Department of Biostatistics
Erasmus University Medical Center

Consultancy Report

Title	Personalized screening for kidney transplant data set
Version	1
Date	12 January 2017
Author	Anirudh Tomer, Dimitris Rizopoulos, Ewout Steyerberg
Collaborators	Jesper Kers, Hessel Peters Sengers

Visiting address: Wytemaweg 80, 3015 CN Rotterdam, Na building (room Na-2418)
Postal address: PO Box 2040, 3000 CA Rotterdam, the Netherlands
Web: http://www.erasmusmc.nl/biostatistiek/

1 Introduction

In this report we present a screening methodology for aperiodically checking the state of kidney transplant grafts in patients. The current standard visiting schedule is same for all patients. The visiting schedule protocol advises checking the serum creatinine and protein-creatinine ratio (pcr) 20 times in the first year of induction in the study and 4 times per year after that. The visits are equally spaced. It is in the interest of both patients and doctors to offer a more personalized care for the patients. A personalized visiting schedule can offer early detection for more severe cases by scheduling more frequent visits. For less severe patients, a personalized schedule offers less financial burden by scheduling infrequent visits. In this report we contrast the fixed schedule with personalized schedules based on the approach proposed by [Rizopoulos et al., 2015]. For creating a personalized schedule for the patients we used serum creatinine as a marker for death censored graft failure. The choice of serum creatinine over pcr was based on the joint analysis of death censored graft failure, serum creatinine and pcr outcomes [Rizopoulos, 2012], where we found serum creatinine to be a stronger marker for death censored graft failure.

For comparing the fixed schedule with personalized schedule, we first fitted a joint model to the Amsterdam kidney transplant data set. The data set contains information about 239 patients who received kidney transplants. Each patient was uniquely identified by their study number('amctx'), which also had a one to one correspondence with patient number ('zis'). While some information such as donor gender, receiver gender, blood pressure etc. were measured only once, serum creatinine and urinary protein-creatinine ratio(pcr) were measured repeatedly over time till the patient either had a transplant failure (death of patient/graft failure), or was not followed up anymore. Using the parameters from the joint model fitted to this data set, we generate a training data set and a test data set. We then train a new model with the training data set and use it to create a personalized schedule for the patients in the test data set. This personalized schedule is then compared with the fixed schedule.

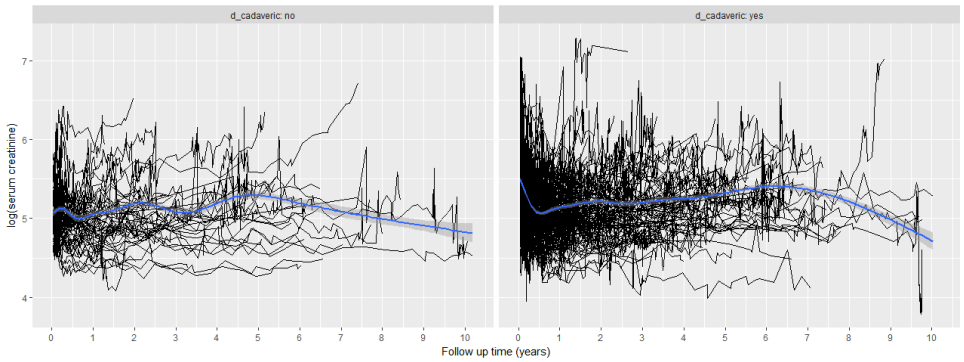
2 Joint analysis of serum creatinine and death censored graft failure

Serum creatinine is an indicator of renal function. Thus it's association with the survival outcome "death censored graft failure" in kidney transplant patients is of interest. Since serum creatinine is an internal covariate, a joint model for time to event and longitudinal data [Rizopoulos, 2012] is most relevant to capture the association between creatinine levels and death censored graft failure. In this section we present the details of the joint model fitted to the Amsterdam kidney transplant data set.

