Transplantation Publish Ahead of Print

DOI: 10.1097/TP.0000000000003209

External Validation of the DynPG for Kidney Transplant Recipients

Rémi Lenain, MSc,[1] Etienne Dantan, PhD,[2] Magali Giral, PhD,[3,4] Yohann Foucher, PhD,[2,5] Özgür Asar, PhD,[6] Maarten Naesens, PhD,[7,8] Marc Hazzan, PhD,[1] and Marie-Cécile Fournier, PhD[2]

[1] Department of Nephrology, Huriez Hospital, Lille University Hospital, Lille, France.

[2] INSERM UMR 1246 - SPHERE, Nantes University, Tours University, Nantes, 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] Centre Hospitalier Universitaire de Nantes, Nantes, France

[6] Department of Biostatistics and Medical Informatics, Acibadem Mehmet Ali Aydinlar University, Istanbul, Turkey

[7] Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium

[8] Department of Nephrology and Renal Transplantation, University Hospitals Leuven, Leuven, Belgium

Financial Disclosure The authors declare no funding was received for this study.

Disclaimer: Yohann Foucher receives speaking honoraria from Biogen and Sanofi. The other authors declare no conflict of interest.

Author Roles

R.L.: Data acquisition, Statistical analysis, Writing - original draft, review and editing. E.D.: Project management, Conceptualization, Methodology, Writing - original draft, review and editing. M.G.: Conceptualization, Providing clinical expertise, Writing - review and editing. Y.F.: Methodology, Writing - review and editing. Ö.A.: Providing statistical expertise, Writing - review and editing. M.N.: Data acquisition, Providing clinical expertise, Writing – review and editing. M.H.: Data acquisition, Providing clinical expertise, Writing – review and editing. M-C.F.: Conceptualization, Methodology, Statistical analysis, Writing - original draft, review and editing.

Correspondence: 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

Abbreviations

AUC: Area Under ROC Curve; BMI: Body Mass Index; CI: Confidence Interval; DynPG: Dynamic prediction of Patient and Graft survival; SCr: Serum Creatinine; SD: Standard Deviation

Abstract

Background. In kidney transplantation, dynamic prediction of patient and kidney graft survival (DynPG) may help to promote therapeutic alliance by delivering personalized evidence-based information about long-term graft survival for kidney transplant recipients. The objective of the current study is to externally validate the DynPG.

Methods. Based on 6 baseline variables, the DynPG can be updated with any new serum creatinine measure available during the follow-up. From an external validation sample of 1637 kidney recipients with a functioning graft at 1-year posttransplantation from 2 European transplantation centers, we assessed the prognostic performance of the DynPG.

Results. As one can expect from an external validation sample, differences in several recipient, donor, and transplantation characteristics compared to the learning sample were observed. Patients were mainly transplanted from deceased donors (91.6% versus 84.8%, $p<0.01$), were less immunized against HLA class I (18.4% versus 32.7%, $p<0.01$) and presented less comorbidities (62.2% for hypertension versus 82.7%, $p<0.01$; 25.1% for cardiovascular disease versus 33.9%, $p<0.01$). Despite these noteworthy differences, the AUC varied from 0.70 (95%CI from 0.64 to 0.76) to 0.76 (95%CI from 0.64 to 0.88) for prediction times at 1 and 6 years posttransplantation respectively, and calibration plots revealed reasonably accurate predictions.

Conclusion. We validated the prognostic capacities of the DynPG in terms of both discrimination and calibration. Our study showed the robustness of the DynPG for informing both the patient and the physician, and its transportability for a cohort presenting different features than the one used for the DynPG development.

Introduction

Prognostic scores can be helpful to involve patients in the management of their disease. In renal transplantation, only time-fixed prognostic scores were available as detailed in the systematic review by Kabore et al.[1] Longitudinal Serum Creatinine (SCr) has already been demonstrated related to graft failure risk from a joint model.[2] Such a methodological approach was specifically developed to correctly consider longitudinal and survival processes in the presence of censoring,[3] and may be used to compute dynamic predictions.[4,5] Recently, we proposed a time-dependent prognostic score to dynamically predict patient and graft survival within 5 years following the prediction time to improve static prognostic scoring systems available in kidney transplantation.[6] Dynamic predictions of patient and graft survival, named DynPG, are defined as updated predictions, whenever any new serum creatinine (SCr) measure becomes available during follow-up. Using 6 additional noninvasive variables (recipient age, graft rank, cardiovascular histories, pretransplantation anti-HLA class I immunization, SCr at 3 months, occurrence of acute rejection in the first year posttransplantation), this score can be easily computed using the proposed online web application available at https://shiny.idbc.fr/DynPG.

Numerous methodological and epidemiological studies have already highlighted the necessity to validate such prognostic scores.[7] In our initial study,[6] we performed an internal validation, consisting of independent patients transplanted more recently in the same DIVAT network than those included in the learning sample. External validation with patient data from other cohorts appears as a supplementary step to demonstrate the validity of an existing tool. Internal validation merely relates to the reproducibility of results, while external validation questions the transportability.[7,8]

The objective of the current study is to externally validate the dynamic prediction of patient and kidney graft survival. From 2 European kidney transplantation centers, we assessed the predictive performance of the DynPG in terms of both calibration and discrimination to promote its use in clinical practice.

