.81	0.0230
Donor weight (kg)	0	74.34 ± 16.11	229	74.81 ± 16.60	0.6002
CIT (hours)	22	17.69 ± 6.81	16	17.19 ± 6.48	0.0875
Time spent on dialysis (years)	18	3.18 ± 3.13	17	3.02 ± 2.68	0.1886
Categorical characteristics - N (%)					
Recipient men	0	2553 (62.06)	0	382 (63.46)	0.5085
Dialysis technique	12		13		0.2146
Preemptive transplantation		435 (10.60)		58 (9.85)	
Peritoneal dialysis		376 (9.17)		67 (11.38)	
Hemodialysis		3291 (80.23)		464 (78.78)	
Relapsing initial disease	2	972 (23.64)	12	145 (24.58)	0.6166
History of diabetes	0	775 (18.84)	0	112 (18.60)	0.8911
History of hypertension	0	3314 (80.55)	0	473 (78.57)	0.2533
History of cardiovascular disease	0	1488 (36.17)	0	223 (37.04)	0.6770
History of dyslipidemia	0	1482 (36.02)	0	209 (34.72)	0.5327
History of neoplasia	0	438 (10.65)	0	72 (11.96)	0.3324
History of urological disease	0	444 (10.79)	0	80 (13.29)	0.0687
HLA A-B-DR incompatibilities > 4	41	677 (16.64)	25	100 (17.33)	0.6690
Pre-transplantation anti-HLA immunization of class I	276	1323 (34.47)	172	161 (37.44)	0.2200
Pre-transplantation anti-HLA immunization of class II	309	1175 (30.88)	171	156 (36.19)	0.0243
Donor men	1	2406 (58.50)	14	367 (62.41)	0.0708
Positive donor HCV serology	0	34 (0.83)	14	5 (0.85)	0.9999
Donation after Cardiac Death	0	238 (5.79)	16	49 (8.36)	0.0148
Donor history of diabetes	0	312 (7.58)	354	28 (11.29)	0.0345
Donor history of hypertension	0	1273 (30.94)	243	82 (22.84)	0.0014
Donor CMV infection	3	2243 (54.56)	14	347 (59.01)	0.0423
Recipient CMV infection	46	2568 (63.13)	16	392 (66.89)	0.0764
Donor CVA death	0	2144 (52.11)	35	284 (50.09)	0.3652
Delayed Graft Function	192	1003 (25.57)	67	142 (26.54)	0.6306
Depleting induction therapy	11	2053 (50.04)	15	386 (65.76)	<0.0001
Expanded Donor Criteria	0	1922 (46.72)	285	94 (29.65)	<0.0001

Figure S1: Distribution of the propensity score according to the High-quality/Medium-quality graft status (Scenario #1) and the Low-quality/Medium-quality graft status (Scenario #2)

