Biometrics 00, 1–11 DOI: 10.1111/0000
January 0000

Joint Models with Multiple Longitudinal Outcomes and a Time-to-Event

Dimitris Rizopoulos

Department of Biostatistics, Erasmus University Medical Center Rotterdam, the Netherlands

email: d.rizopoulos@erasmusmc.nl

Summary: Joint models for longitudinal and survival data have gained a lot of attention the recent years. There have been extended to handle among others multivariate longitudinal data, competing risks and recurrent events, and nowadays there also exist several freely available software packages for their implementation. From the aforementioned extensions, the one that is most practically relevant is the multivariate longitudinal data one. Even though this extension is mathematically straightforward, from a computational viewpoint joint models with multiple longitudinal outcomes remain difficult to fit in practice due to the high number of random effects they require. This difficulty has also hampered to a degree their practical application. Here we present a novel approach that enables fitting such joint models in realistic computing times. The idea behind our approach is to split the estimation in two steps, first to estimate a multivariate mixed model for the longitudinal outcomes, and then use the output of this model to fit the survival submodel. Such two-stage approaches have been previously proposed in the literature and have been shown to be biased. What is different in our approach is a correction we apply in the resulting estimates that transform them to the estimates we would expect to obtain if we would fit the multivariate joint model. This correction is based on importance sampling ideas. Simulation studies have shown that this corrected-two-stage approach works very satisfactorily also in difficult settings.

Key words: Biomarkers; Importance sampling; Multivariate random effects.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

***.

The rest of the paper is organized as follows. ***.

2. Joint Model Specification

We start with a general definition of the framework of multivariate joint models for multiple longitudinal outcomes and an event time. Let $\mathcal{D}_n = \{T_i, T_i^U, \delta_i, \boldsymbol{y}_i; i = 1, \ldots, n\}$ denote a sample from the target population, where we let T_i^* denote the true event time for the i-th subject, T_i and T_i^U the observed event times, and $\delta_i \in \{0, 1, 2, 3\}$ denotes the event indicator, with 0 corresponding to right censoring ($T_i^* > T_i$), 1 to a true event ($T_i^* = T_i$), 2 to left censoring ($T_i^* < T_i$), and 3 to interval censoring ($T_i < T_i^* < T_i^U$). Assuming K longitudinal outcomes we let $\boldsymbol{y}_{ki}$ denote the $n_{ki} \times 1$ longitudinal response vector for the k-th outcome ($k = 1, \ldots, K$) and the i-th subject, with elements y_{kij} denoting the value of the k-th longitudinal outcome taken at time point t_{kij}, $j = 1, \ldots, n_{ki}$.

To accommodate multivariate longitudinal responses of different types in a unified framework, we postulate a generalized linear mixed effects model. In particular, the conditional distribution of $\boldsymbol{y}_{ki}$ given a vector of random effects $\boldsymbol{b}_{ki}$ is assumed to be a member of the exponential family, with linear predictor given by

$$g_k[E\{y_{ki}(t) \mid \boldsymbol{b}_{ki}\}] = \eta_{ki}(t) = \boldsymbol{x}_{ki}^\top(t)\boldsymbol{\beta}_k + \boldsymbol{z}_{ki}^\top(t)\boldsymbol{b}_{ki}, \tag{1}$$

where $g_k(\cdot)$ denotes a known one-to-one monotonic link function, and $y_{ki}(t)$ denotes the value of the k-th longitudinal outcome for the i-th subject at time point t, $\boldsymbol{x}_{ki}(t)$ and $\boldsymbol{z}_{ki}(t)$ denote the design vectors for the fixed-effects $\boldsymbol{\beta}_k$ and for the random effects $\boldsymbol{b}_{ki}$, respectively. The dimensionality and composition of these design vectors is allowed to differ between the multiple outcomes, and they may also contain a mix of baseline and time-varying covariates. To account for the association between the multiple longitudinal outcomes we link their

corresponding random effects. More specifically, the complete vector of random effects $\boldsymbol{b}_i = (\boldsymbol{b}_{1i}, \boldsymbol{b}_{2i}, \ldots, \boldsymbol{b}_{Ki})^\top$ is assumed to follow a multivariate normal distribution with mean zero and variance-covariance matrix $\boldsymbol{D}$. For the survival process, we assume that the risk for an event depends on a function of the subject-specific linear predictor $\eta_i(t)$ and/or the random effects. More specifically, we have

