A Two-Stage Joint Model for Nonlinear Longitudinal Response and a Time-to-Event with Application in Transplantation Studies

Magdalena Murawska [1,*], Dimitris Rizopoulos[1], and Emmanuel Lesaffre[1,2]

November 7, 2011

[1] *Department of Biostatistics, Erasmus University Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands*

[2] *I-Biostat, Catholic University of Leuven, Belgium*

*e-mail: m.murawska@erasmusmc.nl

Abstract

In transplantation studies often longitudinal measurements are collected for important markers prior to the actual transplantation. Using only the last available measurement as a baseline covariate in a survival model for the time to graft failure discards the whole longitudinal evolution. We propose a two-stage approach to handle this type of data sets using all available information. At the first stage we summarize the longitudinal information with nonlinear mixed-effects model, and at the second stage we include the Empirical Bayes estimates of the subject-specific parameters as predictors in the Cox model for the time to allograft failure. To take into account that the estimated subject-specific parameters are included in the model, we use a Monte Carlo approach and sample from the posterior distribution of the random effects given the

observed data. Our proposal is exemplified on a study of the impact of renal resistance evolution on the graft survival.

1. Introduction

Many medical studies involve analyzing responses together with event history data collected for each patient. A well-known and broadly studied example can be found in AIDS research, where CD4 cell counts taken at different time points are related to the time to death. These data need to be analyzed using a joint modeling approach in order to properly take into account the association between the longitudinal data and the event times. The requirement for a joint modeling approach is twofold. Namely, when focus is on the longitudinal outcome, events cause nonrandom dropout that needs to be accounted for in order to obtain valid inferences. When focus is on the event times, the longitudinal responses cannot be simply included in a relative risk model because they represent the output of an internal time-dependent covariate[1].

In this paper, we focus on a setting that shares some similarities with the standard joint modeling framework described above, but also has important differences. In particular, we are interested in the association between longitudinal responses taken before the actual follow-up for the time-to-event has been initiated. This setting is frequently encountered in transplantation studies, where patients in the waiting list provide a series of longitudinal outcomes that may be related to events occurring after transplantation. A standard analysis in transplantation studies is to ignore the longitudinal information and use only the last available measurement as a baseline covariate in a model for the allograft survival. It is however evident that such an approach discards valuable information. An alternative straightforward approach is to put all longitudinal measurements as covariates in the survival

model. Nevertheless, there are several disadvantages with this approach. First, it would require spending many additional degrees of freedom, one for each of the longitudinal measurements. Second, patients with at least one missing longitudinal response need to be discarded, resulting in a great loss of efficiency. Finally, we may encounter multicollinearity problems due to the possibly high correlation between the longitudinal measurements at different time points.

Nowadays, when it comes to measuring the association between a longitudinal marker and an event-time outcome, a standard approach is to use the joint model postulated by Faucett and Thomas [2] and Wulfsohn and Tsiatis [3]. In this setting the longitudinal responses are considered realizations of an endogenous time-dependent covariate (Kabfleish and Prentice [1]), which is measured with error and for which we do not have the complete history of past values available. To account for these features we estimate in the joint modeling framework the joint distribution of the survival and longitudinal processes. Unlike in the multivariate approach, where we have to interpret the estimates for each longitudinal measurement separately, the joint modeling approach allows to get more insight in the nature of the relation between the two analyzed processes since it assumes some underlying process for the longitudinal measures.

However in contrast with the standard joint modeling setting, in our case (i.e., transplantation studies) the longitudinal responses do not constitute an endogenous time dependent variable measured at the same period as the time to event. In particular, since the longitudinal measurements are collected prior to transplantation, occurrence of an event (i.e. graft failure after transplantation) does not cause nonrandom dropout in the longitudinal outcome. Nevertheless, the problem of measurement error still remains. Ignoring the measurement error affects not only the standard errors of the estimates of interest but also it can causes attenuation of the coefficients towards zero [4]. To overcome this we propose a two-stage modeling approach that appropriately summarizes the longitudinal information before the

start of follow-up by means of a mixed effects model and then uses this information to model the time to event by including the Empirical Bayes estimates of the subject specific parameters as predictors in the Cox model. To account for the fact that we include the estimates and not the true values of the parameters, we use a Monte Carlo approach and sample from the posterior distribution of the random effects. The proposed method does not require joint maximization neither fitting the random effects model for each event time as in the two-stage approach of Tsiatis, DeGrutola and Wulfsohn [5]. We compare this approach with the "naive" one when the uncertainty about the estimates from the first step is not taken into account, as well as with the full Bayesian approach. Our approach shares similarities with the two-stage approach of Albert and Shih [6]. They considered a model, in which a discrete event time distribution is modeled as a linear function of the random slope of the longitudinal process estimated from the linear mixed model. The bias from informative dropout was reduced by using the conditional distribution of the longitudinal process given the dropout time to construct the complete data set. To account for the measurement error in the mean of the posterior distribution of the random effects, the variance, that incorporates the error in estimating the fixed effects in the longitudinal model, was used. However we use sampling not to impute missing values and correct for non-random dropout but in order to account for the variability in the predicted longitudinal covariates that are then used in survival model. A method of adjusting for measurement error in covariates, that was used by Albert and Shih, does not apply in our case since it requires the discrete time-to-event and linear model for longitudinal data. The time-to-event could be discretized but still we have a nonlinear model for longitudinal data.

Our research is motivated by data from an international prospective trial on kidney-transplant patients. The study has two arms, where in the first arm donors' kidneys were administered to cold storage, whereas in the second arm they were administered to machine perfusion (MP). The advantage of machine perfusion is

the possibility of measuring different kidney's parameters reflecting the state of the organ. One of the parameters of interest is renal resistance level (RR), which has been measured at 10 minutes, 30 minutes, 1 hour, 2 hours, 4 hours and just before transplantation. Our aim here is to study the association of the renal resistance evolution profile with the risk of graft failure. The time of last measurement was different for different patients and often unknown exactly. However based on the medical consult and visual inspection of the individual profiles the last measurement was chosen to be taken at 6 hours for each patient.

The rest of the paper is organized as follows. Section 2 provides the general modeling framework with the definition of the two submodels for the longitudinal and survival data, respectively. Section 3 describes the estimation methods for the full likelihood and the proposed two-stage approach. In Section 4 we apply the two-stage approach to the renal data. Section 5 contains the setup and the results for the simulation study. Finally, in Section 6 we discuss the proposed methodology.

