A personalized screening strategy to monitor the development of chronic allograft failure in renal transplant recipients

Hessel Peters-Sengers[1,2,*], Anirudh Tomer[3,*], Sandrine Florquin[4], Ewout W. Steyerberg[5,6], Frederike J. Bemelman[1], Dimitris Rizopoulos[3,$], Jesper Kers[4,7,$]

[1]Division of Internal Medicine, Renal Transplant Unit, Academic Medical Center, Amsterdam, The Netherlands
[2]Center for Experimental and Molecular Medicine (CEMM), Academic Medical Center, Amsterdam, The Netherlands
[3]Department of Biostatistics, Erasmus University Medical Center, Rotterdam, The Netherlands
[4]Department of Pathology, Academic Medical Center, Amsterdam, The Netherlands
[5]Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands
[6]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
[7]Van 't Hoff Institute for Molecular Sciences (HIMS), University of Amsterdam, Amsterdam, The Netherlands

[*,$]Authors contributed equally to the study

Introduction

Renal transplantation is the preferred replacement therapy for patients with end-stage-renal-disease (ESRD). Compared to dialysis, renal transplantation improves patient survival, cardiovascular outcomes and also quality of life.[1-4] Short-term graft survival is excellent[5], but unfortunately, long-term graft survival has not benefitted from recent treatment regimens to the same extent[6].

In an effort to stratify patients who are at increased or decreased risk for graft failure, a multitude of studies investigated the value of molecular biomarkers and clinical algorithms for the prediction of long-term graft failure. In clinical practice however, only very few of these markers are being used on a routine basis. The best example of markers, that are practically used in every transplant outpatient clinic, are serum creatinine as a readout for the glomerular filtration rate, and proteinuria as a readout for glomerular and/or tubular dysfunction.[7,8] Although parameters are measured at each outpatient visit, hardly any study determined the dynamic temporal association between serum creatinine and urinary protein content trajectories with the development of graft failure. In fact, the majority of studies, including randomized-controlled trials in kidney transplantation, model renal function under the assumption that renal function declines linearly over time.[9] A recent study by Ferro and colleagues highlighted that the contrary was true. More specifically, 87% of their included patients showed nonlinearity or nonprogression.[10] Patients with subsequent graft failure were more likely to have episodes of rapid progression and less likely to have episodes of nonprogression. Similarly, we might assume that patterns of rapid progression and nonprogression of proteinuria associate with graft failure. These data corroborate on the fact that in renal transplantation, patients have to be continuously monitored and therefore prediction models, contrary to static models, have to be designed as such.

An interesting statistical model to investigate the effect of renal function and urinalysis, on the graft survival is the joint model (JM) for time to event and longitudinal data. For the two longitudinal outcomes, the JM utilizes a multivariate longitudinal submodel, whereas for the event time outcome a relative risk submodel is used. More specifically, the hazard of graft failure at any time depends on a feature of the two longitudinal outcomes at that time. Recently, Rizopoulos and colleagues proposed an application of the joint model to personalize the monitoring, which could be of interest for renal transplant recipients as well.[11] Initially, the concept of personalized screening was investigated in a cohort of patients that had undergone aortic valve allograft replacement and were followed-up with echocardiographic measurements of aortic valve gradients over time (repeated measures submodel) to predict the composite outcome of re-operation and death (survival submodel) with a joint model.[11] Then this joint model was used to create personalized measurement schedules for the biomarkers. These schedules are dynamic in nature. That is, the schedules update as more information about the patient is obtained. Via a simulation study they showed that personalized screening intervals lead to a reduction in number of screening time points compared to a fixed screening interval. In addition, the personalized screening intervals also lead to a smaller absolute error for optimal interval time point.

The reduction in screening time-points when using the personalized screening strategy can be directly translated to a reduction in medical as well as financial burden for patients.[11]

We had two aims in the current study. Firstly, we wanted to create a joint model to predict death-censored graft failure from static baseline clinical data and dynamic longitudinal trajectories of serum creatinine and urinary protein-creatinine ratios. Secondly, we wanted to use the fitted joint model from the previous step to construct a personalized monitoring strategy and compare it with the fixed-term "one size fits all" monitoring strategy that is currently in use.

Methods

To assess the applicability of a personalized monitoring strategy, we performed a single-center retrospective cohort study in the Academic Medical Center (AMC), a tertiary referral hospital in Amsterdam, the Netherlands. The electronic patient database was used to collect all relevant data. All information was processed anonymously according to the code of conduct by the Dutch Medical Scientific Society (FDMSS).[12]

Study population

We screened the records of 239 end-stage renal disease patients that underwent renal transplantation at our institute from June 1, 1996 to October 31, 2009. The inclusion criteria for the study were: age at baseline ≥18 years who had >1 additional serum creatinine (SCr, umol/L) and spot or 24-hour urine collection to calculate the urine protein:creatinine ratio (PCR, g/mol) during follow-up. SCr measurements were available for 239 patients, and PCR measurements were available for 238 patients. Thus, leaving 238 subjects for the multivariate longitudinal model. Last follow-up date was.... Initial immunosuppressive therapy consisted of steroids combined with mycophenolate mofetil or mycophenolic acid and a calcineurin inhibitor, mostly tacrolimus but also cyclosporine. Alternatively, a combination of steroids, tacrolimus, and sirolimus was used. Donor kidneys were acquired through allocation by the Eurotransplant allocation program, Leiden, The Netherlands.[13]

Comment [HPS1]: @Jesper: figuur 1 ziet er niet uit als 'right censoring'. Voor lezer is denk ik beter om te weten wat de uiterste follow-up datum was: tx-date + max follow-up.

