Personalized screening intervals to monitor the development of chronic renal allograft failure: an observational cohort and simulation study

Hessel Peters-Sengers[1,2,*], Anirudh Tomer[3,*], Sandrine Florquin[4], Joris JTH Roelofs[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, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Frederike J Bemelman, professor

[2]Center for Experimental and Molecular Medicine (CEMM), Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Hessel Peters-Sengers, postdoctoral fellow in epidemiology and clinical research

[3]Department of Biostatistics, Erasmus University Medical Center, 's-Gravendijkwal 230, 3015 CE Rotterdam, The Netherlands

Anirudh Tomer, doctoral student in biostatistics, Dimitris Rizopoulos, associate professor in biostatistics

[4]Department of Pathology, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Sandrine Florquin, professor, Joris JTH Roelofs, associate professor, Jesper Kers, postdoctoral fellow.

[5]Department of Public Health, Erasmus University Medical Center, 's-Gravendijkwal 230, 3015 CE Rotterdam, The Netherlands

Ewout W Steyerberg, professor

[6]Department of Biomedical Data Sciences, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

[7]Van 't Hoff Institute for Molecular Sciences (HIMS), University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

[*,$]Authors contributed equally to the study

Correspondence to: J Kers j.kers@amc.uva.nl

Abstract

Objectives:

First, to model the association between longitudinal markers of renal function (serum creatinine; SCr and urinary protein-to-creatinine ratio; PCR) and graft failure in renal transplant recipients with joint models. Second, to calculate personalized screening intervals and compare them to the currently used fixed screening protocol.

Design:

Observational cohort with subsequent simulations study.

Setting:

Single-center cohort from a university hospital performing renal transplantation in The Netherlands.

Participants:

Adult renal transplant recipients of a living or deceased donor kidney with at least 2 SCr and 2 PCR measurements on follow-up**.**

Main outcome measures:

Adjusted hazard ratios (aHRs) for death-censored graft failure (DCGF) based on the joint models. Number of screening intervals, intervention offset and graft failure offset between the personalized and fixed screening approach in the simulation study. Intervention offset is defined as the time difference between the intervention time (the time when the risk of DCGF at the next 6 months ≥ 5%) estimated on the basis of the SCr measurements and the observed intervention time. Graft failure offset is defined as the time difference between estimated intervention time and observed graft failure time.

Results:

There was a median of 45 SCr and 37 PCR measurements per patient. The final joint model showed an aHR of 1.43 (95% credible interval [CI] 1.27-1.59, p<0.001) if the absolute SCr value increased with 25% and an aHR of 1.10 (95% CI 0.99-1.22, p=0.08) when the log SCr slope changed from -0.21 to 0.23, which corresponds to the 1st and 3rd quartiles of the fitted slope. A joint model that included both the SCr and the PCR trajectories did not reveal an increased time-dependent (t)AUC compared to a model with only SCr trajectories ([t]AUC > 0.8 up till 2.5 years). In the simulation study, the personalized screening strategy based on the joint model that included SCr resulted in a median (IQR) of 14 (6.0) versus 29 (8.5) screening intervals as compared to the fixed protocol. Intervention offset was higher for the personalized approach ~~(5.4 [4.4] versus 3.1 [4.1])~~ whereas graft failure offset was comparable (14% versus 12% missed cases with DCGF).

Conclusions:

SCr trajectories rather than PCR trajectories best predict death-censored graft failure after renal transplantation. The joint model-based personalized screening approach reduces the number of interval measurements needed to monitor the risk of DCGF without a loss in prognostic information. The results of this study suggest that a personalized and data-driven screening approach might reduce the number of outpatient visits, physician time burden and healthcare costs.

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 outcome and quality of life [1–4]. Short-term graft survival is excellent [5], but unfortunately long-term graft survival has not benefitted from improvements in treatment regimens to the same extent [5]. Patient management requires continuous monitoring of renal function after transplantation according to local protocols and the choice of outpatient screening time-points is based on expert opinion.