Materials and Methods

Study Population

Two European kidney transplantation centers were considered: the Lille University transplantation center (France) and the Leuven University Hospital (Belgium). Data from the Lille transplantation center were collected from the prospective database CRISTAL (French Agency of Biomedicine) and from local records (CNIL agreement n°2214185) of patients transplanted between January 1st 2007 and December 31th 2017. Leuven is one of the European transplantation centers involved in the EKITE network, a collaborative research project on epidemiology of kidney transplantation.[9] Data from the Leuven transplantation center were collected from the local prospective cohort of patients transplanted between January 1st 2005 and January 31th 2013. A total of 1637 patients (1165 from Lille and 472 from Leuven) met the inclusion criteria defined in Fournier et al.[6] (adult recipients who received a first or second renal graft from a living or heart beating deceased donor, alive with a functioning graft at 1-year posttransplantation and maintained under Tacrolimus and Mycophenolate Mofetil, and did not have any missing data for the variables of the DynPG. All participants gave informed consent for research at the time of transplantation.

Collected data

Most classical risk factors in kidney transplantation were extracted from the database. Donor features included age, gender, last donor Serum Creatinine and deceased (from cerebrovascular death versus other) or living donation. Recipient variables included gender, age, Body Mass Index (BMI), history of diabetes, hypertension, dyslipidemia, cardiovascular diseases, pretransplant dialysis method (hemodialysis, preemptive kidney transplantation or peritoneal dialysis), HLA A-B-DR incompatibilities, and initial nephropathy (classified into relapsing disease or not, as detailed in **Table S1**, **SDC**, http://links.lww.com/TP/B893). Pretransplantation immunization against class I or class II anti-HLA was defined as positive if at least 1 DSA was identified by Luminex Single Antigen Bead technology within 6 months pretransplantation, unless at least 1 DSA was not identified but a later assessment by Luminex screening or other technique (enzyme-linked immunosorbent assay (ELISA) or Complement Dependent Cytotoxicity (CDC)) was positive pretransplantation. Transplantation features were the time between first dialysis and kidney transplantation (data were not available for the Leuven transplantation center) and cold ischemia time. Posttransplantation characteristics included occurrence of at least 1 acute rejection episode during the first year posttransplantation (only treated acute rejections were considered) and longitudinal SCr measures recorded yearly until death with a functioning graft or return to dialysis.

Outcomes

The baseline was 1-year posttransplantation. The endpoint was time to graft failure, which is defined as the first event from return to dialysis, preemptive retransplantation or death with a functioning graft. We considered SCr (μmol/L) evolution through the yearly recorded levels until graft failure. The DynPG was defined as the probability of being graft failure-free over a 5-year prediction window, for each prediction time (landmark time) from 1 to 6 years

posttransplantation. This maximum landmark time at 6 years posttransplantation was retained, since there were only 71 patients still at risk of graft failure at 11 years posttransplantation.

Statistical analyses

The 2 validation samples were pooled to obtain 1 external validation cohort. The external validation and learning samples were compared using t-tests for quantitative variables and chi-square tests for qualitative variables. To study the predictive reliability of the DynPG, we used the same methodology as for the internal validation in the initial study.[6] Prognostic performances were reported according to the TRIPOD recommendations.[10,11] In order to estimate the DynPG prognostic capacities, we used indicators that consider right-censoring. An R2-type curve was used to evaluate global performance.[12] The discriminative capacities were evaluated by the Area Under the ROC Curve (AUC) for dynamic predictions.[13] The calibration was described by comparing predicted values within subgroups (defined from prediction quintiles) to observed graft and patient survival (computed using the Kaplan-Meier estimator). Separate analyses of each transplantation center using the same methodology are available in the Supplemental Digital Content. Additionally, we compared the DynPG prognostic performances with those obtained from a time-fixed score. This latter was obtained given a parametric survival model estimated on the same learning sample used for the DynPG development.[6] This time-fixed score included the time-fixed variables of the DynPG and the single log-transformed SCr measurement available at baseline (Table S2, **SDC**, http://links.lww.com/TP/B893). All analyses were performed using R (v3.3.0) and the JM (v1.4.7), prodlim (v1.6.1), survival (v2.39.2) and timeROC (v0.3) packages.[14-17]

Results

Description of baseline characteristics

Focusing on Table 1, one can note that the learning (n=2749) and external validation (n=1637) samples presented significant differences in recipient, donor and transplantation characteristics. Most importantly, the patients from the external validation sample had higher BMI (25.1 ±4.5 compared to 24.0 ±4.2 kg/m² for the learning sample) and lower cold ischemia time (15.9 ±7.2 compared to 17.8 ±9.8 hours). Patients presented less comorbidities in the external validation sample (62.1% for hypertension compared to 82.6% in the learning sample, 25.1% for cardiovascular disease compared to 33.9% and 25.4% for dyslipidemia compared to 31.3%), and they were less immunized (18.4% for pretransplantation anti-HLA immunization against class I and 18.1% against class II compared to 32.6% and 29.7% respectively). We observed less living donors in the external validation sample (8.4%) compared to the learning sample (15.2%). Finally, the occurrence of acute rejection in the first year posttransplantation was less frequent in the validation sample (12.8%) compared to the learning sample (21.5%).