$$\begin{aligned} h_i(t \mid \mathcal{H}_i(t), \boldsymbol{w}_i(t)) &= \lim_{\Delta t \to 0} \Pr\{t \leqslant T_i^* < t + \Delta t \mid T_i^* \geqslant t, \mathcal{H}_i(t), \boldsymbol{w}_i(t)\}/\Delta t, \quad t > 0 \\ &= h_0(t) \exp\Big[\boldsymbol{\gamma}^\top \boldsymbol{w}_i(t) + \sum_{k=1}^{K} \sum_{l=1}^{L_k} f_{kl}\{\mathcal{H}_{ki}(t), \boldsymbol{w}_i(t), \boldsymbol{b}_{ki}, \boldsymbol{\alpha}_{kl}\}\Big], \end{aligned} \tag{2}$$

where $\mathcal{H}_{ki}(t) = \{\eta_{ki}(s), 0 \leqslant s < t\}$ denotes the history of the underlying longitudinal process up to t, $h_0(\cdot)$ denotes the baseline hazard function, $\boldsymbol{w}_i(t)$ is a vector of exogenous, possibly time-varying, covariates with corresponding regression coefficients $\boldsymbol{\gamma}$. Functions $f_{kl}(\cdot)$, parameterized by vector $\boldsymbol{\alpha}_{kl}$, specifies which components/features of each longitudinal outcome are included in the linear predictor of the relative risk model (Brown, 2009; Rizopoulos and Ghosh, 2011; Rizopoulos, 2012; Rizopoulos et al., 2014). Some examples, motivated by the literature, are (subscripts kl have been dropped in the following expressions but are assumed):

$$\begin{aligned} f\{\mathcal{H}_i(t), \boldsymbol{w}_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha}\} &= \alpha \eta_i(t), \\ f\{\mathcal{H}_i(t), \boldsymbol{w}_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha}\} &= \alpha_1 \eta_i(t) + \alpha_2 \eta_i'(t), \text{ with } \eta_i'(t) = \frac{d\eta_i(t)}{dt}, \\ f\{\mathcal{H}_i(t), \boldsymbol{w}_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha}\} &= \alpha \int_0^t \eta_i(s)\, ds. \end{aligned}$$

These formulations of $f(\cdot)$ postulate that the hazard of an event at time t may be associated with the underlying level of the biomarker at the same time point, the slope of the longitudinal profile at t or the accumulated longitudinal process up to t. In addition, the specified terms from the longitudinal outcomes may also interact with some covariates in the $\boldsymbol{w}_i(t)$. Furthermore, note, that we allow a combination of L_k functional forms per longitudinal outcome. Finally, the baseline hazard function $h_0(\cdot)$ is modeled flexibly using a B-splines approach, i.e.,

$$\log h_0(t) = \sum_{q=1}^{Q} \gamma_{h_0,q} B_q(t, \boldsymbol{v}), \tag{3}$$

where $B_q(t, \boldsymbol{v})$ denotes the q-th basis function of a B-spline with knots $v_1, \ldots, v_Q$ and $\boldsymbol{\gamma}_{h_0}$ the vector of spline coefficients. To avoid the task of choosing the appropriate number and position of the knots, we include a relatively high number of knots (e.g., 15 to 20) and appropriately penalize the B-spline regression coefficients $\boldsymbol{\gamma}_{h_0}$ for smoothness using the differences penalty (Eilers and Marx, 1996).