Measures and outcomes

Included in the database were potential predictors for kidney function and graft failure. Extracted were: donor age, donor gender, donor body mass index (BMI), donor type (living; brain death (DBD); controlled circulatory death (DCD)); number of human leukocyte antigen (HLA) A, B, DR mismatches, cold ischemic time, panel reactive antibodies (PRA) before transplantation, immunosuppressive regiments at 3 months after transplantation, recipient dialysis vintage, recipient blood pressure, recipient age, recipient gender, recipient BMI and recipient cardiovascular diseases. The following data on medication use after transplantation were collected: immunosuppressive regiments (calcineurin inhibitors, prednisone, proliferation inhibitors, induction therapy, mammalian target of rapamycin inhibitors), anti-hypertensives (diuretics, inhibitors of the renin-angiotensin-aldosterone system, beta blockers and calcium channel blockers), statins and the use of antiglycemic medication

or insulin. These were extracted only once within the first year after transplantation, whereas SCr and PCR were measured repeatedly over time till the patient either had a transplant failure (death or graft failure), or was not followed up anymore. If there were multiple SCr measurements per day, we took the mean of the measurements for analysis. We evaluated death-censored graft survival, defined as graft loss leading to dialysis treatment as event and censored for death with a functioning graft. Additionally, we used SCr and PCR to study the longitudinal markers for kidney function. Delayed graft function was defined as need for dialysis treatment within seven days after transplantation.

Data analytic strategy

Our first goal is to check if both SCr and PCR are useful to predict graft failure. To this end, we fitted a JM to the dataset at hand (see supplementary I for detailed specification). The longitudinal submodel for the two biomarkers consisted of additive effects of baseline patient characteristics and effect of time. To accommodate for nonlinear evolution of biomarkers over time, we used b-splines to model the random effect as well the fixed effect of time. The b-spline for both biomarkers consisted of internal knots at 30, 80 and 365 days, and boundary knots at 14 days and 6 years, in the fixed as well as the random effects part. We used log transformed biomarkers in the model to meet the assumption of homoscedasticity of residuals, which was analyzed graphically. In the relative risk submodel, we model the impact of (log transformed) biomarker values and velocity on the risk of graft failure. In addition the impact of cold ischemia time, previous transplantation, HLA mismatches and number of days on dialysis before transplantation, on the risk of graft failure is also modeled. To obviate the issue of overfitting the relative risk submodel, we utilized the Bayesian LASSO shrinkage approach for the coefficients. In addition, we used only those baseline characteristics which were of clinical interest. In both the longitudinal and relative risk model, we standardized the quantitative baseline characteristics to avoid convergence issues.

We fitted the JM which estimates the parameters in the model using the Bayesian methodology.[14,15] Along with the multivariate JM including both biomarkers, we also fitted JM with only SCr longitudinal outcome, and with only PCR longitudinal outcome. This was done to compare the performance of the biomarkers in predicting graft failure. More specifically, we look at the effect size of the association parameters in the JM with both outcomes, ~~and then calculate Area under the curve (AUC)~~ for the three models. For JMs AUC takes a time dependent flavor, that is, AUC is defined per last known visit time and per future time window in which graft failure is to be predicted.[16,17] We calculated it at every six months for a future six month time period. Based on the predictive ability of SCr and PCR, the final JM was chosen to compare the fixed screening schedule with personalized schedule.[11]

For estimation of the personalized schedules, the JM was used to define a patient-specific posterior predictive distribution of time of graft failure, given the observed SCr measurements. The optimal time of the next SCr measurement is the one at which the expected information gained from an extra SCr measurement is maximum. We considered the fixed schedule for measurement of SCr levels, common for all patients, to be 20 times in the first year and every three months thereafter which is the

common clinical protocol used in the AMC. In order to create reasonable predictions, SCr measurements for the first 3 months are taken as per the fixed schedule. Since the SCr measurements were already taken for the kidney transplant patients, we conducted a small simulation in order to demonstrate the efficacy of the personalized schedules. We first assume a population of kidney transplant patients, whose SCr and hazard of graft failure follow a JM with parameters equal to the posterior mean of parameters estimated from the joint model fitted to the kidney transplant dataset. From this population we sample 625 patients, which are further split into a training (575 patients) and test (50 patients) part. ~~Then, iteratively, we scheduled SCr measurements for the test dataset, until the dynamic risk of graft failure became larger than the threshold.[18] The threshold dictates the amount of 'intervention time' at hand between intervention and graft failure. In this simulation we evaluated two thresholds, namely 5% and 2.5% risk of graft loss in 6 months. The difference between the time of intervention due to the schedule (fixed *vs* personalized) and the true intervention time was denoted as intervention offset. Ideally, the intervention offset should be close to zero. In addition, the time at hand between the observed intervention time and the time of graft failure was denoted as the failure offset. Ideally, the failure offset should be as large as possible and far away from zero; number of patients for whom the intervention time was earlier than the true event time should be as high as possible.~~

Continuous variables are presented as mean with standard deviation (SD). Kaplan Meier was used to estimate death-censored graft survival. Results from the JM are presented as regression coefficients with standard deviation and 95% credibility intervals (CIs). Significance levels were set at 5%. Analyses were conducted using R (version 3.4.2)[19] with the GitHub version of the JMbayes package (dated Nov 7, 2017)[16], and survival package (version 2.41)[20].

Results

Sample characteristics

Table 1 shows baseline characteristics of 238 kidney transplant patients and donors. Majority were recipients of deceased donors (74.1%). Mean recipient age was 50.7 (SD 12.7) years, majority firstly transplanted (84.5%), and comorbidities (diabetes 15.5%). In the follow-up period we included a total of ~~13189~~ SCr measurements and 9616 PCR measurements. The median number of repeated SCr and PCR measurements per patient were 45 and 37, respectively. For SCr, 95% of the observations were taken before 6 years, while for PCR they were taken before 5.4 years. The median time between two SCr measurements was 10 days, while for PCR 14 days.

Delayed graft function rate was 32.2%, and higher for deceased donor transplants compared to living donor (42.9% *vs* 1.6%, p<0.001). Figure 1 illustrates death-censored graft survival. At one year, 97.9% still had a functioning graft (95%CI 96.1-99.7), and this was 83.9% (95%CI 78.2-89.6) at 5 years.