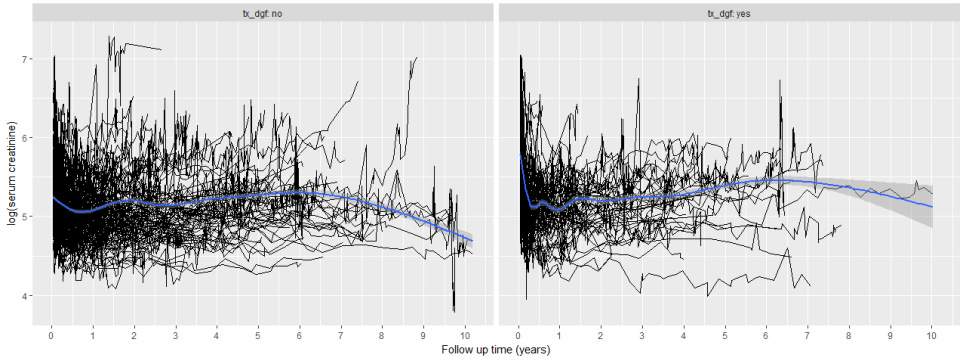
In the data set, for certain patients, there were multiple unequal creatinine measurements at the same time point. Based on the suggestions from AMC, we kept the mean of such measurements. For using the covariate receiver age, we used its value at the first follow up rather than using it as a time varying covariate. This was done because we had another time varying covariate in the model called 'tx_s_days' (days from transplantation until the measurement). We modeled nonlinear evolution of log (creatinine) over time (years from transplantation until the measurement) while controlling for the other covariates present in the data set. For the non linear evolutions we used splines with knots based on graphical analysis of the measurements. The choice of log transforming serum creatinine levels was based on graphical analysis of residuals as well as of the trend plots of the response (Figure 1). It is clear from the observed trend plots that the evolution of serum creatinine depends on both donor deceased status and whether the graft function was delayed or not after transplantation. This was modeled by allowing interaction of the corresponding covariates with time. We now present the covariates for longitudinal submodel.

2.1 Covariates for longitudinal model of serum creatinine

The longitudinal submodel we used for serum creatinine consisted of the following covariates. Interaction between d_cadaveric and tx_s_years, and interaction between tx_dgf and tx_s_years was also modeled. The model also had a random intercept and a spline random effect part with knots at 30 and 70 days.



(a) Observed evolution of log(creatinine) stratified by d_cadaveric (donor deceased or not)



(b) Observed evolution of log(creatinine) stratified by tx_dgf (Delayed graft function after transplantation)

Figure 1: Observed evolution of log(creatinine) for a randomly selected set of patients. X axis is years from transplantation until the measurement.

- rec_age: Receiver age at the first follow up
- rec_gender: Receiver gender
- d_age: Donor age
- tx_dgf: Delayed graft function after transplantation (need for dialysis withing first week after transplantation)
- d_cadaveric: Panel reactive antibody percentage before transplantation
- tx_dm: Diabetes mellitus
- tx_pra: Panel reactive antibody percentage before transplantation
- ah_nr: Number of anti-hypertensive medicaments (at 3 months after transplantation)
- tx_s_years: Natural cubic spline with knots at 30, 70 and 1000 days. In the model we scaled days to years for computational reasons. The knots were chosen after discussion with the medical team. More specifically the choice of knots at 30 days and 70 was based on the reason that these are the time points up to which one sees drop in serum creatinine levels for patients whose donor was deceased.

2.2 Survival submodel for death-censored graft failure

Out of the 238 subjects considered for analysis, only 44 subjects had a failure and rest were censored. Given the large number of censored observations we had a low statistical power to detect significant effects of covariates in a large extensive model. Based on the suggestion of the team at AMC and supplementing it with our own analysis we took the additive effect of the following covariates in our model:

- d_age: Donor age.
- rec_age: Receiver age at the first follow up.
- tx_previoustx: Previous transplantation before the current transplantation.
- d_gender: Donor gender
- rec_bmi: Receiver BMI.
- d_type: Donor type.
- tx_pra: Panel reactive antibody percentage before transplantation.