3. Likelihood and Priors

As explained in Section 1, for the estimation of joint model's parameters we will use a Bayesian approach. The posterior distribution of the model parameters given the observed data is derived under the assumptions that given the random effects, the longitudinal outcomes are independent from the event times, the multiple longitudinal outcomes are independent of each other, and the longitudinal responses of each subject in each outcome are independent. Under these assumptions the posterior distribution is analogous to:

$$p(\boldsymbol{\theta}, \boldsymbol{b}) \propto \prod_{i=1}^{n} \prod_{k=1}^{K} \prod_{j=1}^{n_{ki}} p(y_{kij} \mid \boldsymbol{b}_{ki}, \boldsymbol{\theta})\; p(T_i, T_i^U, \delta_i \mid \boldsymbol{b}_{ki}, \boldsymbol{\theta})\; p(\boldsymbol{b}_{ki} \mid \boldsymbol{\theta})\; p(\boldsymbol{\theta}), \tag{4}$$

where $\boldsymbol{\theta}$ denotes the full parameter vector, and

$$p(y_{kij} \mid \boldsymbol{b}_{ki}, \boldsymbol{\theta}) = \exp\Big\{ \Big[y_{kij} \psi_{kij}(\boldsymbol{b}_{ki}) - c_k\{\psi_{kij}(\boldsymbol{b}_{ki})\} \Big] \Big/ a_k(\varphi) - d_k(y_{kij}, \varphi) \Big\},$$

with $\psi_{kij}(\boldsymbol{b}_{ki})$ and φ denoting the natural and dispersion parameters in the exponential family, respectively, $c_k(\cdot)$, $a_k(\cdot)$, and $d_k(\cdot)$ are known functions specifying the member of the exponential family. For the survival part accordingly we have

$$\begin{aligned} p(T_i, T_i^U, \delta_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = & \Big\{ h_i(T_i \mid \mathcal{H}_i(T_i), \boldsymbol{w}_i(T_i)) \Big\}^{I(\delta_i=1)} \exp\Big\{ - \int_0^{T_i} h_i(s \mid \mathcal{H}_i(s), \boldsymbol{w}_i(s))\, ds \Big\}^{I(\delta_i \in \{0,1\})} \\ \times & \left\{ 1 - \exp\Big\{ - \int_0^{T_i} h_i(s \mid \mathcal{H}_i(s), \boldsymbol{w}_i(s))\, ds \Big\} \right\}^{I(\delta_i=2)} \\ \times & \left\{ \exp\Big\{ - \int_0^{T_i} h_i(s \mid \mathcal{H}_i(s), \boldsymbol{w}_i(s))\, ds \Big\} - \exp\Big\{ - \int_0^{T_i^U} h_i(s \mid \mathcal{H}_i(s), \boldsymbol{w}_i(s))\, ds \Big\} \right\}^{I(\delta_i=3)} \end{aligned} \tag{5}$$

where $I(\cdot)$ denotes the indicator function. The integral in the definition of the cumulative hazard function does not have a closed-form solution, and thus a numerical method must be employed for its evaluation. Standard options are the Gauss-Kronrod and Gauss-Legendre quadrature rules.

For the parameters of the longitudinal outcomes we use standard default priors. More specifically, independent normal priors with zero mean and variance 1000 for the fixed effects and inverse Gamma priors for scale parameters. The covariance matrix of the random effects is parameterized in terms of a correlation matrix $\boldsymbol{\Omega}$ and a vector of $\boldsymbol{\sigma}_d$. For the correlation matrix $\boldsymbol{\Omega}$ we use LKJ-Correlation prior proposed by Lewandowski et al. (2009) with parameter $\zeta = 1.5$. For each element of $\boldsymbol{\sigma}_d$ we use a half-Student's t prior with 3 degrees of freedom. For the regression coefficients $\boldsymbol{\gamma}$ of the relative risk model we assume independent normal priors with zero mean and variance 1000. The same prior is also assumed for the vector of association parameters $\boldsymbol{\alpha}$. However, when $\boldsymbol{\alpha}$ becomes high dimensional (e.g., when several functional forms are considered per longitudinal outcome), we opt for a global-local ridge-type shrinkage prior, i.e., for the s-th element of $\boldsymbol{\alpha}$ we assume

$$\alpha_s \sim \mathcal{N}(0, \tau\psi_s), \quad \tau^{-1} \sim \text{Gamma}(0.1, 0.1), \quad \psi_s^{-1} \sim \text{Gamma}(1, 0.01).$$