As described in the Supplemental Digital Content (Table S3, **SDC**, http://links.lww.com/TP/B893), patients from the Leuven transplantation center were older (53.0 years versus 50.4 years, p=0.0007), received their graft from younger donors (47.9 year versus 51.3 years, p<0.0001) and had a shorter CIT (13.6h versus 16.9h) compared to the patients from the Lille transplantation center. They were also from an older transplantation period (35.8% of patients transplanted before 2008 versus 8.3%, p<0.0001), possibly explaining the higher frequency of acute rejection episodes in the first posttransplantation year (22.7% versus 8.8%, p<0.0001) and the lower preemptive transplantation frequency (3.2% versus 11.5%, p<0.0001).

Follow-up description

Among the 1637 included patients, 153 (100 in Lille and 53 in Leuven) patients returned to dialysis, none were preemptively retransplanted and 187 (89 in Lille and 98 in Leuven) died with a functioning graft. The median follow-up time for the whole external validation was 6.05 years (1st quartile 3.04; 3rd quartile 9.02). The number of SCr measurements during the follow-up was between 1 and 6 with a median value of 4 (1st quartile 3; 3rd quartile 6). Figure S1 (**SDC**, http://links.lww.com/TP/B893) describes the Serum Creatinine distribution from 1 to 6 years posttransplantation. The patient and graft survival probability at 8-year posttransplantation was 72.9% (95%CI from 70.1% to 75.8%), whereas it was 71.8% (95%CI from 69.3% to 74.5%) in the learning sample (Figure 1, p=0.8400). We observed a longer median follow-up for the Leuven patients of 8.83 years (1st quartile 6.62; 3rd quartile 10.61) compared to the Lille patients with a median of 4.99 years (1st quartile 2.97; 3rd quartile 8.00). Additionally, we did not observe any significant difference in patient and graft survival probabilities between patients from the Lille and Leuven transplantation centers (Figure S2, **SDC**, http://links.lww.com/TP/B893, p=0.1900).

Prognostic capacities

While the overall (discrimination and calibration) prognostic capacities for predicting patient and graft survival seemed relatively independent from the landmark times for making predictions (R^2 values were about 14%) in Fournier et al.,[6] they increased in the external validation at successive transplantation anniversaries up to 4 years. More precisely, the R^2 values ranged from 8.9% (95%CI from 2.0% to 15.9%) at 1-year posttransplantation to 19.4% at 4-years posttransplantation (95%CI from 6.6% to 33.2%) (Part A - Figure 2). In Fournier et al., the discriminative capacities increased through prediction times (AUC from 0.72 at 1-year to 0.76 at 6-years posttransplantation). In comparison, in this external validation, the AUC values increased from 0.70 (95%CI from 0.64 to 0.76) at 1-year posttransplantation to 0.76

(95%CI from 0.64 to 0.88) at 6-years posttransplantation (Part B - Figure 2). As illustrated in Figure 3, we may consider the predictions for the 3 intermediate quintiles as reasonably calibrated until 4-years posttransplantation. In the Supplemental Digital Content, we compared the DynPG prognostic performances with the ones of the time-fixed score. The R^2 values of the time-fixed score fluctuated around 12% (Figure S3A, **SDC**, http://links.lww.com/TP/B893), while R^2 values of the DynPG increased with the landmark times. The AUC values were also lower for the time-fixed score compared to the DynPG (Figure S3B, **SDC**, http://links.lww.com/TP/B893). These differences illustrated the interest of considering the longitudinal SCr evolution to predict the patient and graft survival dynamically.

In the Supplemental Digital Content, we also provided separate analyses of the DynPG for the Lille and Leuven transplantation centers. Few patients in Lille are still at risk at 6-years posttransplantation (Figure S2, **SDC**, http://links.lww.com/TP/B893). The prognostic performances for the whole validation cohort were essentially supported by the Leuven patients for the higher landmark times at 6 years posttransplantation. We may therefore reasonably considered the prognostic performances of the Lille transplantation center up until 5-years posttransplantation (with 73 patients still at-risk at the end of the prediction window). The overall prognostic capacities increased for R2 from 9.6% (95%CI from 1.4% to 17.7%) at 1-year posttransplantation to 22.1% at 5-years posttransplantation (95%CI from 0% to 52.5%) (Figure S4A, **SDC**, http://links.lww.com/TP/B893). The AUC values increased from 0.69 (95%CI from 0.61 to 0.77) at 1-year posttransplantation to 0.80 at 5-years posttransplantation (95%CI from 0.68 to 0.91) (Figure S4B, **SDC**, http://links.lww.com/TP/B893). The calibration seems reasonable up until 4-years posttransplantation (Figure S5, **SDC**, http://links.lww.com/TP/B893). For the Leuven transplantation center, the overall prognostic capacities seemed relatively stable along the prediction times with R2 values around 10%

(Figure S6A, **SDC**, http://links.lww.com/TP/B893), while the discriminative capacities tended to increase until an AUC at 0.73 at 6-years posttransplantation (95%CI from 0.60 to 0.86) (Figure S6B, **SDC**, http://links.lww.com/TP/B893). This is in line with the calibration plot which showed overestimated predictions for high landmark times (Figure S7, **SDC**, http://links.lww.com/TP/B893). This probably explains the decreasing performances in R2 and calibration after 4 years posttransplantation on the whole validation cohort.

Description of the predicted probabilities of being graft failure-free

From the predicted probabilities of being graft failure-free over a 5-year window calculated at landmark times from 1 to 6 years posttransplantation, we described the proportion of patients having a DynPG higher than 90%, between 80% and 90% or lower than 80% (Figure 4). The proportion of at-risk patients having a predicted probability higher than 90% ranged from 54% at 1-year posttransplantation to 45% at 6-years posttransplantation. Over these years, a considerable proportion of patients can be considered as having an encouraging and reassuring future.