2.3 Joint model for time to event analysis of death-censored graft failure

Using the longitudinal and survival submodels described above we fitted a joint model. In the joint model the association between the survival and longitudinal outcomes was considered via the value and slope of log(creatinine). Since we had very few subjects who had events, to estimate the parameters in the survival submodel we used a Bayesian ridge approach. Table 1 shows the parameter estimates for the survival submodel in this joint model. It can be seen that among the baseline covariates only 'rec_bmi' is significant. However we still keep others in the model as they have clinical relevance. The value and slope of log(creatinine) were both found to be significant. To interpret the associations in the survival submodel, let us imagine a patient whose serum creatinine levels at some time point are 150 umol/L. If the patient instead had twice of these serum creatinine levels at that time point, i.e. 300 umol/L, then his hazard of failure had been 1.7 times more (95% CI [1.5, 2.0]). This however is valid only if all other covariate, including the slope of log(creatinine) remained the same for both patients.

	Mean	Std.Dev	2.5%	97.5%	P
rec_age	0.012	0.011	-0.009	0.034	0.256
d_age	0.008	0.011	-0.013	0.031	0.458
tx_previoustx: yes	0.694	0.516	-0.054	1.645	0.202
d_gender: Male	0.174	0.276	-0.063	0.896	0.518
rec_bmi	0.076	0.039	0.002	0.147	0.048
tx_pra	0.007	0.007	-0.007	0.021	0.328
I(tx_dial_days/365)	-0.009	0.026	-0.068	0.039	0.744
log(creatinine): Value	1.707	0.255	1.237	2.233	0.000
log(creatinine): Slope	0.211	0.113	0.005	0.436	0.032

Table 1: Parameter estimates for the survival submodel in the joint model with associations for both log(creatinine) and its slope.

Table 2 shows the parameter estimates for longitudinal submodel. As we can see in Table 2, almost all covariates are strongly related with log (serum creatinine). The interpretation for 'tx_dm' is that at any given time point, a patient with diabetes who takes OAM or insulin will have log(creatinine) level lower by 0.113 units if he/she weren't taking OAM or insulin. Interpretation for other categorical and continuous covariates can be done in the same way because the model has only the additive effects of the aforementioned covariates. The fitted evolution of log(creatinine) for a hypothetical kidney transplant recipient can be seen in Figure 2. This hypothetical patient has covariates levels equal to the median values of the covariates for patients in the data set at hand. Because of the interaction of follow up time with 'd_cadaveric' and 'tx_dgf' the evolution is plotted after stratifying for the corresponding covariates.

3 Personalized schedule

The goal of personalized screening is multi-fold: early detection, assessing the health of patients periodically, reducing tangible as well as intangible burden on the patients and having a schedule tailored for the medical condition of every patient. The survival event of interest here is death