Results of the Joint Models

Out of 239 patients, we use the data of only those 238 patients for whom both PCR and SCr data was available. Table 2 and Table 3 summarizes the regression coefficients of the longitudinal submodel for SCr and PCR, respectively. Since the quantitative variables are standardized, the effect sizes correspond to one standard deviation increase. Significantly associated with evolution of SCr were donor age, donor type, recipient age, recipient gender, recipient diabetes, recipient anti-hypertensive medicament use, and delayed graft function. Only donor age and recipient BMI were significantly associated with evolution of PCR. For interpretation, Figure 2 and 3 shows the fitted evolution of SCr and PCR, respectively, from time of transplantation according to a female recipient of 50.7 years old of a first living donor kidney aged 49.7 years, and at the median of other variables. Table 4 summarizes the hazard ratios for the submodel of death-censored graft failure. The SCr levels were strongly associated with the hazard of graft failure: for a given patient at any time point, if the SCr levels increased with 25% and other variables remained the same, the ~~hazard ratio of graft failure increased 1.43 times~~ (adjusted HR 1.43, 95%CI 1.27-1.59, p<.001). Slope of SCr did not reach significance, interpreted as for patients having the same value for SCr and keeping other variables constant, the hazard for death-censored graft failure ~~increased 1.01 times~~ if the slope of the log SCr values increased from -0.2134 to 0.2346 (1st and 3rd quartiles of fitted slope of log SCr; adjusted HR 1.095, 95%CI 0.99 to 1.22, p=0.082. PCR (both value and velocity) were not significantly associated with graft failure. Including the longitudinal SCr and PCR, previous transplantation, HLA mismatches, CIT, and dialysis vintage were also not significantly associated. Figure 3 shows dynamic predictions of both graft survival and evolution of SCr of 2 randomly chosen transplant patients. To further verify if PCR were required in the model in the presence of both log SCr levels and velocity, we fitted two more JMs. In the first JM, only SCr measurements (value and velocity) were included, and in the second JM only PCR measurements (value and velocity). Table 5 and Figure 4 ~~summarizes the predictive ability with tAUCs of the longitudinal markers of the different JMs~~. ~~Both SCr and PCR in the JM~~ performed the same as the model with only SCr to discriminate between patients who experienced graft failure. Therefore, the JM including only SCr measurements was chosen for personalized screening analyses, including SCr velocity although not significant. Except for year 3, all the tAUC of SCr (including value and velocity) were >0.8, indicating good discrimination.

Comment [Office2]: Dit is wel gek

Results of the personalized vs fixed schedules of serum creatinine screening

Figure 5 A-C show the boxplot of the observed number of SCr measurements, intervention offset, and failure offset, respectively, considering a threshold of 5% graft failure risk in 6 months. The median number of SCr measurements for the personalized schedule was lower compared with the fixed schedule (14 *vs* 29, respectively). Also the inter quartile range (IQR) for number of SCr measurements was lower for the personalized compared to the fixed schedule (6 *vs* 8.5, respectively). The median absolute intervention offset for personalized schedules was 4.5 months, and higher compared with 3.1 months for fixed schedule. The higher positive intervention offset at the median can be interpreted as taking a slightly higher risk when applying the personalized schedule. The inter quartile range of the intervention offset, the accuracy to predict the time of intervention, was comparable between

personalized and fixed schedule (4.4 *vs* 4.1, respectively). The graft failure offset denoted that in 12% of the times the graft failure was not detected for the test patients when fixed schedule is used, which was comparable with the personalized schedule at a rate of 14%. In these cases, the schedule was not on time to prevent graft failure. The inter quartile range of the graft failure offset, ~~the accuracy of avoiding time of failure~~, was comparable between personalized and fixed schedule (30.5 *vs* 33.5, respectively). Standard deviation of graft failure offset was comparable between personalized and fixed schedule (4.1 *vs* 4.5, respectively). A more conservative threshold of 2.5% graft failure risk in 6 months resulted in less overshoot of the true graft failure time (6%), however at the expense that it exceeded the true intervention time more often than the fixed schedule (see supplemental file II).

Discussion

The joint model enabled us to study the trajectory of longitudinal SCr and PCR and to specify their association with the risk of graft failure and optimize the screening visits. We demonstrate that SCr has better discriminative ability for risk of graft failure than PCR. Static baseline clinical data was associated with evolution of SCr, and if included in the model for death-censored graft failure, donor and transplant characteristics were not associated anymore. Nephrologists routinely supervise both the current SCr and PCR level and their increase. Indeed, our results suggest not only the current value of SCr is important but also how rapid the rate of increase to this SCr value was. Also we tested our "one size fits all" screening approach. Our goal was to compare an empirical personalized screening schedule based on the join model with the currently used fixed schedule, consisting of 20 SCr measurements in the first year after transplantation and hereafter every 3 months, which is mainly based on prior expert opinion. With the joint model, ~~that~~ is inherently patient specific, we show that a personalized screening approach may result in obtaining less SCr measurements while the time to intervene and overcome the risk for graft failure was comparable with the fixed schedule.

The framework of joint models allows one to tailor screening to the needs of individual patients and dynamically adapt during follow-up. The fixed and frequent schedules are often burdensome for the transplant patients. Patients who remain relatively stable after transplantation may not require frequent measurement of SCr in the first year. On the other hand, patients for whom the graft function deteriorates faster after the first year, a frequent schedule of SCr may be required to determine the best moment for - if possible - intervention. In this regard, instead of a common fixed schedule for all patients, we propose using a different schedule for every patient that is based on their prior SCr trajectory. To test whether the personalized schedule is beneficial for the patient, ~~we used scheduled SCr measurements and determined the intervention time at which the 6-month dynamic risk of graft failure of the patient becomes larger than a certain threshold. According to this threshold and the intervention time,~~ we then calculate the time difference between the ~~observed~~ intervention time according to the schedule and the true intervention time (intervention offset), and the time difference between the observed intervention time and the time of graft failure (graft failure offset). To put it simply, intervention offset calculates ‘how accurate are we with regards to our stopping rule of not

crossing a dynamic 6 month risk threshold of 5%?', and graft failure offset calculates ''how much do we overshoot or undershoot the true graft failure time?'. For scheduling there are two levels of uncertainty: not knowing true failure time and not knowing true SCr measurements. The personalized schedule should be accurate to predict the time of intervention as well as avoiding time of failure, depending on the threshold for graft failure.