2.   Joint Modeling Framework

Let $Y_i(u)$ denote the longitudinal profiles for individual i, $i = 1, 2, \ldots, N$. We assume that $Y_i(u)$ are collected for the ith individual prior to the specified time t_i, $u \in (0, t_i)$. Let $t = 0$ denote the time of the first longitudinal measurement and t_i - the time of the last collected measurement. t_i can be different for different individuals and we denote by m_i the number of longitudinal measurements for subject i collected until time t_i and by u_{ij} the time of jth measurement. Denote by $T_i^* \geq t_i$ the true survival time for individual i. Since the survival time is right censored we observe only $T_i = \min(T_i^*, C_i)$, where $C_i \geq t_i$ is the censoring time with the failure indicator Δ_i, which equals to 1 if the failure is observed and 0 otherwise, i.e. $\Delta_i = I(T_i \leq C_i)$ with $I(\cdot)$ denoting the indicator function. We will assume that censoring is independent of all other survival and covariate information. In addition

we assume that the observed longitudinal responses $Y_i(u)$ are measured with error (i.e. biological variation) around the true longitudinal profile $W_i(u)$, i.e.,

$$Y_i(u) = W_i(u) + \varepsilon_i(u), \text{with } \varepsilon_i(u) \sim N(0, \sigma^2) \text{ and } \text{cov}(\varepsilon_i(u), \varepsilon_i(u')) = 0, u' \neq u. \quad (1)$$

We will consider the longitudinal response that exhibit a nonlinear profiles in time. Therefore, we approximate $W_i(u)$ by means of a nonlinear mixed effects model:

$$W_i(u) = f(u; \boldsymbol{\phi}_i), \text{with } \boldsymbol{\phi}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{\alpha}_i, \quad (2)$$

where $f(\cdot)$ is a nonlinear function, parameterized by the vector $\boldsymbol{\phi}_i$. The parameters $\boldsymbol{\phi}_i$ control the shape of the nonlinear function and subscript i denotes that each subject may have its own nonlinear evolution in time in the family $f(\cdot; \boldsymbol{\phi})$. For the subject-specific parameter $\boldsymbol{\phi}_i$ we assume a standard mixed model structure with $\boldsymbol{X}_i$ denoting the fixed effects design matrix with corresponding regression coefficients $\boldsymbol{\beta}$, $\boldsymbol{Z}_i$ the random effects design matrix and $\boldsymbol{\alpha}_i$ the random effects. The random effects $\boldsymbol{\alpha}_i$ are assumed to be independent and normally distributed with mean zero and variance-covariance matrix $\boldsymbol{D}$.

For the event process we postulate the standard relative risk model of the form:

$$\lambda_i(t) = \lambda_0(t) \exp(\boldsymbol{\gamma}^T \boldsymbol{\phi}_i), \quad (3)$$

where $\lambda_i(t)$ is the hazard function, $\lambda_0(t)$ is the baseline hazard, which can be modeled parametrically or left completely unspecified. The subject specific parameters $\boldsymbol{\phi}_i$ summarize the longitudinal evolutions of the response for each subject, and therefore coefficients $\boldsymbol{\gamma}$ measure the strength of the association between the different characteristics of the underlying subject-specific nonlinear evolution of the longitudinal profiles and the risk for an event. Within the formulation of the two submodels (2) and (3) the same random effects now account for both the association between the longitudinal and event outcomes, and the correlation between the

repeated measurements in the longitudinal process.

In the particular transplantation setting that will be analyzed in the following study $Y_i(u)$ are the renal resistance level measurements collected for the ith donor prior to the transplantation time t_i and the same index i is used to denote the allograft transplanted to the ith patient. Time $t = 0$ represents the time that the kidney is removed from the donor and put in cold storage or in a perfusion machine.

3. Estimation

3.1. Full likelihood framework: Bayesian approach

In the standard joint modeling framework the estimation is typically based on maximum likelihood or Bayesian methods (MCMC). This proceeds under the following set of conditional independence assumptions:

$$p(T_i, \Delta_i, \boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}) = p(T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_t)p(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y)$$
$$p(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) = \prod_{j=1}^{m_i} p(Y_i(u_{ij}) \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y). \tag{4}$$

In particular, we assume that given the random effects the longitudinal process is independent from the event times, and moreover, the longitudinal measurements are independent from each other.

Maximum likelihood methods use the joint likelihood and maximize the log-likelihood function $l_i(\boldsymbol{\theta}) = \sum_i \log p(T_i, \Delta_i, \boldsymbol{Y_i}; \boldsymbol{\theta})$. This requires numerical integration and optimization, which makes the fit difficult, especially in high-dimensional random effects settings. Standard options for numerical integration are Gaussian quadrature, Laplace approximation or Monte Carlo sampling ([7],[8]). Maximization of the approximated log-likelihood is based on an EM algorithm ([3], [9], [10], [5], [11]). Several authors proposed a Bayesian approach (MCMC)([2], [12], [13]). Bayesian estimation, that generalizes a joint model for the case with multivariate

longitudinal data, has been discussed by Ibrahim Chen and Sinha [14]. Brown and Ibrahim [15] considered semiparametric model relaxing the distributional assumption for the random effects. In most papers the longitudinal submodel is a linear mixed effectx model. Joint models with nonlinear mixed-effects submodels have been less studied in the literature [16]. Nonlinear mixed models are more common in pharmacokinetics and pharmacodynamics, where they are jointly modeled with non-random dropout using NONMEM software. Several authors considered a Bayesian approach with a nonlinear mixed model and informative missingness ([17], [18]).

Here we will proceed under the Bayesian paradigm to estimate the model parameter. Under the conditional independence assumption (4) the posterior distribution of the parameters and the latent terms, conditional on the observed data, are derived as:

$$
p(\boldsymbol{\theta}, \boldsymbol{\alpha}_i \mid T_i; \Delta_i; \boldsymbol{Y}_i) \propto \prod_{i=1}^{N} \prod_{j=1}^{m_i} \{ p(Y_i(u_{ij}) \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) \} \, p(T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_t)
$$

$$
p(\boldsymbol{\alpha}_i; \boldsymbol{\theta}_\alpha) p(\boldsymbol{\theta}_y, \boldsymbol{\theta}_t, \boldsymbol{\theta}_\alpha),
$$

$$(5)$$

where $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_y^T, \boldsymbol{\theta}_t^T, \boldsymbol{\theta}_\alpha^T)$ is a vector of parameters from the longitudinal and survival models and the vector of the random effects, respectively and $p(\cdot)$ denotes the appropriate probability density function. The likelihood contribution for the ith subject conditionally on the random terms is given by:

$$
p(\boldsymbol{Y_i}, T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}) = p(\boldsymbol{Y_i} \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) p(T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_t)
$$

$$
= \left[\lambda_0(T_i) \exp\{\boldsymbol{\gamma}^T \boldsymbol{\phi}_i(\boldsymbol{\alpha}_i)\} \right]^{\Delta_i} \exp \left[- \int_0^{T_i} \lambda_0(t) \exp\{\boldsymbol{\gamma}^T \boldsymbol{\phi}_i(\boldsymbol{\alpha}_i)\} dt \right]
$$

$$
\frac{1}{(2\pi\sigma^2)^{m_i/2}} \exp \left[- \sum_{j=1}^{m_i} \frac{\{W_i(u_{ij}, \boldsymbol{\alpha}_i) - Y_i(u_{ij})\}^2}{2\sigma^2} \right].
$$

$$(6)$$

The baseline hazard can be assumed of a specific parametric form, e.g. the Weibull hazard. For the priors of the model parameters we make standard assumptions following Ibrahim et al. [14]. In particular, for the regression coefficients $\boldsymbol{\beta}$ of the longitudinal submodel and for the coefficients $\boldsymbol{\gamma}$ of survival submodel we used multivariate normal priors. For variance-covariance matrices we assumed an inverse Wishart distribution and for the variance-covariance parameters we took as a prior an inverse-gamma. For all parameters the vague priors have been chosen.