	Mean	Std. Dev	2.5%	97.5%	P
Intercept	4.736	0.164	4.419	5.055	<0.000
rec_age: At first follow up	-0.005	0.002	-0.008	-0.002	0.002
rec_gender: Male	0.196	0.041	0.120	0.276	<0.000
d_age	0.006	0.002	0.003	0.009	0.002
tx_pra	0.002	0.001	-4.809 $\times 10^{-4}$	0.005	0.096
ah_nr	0.036	0.020	-0.005	0.075	0.076
tx_dm: Yes	-0.113	0.059	-0.223	0.001	0.058
Spline (0 to 30 days)	-0.060	0.123	-0.289	0.186	0.612
Spline (30 to 70 days)	0.220	0.109	-0.013	0.452	0.060
Spline (70 to 1000 days)	0.296	0.277	-0.246	0.865	0.290
Spline (1000 days to 2190 days)	0.282	0.290	-0.290	0.847	0.330
d_cadaveric: Yes	0.852	0.150	0.549	1.123	<0.000
tx_dgf: Yes	0.686	0.139	0.421	0.955	<0.000
Spline (0 to 30 days) × d_cadaveric (Yes)	-0.680	0.152	-0.960	-0.376	<0.000
Spline (30 to 70 days) × d_cadaveric (Yes)	-0.464	0.138	-0.729	-0.173	<0.000
Spline (70 to 1000 days) × d_cadaveric (Yes)	-1.195	0.349	-1.894	-0.539	0.002
Spline (1000 to 2190 days) × d_cadaveric (Yes)	-0.124	0.372	-0.861	0.584	0.750
Spline (0 to 30 days) × tx_dgf (Yes)	-0.703	0.142	-0.980	-0.425	<0.000
Spline (30 to 70 days) × tx_dgf (Yes)	-0.505	0.136	-0.769	-0.238	<0.000
Spline (70 to 1000 days) × tx_dgf (Yes)	-1.373	0.339	-2.024	-0.684	<0.000
Spline (1000 to 2190 days) × tx_dgf (Yes)	-0.776	0.366	-1.525	-0.065	0.028
σ	0.201	0.001	0.199	0.204	<0.000

Table 2: Parameter estimates for the longitudinal submodel. The outcome here is log(creatinine).

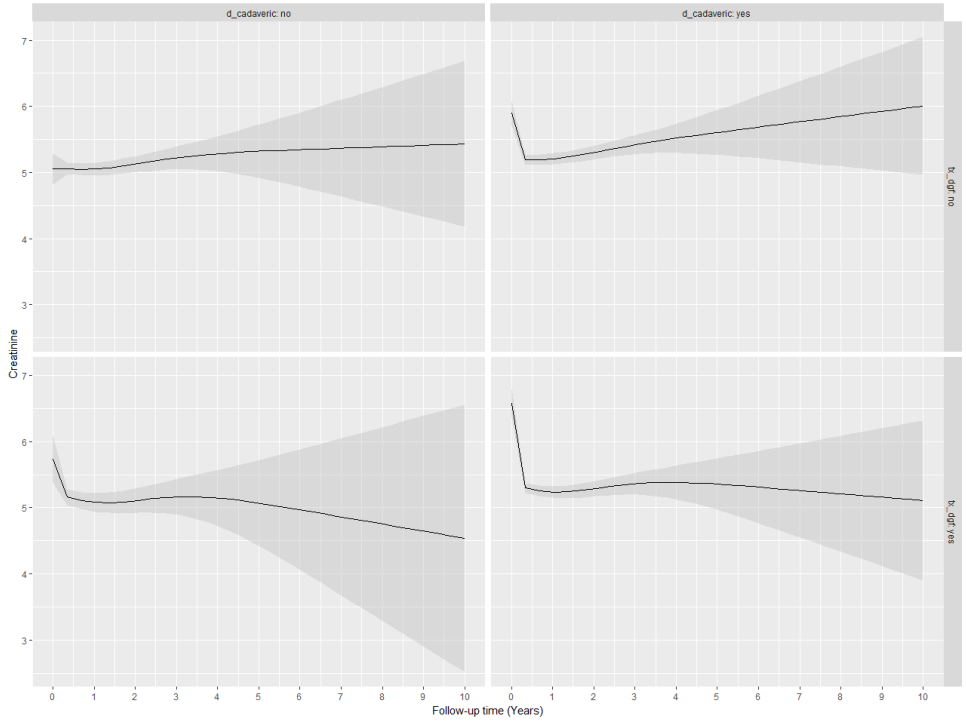


Figure 2: Fitted longitudinal profile of a patient with rec_gender = Male, rec_age = 53 years, d_age=52 years, d_bmi = 24, ah_nr = 2, tx_pra = 0, tx_dm = No

censored graft failure. Based on expert advice, we decided that the first time the 6 months dynamic survival probability falls below 87.5%, the personalized screening algorithm will stop and the patient will be taken out of the study. Let t_j^{s*} denote this time. The dynamic survival probability for patient j at time $u > t$, given that the patient has survived till time t and has given serum

creatinine measurements $\mathcal{Y}_j(t)$ is given by:

$\pi(u|t) = Pr(T_j^* \geq u | T_j^* > t, \mathcal{Y}_j(t), D_n; \theta^*)$, where θ^* denotes the true parameter values.

Borrowing the notation from [Rizopoulos et al., 2015], let t denote the time up to which we know that the patient j has not had the event of interest. The next visit of the patient is to be scheduled at a future time $u > t$. However we would not want to exceed time t_j^{s*}. Thus if $\pi(u+6|t) \leq 0.875$ then we take the patient out of the study. The ideal time $u < t_j^{s*}$ is the one where we get to know the most about the failure time distribution of the patient in question. Based on these goals we propose choosing a time u which maximizes the following utility function.

$$U(u|t) = \lambda_1 E\Big\{ \log \frac{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n)}{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), D_n)} \Big\} - \lambda_2 I(u > t_j^{s*}) \tag{1}$$

where the expectation is taken w.r.t. the joint predictive distribution $p(T_j^*, y_j(u) | T_j^* > t, \mathcal{Y}_j(t), D_n)$. λ_1 and λ_2 are two constants used as weights for the two components of the utility function. An interesting aspect of this utility function is that the first part in equation 1 is an analogue of Kullback Leibler divergence between two density functions of time to death censored graft failure; one conditional on survival upto time t as well as historical record of serum creatinine $\mathcal{Y}_j(t)$ upto time t, while the second density is condition on the same information supplemented with new serum creatinine measurement $y_j(u)$ at time u. i.e. It is estimate of amount of information gained solely by taking a new serum creatinine measurement $y_j(u)$ at time u. The second part of the utility function is the penalty term. The choice of constants λ_1 and λ_2 is not clear. This can however be tackled by optimizing the first part in equation 1 under the constraint $u \leq t_j^s$, as shown by [Rizopoulos et al., 2015].

An interesting modification of the utility function in equation 1 is given in equation 2. The first part in this utility function corresponds to entropy or expected information at time u. The information comes from two sources, one by knowing that the patient survived till time u and secondly from the serum creatinine measurement at time u.

$$U(u|t) = \lambda_1 E\big\{ \log p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n) \big\} - \lambda_2 I(u > t_j^{s*}) \tag{2}$$

As the difference u and t increases the amount of uncertainty in the measurement of $y_j(u)$ also increases. This increases the amount of uncertainty in the information available at time u. i.e. Even though the entropy or KL divergence at time point $u+5$ (say) may be more than u but the uncertainty in the information is also higher. This can be tackled by considering standardized versions of these quantities. Following two utility functions illustrate this.

$$U(u|t) = \lambda_1 \frac{E\Big\{ \log \frac{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n)}{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), D_n)} \Big\}}{S.D.\Big\{ \log \frac{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n)}{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), D_n)} \Big\}} - \lambda_2 I(u > t_j^{s*}) \tag{3}$$

$$U(u|t) = \lambda_1 \frac{E\big\{ \log p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n) \big\}}{S.D.\big\{ \log p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n) \big\}} - \lambda_2 I(u > t_j^{s*}) \tag{4}$$

In all of the utility functions mentioned so far, the value of t_j^{s*} in the penalty part is also not known in advance. We propose using $\underset{t_j^s \epsilon (t,\infty)}{\arg} Pr(T_j^* \geq t_j^s + 6 | T_j^* > t_j^s, \mathcal{Y}_j(t), D_n; \theta^*) = 0.875$. An alternate idea is to use a penalty $\lambda_2 I(u > t_j^{0.875})$ where, $\underset{t_j^{0.875} \epsilon (t,\infty)}{\arg} \pi(t_j^{0.875} | t) = 0.875$. i.e. $t_j^{0.875}$ is the time at which survival probability is 87.5%, conditional on the fact that the patient has survived till time t.