Given a liberal threshold of 5% risk of graft failure in 6 months, our simulation resulted in a dramatic reduction of the number of visits (~50%) while intervention offset and graft failure offset were comparable with the fixed schedule. This ~50% reduction in screening is directly associated with a reduction in patient management costs, physician time and it will also aid to a higher quality of life for transplant recipients due to a decrease in scheduled hospital visits. When we extrapolate our results to show the potential of the personalized screening approach, assuming a fixed screening approach that is similar to our hospital, an estimated of ~$500 per screening and the total number of transplanted patients in the Netherlands[21], the personalized screening could reduce annual costs by >$14.500.000. Considering the WHO 2015 worldwide kidney transplantation activity, based on the Global Observatory on Donation and Transplantation (GODT) data, produced by the WHO-ONT collaboration, the personalized screening could reduce costs by >$630.000.000 anually worldwide. The next step would be to increase sample size to include more risk factors for graft survival which could facilitate the personalized screening approach. As we included well-known risk factors for graft failure, other biomarkers of interest such as graft histology or longitudinal genomic data could theoretically be introduced. Finally, our findings have to be externally validated in other observational cohorts, and in a non-inferiority randomized controlled trial, preferably in multiple centers with different fixed screening approaches (current clinical practice).

Comment [HPS3]: Wijze van berekening:

In 2016: 10.797 prevalentie transplantatie patienten. 1000 mensen/jaar getransplanteerd. Dus incidentie is 1000, resterend is 10.797-1000 die 4/jaar worden gescreend = 9.797 * 4 = 39188
1000*20 screens = 20000
Totaal aantal screening = 59188
Totaal kosten= 59188*500= 29594000
Gedeeld door 2 (50% reductie)
= 14797000

Numerous studies have addressed progressively worsening kidney function in a fixed time-window with the risk of graft failure, mostly evaluating linear trajectories of eGFR[22-25], but some included also nonlinearity of progression[10,26] Only two studies included nonlinear trajectories in a joint model with graft survival which allows for dynamic predictions.[27,28] To understand more precisely how SCr and PCR measurements change over time and how those changes impact graft survival, joint models efficiently uses all available data. We show which clinical factors associate with evolution of SCr and PCR leading to graft failure, whereas other clinical factors may associate directly with graft failure. We corroborate the findings of Fournier and colleagues who concluded that the graft failure risk depended on both the current value and slope of the SCr.[27] We observed that a majority of clinical data at time of transplantation were associated with longitudinal SCr, and less with longitudinal PCR. Once longitudinal SCr and PCR were included in the graft survival model, none of the clinical data at time of transplantation were significantly associated with death-censored graft failure. This could be explained by that the static baseline risk factors are not directly associated with graft failure, which is in line with previous study.[27] In our study common risk factors for graft loss such as previous transplantation, cold ischemic time, HLA mismatches, and recipient dialysis vintage were also not significantly associated with graft failure once we included the SCr trajectories in the joint model. Compared to the study by

Fournier and colleagues, our study design differs in two important aspects. First we were interested in risk factors for the graft, and therefore modelled death-censored graft failure instead of composite outcome of graft failure. Secondly, for clinical reasons, we were also interested in predictive ability including the first year after transplantation as well, and not only after the first year in more stable transplanted patients. It may also be that the relatively low number of graft failures increased the risk for type II error in finding predictors for graft failure beside including longitudinal SCr and PCR.

To our knowledge, we are the first to compare the predictive ability of SCr against PCR in renal transplantation, both considered markers for kidney function decline. We chose to include SCr instead of calculating eGFR to reduce the risk of overcorrection since recipient age, and gender were already included in the longitudinal SCr and PCR mixed models. Our results suggest that there is no additive predictive value for graft failure of a joint model that includes longitudinal PCR with SCr compared with SCr alone. The KDIGO guidelines provide suggestions for the effective control of serum creatinine and proteinuria, whereas SCr is recommended to measure more frequently then PCR.[29,30] For measuring urine protein, it is recommended within one month after transplantation as an initial value, and every 3 months during the first year, and annually thereafter.[29] For SCr, it is recommended to measure daily within the first week after transplantation, 2-3 times weekly within 2-4 weeks, then once a week within 3 months, then every two weeks within 4-6 months, hereafter monthly within 7-12 months, and every 2-3 months, thereafter.[29] Next step is to validate the added predictive value of PCR measurements in other centers, by using dynamic modelling. We should acknowledge that in patients with a primary proteinuric disease that could recur in the graft, especially primary focal and segmental glomerulosclerosis, urinary PCR should always be incorporated in the outpatient screening approach. The variation in SCr measurements in the clinical KDIGO guideline already reflect stability of the graft: if stable, measurements could be reduced by some level. Depending on the clinical transplant data and follow-up measurements, we show a personalized and data-driven approach when to plan the next visit.

Some limitations should be considered as well. Results should be interpreted with care, and externally validated given the relatively low number of events in the survival model. Also, we chose to log transform the slope of SCr to meet model assumptions, however back transformation is not straightforward like the value of SCr, and therefore difficult to interpret. We should note that due to the timing of each next visit depends on the collected longitudinal SCr and PCR of each patient. Therefore we included ~~13189~~ SCr measurements and 9616 PCR measurements of only 238 transplant patients in our tertiary referral transplant hospital. Most large transplant registries include monthly or yearly SCr measurements. Also, increasing the sample size may lead to heavy computational burden. As a quid pro quo, the joint model, which takes into account the measurement error of longitudinal SCr, overcomes the disadvantages of the Cox model with time-dependent covariates, which has been shown to underestimate the hazard ratios of SCr and graft failure.[28]

Up to now, only few studies used joint models to evaluate the predictive value of longitudinal measurements of renal function. A greater understanding of renal function trajectory has important implications with regard to the clinical care of transplant patients. SCr has better discriminative ability for risk of graft failure than PCR. Static baseline clinical data was associated with evolution of SCr, and if included in the model for death-censored graft failure, donor and transplant characteristics were not significantly associated anymore. Based on the joint model it is possible to select the optimal time point to plan the next measurement, resulting in a tailored approach for the patient.