The implementation of the Cox and piecewise constant hazard models is typically based on the counting process notation introduced by Andersen and Gill [19] and formulated by Clayton [20]. In particular we treat the counting process increments $dN_i(t)$ in the time interval $[t, t + \Delta t]$ as independent Poisson random variables with means $\Lambda_i(t)dt$:

$$\Lambda_i(t)dt = \omega_i(t) \exp(\boldsymbol{\gamma}^T \boldsymbol{\phi}_i) d\Lambda_0(t), \tag{7}$$

where $\omega_i(t)$ is an observed process taking the value 1 if subject i is observed at time t and 0 otherwise, $d\Lambda_0(t)$ is the increment in the integrated baseline hazard function occurring during the time interval $[t, t + \Delta t]$. Since the conjugate prior for the Poisson mean is the gamma distribution, we assume the conjugate independent increments prior suggested by Kalbfleisch [21], namely:

$$d\Lambda_0(t) \sim \text{Gamma}(c * d\Lambda_0^*(t), c), \tag{8}$$

where $d\Lambda_0^*(t)$ is a prior mean hazard with c being a scaling parameter representing the "strength" of our prior beliefs. The gamma prior was also chosen for the baseline risk parameter of the Weibull distribution in parametric survival model. Alternatively to implement the Cox model in a fully Bayesian approach one may use the "multinomial-Poisson trick" described in the OpenBUGS manual that is equivalent to assuming independent increments in the cumulative hazard function. The increments are treated as failure times and noninformative priors are given for their

logarithms. Analogically to the Cox model a piecewise constant hazard model was implemented. We have modeled baseline hazard using a step function with 3 quantiles t_1, t_2 and t_3 as changing points assuring the same number of events in-between. Let t_0 denote the start of the follow up, t_4 the maximum censoring time and $d\Lambda_{0k}(t)$ the increment in the integrated baseline hazard function occurring during the time interval $[t_k, t_{k+1}], k = 0, 1, 2, 3$. Then for different intervals we specify a separate prior hazard mean $d\Lambda_0^*(t)$ and:

$$d\Lambda_{0k}(t) \sim \text{Gamma}(c * d\Lambda_{0k}^*(t), c). \qquad (9)$$

Similarly as for the Cox model the results were not sensitive with respect to the choice of the hyperparameters as long as the priors were sufficiently diffuse. The above nonparametric approach can be criticized as having the independent priors for the hazard distribution. However as suggested by Kalbfleisch [21] a mixture of gammma priors can be considered as an alternative. Moreover in a piecewise constant model one can also include change points as unknown parameters in the model as proposed in a Bayesian context by Patra and Dey [22] and applied by Cassellas [23].

In order to assess convergence for the full Bayesian model standard MCMC diagnostic plots were used. The burn-in size was set to 10000 iterations, which was chosen based on the visual inspection of the trace plots, and confirmed by the the Raftery and Lewis diagnostics. The same number of iterations were used for constructing the summary statistics. Based on the autocorrelation plots we have chosen every 30th iteration. Therefore in total to obtain 10000 iterations for the final inference 300000 iterations were required after the burn-in part. Additionally we run a second parallel chain and used Gelman and Rubin diagnostic plots to asses the convergence.

3.2. Two-Stage approach

As mentioned in Section 1, the longitudinal measurements in our setting do not constitute an internal time-dependent covariate, since the events took place after the last longitudinal measurement was collected. In particular, since events do not cause nonrandom dropout, the event process does not carry any information for the longitudinal outcome. Mathematically this means that information for the random effects $\boldsymbol{\alpha_i}$ is actually only coming from the longitudinal responses, that is:

$$p(\boldsymbol{\alpha}_i \mid Y_i(u_{ij}); T_i; \Delta_i; \boldsymbol{\theta}_y) = p(\boldsymbol{\alpha}_i \mid Y_i(u_{ij}); \boldsymbol{\theta}_y) \tag{10}$$

Thus, we can avoid the computational complexity of the full likelihood framework presented in Section 3.1 by employing a two-stage approach. More specifically: At Stage I: we obtain $\hat{\boldsymbol{\theta}}_y$ by maximizing the log-likelihood:

$$l_y(\boldsymbol{\theta}_y) = \sum_{i=1}^{N} \int p(\boldsymbol{Y_i} \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) p(\boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) d\boldsymbol{\alpha}_i$$

This requires numerical integration and we use a Gaussian quadrature for that purpose. Then we obtain the corresponding empirical Bayes estimates:

$$\hat{\boldsymbol{\alpha}}_i = \arg\max_{\alpha} \left[\log p(Y_i \mid \boldsymbol{\alpha}; \hat{\boldsymbol{\theta}}_y) + \log p(\boldsymbol{\alpha}; \hat{\boldsymbol{\theta}}_y) \right]$$

and compute the predictions:

$$\hat{\boldsymbol{\phi}}_i = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\hat{\boldsymbol{\alpha}}_i.$$

At Stage II we fit the relative risk model:

$$\lambda_i(t) = \lambda_0(t) \exp\left(\boldsymbol{\gamma}^T \hat{\boldsymbol{\phi}}_i\right).$$

However, a potential problem in the above is that $\hat{\boldsymbol{\phi}}_i$ is not the true subject-specific parameters but rather an estimate with a standard error. If we ignore this measurement error, the regression coefficients $\boldsymbol{\gamma}_i$ will be possibly attenuated. To overcome

this problem we propose here a sampling approach to account for the variability in $\hat{\boldsymbol{\phi}}_i$, very close in spirit to the Bayesian approach of Section 3.1. In particular, we use the following sampling scheme:

Step 1: simulate $\boldsymbol{\theta}_y^{(m)} \sim N(\hat{\boldsymbol{\theta}}_y, \widehat{\text{var}}(\hat{\boldsymbol{\theta}}_y))$

Step 2: simulate $\boldsymbol{\alpha}_i^{(m)} \sim \left[\boldsymbol{\alpha}_i \mid \boldsymbol{Y_i}, \boldsymbol{\theta}_y^{(m)} \right]$

Step 3: calculate $\boldsymbol{\phi}_i^{(m)} = \boldsymbol{X}\boldsymbol{\beta}^{(m)} + \boldsymbol{Z}_i\boldsymbol{\alpha}_i^{(m)}$ and fit the relative risk model $\lambda_i(t) = \lambda_0(t)\exp\{\boldsymbol{\gamma}^T\boldsymbol{\phi}_i^{(m)}\}$ from which $\hat{\boldsymbol{\theta}}_{\boldsymbol{t}}^{(\boldsymbol{m})} = \hat{\boldsymbol{\gamma}}^{(m)}$ and $\widehat{\text{var}}(\hat{\boldsymbol{\theta}}_t^{(m)})$ are kept.