3.1 Checking efficacy of the personalized screening methods

To check the potential of personalized schedules in contrast with fixed schedules we check the following:

- Number of visits each method takes to get as close to t_j^{s*} as possible.

- The difference between t_j^{s*} and the observed first time at which the 6 months dynamic survival probability falls below 87.5%.

Using the parameters from the joint model fitted to Amsterdam kidney transplant data set we generated longitudinal profiles and failure times for 611 patients. We also generated 13 patients with only longitudinal profiles. First we fitted a joint model with the 611 patients from the training data set and then we used the parameters from that model to create a personalized schedule for the 13 patients with only longitudinal profiles/serum creatinine measurements. The median and mean failure time for the patients in training data set were 3.7 years and 3.6 years respectively, while the first and third quartile were at 2 and 5 years respectively. It is to be noted that in the Amsterdam kidney transplant data set, the median and mean failure times were 3.6 and 3.9 years. The first and third quartile were at 1.7 and 5.8 years respectively.

Since we simulate the profiles of the patients we know the true random effect b_j for them. To compute t_j^{s*} for these patients we propose the following approaches.

1. $\underset{t_j^s \epsilon (0,\infty)}{\arg} Pr(T_j^* \geq t_j^s + 6 | T_j^* > t_j^s, b_j, D_n; \theta^*) = 0.875$
2. $\underset{t_j^s \epsilon (0,\infty)}{\arg} Pr(T_j^* \geq t_j^s + 6 | T_j^* > t_j^s, M_j(t), D_n; \theta^*) = 0.875$, where $M_j(t)$ corresponds the true longitudinal profile of the patient; i.e. without measurement error.

In the second approach the random effects are estimated from the true longitudinal profile observed up to time t. This is an extra source of variation which does not exist in first approach. When we generate visiting times using personalized screening approach, we estimate random effects from the observed data while comparing the stopping time from personalized or fixed schedule against t_j^{s*}. Here we have three sources of variation: one from the measurement error in observed profile, second from the random effects estimated from the observed data, third from the personalized screening/fixed screening method itself. Thus comparing results from personalized/fixed screening to t_j^{s*} computed from the second approach allows isolating out the true variation in results due to the different scheduling methods. Indeed there is little variation due to measurement error however we found that it does not impact the results a lot. Thus in this report we calculate t_j^{s*} using the second approach. To facilitate the computation for t_j^{s*}, the entire time scale is discretized on a fine scale.

3.2 Personalized vs. Fixed schedule

As mentioned earlier the fixed schedule mandates measurement of serum creatinine 20 times in the first year of induction in the study and 4 times per year after that. For the personalized schedule we considered the following 3 utility functions.

$$U_1(u|t) = \lambda_1 E\left\{ \log \frac{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n)}{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), D_n)} \right\} - \lambda_2 I(u > t_j^{0.875}) \tag{5}$$