Discussion

In this study, we externally validated the Dynamic prediction of Patient and Graft survival (DynPG) for kidney transplant recipients initially proposed by Fournier et al. as a patient information tool.[6] From a prospective observational cohort outside the DIVAT consortium drawn from 2 European transplantation centers, we observed good prognostic capacities in terms of both discrimination and calibration.

This external validation sample had some noteworthy differences compared to the learning sample. Firstly, the proportion of living donors was 2-fold lower than in the learning DIVAT cohort. This partly explains the differences in terms of HLA A-B-DR incompatibilities. Secondly, transplanted patients in the external validation sample appear more likely to be selected before transplantation as they had less comorbidities and were less immunized. The

lower proportion of acute allograft rejection in the first year posttransplantation may result from the differences already highlighted. Despite these variations, the DynPG tool exhibited reasonable calibration for intermediate predictions. To deliver individual-specific information, the DynPG should be used with caution for extreme predictions that can be slightly overestimated or underestimated, and thus cause unwarranted anxiety. The DynPG also presented good AUC values, improving over the landmark times. Even if the calibration could be better, one can accept the relevant accuracy of the DynPG to predict the probability of being graft failure-free. In particular, the good discrimination of the DynPG allows to correctly order the patients given their risk of graft-failure. The DynPG may be used to indicate to the patients to which class (quintile) of risk they belong to.

Nevertheless, the utility of the DynPG cannot be only appreciated by its prognostic capacities. In 2012, we proposed the TELEGRAFT randomized clinical trial in which the follow-up frequency is adapted given the risk defined by the KTFS, a time-fixed score predicting the return to dialysis.[18,19] The last patients being followed-up, the analyses will be conducted in the near future. Only the results of this trial will demonstrate the possible clinical utility of the KTFS. In case of positive results, we would consider extending dynamic follow-up frequency. Additionally, the aim of our dynamic prediction tool is to promote therapeutic alliance in clinical practice by delivering personalized and evidence-based information. We are convinced that the graphic illustration included in the online application (https://shiny.idbc.fr/DynPG) is an opportunity to favor a collaborative process between practitioners and kidney transplant recipients. In the case of a poor predicted transplantation outcome, Sheu et al. remind us of the necessity to provide early information and to anticipate a potential return to dialysis.[20] Several studies have shown that immunosuppression adherence decreases over time in kidney transplantation.[21–23] By using personalized long-term graft survival information we expect therapeutic adherence to be reinforced, better involvement of

the patients in their care pathway and in future crucial therapeutic choices for delaying graft failure. Favorable long-term prognosis can lead to a personalized follow-up care schedule including longer delays between successive visits or video conference based meetings.[19] If the predicted probability of being graft failure-free over a 5-year window appears high, it can play an important part in removing anxieties about an uncertain future.[24] An important proportion of at-risk patients would appreciate having this reassuring information regarding their high predicted probability of being graft failure-free. This may strengthen their involvement in personal and/or professional activities and contribute to their improved quality of life. We therefore think that the DynPG could integrate therapeutic education programs as a synthetic information tool for patients at low risk of graft failure, which concerns the major part of at-risk patients. We are currently developing a research program to evaluate the possible impact of such information on patients' feelings and experience of the disease.

Several limitations may be highlighted regarding our observational study. We performed an external validation from 1 French transplantation center and 1 Belgium transplantation center. The Leuven sample size may nevertheless be considered as small compared to the Lille center. Even though the Lille Hospital is a French transplantation center, our external validation is relevant because each center has specific features and practices as highlighted by the numerous differences in patient characteristics discussed above. Steyerberg et al. clearly explained that these observed differences can be viewed as key advantages to assess the robustness and the transportability of the predictive tools.[7] While this external validation study brings interesting results, we are also aware that the inherent limitations of the DynPG may be an obstacle to convincing the kidney transplant community of its potential. Therefore, we follow methodological developments to improve the DynPG. Notably, the DynPG was developed from only 1 longitudinal marker, while others such as proteinuria or posttransplantation immunization could provide important valuable information for improving

prognosis. Recently, the iBox score has been proposed to predict kidney graft failure and can be calculated using histological and immunological measurements obtained at the punctual posttransplantation time of allograft biopsy.[25] While the debate surrounding the clinical utility of kidney biopsy is still open,[26] it may be relevant to further investigate in dynamic context the predictive performance of the whole trajectory of such histological and immunological markers in addition to longitudinal clinical markers as serum creatinine or proteinuria. Using only the SCr evolution is an important limitation, since the dynamic predictions over landmark times can only be a reflection of SCr modification. The DynPG will not be sensitive to any clinical event such as cancer or cerebrovascular injury where there is no effect on SCr evolution, and should therefore be used with caution in such clinical situations. Note that any intermediate clinical event with SCr changes (acute rejection, infection) or without changes will probably result in the patient visiting their nephrologist which would correspond to an informative follow-up time instead of the routinely programmed visits used in the DynPG. Of note, the DynPG is validated for longitudinal SCr measures collected at routine and not informative visits. Considering multiple longitudinal markers, the time-dependent occurrence of intermediate clinical events and informative delay between visits in joint models for dynamic prediction are a current statistical research question that would considerably improve our proposal.[27] Moreover, the DynPG is obtained from joint modelling of longitudinal and survival data assuming Gaussian random-effects and error terms in the longitudinal submodel. This assumption may be too restrictive in the presence of longitudinal outliers.[28] In ongoing methodological work which delivers interesting preliminary results to improve calibration and discrimination, we proposed to relax the Gaussian assumption by defining a robust joint model with a more flexible distribution for random effects and error terms.[29-31] The choice of the outcome of interest (first event between return to dialysis, preemptive retransplantation or death with a functioning graft) also raises questions regarding its practical use and a future