Steps 1-3 are repeated $m = 1, \ldots, M$ times.

Step 1 takes into account the variability of the MLEs, and Step 2- the variability of $\boldsymbol{\alpha_i}$. Moreover, because the distribution in Step 2 is not of a standard form, we use a independence Metropolis-Hastings algorithm to sample from it with multivariate t-proposal density centered at an Empirical Bayes estimates $\hat{\boldsymbol{\alpha}}_i$, covariance matrix $\widehat{\text{var}}(\hat{\boldsymbol{\alpha}}_i)$ and df=4. The low number of degrees of freedom was chosen to ensure that the proposal density has heavy tails to provide sufficient coverage of the target density $[\boldsymbol{\alpha}_i \mid \boldsymbol{Y}_i, \boldsymbol{\theta}_y]$. The variance-covariance matrix estimated from the nonlinear mixed model was additionally scaled by some parameter *Scale*. The tuning parameter allows to control the acceptance rate through the range of the proposed distribution. If the range is too narrow, the proposed values will be close to the current ones leading to low rejection rate. On the contrary if the range is too large, the proposed values are far away from the current ones leading to high rejection rate. We chose the acceptance rate to be 0.5 following Carlin [24] that suggests a desirable acceptance rates of Metropolis-Hastings algorithms to be around 1/4 for the dependence (random walk) M-H version and 1/2 for the independent M-H. Roberts et al.[25] recommended to use the acceptance rate close to 1/4 for high dimensions and 1/2 for the models with dimensions 1 or 2. They discussed this issue in the context of the random walk proposal density. The authors showed that if

the target and proposal densities are normal, then the scale of the latter should be tuned so that the acceptance rate is approximately 0.45 in one-dimensional problems and approximately 0.23 as the number of dimensions approaches infinity, with the optimal acceptance rate being around 0.25 in as low as six dimensions. In our case an independence version of Metropolis-Hastings algorithm is applied. The proposal density in the algorithm does not depend on the current point as in the random-walk Metropolis algorithm. Therefore for this sampler to work well, we want to have a proposal density that mimics the target distribution and have the acceptance rate be as high as possible. In order to obtain a desirable acceptance rate one needs to run a sampling algorithm for a number of iterations for a given data set and compute an acceptance rate and then repeat the procedure changing the tuning parameter until the chosen acceptance rate is obtained. Usually a small number of iterations (100-500) is sufficient for the purpose of calibration. More details about the Metropolis-Hastings acceptance-rejection procedure can be found in the supplementary material. A final estimate of $\boldsymbol{\theta}_t$ is obtained using the mean of the estimates from all M iterations:

$$\bar{\bar{\boldsymbol{\theta}}}_t = \sum_{m=1}^{M} \hat{\boldsymbol{\theta}}_t^m \Big/ M. \tag{11}$$

To obtain the SE of $\bar{\bar{\boldsymbol{\theta}}}_t$ we use the variance-covariance matrix $\boldsymbol{V}$:

$$\hat{\boldsymbol{V}} = \hat{\boldsymbol{W}} + (M+1)\hat{\boldsymbol{B}} \Big/ M, \tag{12}$$

where $\hat{\boldsymbol{W}}$ is the average within-iteration variance and $\hat{\boldsymbol{B}}$ is the between-iteration variance, i.e.,

$$\hat{\boldsymbol{W}} = \sum_{m=1}^{M} \hat{\boldsymbol{U}}^m \Big/ M,$$

and

$$\hat{\boldsymbol{B}} = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\boldsymbol{\theta}}_t^m - \bar{\bar{\boldsymbol{\theta}}}_t)(\hat{\boldsymbol{\theta}}_t^m - \bar{\bar{\boldsymbol{\theta}}}_t)^T \tag{13}$$

$\hat{\boldsymbol{U}}^m$ represents a variance-covariance matrix estimated in iteration m for $\hat{\boldsymbol{\gamma}}^m$.

4. Analysis of the RR Data

4.1. Models' specification.

We apply the proposed two-stage procedure and a fully Bayesian approach to the transplantation study introduced in Section 1. The data was taken from an international prospective trial on 337 kidney pairs, that aimed to compare two different types of storage, namely cold storage and machine perfusion (MP). Here we focus on the second arm. Our main outcome of interest is graft survival time censored after 1 year. At the end of the study only 26 graft failures were observed. The renal resistance level (RR) was expected to be an important risk factor for graft failure. It was measured using the perfusion machine at the moment of taking the organ out from a donor ($t = 0$), and thereafter at 10 minutes, 30 minutes, 1 hour, 2 hours, 4 hours and just before transplantation. As mentioned in the Section 1, the time of last measurement was different for different patients and sometimes unknown. However there was a clear asymptote visible from the individual profiles that was reached after about 5 hours by each patient. Based on that behavior and after the medical consult we chose the last measurement to be taken at 6 hours for each patient. Other variables of interest include the age of the donor, donor's region (3 countries considered) and donor's type (heart-beating or non-heart-beating).

The individual profiles of 50 randomly selected kidney donors are presented in Figure 1. This plot confirms the biological expectation that allografts exhibit their highest renal resistance levels just after being extracted from the donor. Thereafter they show a smooth decrease in RR until they reach an asymptote above zero indication that there is no "perfect flow" through the kidney. Furthermore, we observe that the initial RR level, the rate of decrease as well as the final RR level differ from donor to donor. Additional descriptive plots for our data are presented in the supplementary material.

In the first step of our analysis we aim to describe the evolution of the renal
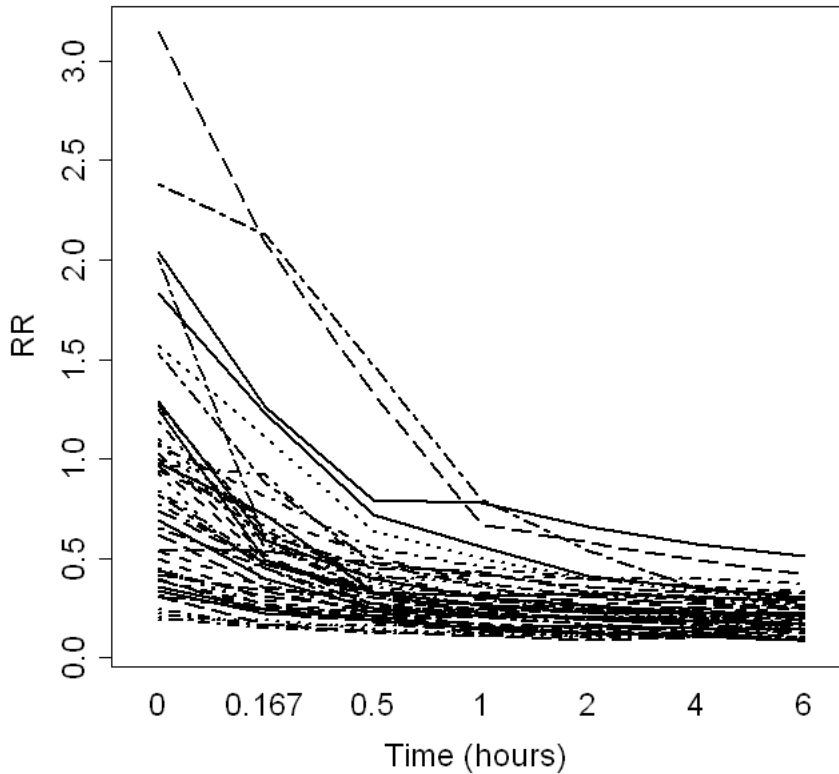
Fig. 1: *Individual profiles of renal resistance level for 50 sampled donors*