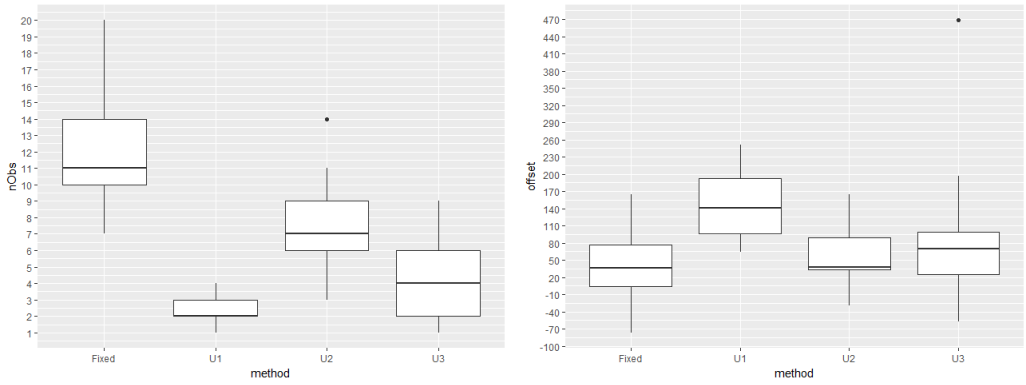
$$U_2(u|t) = \lambda_1 \frac{E\left\{ \log \frac{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n)}{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), D_n)} \right\}}{S.D.\left\{ \log \frac{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n)}{p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), D_n)} \right\}} - \lambda_2 I(u > t_j^{s*}) \tag{6}$$

$$U_3(u|t) = \lambda_1 \frac{E\{ \log p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n) \}}{S.D.\{ \log p(T_j^* | T_j^* > u, \mathcal{Y}_j(t), y_j(u), D_n) \}} - \lambda_2 I(u > t_j^{s*}) \tag{7}$$

The other utility functions did not perform well enough and so for brevity we avoided presenting them here. To be robust against outliers we used median and median absolute deviation instead of mean and standard deviation, while computing the first part in the utility functions U_1, U_2 and U_3. The personalized schedule requires a certain number of measurements in the longitudinal history, to make accurate predictions. We chose to measure patients according to the fixed schedule for the first 15 visits. The 15th visit corresponded to a period of 9 months. The next visit at a new

time u is scheduled at the time which maximizes the utility functions.

Figure 3a shows the boxplot of the number of visits taken by each of the scheduling methods before the 6 months survival probability goes below 87.5%. Figure 3b shows the boxplot of the offset for each of the methods. The offset is equal to the difference between the visit time at which the 6 months dynamic survival probability falls below 87.5% and t_j^{s*}. It seems clear that methods which schedule more visits tend to have a lower offset. Personalized scheduling method based on utility function U_1 takes the least number of visits with the median being 2 visits, however it always overshoots the true stopping time with the median overshooting time being 4.7 months. In contrast the other 3 methods also undershoot the target time t_j^{s*}. i.e. the offsets are negative. To make a comparison within the 3 remaining methods we check the absolute offset for each of them. Fixed schedule has a median absolute offset of 40 days and median number of visits after first 15 compulsory visits are 11. The personalized schedule based on utility function U_2 has a median absolute offset of 38 days and median number of visits are 7. Thus in terms of offset it performs as worse as the fixed schedule however takes 4 less visits in comparison. Personalized schedule based on utility function U_3 has a median absolute offset of 70 days and median number of visits are 4. Interestingly in majority of the 13 test cases, the personalized screening technique based on equation utility function U_2 takes first a big leap to a time closer to the true stopping time and then schedules visits very frequently. On the contrary the fixed schedule also has visits in the time period which is not closer to t_j^{s*}.



(a) nObs = Total number of visits before stopping - 15

(b) Offset = stopping time based on scheduling approach - t_j^{s*}

Figure 3: Number of observations and offset for each of the scheduling methods.

Although the results we presented are not based on a full fledged simulation study, these results do indicate that there is potential in personalized methods to replace fixed schedules. More importantly, personalized schedules offer a diverse range of options in terms of their accuracy of being closer to the true stopping time, or scheduling less visits which is more comfortable for patients.

References

[Rizopoulos, 2012] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC Press.

[Rizopoulos et al., 2015] Rizopoulos, D., Taylor, J. M., Van Rosmalen, J., Steyerberg, E. W., and Takkenberg, J. J. (2015). Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics*, page kxv031.