References

1. Snoeijs MG, Schaubel DE, Hene R, et al. Kidneys from Donors after Cardiac Death Provide Survival Benefit. *Journal of the American Society of Nephrology.* 2010;21(6):1015-1021.
2. Orandi BJ, Luo X, Massie AB, et al. Survival Benefit with Kidney Transplants from HLA-Incompatible Live Donors. *New Engl J Med.* 2016;374(10):940-950.
3. Sorensen VR, Heaf J, Wehberg S, Sorensen SS. Survival Benefit in Renal Transplantation Despite High Comorbidity. *Transplantation.* 2016;100(10):2160-2167.
4. Liem YS, Bosch JL, Hunink MGM. Preference-based quality of life of patients on renal replacement therapy: A systematic review and meta-analysis. *Value Health.* 2008;11(4):733-741.
5. Stegall MD, Gaston RS, Cosio FG, Matas A. Through a Glass Darkly: Seeking Clarity in Preventing Late Kidney Transplant Failure. *Journal of the American Society of Nephrology.* 2015;26(1):20-29.
6. Sellares J, de Freitas DG, Mengel M, et al. Understanding the Causes of Kidney Transplant Failure: The Dominant Role of Antibody-Mediated Rejection and Nonadherence. *Am J Transplant.* 2012;12(2):388-399.
7. Hariharan S, McBride MA, Cherikh WS, Tolleris CB, Bresnahan BA, Johnson CP. Post-transplant renal function in the first year predicts long-term kidney transplant survival. *Kidney international.* 2002;62(1):311-318.
8. Roodnat JL, Mulder PGH, Rischen-Vos J, et al. Proteinuria after renal transplantation affects not only graft survival but also patient survival. *Transplantation.* 2001;72(3):438-444.
9. Streja E, Goldstein L, Soohoo M, Obi Y, Kalantar-Zadeh K, Rhee CM. Modeling longitudinal data and its impact on survival in observational nephrology studies: tools and considerations. *Nephrol Dial Transpl.* 2017;32:77-83.
10. Ferro CJ, Hodson J, Moore J, et al. Bayesian Analysis of Glomerular Filtration Rate Trajectories in Kidney Transplant Recipients A Pilot Study. *Transplantation.* 2015;99(3):533-539.
11. Rizopoulos D, Taylor JMG, Van Rosmalen J, Steyerberg EW, Takkenberg JJM. Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics.* 2016;17(1):149-164.
12. Federation of Dutch Medical Scientific Societies (FDMSS). Human Tissue and Medical Research: Code of Conduct for responsible use (2011). 2015; https://www.federa.org/sites/default/files/images/print_version_code_of_conduct_english.pdf. Accessed October 1, 2016.
13. Eurotransplant International Foundation. Eurotransplant Manual Chapter 4: ET Kidney Allocation System (ETKAS). 2016;5.5.
14. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R.*: CRC Press.; 2012.
15. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Stat Sinica.* 2004;14(3):809-834.
16. Rizopoulos D. The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC. *Journal of Statistical Software.* 2016;72(7):1-46.

17. Rizopoulos D, Molenberghs G, Lesaffre EMEH. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical J.* 2017;59(6):1261-1276.
18. Rizopoulos D. Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data. *Biometrics.* 2011;67(3):819-829.
19. R Core Team. R: A Language and Environment for Statistical Computing. 2016.
20. Therneau T. A Package for Survival Analysis in S. *version 2.41* 2017; https://CRAN.R-project.org/package=survival.
21. Nefrovisie. Nierfunctievervangende behandeling in Nederland. Jaarboek 2016. 2017.
22. Gera M, Slezak JM, Rule AD, Larson TS, Stegall MD, Cosio FG. Assessment of changes in kidney allograft function using creatinine-based estimates of glomerular filtration rate. *Am J Transplant.* 2007;7(4):880-887.
23. Gill JS, Tonelli M, Mix CH, Pereira BJG. The change in allograft function among long-term kidney transplant recipients. *Journal of the American Society of Nephrology.* 2003;14(6):1636-1642.
24. Park WD, Larson TS, Griffin MD, Stegall MD. Identification and Characterization of Kidney Transplants With Good Glomerular Filtration Rate at 1 Year But Subsequent Progressive Loss of Renal Function. *Transplantation.* 2012;94(9):931-939.
25. Kaplan B, Schold J, Meier-Kriesche HU. Poor predictive value of serum creatinine for renal allograft loss. *Am J Transplant.* 2003;3(12):1560-1565.
26. Clayton PA, Lim WH, Wong G, Chadban SJ. Relationship between eGFR Decline and Hard Outcomes after Kidney Transplants. *Journal of the American Society of Nephrology.* 2016;27(11):3440-3446.
27. Fournier MC, Foucher Y, Blanche P, Buron F, Giral M, Dantan E. 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.
28. Asar O, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *Int J Epidemiol.* 2015;44(1):334-344.
29. Kasiske BL, Zeier MG, Chapman JR, et al. KDIGO clinical practice guideline for the care of kidney transplant recipients: a summary. *Kidney international.* 2010;77(4):299-311.
30. Eckardt KU, Kasiske BL. KDIGO Clinical Practice Guideline for the Care of Kidney Transplant Recipients. *Am J Transplant.* 2009;9:S1-S155.

TABLES

Table 1	
Characteristics of the transplant cohort (n=239)	
Donor	**Mean (SD) / N (%)**
Age (years)	49.7 (12.7)
Gender (Female)	
BMI	25.1 (4.4)
Donor type	
- Living unrelated	38 (15.9%)
- Living related	24 (10.0%)
- DBD	98 (41.0%)
- DCD	79 (33.1%)
Transplant	
Cold ischemia time (hours)	14.8 (8.7)
Panel Reactive Antibodies (%)	
- 0%	181 (75.7%)
- >1%	58 (24.3%)
HLA A, B, DR mismatches	
- 0 mismatch	28 (11.7%)
- 1 mismatch	16 (6.7%)
- 2 mismatches	49 (20.5%)
- 3 mismatches	73 (30.5%)
- 4 mismatches	36 (15.1%)
- 5 mismatches	28 (11.7%)
- 6 mismatches	9 (3.8%)
Recipient	
Age (years)	50.7 (13.1)
Gender (Female)	102 (42.7%)
BMI	25.4 (4.3)
Previous transplantation (yes)	37 (15.5%)
Dialysis vintage (years)	3.7 (3.5)
Diabetes (yes)	38 (15.9%)
Cardiovascular events before Tx (yes)	91 (38.1%)
Number of anti-hypertensives	
- 0	33 (13.8
- 1	79 (33.1
- 2	85 (35.6
≥ 3	42 (17.7