resistance level in time. Motivated by both biological expectation and Figure 1 we postulate the following nonlinear function:

$$f(t) = \phi_1 + \phi_2 e^{-\phi_3 t}, \tag{14}$$

where ϕ_1 is a lower asymptote, $\phi_1 + \phi_2$ is an initial value for t=0, and ϕ_3 is the rate of decrease from ϕ_2 to ϕ_1 (see also Figure 2 in Supplementary material).

To accommodate for the shapes of RR evolutions observed in Figure 1, we need to constraint ϕ_1, ϕ_2 and ϕ_3 to be positive. Moreover, in order to allow for individual donor effects, we use the following formulation:

$$Y_i(t) = W_i(t) + \varepsilon(t), \text{with}$$

$$W_i(t) = f_i(t) = \exp(\phi_{1i}) + \exp(\phi_{2i})e^{-\exp(\phi_{3i})t},$$

where

$$\phi_1 = \beta_{10} + \beta_{11} Donor\,Age + \beta_{12} Donor\,Type + \beta_{13} Donor\,Reg1 + \beta_{14} Donor\,Reg2 + \alpha_1$$

$$\phi_2 = \beta_{20} + \beta_{21} Donor\,Age + \beta_{22} Donor\,Type + \beta_{23} Donor\,Reg1 + \beta_{24} Donor\,Reg2 + \alpha_2$$

$$\phi_3 = \beta_{30} + \beta_{31} Donor\,Age + \beta_{32} Donor\,Type + \beta_{33} Donor\,Reg1 + \beta_{34} Donor\,Reg2 + \alpha_3,$$

and $\boldsymbol{\alpha_i} \sim N(0, D)$, $\varepsilon(t) \sim N(0, \sigma^2)$ with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and $\text{cov}(\alpha_i, \varepsilon(t)) = 0$. In the second step the predicted parameters (ϕ_1, ϕ_2, ϕ_3) summarizing the RR evolution of the nonlinear mixed model are included in the graft survival model. The final model for graft survival was of the form:

$$\lambda_i(u) = \lambda_0(u) \exp\left(\gamma_1 \hat{\phi}_{1i} + \gamma_2 \hat{\phi}_{2i} + \gamma_3 \hat{\phi}_{3i}\right).$$

To investigate the impact of ignoring that the covariate $\hat{\boldsymbol{\phi}}_i$ is measured with error, we compared the naive approach in which we ignored this measurement error and our proposal that accounts for the uncertainty in $\hat{\boldsymbol{\phi}_i}$ via Monte Carlo sampling. We used Metropolis-Hastings algorithm with independent t-proposal and acceptance rate around 50% for the reason given in Section 3.2. We simulated $M = 10000$ samples with additional initial step of the scaling parameter calibration in order to achieve the desirable acceptance rate. The final estimates (and SE) of the parameters associated with RR covariates were calculated using the formulas described in the Section 2.3. We compared the results from the two-stage procedure with the estimates obtained from the fully Bayesian joint model fitted for the data using the priors specified in Section 3.1.

The analysis was performed using R Statistical Software. Packages *survival* and *nlme* were used for the two submodels fit and a separate code was written by the

first author for the sampling part. The fully Bayesian model was fitted using Open-BUGS software with the priors specified in Section 3.1. In particular, for the $p \times p$ variance-covariance matrices of multivariate normal priors we used inverse Wishart distribution with p degrees of freedom. For the variance-covariance parameter of the normal longitudinal response we took as a prior an inverse-Gamma(10^{-3},10^{-3}). For the baseline risk parameter of the Weibull distribution in survival submodel a Gamma(10^{-3},10^{-3}) prior was used. To analyze the data using the fully Bayesian Cox model described in Section 3.1 we chose the scaling parameter c in a gamma prior for the independent increments to be equal 0.001 and a prior mean $d\Lambda_0^*(t) = 0.1$. We did not observe any substantial difference for the different values of parameter c as long as c was small enough to keep the prior noninformative. We do not recommend too small values of the scaling parameter c as they can lead to the computation problems. Analogically we have chosen gamma priors for the piecewise constant hazard model. The code for the Bayesian full joint model as well as the R codes for the sampling two-stage procedure are available from the authors on request.

4.2. Results

The results for the nonlinear mixed model are presented in Table 1, for the two-stage approach and in supplementary material, for the full Bayesian approach with Weibull survival model. The results for the longitudinal part for the full Bayesian approach with Cox and piecewise constant hazard models were similar (not presented). It can be observed, based on the the two-stage model results, that only Donor Age had a significant impact on the RR asymptote. Donor Type and Region had a significant impact on the steepness parameter. However we keep all the covariates in the model for the purpose of prediction for the second stage. The mean RR profiles for Heart-Beating and Non-Heart-Beating donors from different regions together with fitted values based on the obtained nonlinear mixed model are given in Supplement.

In the second step of the analysis we applied at first the naive approach and used

Table 1: *Parameter estimates, standard errors and 95 % confidence intervals from the non-linear mixed model for RR*

Effect	Parameter	Estimate	SE	(95%CI)
Fixed effects				
ϕ_1				
Constant	β_{10}	2.838	0.094	(2.654; 3.022)
Donor Age	β_{11}	0.005	0.002	(0.001; 0.009)
Donor Type (HB vs NHB)	β_{12}	-0.102	0.068	(-0.235; 0.031)
Donor Region 1 vs 3	β_{13}	-0.078	0.065	(-0.205; 0.049)
Donor Region 2 vs 3	β_{14}	-0.072	0.072	(-0.213; 0.069)
ϕ_2				
Constant	β_{20}	3.510	0.211	(3.096; 3.924)
Donor Age	β_{21}	0.004	0.004	(-0.004; 0.012)
Donor Type (HB vs NHB)	β_{22}	-0.064	0.154	(-0.365; 0.238)
Donor Region 1 vs 3	β_{23}	-0.107	0.147	(-0.395; 0.181)
Donor Region 2 vs 3	β_{24}	0.033	0.163	(-0.286; 0.352)
ϕ_3				
Constant	β_{30}	1.010	0.186	(0.645; 1.375)
Donor Age	β_{31}	0.003	0.003	(-0.003; 0.009)
Donor Type (HB vs NHB)	β_{32}	0.402	0.130	(0.147; 0.657)
Donor Region 1 vs 3	β_{33}	-0.284	0.125	(-0.529; -0.039)
Donor Region 2 vs 3	β_{34}	-0.032	0.138	(-0.302; 0.238)
Random effects				
$se(\alpha_1)$	d_{11}	0.396		
$se(\alpha_2)$	d_{22}	0.955		
$se(\alpha_3)$	d_{33}	0.572		
$cov(\alpha_1, \alpha_2)$	d_{12}	0.257		
$cov(\alpha_1, \alpha_3)$	d_{13}	-0.053		
$cov(\alpha_2, \alpha_3)$	d_{23}	0.023		
$se(\varepsilon_{ij})$	σ	7.507		