The global smoothing parameter τ has sufficiently mass near zero to ensure shrinkage, while the local smoothing parameter ψ_s allows individual coefficients to attain large values. Other options of shrinkage or variable-selection priors could be used as well (Andrinopoulou and Rizopoulos, 2016). Finally, the penalized version of the B-spline approximation to the baseline hazard is specified using the following hierarchical prior for $\boldsymbol{\gamma}_{h_0}$ (Lang and Brezger, 2004):

$$p(\boldsymbol{\gamma}_{h_0} \mid \tau_h) \propto \tau_h^{\rho(K)/2} \exp\Big(-\frac{\tau_h}{2}\boldsymbol{\gamma}_{h_0}^\top \boldsymbol{K} \boldsymbol{\gamma}_{h_0}\Big),$$

where τ_h is the smoothing parameter that takes a Gamma$(1, \tau_{h\delta})$ prior distribution, with a hyper-prior $\tau_{h\delta} \sim \text{Gamma}(10^{-3}, 10^{-3})$, which ensures a proper posterior distribution for

$\boldsymbol{\gamma}_{h_0}$ (Jullion and Lambert, 2007), $\boldsymbol{K} = \Delta_r^\top \Delta_r + 10^{-6}\mathbf{I}$, with Δ_r denoting the r-th difference penalty matrix, and $\rho(\boldsymbol{K})$ denotes the rank of $\boldsymbol{K}$.

4. Corrected Two-Stage Approach

4.1 *Importance sampling correction*

Carrying out a full Bayesian estimation of the multivariate joint model is straightforward, either using Markov chain Monte Carlo (MCMC) or Hamiltonian Monte Carlo (HMC). However, from a computational viewpoint, this estimation becomes very challenging due to the high number of the involved random effects, and the requirement for numerical integration for the calculation of the density of the survival outcome (5). This limitation has hampered the use of multivariate joint models in practice.

To overcome this practical deadlock, an alternative solution is to utilize a two-stage approach. This would entail separately fitting the longitudinal and survival outcomes. More specifically, under the Bayesian framework, we would have the following two stages:

S-I: We fit a multivariate mixed model in the longitudinal outcomes using either MCMC or HMC, and we obtain a sample $\{\boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)}; m = 1, \ldots, M\}$ of size M from the posterior,

$$p(\boldsymbol{\theta}_y, \boldsymbol{b} \mid \boldsymbol{y}) \propto \prod_{i=1}^{n} \prod_{k=1}^{K} \prod_{j=1}^{n_{ki}} p(y_{kij} \mid \boldsymbol{b}_{ki}, \boldsymbol{\theta})\; p(\boldsymbol{b}_{ki} \mid \boldsymbol{\theta})\; p(\boldsymbol{\theta}_y),$$

where $\boldsymbol{\theta}_y$ denotes the subset of the parameters that are involved in the definition of the longitudinal submodels (including the parameters in the random-effects distribution).

S-II: Utilizing the sample from Stage I, we obtain a sample for the parameters of the survival submodel $\{\boldsymbol{\theta}_t^{(m)}; m = 1, \ldots, M\}$ from the corresponding posterior distribution,

$$p(\boldsymbol{\theta}_t \mid T, \delta, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)}) \propto \prod_{i=1}^{n} p(T_i, \delta_i \mid \boldsymbol{\theta}_t, \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)})\; p(\boldsymbol{\theta}_t),$$

where $\boldsymbol{\theta}_t$ denotes the subset of the parameters that are involved in the definition of the survival submodel.

This two-stage procedure essentially entails the same number of iterations as the full Bayesian estimation of the multivariate joint model. The computational benefits stem from the fact that in Stage I we do need to numerically integrate the survival submodel density function.

However, even though this approach greatly reduces the computational burden, there has been considerable work in the literature demonstrating that it produces biased estimates even in the simpler case of univariate joint models (see Tsiatis and Davidian, 2004; Rizopoulos, 2012, and references therein). This bias is caused from the fact that we do not work with the full joint distribution that produces estimates of $\boldsymbol{\theta}_y$ and $\boldsymbol{b}$ that are corrected for informative dropout due to the occurrence of an event.