goal would be to predict which event will occur first. In a competing risk framework, it may be not easy to solution it and vertical modelling may be a relevant approach.[32] Additional perspectives of DynPG validation will be relevant in the future. Whilst the DynPG has been developed and validated for living or heart beating deceased donors, it would be interesting to validate its prognostic performances for DCD organs (e.g. Maastricht classification category III that could increase in the future). Furthermore, it would be beneficial to validate the DynPG on a cohort presenting a longer follow-up, since the prognostic performances might be debatable for the high landmark times.

Conclusion

In conclusion, the DynPG showed good performance in dynamically predicting patient and graft survival on a 5-year prediction window in this external validation sample of kidney transplant recipients. While this study confirmed the prognostic capacities of the DynPG, further studies are required to assess its utility in clinical practice. Such studies should investigate the clinical impact of the information brought by the DynPG and given to the patients on therapeutic adherence, shared decision making and self-management. This kind of tool could contribute to patient centered care by replacing the patient as the main actor of their health.

Acknowledgments

We wish to thank the clinical research assistant (S. Gomis) of the French transplantation center of Lille University Hospital for his role in data collection, the datamanager (M. Coemans) of the Leuven transplantation center for his role in preparing the data and the DIVAT project manager coordinator (C. Kerleau) for her help in the data transfer agreement.

References

1. Kaboré R, Haller MC, Harambat J, et al. Risk prediction models for graft failure in kidney transplantation: a systematic review. *Nephrol Dial Transplant*. 2017;32(suppl_2):ii68–ii76.
2. Fournier MC, Foucher Y, Blanche P, et al. A joint model for longitudinal and time-to-event data to better assess the specific role of donor and recipient factors on long-term kidney transplantation outcomes. *Eur J Epidemiol*. 2016:31(5):469–479.
3. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Boca Raton, FL: CRC Press; 2012.
4. Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Comput Stat Data Anal*. 2012;56(3):491–501.
5. Proust-Lima C, Blanche P. Dynamic Predictions. In: *Wiley StatsRef Stat Ref Online*. Hoboken, NJ; John Wiley & Sons, Ltd: 2015.
6. Fournier MC, Foucher Y, Blanche P, et al. Dynamic predictions of long-term kidney graft failure: an information tool promoting patient-centred care. *Nephrol Dial Transplant*. 2019;34(11):1961–1969.
7. Steyerberg E. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY; Springer-Verlag New York: 2009.
8. Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683–690.
9. Lorent M, Foucher Y, Kerleau K, et al. The EKiTE network (epidemiology in kidney transplantation - a European validated database): an initiative epidemiological and translational European collaborative research. *BMC Nephrol*. 2019;20(1):365.

10. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55–63.
11. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1–W73.
12. Fournier M-C, Dantan E, Blanche P. An R^2-curve for evaluating the accuracy of dynamic predictions. *Stat Med*. 2017;37(7):1125–1133.
13. Blanche P, Proust-Lima C, Loubère L, et al. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*. 2015;71(1):102–113.
14. *R: A Language and Environment for Statistical Computing* [computer program]. Vienna, Austria; R Development Core Team: 2010. Available at http://www.R-project.org/.
15. Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. *J Stat Softw*. 2010;35(9):1–33.
16. Therneau T, Grambsch P. *Modeling Survival Data: Extending the Cox Model*. New York, NY; Springer-Verlag New York: 2000.
17. Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med*. 2013;32(30):5381–5397.
18. Foucher Y, Daguin P, Akl A, et al. A clinical scoring system highly predictive of long-term kidney graft survival. *Kidney Int*. 2010;78(12):1288–1294.

19. Foucher Y, Meurette A, Daguin P, et al. A personalized follow-up of kidney transplant recipients using video conferencing based on a 1-year scoring system predictive of long term graft failure (TELEGRAFT study): protocol for a randomized controlled trial. *BMC Nephrol*. 2015;16:6.

20. Sheu J, Ephraim PL, Powe NR, et al. African American and non-African American patients' and families' decision making about renal replacement therapies. *Qual Health Res*. 2012;22(7):997–1006.

21. Dharancy S, Giral M, Tetaz R, et al. Adherence with immunosuppressive treatment after transplantation: results from the French trial PREDICT. *Clin Transplant*. 2012;26(3):E293–E299.

22. Tsapepas D, Langone A, Chan L, et al. A longitudinal assessment of adherence with immunosuppressive therapy following kidney transplantation from the Mycophenolic Acid Observational REnal Transplant (MORE) study. *Ann Transplant*. 2014;19:174–181.

23. Ichimaru N, Kakuta Y, Abe T, et al. Treatment adherence in renal transplant recipients: a questionnaire survey on immunosuppressants. *Transplant Proc*. 2008;40(5):1362–1365.