the estimates $\hat{\phi}_1, \hat{\phi}_2$ and $\hat{\phi}_3$ from the nonlinear mixed model as fixed covariates in the final Cox models for graft survival. Table 2 presents the results for the survival submodel for the all approaches, namely the plug-in method, two-stage approach and the fully Bayesian model. For the fully Bayesian approach the results for the parametric Weibull model together with Cox and piecewise constant hazard models are listed. The results from Table 2 indicate that only the RR asymptote could have a significant impact on graft survival.

We observe that the estimates are close or almost identical as in plug-in model. SE of the Cox regression coefficients for the model with sampling are greater than SE from the plug-in model in Table 2 (a), especially for the parameter ϕ_3. The increase in SE is somewhat the expected and is caused by the additional variability in covariates captured by the sampling approach. The fully Bayesian model produces similar results to our semi-Bayesian sampling model with somewhat lower SE. We do not observe substantial difference between fully parametric and nonparametric models in a fully Bayesian approach. Since in the analyzed real data the number of events is small the fully Bayesian Cox and piecewise constant hazard Bayesian models were expected to produce similar results. We did not observe any substantial difference for the different values of hyper parameters.

Even though it is hard to compare exactly the computational time for the two approaches, the rough estimation of the total computational time needed to estimate and assess the convergence (2 chains) of the full Bayesian model was about 21.6 hours and depended on the implemented survival model. A similar computational time was needed for the full Bayesian model with the Cox survival model and piecewise constant hazard model with a slightly more time needed for the parametric Weibull model. For the two-stage approach the total computational time was about 10 hours using the Intel(R) Core(TM)2 Duo T9300 2.5 GHz and 3.5 GB RAM.

Table 2: *Parameter estimates, SE and 95 % confidence/credibility intervals from proportional hazards Cox model for graft survival for plug-in method (a), sampled covariates (b) and fully Bayesian approach (c, d, e)*

(a)

		Graft Survival - Plug-in		
Effect	Parameter	log(HR)	SE	(95%CI)
$\exp(\phi_1)$	γ_1	0.052	0.022	(0.009; 0.095)
$\exp(\phi_2)$	γ_2	-0.005	0.005	(-0.015; 0.005)
$\exp(\phi_3)$	γ_3	0.053	0.158	(-0.257; 0.363)

(b)

		Graft Survival - Sampling Two-Stage		
Effect	Parameter	log(HR)	SE	(95%CI)
$\exp(\phi_1)$	γ_1	0.053	0.024	(0.006; 0.100)
$\exp(\phi_2)$	γ_2	-0.006	0.008	(-0.022; 0.010)
$\exp(\phi_3)$	γ_3	0.055	0.185	(-0.308; 0.418)

(c)

		Graft Survival - Fully Bayesian - Weibull		
Effect	Parameter	log(HR)	SE	(95%HPD)
$\exp(\phi_1)$	γ_1	0.058	0.023	(0.013; 0.103)
$\exp(\phi_2)$	γ_2	-0.005	0.008	(-0.020; 0.011)
$\exp(\phi_3)$	γ_3	0.056	0.180	(-0.299; 0.411)

(d)

		Graft Survival - Fully Bayesian - Cox		
Effect	Parameter	log(HR)	SE	(95%HPD)
$\exp(\phi_1)$	γ_1	0.056	0.023	(0.010; 0.101)
$\exp(\phi_2)$	γ_2	-0.006	0.008	(-0.022; 0.010)
$\exp(\phi_3)$	γ_3	0.055	0.171	(-0.280; 0.390)

(e)

		Graft Survival - Fully Bayesian - Piecewise constant hazard		
Effect	Parameter	log(HR)	SE	(95%HPD)
$\exp(\phi_1)$	γ_1	0.054	0.024	(0.007; 0.102)
$\exp(\phi_2)$	γ_2	-0.005	0.009	(-0.022; 0.012)
$\exp(\phi_3)$	γ_3	0.054	0.179	(-0.297; 0.405)

5. Simulations

5.1. Design

We have conducted a number of simulations to investigate the performance of our proposed two-stage method. In particular, we compared the plug-in method which uses the Empirical Bayes estimates $\hat{\phi}_i$ from the nonlinear mixed model and ignores the measurement error, the two-stage Monte Carlo sampling approach that accounts for the variability in $\hat{\phi}_i$ and the fully Bayesian approach. In the fully Bayesian approach only the parametric Weibull model was considered.

For the longitudinal part the data were simulated for 500 patients from model (15) with $\phi_{1i} = \beta_{10} + \alpha_{1i}$, $\phi_{2i} = \beta_{20} + \alpha_{2i}$ and $\phi_{3i} = \beta_{30} + \alpha_{3i}$, $\boldsymbol{\alpha}_i \sim N(0, \boldsymbol{D})$, $Y \sim N(f(t), \sigma^2)$. The variance-covariance matrix $\boldsymbol{D}$ of the random effects was chosen to be $\boldsymbol{D} = \text{vech}(0.6, 0.01, -0.01, 0.6, 0.01, 0.3)$. We kept 7 measurement points as in the original analyzed data set as well as the nonequal distances between them. The variance of the measurement error σ^2 was chosen to be 0.25, 1 and 25. Survival times were simulated for each patient using the exponential model with the rate parameter equal $\exp(\lambda)$, where λ:

$$\lambda = \gamma_1 \exp(\phi_1) + \gamma_2 \exp(\phi_2) + \gamma_3 \exp(\phi_3).$$

We considered scenarios with fixed coefficients $\gamma_1 = 0.5, \gamma_2 = 0.5$ and $\gamma_3 = -0.2$. The censoring mechanism was simulated independently using an exponential distribution $Exp(\lambda_C)$. Parameter λ_C was changed in order to control proportion of censored observations. Different scenarios with 40% and 20% of censoring were examined . For each simulated data set we fitted four survival models, namely the gold standard model that uses the true simulated values ϕ_i, the plug-in model, the sampling model and fully Bayesian joint model. Neither nonparametric Cox nor piecewise constant hazard model were considered in a fully Bayesian approach since we have simulated the data from the parametric exponential model and wanted to compare the proposed two-stage approach with the "best" fully Bayesian model. All

the prior settings, size of burn-in, number of iterations etc. for the fully Bayesian model were the same as for the real data analysis.