To overcome this issue, we propose here to correct the estimates we obtain from the two-stage approach using importance sampling weights (Press et al., 2007, Section 7.9). In particular, we consider that the realizations $\{\boldsymbol{\theta}_t^{(m)}, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)}; m = 1, \ldots, M\}$ we have obtained from the two-stage approach can be considered as a sample from the full posterior of the multivariate joint models by applying the weights,

$$w^{(m)} = \frac{p(\boldsymbol{\theta}_t^{(m)} \mid T, \delta, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)})\; p(\boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)} \mid \boldsymbol{y}, T, \delta)}{p(\boldsymbol{\theta}_t^{(m)} \mid T, \delta, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)})\; p(\boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)} \mid \boldsymbol{y})}. \tag{6}$$

The numerator is the posterior distribution of the multivariate joint model and the denominator the posterior distributions from the two stages. As also stated above, from (6) we observe that difference between fitting the full joint model versus the two-stage approach comes from the second term in the numerator and denominator. By expanding these two terms we obtain

$$\frac{p(\boldsymbol{\theta}_y^{(m)}, \boldsymbol{b} \mid \boldsymbol{y}, T, \delta)}{p(\boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)} \mid \boldsymbol{y})} \;\propto\; \frac{\prod_i p(\boldsymbol{y}_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)})\; p(T_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)})\; p(\boldsymbol{b}_i^{(m)} \mid \boldsymbol{\theta}_y^{(m)})\; p(\boldsymbol{\theta}_y^{(m)})}{\prod_i p(\boldsymbol{y}_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)})\; p(\boldsymbol{b}_i^{(m)} \mid \boldsymbol{\theta}_y^{(m)})\; p(\boldsymbol{\theta}_y^{(m)})}$$

$$= \prod_i p(T_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}) = \prod_i \int p(T_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{\theta}_t)\, d\boldsymbol{\theta}_t.$$

The resulting weights involve a marginal likelihood calculation, which we perform using a Laplace approximation, namely

$$\varpi^{(m)} = \exp\Big[\frac{q\log(2\pi) - \log\{\det(\widehat{\Sigma}^{(m)})\}}{2} + \log\{p(T_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}, \widehat{\boldsymbol{\theta}}_t^{(m)})\}\Big],$$

$$\tilde{w}^{(m)} = \varpi^{(m)} \Big/ \sum_{m=1}^{M} \varpi^{(m)},$$

where $\widehat{\boldsymbol{\theta}}_t^{(m)} = \arg\max_{\boldsymbol{\theta}_t}[\log\{p(T_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}, \widehat{\boldsymbol{\theta}}_t)\}]$, $\det(A)$ denotes the determinant of matrix A, and $\widehat{\Sigma}^{(m)} = -\partial^2 \log\{p(T_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}, \widehat{\boldsymbol{\theta}}_t)\}/\partial\boldsymbol{\theta}_t^\top \partial\boldsymbol{\theta}_t|_{\theta_t = \hat{\theta}_t^{(m)}}$. The extra computational burden of performing this Laplace approximation is minimal in practice. This is due to the fact that good initial values can be provided from one iteration m to the next $m+1$, which substantially reduces the number of required optimization iterations for finding $\widehat{\boldsymbol{\theta}}_t^{(m)}$ (i.e., $\widehat{\boldsymbol{\theta}}_t^{(m)}$ is provided as an initial value to find $\widehat{\boldsymbol{\theta}}_t^{(m+1)}$).

4.2 *Performance*

To evaluate whether the introduction of the importance sampling weights alleviates the bias the simple two-stage approach (i.e., without the weights), we perform a 'proof-of-concept' simulation study. In particular, we compare the proposed corrected two stage approach with the simple two-stage approach and the full multivariate joint model in the case of two continuous longitudinal outcomes. The specific details behind this simulation setting are given in Web Section 1. The results from 500 simulated datasets are presented in Web Figures 1–3, and in Figure 1.

[Figure 1 about here.]

Web Figure 1 shows boxplots with the computing times required to fit the joint model under each of the three approaches. As expected, the full multivariate joint model took substantially more computing time to fit compared to the two-stage approach. In addition,

as we argued above, calculating the importance sampling weights had minimal computing cost. Web Figures 2 and 3 show boxplots of posterior means from the 500 datasets for the parameters of the longitudinal submodels. We observe that all approaches provide very similar results with minimal bias. Figure 1 shows the corresponding boxplots of posterior models for the parameters of the survival submodel. As expected, the full multivariate joint model returns unbiased results. Similarly, as has been previously reported in the literature, the simple two-stage approach exhibits considerable bias. However, even though theoretically the importance sampling weights should make the posterior means obtained from the simple two-stage approach to come much closer to the full multivariate joint model, we observe that they do not alleviate any of this bias.