Table 2					
Parameter estimates for the longitudinal model for SCr					
Variable	**Mean**	**Std. Dev**	**2.5%**	**97.5%**	**P**
Intercept	5.226	0.080	5.064	5.378	<0.000
Receiver age	-0.063	0.022	-0.107	-0.019	0.010
Donor age	0.083	0.020	0.045	0.119	<0.000
Donor BMI	-0.011	0.021	-0.054	0.028	0.612
Receiver BMI	0.018	0.023	-0.025	0.060	0.420
#HLA mismatches between donor and recipient	0.020	0.022	-0.022	0.065	0.342
Panel reactive antibody percentage	0.048	0.027	-0.008	0.100	0.082
#Anti-hypertensive medicaments	0.040	0.020	0.001	0.080	0.048
Cold ischemia time	0.029	0.035	-0.039	0.102	0.390
#Days on dialysis before transplant	0.015	0.029	-0.042	0.071	0.580
Receiver gender: Male	0.197	0.042	0.111	0.276	<0.000
Previous transplant: Yes	0.016	0.064	-0.115	0.141	0.786
Donor gender: Male	0.053	0.042	-0.027	0.136	0.198
Delayed graft function: Yes	0.118	0.049	0.025	0.216	0.006
Diabetes Mellitus: Yes	-0.103	0.059	-0.217	0.012	0.076
Cardiovascular events before transplantation: Yes	-0.047	0.043	-0.129	0.044	0.272
Deceased donor: Yes	0.163	0.082	0.004	0.313	0.044
Spline: visit time [0.039, 0.082] years	-0.440	0.041	-0.517	-0.358	<0.000
Spline: visit time [0.082, 0.219] years	-0.182	0.053	-0.284	-0.081	<0.000
Spline: visit time [0.219, 1] years	-0.545	0.081	-0.712	-0.395	<0.000
Spline: visit time [1, 6] years	0.007	0.083	-0.155	0.176	0.946
σ	0.190	0.001	0.187	0.192	

Table 3					
Parameter estimates for the longitudinal model for PCR					
Variable	**Mean**	**Std. Dev**	**2.5%**	**97.5%**	**P**
Intercept	3.731	0.179	3.398	4.083	<0.000
Receiver age	0.030	0.052	-0.066	0.138	0.604
Donor age	0.209	0.047	0.118	0.301	<0.000
Donor BMI	-0.019	0.051	-0.121	0.084	0.716
Receiver BMI	-0.116	0.050	-0.219	-0.021	0.014
#HLA mismatches between donor and recipient	-0.013	0.049	-0.112	0.086	0.776
Panel reactive antibody percentage	0.047	0.061	-0.066	0.166	0.446
#Anti-hypertensive medicaments	0.056	0.047	-0.03	0.147	0.208
Cold ischemia time	0.062	0.082	-0.097	0.211	0.468
#Days on dialysis before transplant	0.006	0.066	-0.120	0.134	0.952
Receiver gender: Male	-0.026	0.094	-0.207	0.166	0.798
Previous transplant: Yes	0.035	0.149	-0.241	0.332	0.816
Donor gender: Male	0.114	0.096	-0.079	0.303	0.228
Delayed graft function: Yes	0.043	0.118	-0.174	0.275	0.740
Diabetes Mellitus: Yes	0.153	0.135	-0.124	0.396	0.256
Cardiovascular events before transplantation: Yes	-0.016	0.106	-0.221	0.199	0.890
Deceased donor: Yes	0.144	0.193	-0.246	0.509	0.462
Spline: visit time [0.039, 0.082] years	-0.821	0.090	-0.989	-0.638	<0.000
Spline: visit time [0.082, 0.219] years	-0.578	0.131	-0.838	-0.304	<0.000
Spline: visit time [0.219, 1] years	-0.898	0.160	-1.218	-0.587	<0.000
Spline: visit time [1, 6] years	0.460	0.234	0.015	0.927	0.036
σ	0.479	0.004	0.472	0.486	

:.

Table 4					
Relative risk submodel estimates for the log hazard ratio and 95% credible interval					
Variable	**Mean**	**Std. Dev**	**2.5%**	**97.5%**	**P**
Previous transplant: Yes	0.305	0.339	-0.099	0.986	0.352
#HLA mismatches between donor and recipient	0.048	0.093	-0.144	0.269	0.620
Cold ischemia time	-0.051	0.105	-0.277	0.133	0.644
#Days on dialysis before transplant	-0.013	0.102	-0.251	0.178	0.934
log PCR	0.145	0.125	-0.056	0.431	0.188
slope(log PCR)	0.021	0.058	-0.076	0.145	0.828
log SCr	1.599	0.241	1.067	2.063	<0.000
slope(log SCr)	0.203	0.123	-0.017	0.443	0.082

Table 5						
Area under curve characteristics for the JMs fitted to the kidney transplant data set						
Biomarkers	**Year 0.5**	**Year 1**	**Year 1.5**	**Year 2**	**Year 2.5**	**Year 3**
Both SCr and PCR	0.845	0.927	0.915	0.828	0.953	0.686
Only SCr	0.801	0.901	0.918	0.866	0.957	0.692
Only PCR	0.844	0.858	0.755	0.640	0.825	0.473

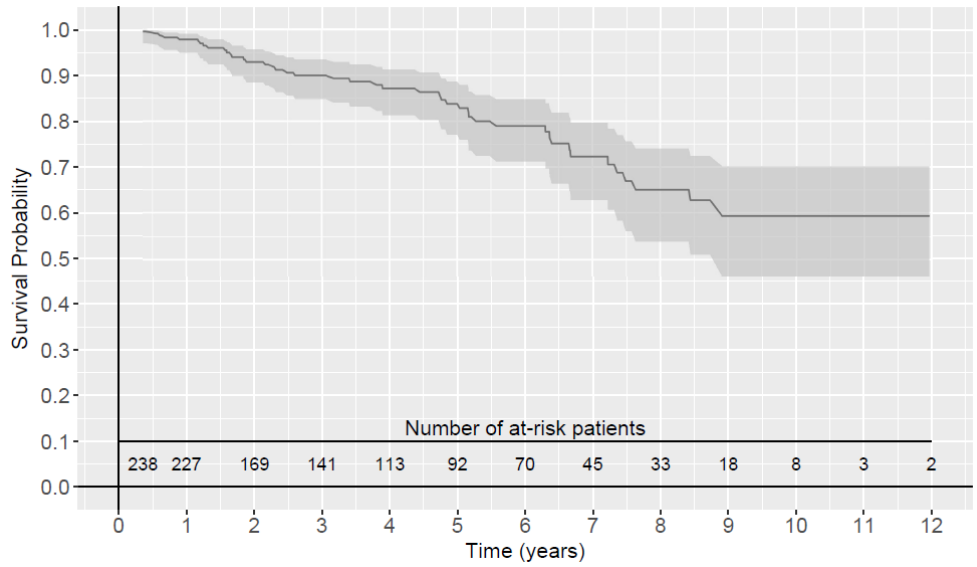

Figure 1: Death-censored graft survival and 95% CI using Kaplan Meier.