24. Jamieson NJ, Hanson CS, Josephson MA, et al. Motivations, challenges, and attitudes to self-management in kidney transplant recipients: a systematic review of qualitative studies. *Am J Kidney Dis*. 2016;67(3):461–478.

25. Loupy A, Aubert O, Orandi BJ, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. *BMJ*. 2019;366:l4923.

26. Couvrat-Desvergnes G, Foucher Y, Le Borgne F, et al. Comparison of graft and patient survival according to the transplantation centre policy for 1-year screening biopsy among stable kidney recipients: a propensity score-based study. *Nephrol Dial Transplant*. 2019;34(4):703–711.

27. Hickey GL, Philipson P, Jorgensen A, et al. joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC Med Res Methodol*. 2018;18(1):50.

28. Pinheiro JC, Liu C, Wu YN. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *J Comput Graph Stat*. 2001;10(2):249–276.

29. Lange KL, Little RJA, Taylor JMG. Robust statistical modeling using the t distribution. *J Am Stat Assoc*. 1989;84(408):881–896.

30. Baghfalaki T, Ganjali M, Berridge D. Robust joint modeling of longitudinal measurements and time to event data using normal/independent distributions: a Bayesian approach. *Biom J*. 2013;55(6):844–865.

31. Huang X, Li G, Elashoff RM. A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements. *Stat Interface*. 2010;3(2):185–195.

32. Nicolaie MA, van Houwelingen HC, Putter H. Vertical modelling: analysis of competing risks data with missing causes of failure. *Stat Methods Med Res*. 2015;24(6):891–908.

Figure Legends

Figure 1: Patient and graft survival from the Kaplan-Meier estimator and their corresponding 95%CI according to the external validation sample (solid line) and the learning sample from Fournier et al. (2019) (dashed line) (Log-Rank test: p=0.8400).

Figure 2: Prognostic capacities of the dynamic predictions (n=1637) estimated for prediction times from 1 to 6-years posttransplantation for a given horizon window of 5 years, R^2 supplied global performance (Part A) while Area under ROC curve (AUC) appraised discrimination accuracy (Part B). Estimations are drawn in solid lines and the corresponding 95% confidence interval is drawn in dashed lines.

Figure 3: Calibration plot of dynamic predictions on the validation sample (n=1637) for prediction times from 1 to 6-years posttransplantation. Mean predicted risks and observed risks (Kaplan-Meier) of being graft failure-free over a 5-year window are displayed for each subgroup, defined from prediction quantiles.

Figure 4: Proportion of patients having a predicted probability of being graft failure-free over a 5-year window of prediction higher than 90%, between 80% and 90% or lower than 80% for prediction times from 1 to 6-years posttransplantation.

Tables

Table 1: Description of recipients, donors and transplantation characteristics according to the learning sample (n=2749) and the external validation sample (n=1637).

	Learning sample (n=2749)		External validation sample (n=1637)		p-value
	NA	estimations	NA	estimations	
Quantitative characteristics : mean ± SD					
Recipient age (years)	0	49.71 ± 13.59	0	51.17 ± 13.68	0.0006
Recipient BMI (kg/m²)	10	23.99 ± 4.24	3	25.14 ± 4.53	< 0.0001
Donor age (years)	1	50.74 ± 15.52	0	50.30 ± 15.32	0.3560
Last donor SCr (µmol/L)[a]	25	89.91 ± 52.77	232	84.57 ± 54.78	0.0027
Cold ischemia time (hours)	10	17.76 ± 9.79	63	15.94 ± 7.16	< 0.0001
Time spent on dialysis (years)	36	2.98 ± 3.08	457	3.13 ± 4.43	0.2869
3-months SCr (µmol/L)	38	138.30 ± 53.38	0	141.81 ± 50.50	0.0296
6-months SCr (µmol/L)	75	136.64 ± 53.18	40	134.23 ± 48.17	0.1288
Categorical characteristics : N (%)					
Recipient men	0	1674 (60.89)	0	1013 (61.88)	0.5165
Second transplantation	0	474 (17.24)	0	245 (14.97)	0.0489
Dialysis technique	4		32		< 0.0001
Preemptive transplantation		342 (12.46)		145 (9.03)	
Hemodialysis		2191 (79.82)		1211 (75.45)	
Peritoneal dialysis		212 (7.72)		249 (15.51)	
Relapsing initial disease	0	799 (29.07)	0	490 (29.93)	0.5418
History of diabetes	0	319 (11.60)	5	241 (14.77)	0.0024
History of hypertension	0	2272 (82.65)	15	1008 (62.15)	< 0.0001
History of cardiovascular disease	0	933 (33.94)	0	410 (25.05)	< 0.0001
History of dyslipidemia	0	860 (31.28)	131	382 (25.37)	< 0.0001
More than 4 HLA A-B-DR incompatibilities	7	350 (12.76)	2	320 (19.57)	< 0.0001
Pretransplantation anti-HLA immunization of class I	66	876 (32.65)	0	301 (18.39)	< 0.0001
Pretransplantation anti-HLA immunization of class II	87	792 (29.75)	15	294 (18.13)	< 0.0001
Donor men	8	1545 (56.37)	0	919 (56.14)	0.8835
Donor vital status	0		0		< 0.0001
Living donor		424 (15.21)		138 (8.43)	
Cerebrovascular donor death		1309 (47.74)		786 (48.01)	
Non cerebrovascular donor death		1016 (37.05)		713 (43.56)	
Acute rejection episode(s) during the first year	0	591 (21.50)	0	210 (12.83)	< 0.0001
Transplanted before 2008	0	2091 (39.17)	0	266 (16.25)	< 0.0001

Abbreviations: BMI Body Mass Index; HLA Human Leucocyte Antigen; NA: Not Available (missing data); SCr Serum Creatinine; SD Standard Deviation
[a]Living donor last SCr were all missing in Lille center (96 patients). In Leuven center, 32 missing data among 135 were living donors.