5. Corrected Two-Stage Approach with Random Effects

5.1 *Importance sampling correction with random effects*

The fact that the importance sampling weights do not work seems unexpected, because according to Web Figures 2 and 3 the corrected two-stage (actually even the simple two-stage approach) unbiasedly estimates the fixed effects and the variance components. However, a deeper investigation into the results obtained by the corrected two-stage approach versus the 'gold-standard' results from the multivariate joint model shows that there is a considerable difference in the posterior of the random effects. This is depicted in Figure 2.

[Figure 2 about here.]

The data have been simulated such that high values for the longitudinal outcome y_1 and low values for outcome y_2 are associated with a higher hazard of an event. From Figure 2 we observe that ***.

5.2 *Performance*

6. Analysis of the Aortic Valve Dataset

We return to the Aortic Valve dataset introduced in Section 1. Our aim is to use the existing data to build joint models that can be utilized for planning the visiting patterns of future patients from the same population. In our study, a total of 77 (27%) patients received a sub-coronary implantation (SI) and the remaining 208 patients a root replacement (RR). These patients were followed prospectively over time with annual telephone interviews and biennial standardized echocardiographic assessment of valve function until July 8, 2010. Echo examinations were scheduled at 6 months and 1 year postoperatively, and biennially thereafter, and at each examination, echocardiographic measurements were taken. By the end of follow-up, 1262 assessments have been recorded, with an average of 4.3 measurements per patient (s.d. 2.4 measurements), 59 (20.7%) patients had died, and 73 (25.6%) patients required a re-operation on the allograft. Here we are interested in the composite event re-operation or death, which was observed for 125 (43.9%) patients.

7. Discussion

In this paper we have presented ***.

Finally, to facilitate estimation of multivariate joint models, the estimation approach presented in this paper has been implemented in functions `cvDCL()` and `dynInfo()`, available in package the **JMbayes** (version 0.8-0) for the R programming language (freely available from the Comprehensive R Archive Network at `http://cran.r-project.org/package=JMbayes`). A working example of how these functions should be used can be found in the supplementary material.

Acknowledgements

The author ***.

Supplementary Materials

Web Appendix A, referenced in Section ***, is available with this paper at the Biometrics website on Wiley Online Library.

References

Andrinopoulou, E. and Rizopoulos, D. (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Statistics in Medicine* **35,** 4813–4823.

Brown, E. (2009). Assessing the association between trends in a biomarker and risk of event with an application in pediatric HIV/AIDS. *The Annals of Applied Statistics* **3,** 1163–1182.

Eilers, P. and Marx, B. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science* **11,** 89–121.

Jullion, A. and Lambert, P. (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive bayesian p-splines models. *Computational Statistics and Data Analysis* **51,** 2542–2558.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics* **13,** 183–212.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* **100,** 1989–2001.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, New York, 3rd edition.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data, with Applications in R*. Chapman & Hall/CRC, Boca Raton.

Rizopoulos, D. and Ghosh, P. (2011). A Bayesian semiparametric multivariate joint model

for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine* **30,** 1366–1380.

Rizopoulos, D., Hatfield, L., Carlin, B., and Takkenberg, J. (2014). Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *Journal of the American Statistical Association* **109,** 1385–1397.

Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* **14,** 809–834.

Received January 0000. *Revised January* 0000. *Accepted January* 0000.