In an effort to stratify patients who are at increased or decreased risk for irreversible graft failure, a multitude of studies investigated the value of clinical algorithms and molecular biomarkers measured at one time-point after transplantation for the prediction of long-term graft failure, but none are used to monitor renal transplant recipients as of yet. The Kidney Disease: Improving Global Outcome (KDIGO) guideline on the management of renal transplant recipients suggests screening serum creatinine daily the first 7 weeks after transplantation, 2-3 times per week for week 2-4, weekly for months 2-3, every 2 weeks for months 4-6, monthly for months 7-12 and every 2-3 months thereafter [6]. Similarly, the KDIGO guideline suggests screening proteinuria at least once in the first month, every 3 months during the first year and annually thereafter [6]. Although both markers are used in routine transplant recipient outpatient management worldwide, the quality of the evidence for the abovementioned screening suggestions is low (for serum creatinine) to very low (for proteinuria), corresponding to GRADE levels C and D respectively [6]. Only very few studies have looked at the temporal dynamics of the serum creatinine (SCr) and total urinary protein-to-creatinine ratio (PCR) trajectories after transplantation as a predictor for irreversible 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 [7]. A recent study by Ferro and colleagues highlighted that the contrary was true. More specifically, 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.

In the current study, we had two aims. Firstly, we wanted to understand the dynamic relationship between longitudinal indicators of renal function (SCr and PCR) and irreversible graft failure. Secondly, we wanted to optimize monitoring based on personalized risk estimates and compare such personalized screening with the fixed-term one-size-fits-all protocol that is currently being used in our hospital. For both aims, we developed joint models (JMs) that combine information from longitudinal and survival data into one model [8–12]. In a previous study, the concept of personalized screening based on a JM 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 to predict the composite outcome of re-operation and death [13]. The JM was then used to create personalized screening schedules for echocardiography tailored to the

risk of the individual patient. Such a personalized medicine approach could potentially lead to a lower medical and financial burden in renal transplant patients with stable transplant function without losing important information on patients at risk for irreversible graft failure that would require timely intervention.

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) [14].

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 SCr (in umol/L) and spot or 24-hour urine collection to calculate the PCR (in 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 April 29, 2014. Initial immunosuppressive therapy consisted of steroids combined with mycophenolate mofetil or mycophenolic acid and a calcineurin inhibitor, mostly tacrolimus, but also cyclosporine A. Alternatively, a combination of steroids, tacrolimus, and sirolimus was used. Donor kidneys were acquired through allocation by the Eurotransplant allocation program, Leiden, The Netherlands [15].

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, 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 Appendix A 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 [16,17]. 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 [18]~~. 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 estimated the parameters in the JM using Bayesian methodology fitted to the dataset. [8,10,19]. 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 looked at the effect size of the association parameters in the JM with both outcomes, and then calculate area-under-the-curves (AUC) for the three models. For JMs, the 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 [10,11]. 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 [13].

To create the personalized schedules, the fitted 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. For the two sets of patients we generated a true graft failure time, and a time of intervention. The true time of intervention is the time at which the patient is taken out of surveillance and treated, so that graft failure is avoided. The intervention time is the time at which a

patient's true dynamic risk of death-censored graft failure is above a certain threshold [13]. We used two thresholds in this paper, namely 5% and 2.5% in six-month period. Higher risk thresholds give us intervention time closer to true graft failure times.

We fitted the JM to the training patients and used it to create a personalized schedule of SCr measurement for each of the test patients. We then conducted hypothetical SCr measurements for the test patients according to the two schedules. The schedules estimate the dynamic risk of graft failure, and therefore the generated intervention times are not equal to the true intervention time. We define the difference between estimated intervention time and true intervention time as intervention offset. The better schedule will be the one for whom this difference is zero. That is, the schedule with only a few set of observations matches the entire profile of the patient. In addition, since we want to avoid graft failure we define the difference between estimated intervention time and graft failure time as failure offset. Ideally this difference should be less than zero for the majority of the patients, and the time difference should be enough to give immediate treatment. 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) with the GitHub version of the JMbayes package (dated Nov 7, 2017) [10], and survival package (version 2.41).