5.2. Results

In Table 3 we present the average results for 200 simulations of different scenarios are presented. The results from our sampling model were very close to the results obtained for the fully Bayesian model with slightly smaller bias and RMSE for the fully Bayesian approach. That was due to the better estimation of random effects variability in fully Bayesian approach. The plug-in method produced the biggest bias that sometimes with combination with the small variability of the estimates around the biased mean resulted in larger RMSE than in sampling approach. As the measurement error or the percentage of censored observations increased, the estimates of survival submodel were more biased with larger RMSE for all approaches. The increase in bias was more severe when the measurement error variance was increased rather than when the percentage of to censoring was higher. This bias was however decreased when the number of repeated measures per individual was increased. This has to do with the amount of information that is available in the data for the estimation of $\hat{\phi}_i$. As it is known from the standard mixed models literature [26], the degree of shrinkage in the subject-specific predicted values is proportional to σ and inversely proportional to n_i and σ_α. To compare the relation between variance of the random effects and variance of the measurement error, one can use intra class correlation (ICC) defined as the proportion of the total variability that is explained by the clustering with a given random effect. ICC was similar for the simulated and the real data for the biggest σ and increased in a simulation data as σ decreased.

Since the calculations for the simulation study were highly computationally intensive we have used the cluster with about 20 nodes with AMD Quad-Core Opteron 835X, 4 x 2GHz and 16GB RAM per node. The analysis for the the 200 simulated

data sets for a single scenario took about 65.5 hours using the Bayesian approach and 31.2 hours using the two-stage approach.

6. Discussion

We have proposed a two-stage method that can be used in a joint analysis of longitudinal and time to event data when the longitudinal data are collected before the start of follow-up for survival and the interest is in estimation of the impact of longitudinal profiles on survival. The modeling strategy is based on specification of two separate submodels for the longitudinal and time to event data. First the longitudinal outcome is modeled using a random effects model. Then the survival outcome is modeled using the Empirical Bayes estimates of the subject specific effects from the first stage. The variability of the estimates from the first stage is properly taken into account using a Monte Carlo approach by sampling from the posterior distribution of the random effects given the data.

As it was demonstrated, ignoring the additional variability of the subject-specific estimates when modeling survival leads to some bias, and in particular, attenuates the regression coefficients towards zero [4]. That was also confirmed by our simulation study. In comparison with the fully Bayesian approach, the proposed partially Bayesian method produced similar results with substantially less number of iterations required. This is due to the fact that sampling was conducted already around the EB estimates and there is no needed for a burn-in part as in a standard fully Bayesian approach. We used 10000 iterations per subject, which was about the size of burn-in needed in the fully Bayesian models. No thinning was used in our approach, based on the visual inspection of the trace plots. Though it is hard compare the fully Bayesian approach and the two-stage approach with respect to the computational time precisely, the rough approximation of the total computational time required for the two-stage approach was about half in comparison with the fully

Table 3: *Bias and Residual Mean Squared Error (RMSE) for the method with true ϕ_i (GS), Empirical Bayes estimates $\hat{\phi}_i$ (Plug-in), sampled ϕ_i and fully Bayesian approach*

		7 time points				
% censoring		20			40	
σ			0.5			
	γ_1	γ_2	γ_3	γ_1	γ_2	γ_3
GS	0.00(0.04)	-0.02(0.03)	0.01(0.03)	-0.01(0.04)	0.02(0.04)	-0.02(0.04)
plug-in	-0.05(0.06)	-0.04(0.05)	0.06(0.07)	-0.08(0.09)	-0.04(0.05)	0.12(0.12)
sampling	-0.04(0.05)	0.03(0.08)	0.02(0.07)	-0.05(0.11)	-0.02(0.06)	0.03(0.10)
Bayesian	-0.03(0.04)	-0.02(0.04)	0.01(0.02)	-0.01(0.04)	-0.02(0.04)	0.02(0.07)
σ			1			
GS	0.04(0.05)	0.04(0.07)	-0.03(0.07)	-0.05(0.09)	-0.04(0.06)	-0.03(0.05)
plug-in	-0.07(0.08)	-0.08(0.09)	0.07(0.09)	-0.10(0.12)	-0.08(0.09)	0.08(0.11)
sampling	-0.07(0.09)	-0.06(0.10)	-0.02(0.11)	-0.05(0.12)	0.05(0.11)	-0.03(0.12)
Bayesian	0.01(0.03)	0.05(0.06)	-0.03(0.07)	0.05(0.06)	0.04(0.06)	-0.04(0.07)
σ			5			
GS	0.04(0.06)	0.05(0.06)	0.04(0.08)	0.05(0.10)	0.01(0.05)	-0.02(0.06)
plug-in	-0.09(0.10)	-0.10(0.11)	0.08(0.11)	-0.20(0.22)	-0.21(0.22)	0.14(0.18)
sampling	0.08(0.13)	0.06(0.12)	-0.05(0.12)	0.07(0.14)	-0.05(0.13)	-0.11(0.18)
Bayesian	0.09(0.10)	0.05(0.09)	-0.09(0.10)	-0.09(0.10)	0.08(0.12)	-0.12(0.18)
		14 time points				
% censoring		20			40	
σ			0.5			
	γ_1	γ_2	γ_3	γ_1	γ_2	γ_3
GS	-0.03(0.03)	0.00(0.02)	-0.02(0.03)	0.02(0.03)	-0.03(0.04)	0.02(0.04)
plug-in	-0.02(0.03)	-0.03(0.04)	0.05(0.07)	-0.02(0.04)	-0.03(0.04)	0.05(0.06)
sampling	0.03(0.04)	0.02(0.06)	0.02(0.07)	0.02(0.04)	0.04(0.05)	0.02(0.08)
Bayesian	-0.03(0.04)	-0.02(0.04)	-0.02(0.04)	0.02(0.04)	0.03(0.04)	-0.05(0.06)
σ			1			
GS	-0.03(0.04)	-0.03(0.04)	-0.01(0.03)	0.00(0.03)	-0.02(0.04)	0.05(0.06)
plug-in	-0.09(0.06)	-0.05(0.06)	0.06(0.07)	-0.02(0.04)	-0.04(0.05)	0.11(0.11)
sampling	0.04(0.08)	0.02(0.08)	-0.02(0.07)	-0.02(0.04)	-0.02(0.08)	0.04(0.09)
Bayesian	-0.03(0.04)	0.04(0.05)	-0.03(0.05)	0.02(0.04)	0.03(0.05)	0.06(0.07)
σ			5			
GS	-0.03(0.04)	-0.03(0.04)	0.01(0.04)	-0.01(0.04)	-0.02(0.04)	0.05(0.06)
plug-in	-0.05(0.06)	-0.10(0.11)	0.07(0.09)	-0.10(0.11)	-0.09(0.10)	0.11(0.12)
sampling	0.04(0.09)	0.04(0.11)	-0.05(0.11)	0.07(0.12)	0.05(0.11)	-0.06(0.16)
Bayesian	0.03(0.05)	0.03(0.08)	-0.05(0.10)	0.02(0.04)	0.06(0.10)	-0.09(0.14)

Bayesian approach. The fully Bayesian approach provided similar results with the two-stage approach for the special setting we have considered here. However fitting a fully Bayesian model was a bit of "overdone" in the sense that by design the longitudinal data could not be affected by the survival. Since in many transplantation studies the longitudinal data are collected before the start of follow-up for survival, therefore using our method in that cases seems to be more appropriate than using a fully Bayesian approach. We recommend the proposed approach not only for the particular transplantation studies but in any setting that shares the similarity of the separated follow-up periods for the two analyzed endpoints. That is for example when the event process does not carry any information for the longitudinal outcome and the condition (10) from Section 3.2 holds. The simulation results indicate that even if the data come from the real joint setting in which (10) may not hold, the proposed two-stage procedure can be comparable to the fully Bayesian approach.