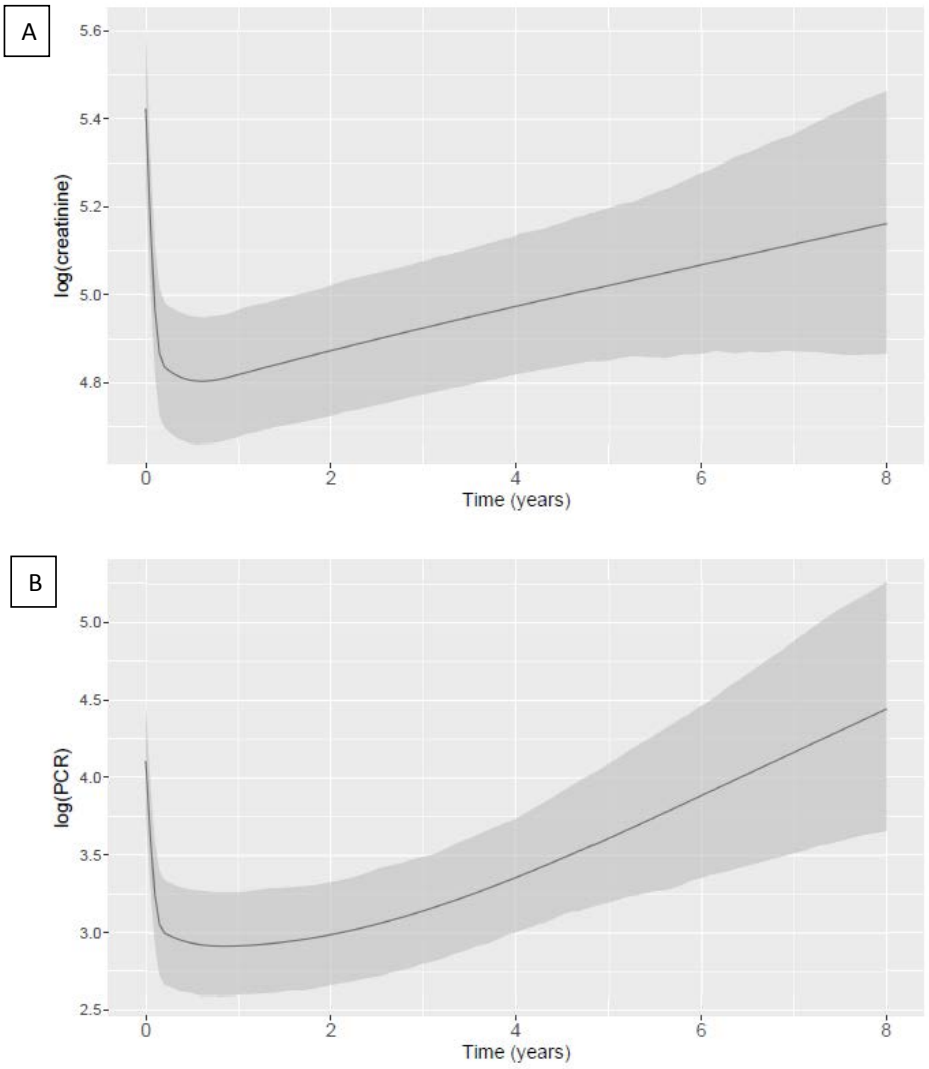

Figure 2: Fitted longitudinal evolution of SCr (A) and PCR (B) with 95% CI, corresponding to a female recipient aged 50.7 years, BMI 25.4, first transplantation, no diabetes, no history of any cardiovascular events, of a living female donor aged 49.7 years, with 3 mismatches on HLA A, B, DR, with 14.8 hours of CIT, 4.8% of Panel Reactive Antibodies, and 3.7 years of dialysis vintage.

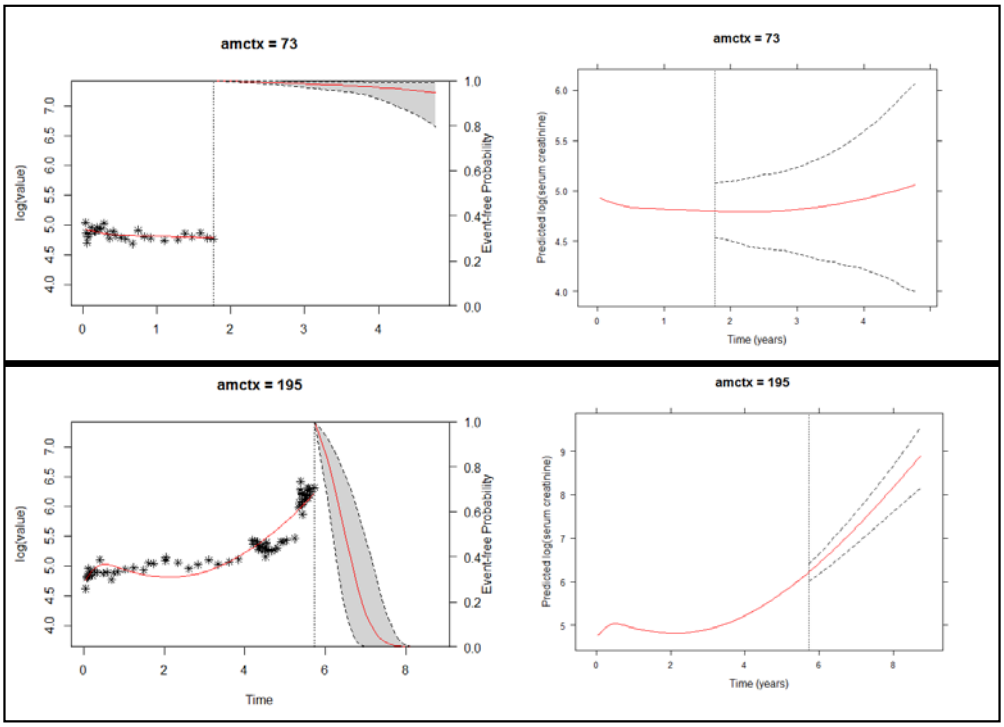

Figure 3: Dynamic predictions of death-censored survival probabilities and dynamic predictions of log(creatinine) up to 3 years after loss of follow up.