Results

Sample characteristics

Table 1 shows baseline characteristics of 239 kidney transplant patients and donors. Most were recipients of deceased donors (177 (74.1%). Mean recipient age was 51 (SD 13) years, and majority firstly transplanted (84.5%). 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 (43% *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. Appendix B 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 A and B (Appendix B) show 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. Appendix B summarizes the hazard ratios for the survival 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<0.001). The 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 -0.21 to 0.23 (1st and 3rd quartiles of the fitted slope of log SCr; adjusted HR 1.10, 95% CI 0.99 to 1.22, p=0.082. PCR (both value and velocity) were not significantly associated with graft failure. In a model including the longitudinal trajectories of SCr and PCR, the parameters previous transplantation, HLA mismatches, CIT, and dialysis vintage were also not significantly associated with graft failure. Figure 3 shows dynamic predictions of both graft survival and evolution of SCr of the same transplant patient according to different follow-up times. To further verify if the PCR was 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 (Appendix B.1), and in the second JM only PCR measurements (value and velocity). Table 2 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. Except for year 3, all the AUC estimates of SCr (including value and velocity) were >0.8, indicating good discrimination.

Results of the personalized vs fixed schedules of serum creatinine screening

Figure 4 A to C (Appendix C) show the boxplots 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 the 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 IQR 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 the fixed schedule was 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 IQR of the graft failure offset, the accuracy of avoiding time of failure, was comparable between the 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 Appendix C).

Discussion

The joint model enabled us to study the dynamic trajectory of 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 were associated with evolution of SCr, and if included in the model for death-censored graft failure, donor and transplant characteristics were not associated with graft outcome 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. We compared an empirical personalized screening schedule based on the fitted JM with the one-size-fits-all fixed schedule that is currently used in our hospital, consisting of 20 SCr measurements in the first year after transplantation and hereafter every 3 months, which is already less stringent than the screening schedule as proposed by the KDIGO guidelines [6]. With the JM, that is inherently patient specific, we show that a personalized screening approach may result in obtaining less SCr measurements while the same information to predict the risk of irreversible graft failure remains the same. Therefore, the framework of joint models allows one to tailor screening to the needs of individual patients and adapt during follow-up. The fixed and frequent schedules are often burdensome for the transplant patients. Patients who have a stable allograft function after transplantation may not require frequent measurement of SCr. 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 (e.g. by timely planning of a transplant biopsy).

Numerous studies have modelled progressively worsening kidney function in a fixed time-window using linear trajectories to evaluate the risk of graft failure [20–23]. This approach less frequently also included nonlinearity of progression [24,25]. Only five studies included nonlinear renal function trajectories in a joint model with renal outcome: four in renal transplantation [26–28,17] and one in native chronic kidney disease [19]. In line with our findings, these studies all showed that both the current SCr value as well as the SCr slope associated with irreversible renal (allograft) outcome. Of these studies, only the study by Rizopoulos et al. investigated the added value of proteinuria (as a binary measure) in renal transplantation [17]. They showed in multivariate joint modeling that eGFR, proteinuria (as a binary measure) and hematocrit trajectories associated with graft outcome. Extending on this, we showed in the current study that the joint model that only included SCr trajectories had similar time-dependent discriminative power as the joint model that included both SCr and PCR trajectories. Although we postulate that PCR does not add discriminative value in the joint model that included SCr trajectories, we acknowledge that this might be the case in patients with a recurrence of primary proteinuric renal disease after transplantation (e.g. primary focal and segmental glomerulosclerosis).

To the best of our knowledge, the current study is the first in nephrology to use joint model estimates to tailor the SCr screening schedule to the inidivudual renal transplant recipient. Our statistical simulation study resulted in a nearly 50% reduction in the number of necessary visits. This 50% reduction in screening intervals can be directly translated to 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 national and global potential of this personalized screening approach, assuming a fixed screening approach that is similar to our hospital, an estimated €500 per screening and the prevalence and incidence of number of transplanted patients in the Netherlands [29], the personalized screening could reduce annual costs by more than €14.500.000 in The Netherlands. 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 [30], the personalized screening could reduce costs by more than €630.000.000 annually worldwide. We have to acknowledge that the fixed screening protocol is a guideline. In daily practice, treating physicians personalize the screening intensity according to prior knowledge on the individual patient (expert opinion-based personalized screening).