Since the sampling in the proposed method relies strongly on the results of the first part, the accurate estimation of all parameters of nonlinear mixed model is a key feature and should be performed carefully. This can be a problematic when the deviation from normality of the random effects is suspected. However it was shown that even for the non-normal random effects one can still use a standard software such as *nlmixed* in SAS with just a small change in a standard program code. In such cases the probability integral transformation (PIT) proposed by Nelson et al.[27] can be used or the reformulation of the likelihood proposed by Liu and Yu [28]. An alternative is fitting a Bayesian model only to estimate the longitudinal submodel in the first stage, instead of the likelihood methods. Fitting nonlinear mixed models using Bayesian standard software can be however problematic due to the high nonlinearity in random effects that is caused both by the nonlinear function of the longitudinal profiles and by the possible restrictions on parameters [29].

In comparison with the two-stage approach proposed by Tsiatis, DeGruttola and Wulfsohn [5] our method is less computationally intensive since it does not require

fitting as many mixed models as there are event times in the data. An alternative, that is somewhat simpler to implement and does not require any assumption about the distribution on the underlying random effects, is the conditional score approach proposed by Tsiatis and Davidian [11]. However this method is less efficient than the methods based on likelihood. The focus in the discussed approaches is on the association between the longitudinal and event time processes. However in transplantation studies when the two follow-up periods for longitudinal and survival outcomes are often separated the interest is rather in making an inference on the marginal event-time distribution. This is similar to the Bayesian approach proposed by Xu and Zeger [12], that uses the longitudinal data as auxiliary information or surrogate for time-to-event data. Our approach is particulary useful in this setting. Since each of the two submodels is fitted separately, a standard software can be used to implement our method with just a small part of additional programming for Monte Carlo sampling.

Another advantage of the proposed two-stage method is that it can be easily generalized from survival to other types of models as it was applied for the binary Delayed Graft Failure (DGF) indicator (results not shown) in the analysis of the renal data. For that purpose in the second step of the two-stage procedure the survival model was replaced by the logistic regression model for the DGF indicator. The first stage of the proposed approach could be also modified allowing for other types of longitudinal response and other types of mixed models. Therefore instead of using a nonlinear mixed model a linear mixed model or generalized linear mixed model (GLMMs) can be considered depending on the type and the shape of the longitudinal response. In the presented real data example we have chosen the three parameters that described the evolution of the longitudinal response. However for the particular question of interest one can easily choose the most convenient parametrization for the longitudinal model and use the selected parameters to analyze the non-longitudinal response in the second stage.

Acknowledgements

References

1. Kabfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data.* John Wiley & Sons: New York, 2002.

2. Faucett C, Thomas D. Simultaneously modelling censored survival data and repeatedly measured covariates. *Statistics in Medicine* 1996; **15**:1460–1461.

3. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**:330–339.

4. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 1982; **69**:331–342.

5. Tsiatis A, DeGruttola V, Wulfsohn M. Modeling the relationship of survival to longitudinal data measured with error: Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association* 1995; **90**:27–37.

6. Albert PS, Shih JH. On estimating the relationship between longitudinal measurments and time-to-event data using a simple two-stage procedure. *Biometrics* 2009; **66**:983–991.

7. Rizopoulos D, Verbeke G, Lesaffre E. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B* 2009; **71**:637–654.

8. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **4**:465–480.

9. DeGruttola V, Tu X. Modeling progression of CD4 lymphocyte count and its relationship to survival time. *Biometrics* 1994; **50**:1003–14.

10. Self S, Pawitan Y. *Modeling a marker of disease progression and onset of disease.* In AIDS Epidemiology: Methodological Issues (N.P. Jewell, K. Dietz, and V.T. Farewell, eds.: Birkhauser, Boston, 1992.

11. Tsiatis A, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* 2001; **88**:447–58.

12. Xu J, Zeger S. Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* 2001; **50**:375–387.

13. Wang Y, Taylor J. Jointly modeling longitudinal and event time data with application to acquired immunodeciency syndrome. *J. A. Statist. Assoc.* 2001; **96**:895–905.

14. Ibrahim JG, Chen M, Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica* 2004; **14**:863–883.

15. Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* 2003; **59**:221–228.

16. Wu L, Hu XJ, Wu H. Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data. *Biostatistics* 2008; **9**:308–320.

17. Hu C, Sale ME. A joint model for nonlinear longitudinal data with informative dropout. *Journal of Pharmacokinetics and Pharmacodynamics* 2008; **30**:83–103.

18. Kaciroti NA, Raghunathan ER, Schork MA, Clark MC. A Bayesian model for longitudinal count data with non-ignorable dropout. *J R Stat Soc Ser C (Appl Stat)* 2008; **57**:521–534.

19. Andersen P, Gill R. Cox's regression for counting processes: A large sample study. *Annals of Statistics* 1982; **10**:1100–1120.

20. Clayton D. Bayesian analysis of frailty models. Tech. report, Medical Research Council Biostatistics Unit. Cambridge, 1994.

21. Kalbfleisch J. Non-parametric Bayesian analysis of survival time data. *J Roy Statist Soc B* 1978; **40**:214.

22. Patra K, Dey DK. A general class of change point and change curve modelling for life data. *Inst. Stat. Math.* 2002; **54**:517–530.

23. Casellas J. Bayesian inference in a piecewise Weibull proportional hazards model with unknown change points. *J.Anim Breed. Genet.* 2007; **124**:176–184.

24. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman & Hall: New York, 2000.

25. Roberts GO, Gelman A, Gilks WR. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* 1997; **7**:110–120.

26. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data.* Springer-Verlag: New York, US, 2000.

27. Nelson KP, Lipsitz SR, Fitzmaurice GM, Ibrahim J, Parzen M, Strawderman R. Using the probability integral transformation for non-normal random effects in non-linear mixed models. *J. Comput. Graph. Statist.* 2004; **15**:3957.

28. Liu L, Yu Z. A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine* 2008; **27**:3105–3124.

29. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurment Data.* Chapman and Hall: New York, US, 1998.