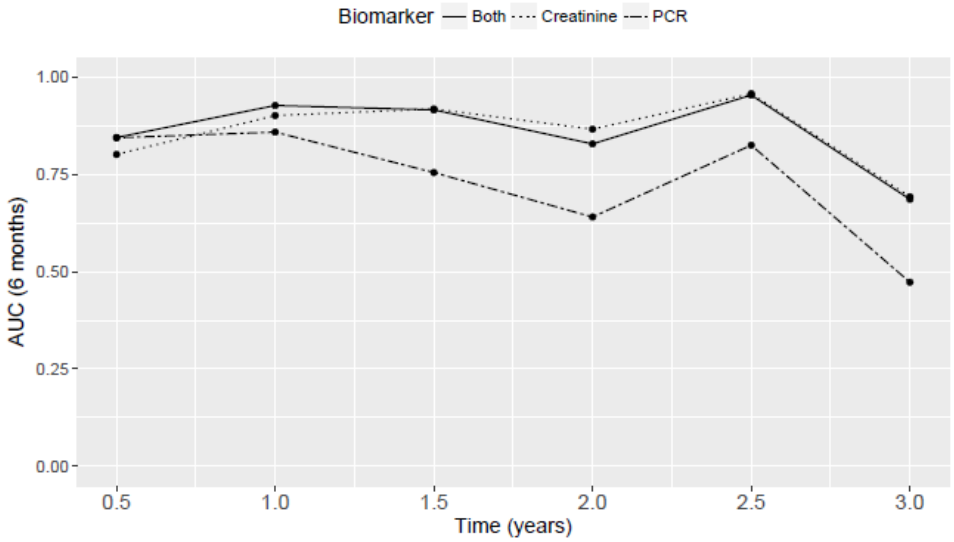

Figure 4: Time-dependent area under the receiver operating curves (AUC) for the joint models.

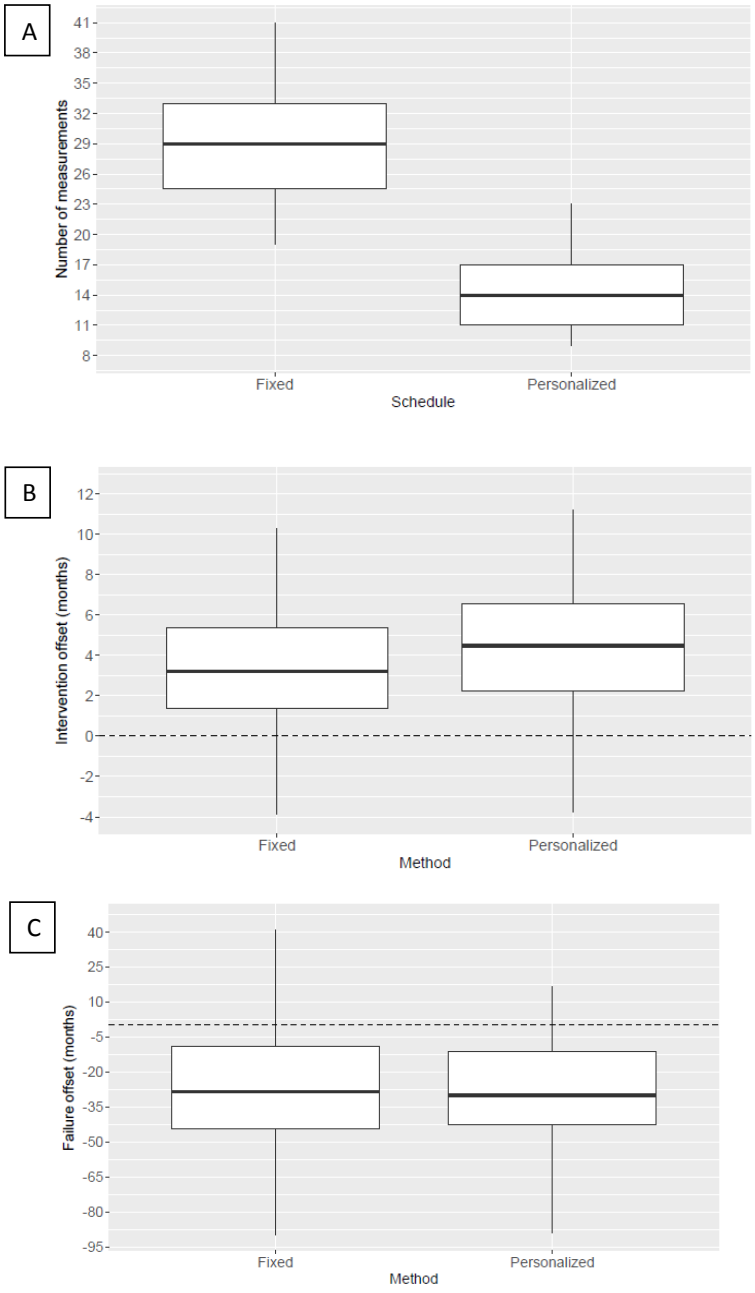

Figure 5: Box plots of the number of scheduled SCr measurements (A), intervention offset (B), and the graft failure offset (C). Fixed schedules were compared with personalized schedules, depended on the dynamic predictions of the joint model. The threshold was set at 5% risk of graft failure per 6 months. The zero offset mark (for B and C) is displayed with the dashed line.