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 to increase specificity for the underlying disease process leading to irreversible graft failure. Finally, and most importantly, 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 protocols. This will allow us to compare our evidence-based personalized screening approach back-to-back to expert opinion-based personalized screening.

Acknowledgements

Dr. Kers is supported by a grant from the Dutch Kidney Foundation (grant number 17OKG23). Hyperlinks to all source codes for the joint model fits and the simulation study can be found in Appendix D.

References

1 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*;**11**:733–41. doi:10.1111/j.1524-4733.2007.00308.x

2 Snoeijs MG, Schaubel DE, Hené R, *et al.* Kidneys from donors after cardiac death provide survival benefit. *J Am Soc Nephrol* 2010;**21**:1015–21. doi:10.1681/ASN.2009121203

3 Orandi BJ, Luo X, Massie AB, *et al.* Survival Benefit with Kidney Transplants from HLA-Incompatible Live Donors. *N Engl J Med* 2016;**374**:940–50. doi:10.1056/NEJMoa1508380

4 Sørensen VR, Heaf J, Wehberg S, *et al.* Survival Benefit in Renal Transplantation Despite High Comorbidity. *Transplantation* 2016;**100**:2160–7. doi:10.1097/TP.0000000000001002

5 Stegall MD, Gaston RS, Cosio FG, *et al.* Through a glass darkly: seeking clarity in preventing late kidney transplant failure. *J Am Soc Nephrol* 2015;**26**:20–9. doi:10.1681/ASN.2014040378

6 Kidney Disease: Improving Global Outcomes (KDIGO) Transplant Work Group. KDIGO clinical practice guideline for the care of kidney transplant recipients. *Am J Transplant* 2009;**9 Suppl 3**:S1-155. doi:10.1111/j.1600-6143.2009.02834.x

7 Streja E, Goldstein L, Soohoo M, *et al.* Modeling longitudinal data and its impact on survival in observational nephrology studies: tools and considerations. *Nephrol Dial Transplant* 2017;**32**:ii77-ii83. doi:10.1093/ndt/gfx015

8 Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press 2012.

9 Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Stat Sin* 2004;**14**:809–34.

10 Rizopoulos D. The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC. *J Stat Softw* 2016;**72**. doi:10.18637/jss.v072.i07

11 Rizopoulos D, Molenberghs G, Lesaffre EMEHEH. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biom J* 2017;**59**:1261–76. doi:10.1002/bimj.201600238

12 Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011;**67**:819–29. doi:10.1111/j.1541-0420.2010.01546.x

13 Rizopoulos D, Taylor JMG, Van Rosmalen J, *et al.* Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics* 2016;**17**:149–64. doi:10.1093/biostatistics/kxv031

14 Federation of Dutch Medical Scientific Societies. 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 1 Oct 2016).

15 Eurotransplant International Foundation. *Eurotransplant Manual Chapter 4: ET Kidney Allocation System (ETKAS).* 2016.

16 Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal

biomarkers and survival. *Biometrics* 2005;**61**:64–73. doi:10.1111/j.0006-341X.2005.030929.x

17 Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Stat Med* 2011;**30**:1366–80. doi:10.1002/sim.4205

18 Andrinopoulou E-R, Rizopoulos D. Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Stat Med* 2016;**35**:4813–23. doi:10.1002/sim.7027

19 Asar Ö, Ritchie J, Kalra PA, *et al.* Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *Int J Epidemiol* 2015;**44**:334–44. doi:10.1093/ije/dyu262

20 Gera M, Slezak JM, Rule AD, *et al.* Assessment of changes in kidney allograft function using creatinine-based estimates of glomerular filtration rate. *Am J Transplant* 2007;**7**:880–7. doi:10.1111/j.1600-6143.2006.01690.x