Figure 1

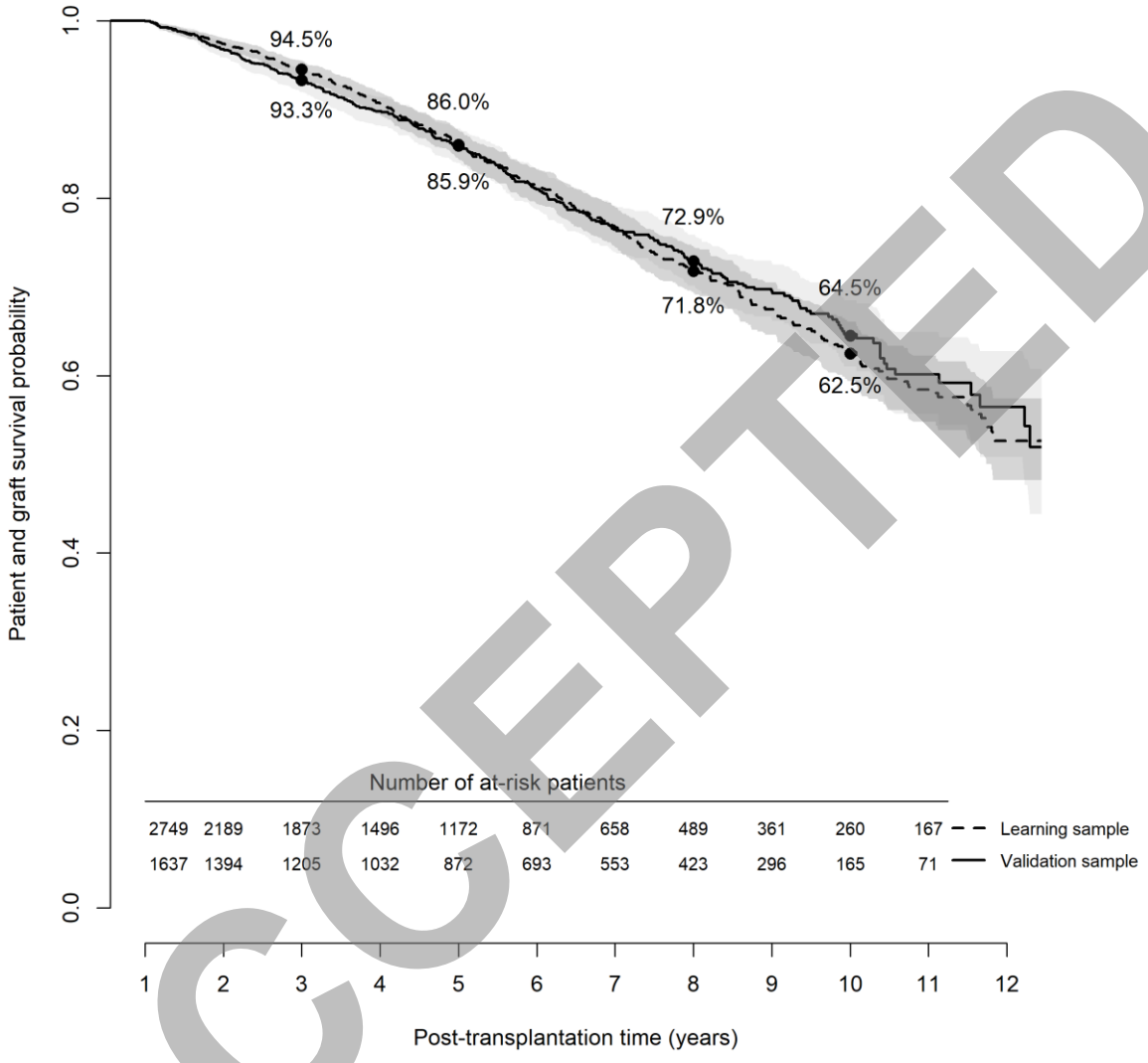

Figure 2