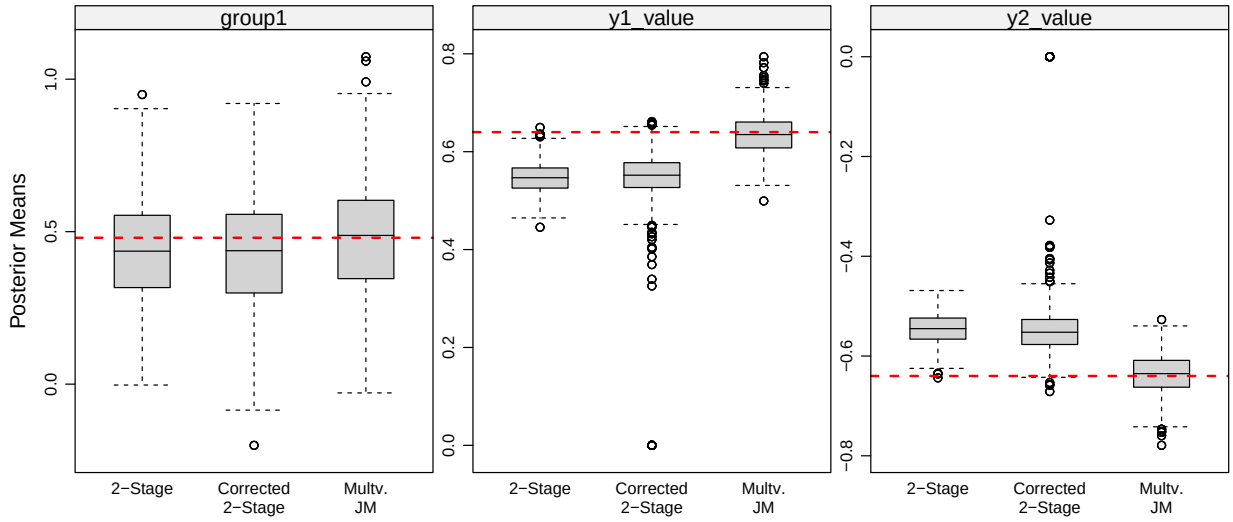
group1
y1_value
y2_value
Posterior Means
0.0
0.5
1.0
0.2
0.4
0.6
0.8
−0.8
−0.6
−0.4
−0.2
2−Stage
Corrected 2−Stage
Multv. JM

Figure 1. ***.

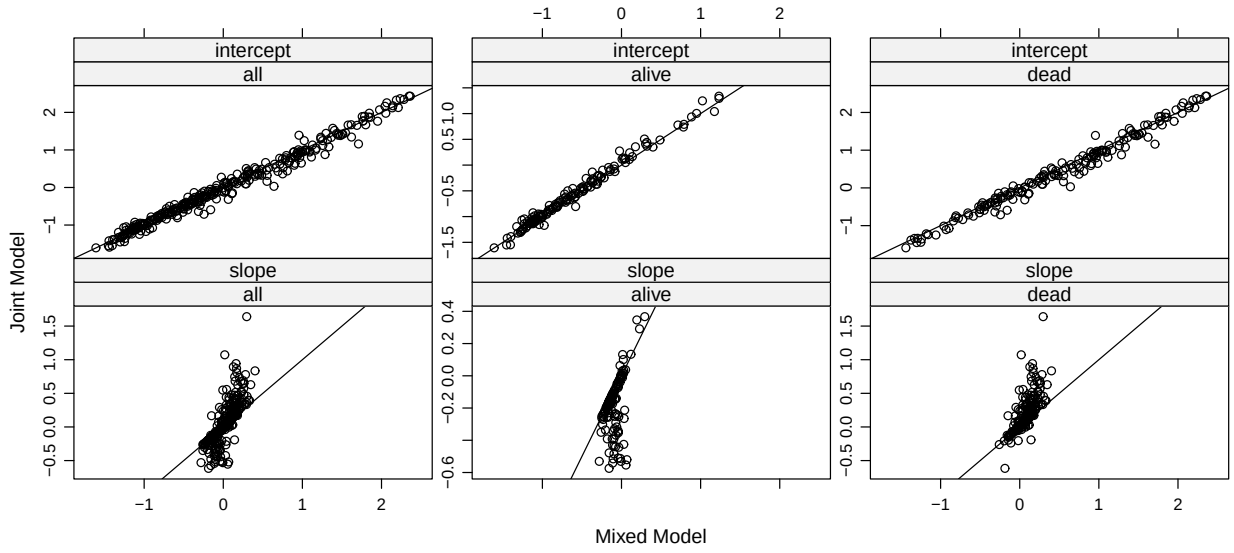
intercept
all
slope
all
intercept
alive
slope
alive
intercept
dead
slope
dead
Joint Model
Mixed Model

Figure 2. ***.