21 Gill JS, Tonelli M, Mix CH, *et al.* The change in allograft function among long-term kidney transplant recipients. *J Am Soc Nephrol* 2003;**14**:1636–42.http://www.ncbi.nlm.nih.gov/pubmed/12761266

22 Park WD, Larson TS, Griffin MD, *et al.* Identification and characterization of kidney transplants with good glomerular filtration rate at 1 year but subsequent progressive loss of renal function. *Transplantation* 2012;**94**:931–9. doi:10.1097/TP.0b013e3182692b70

23 Kaplan B, Schold J, Meier-Kriesche H-U, *et al.* Poor predictive value of serum creatinine for renal allograft loss. *Am J Transplant* 2003;**3**:1560–5. doi:10.1046/j.1600-6135.2003.00275.x

24 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**:533–9. doi:10.1097/TP.0000000000000377

25 Clayton PA, Lim WH, Wong G, *et al.* Relationship between eGFR Decline and Hard Outcomes after Kidney Transplants. *J Am Soc Nephrol* 2016;**27**:3440–6. doi:10.1681/ASN.2015050524

26 Fournier M-CC, 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**:469–79. doi:10.1007/s10654-016-0121-2

27 Garre FG, Zwinderman AH, Geskus RB, *et al.* A joint latent class changepoint model to improve the prediction of time to graft failure. *J R Stat Soc Ser A* 2008;**171**:299–308. doi:10.1111/j.1467-985X.2007.00514.x

28 Younespour S, Rahimi Foroushani A, Maraghi E, *et al.* Longitudinal Serum Creatinine Levels in Relation to Graft Loss Following Renal Transplantation: Robust Joint Modeling of Longitudinal Measurements and Survival Time Data. *Nephrourol Mon* 2016;**8**:e39292. doi:10.5812/numonthly.39292

29 Nefrovisie. Nierfunctievervangende behandeling in Nederland. 2017.

30 WHO-ONT. Global Observatory on Donation and Transplantation. 2015.http://www.transplant-observatory.org/data-charts-and-tables/ (accessed 25 Mar 2018).

Table 1. Characteristics of the transplant cohort (n=239)

Donor	**Mean (SD) / N (%)**
Age (years)	49.7 (12.7)
Body mass index (kg/m^2)	25.1 (4.4)
Donor type	
- Living unrelated	38 (15.9%)
- Living related	24 (10.0%)
- Deceased brain death (DBD)	98 (41.0%)
- Deceased cardiac death (DCD)	79 (33.1%)
Transplantation	
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%)
Body mass index (kg/m^2)	25.4 (4.3)
Previous transplantation	37 (15.5%)
Dialysis vintage (years)	3.7 (3.5)
Diabetes	38 (15.9%)
Cardiovascular events before transplantation	91 (38.1%)
Number of anti-hypertensives	
- 0	33 (13.8%)
- 1	79 (33.1%)
- 2	85 (35.6%)
≥ 3	42 (17.7%)

SD, standard deviation; HLA, human leukocyte antigen.

Table 2. Area under curve characteristics for the joint models 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	0.64	0.825	0.473

SCr, serum creatinine; PCR, urinary protein-creatinine ratio.

Figure Legends

Figure 1: Death-censored graft failure and 95% confidence intervals.
The curve and confidence bands were calculated with a Kaplan-Meier analysis.

Figure 2: Fitted longitudinal evolution of serum creatinine and urinary protein-creatinine ratio measurements.
Fitted longitudinal evolution of SCr **(A)** and PCR **(B)** with 95% confidence intervals corresponding to a female recipient aged 51 years, BMI 25, first transplantation, no diabetes, no history of any cardiovascular events, of a living female donor aged 50 years, with 3 mismatches on HLA A, B and DR, with 15 hours of cold ischemia time, 5% panel reactive antibodies, and a 4 year history of dialysis.

Figure 3: Dynamic predictions based on a joint model in an example patient.
Dynamic prediction of death-censored graft survival probabilities of one an example patient, **(A)** using log SCr values up to 5 years and **(B)** using all available log SCr values.

Figure 4: Simulation results comparing a fixed versus a personalized screening approach based on serum creatinine measurements.
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.