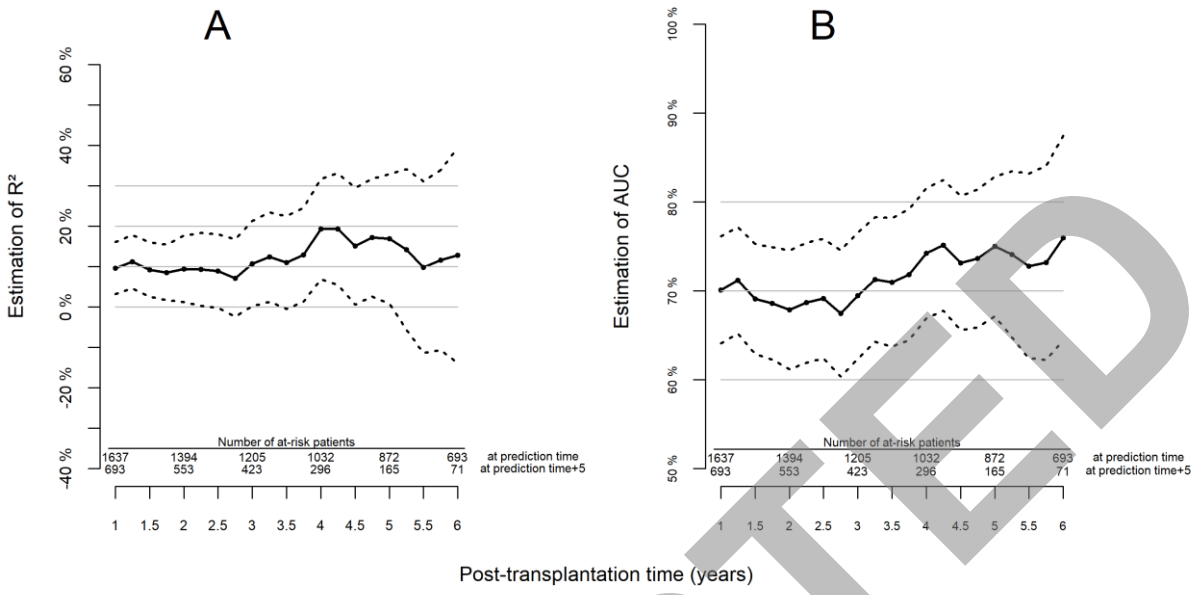

Figure 3

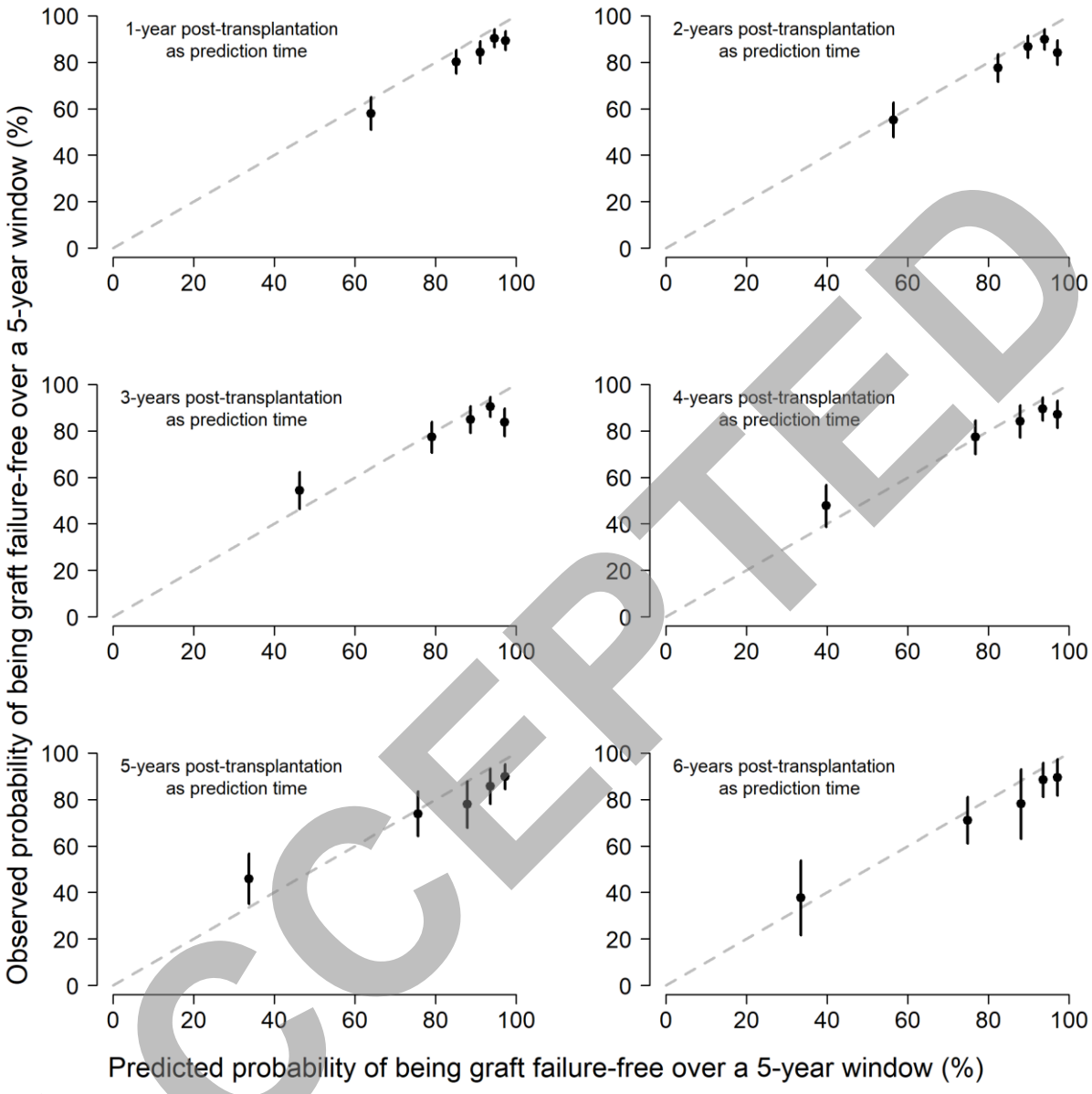

Figure 4