Joint models with multiple longitudinal outcomes and a time-to-event outcome: a corrected two-stage approach

Katya Mauff[1] · Ewout Steyerberg[2,3] · Isabella Kardys[4] · Eric Boersma[4] · Dimitris Rizopoulos[1]

Abstract

Joint models for longitudinal and survival data have gained a lot of attention in recent years, with the development of myriad extensions to the basic model, including those which allow for multivariate longitudinal data, competing risks and recurrent events. Several software packages are now also available for their implementation. Although mathematically straightforward, the inclusion of multiple longitudinal outcomes in the joint model remains computationally difficult due to the large number of random effects required, which hampers the practical application of this extension. We present a novel approach that enables the fitting of such models with more realistic computational times. The idea behind the approach is to split the estimation of the joint model in two steps: estimating a multivariate mixed model for the longitudinal outcomes and then using the output from this model to fit the survival submodel. So-called two-stage approaches have previously been proposed and shown to be biased. Our approach differs from the standard version, in that we additionally propose the application of a correction factor, adjusting the estimates obtained such that they more closely resemble those we would expect to find with the multivariate joint model. This correction is based on importance sampling ideas. Simulation studies show that this corrected two-stage approach works satisfactorily, eliminating the bias while maintaining substantial improvement in computational time, even in more difficult settings.

Keywords Biomarkers · Importance sampling · Multivariate random effects

1 Introduction

Joint models for longitudinal and survival data have become a valuable asset in the toolbox of modern data scientists. After the seminal papers of Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997), several extensions of these models have been proposed in the literature. These include, among others, flexible specification of the longitudinal model (Brown et al. 2005a), consideration of competing

risks (Elashoff et al. 2008; Andrinopoulou et al. 2014) and multistate models (Ferrer et al. 2016), and the calculation of dynamic predictions (Proust-Lima and Taylor 2009; Rizopoulos 2011; Rizopoulos et al. 2014; Andrinopoulou and Rizopoulos 2016; Rizopoulos et al. 2017; Andrinopoulou et al. 2018). A particularly useful and practical extension is the one which allows for the inclusion of multiple longitudinal outcomes (Rizopoulos and Ghosh 2011a; Chi and Ibrahim 2006; Brown et al. 2005b; Lin et al. 2002). In medical settings in particular, data collection is likely to be complex: while the standard joint model allows us to determine the association between a survival outcome and a single longitudinal outcome (biomarker), there are more often than not multiple biomarkers relevant to the event of interest. Extending the univariate joint model to accommodate these multiple longitudinal outcomes allows us to incorporate more information, improving prognostication and enabling us to better make sense of the complex underlying nature of the disease dynamics. A motivating example of this is the Bio-SHiFT cohort study; a prospective observational study conducted in the Netherlands on chronic heart failure (CHF) patients.

✉ Katya Mauff
  k.mauff@erasmusmc.nl

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

[2] Center for Medical Decision Making, Department of Public Health, Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands

[3] Department of Biomedical Data Sciences, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands

[4] Department of Cardiology, Erasmus University Medical Center, 3000 CA Rotterdam, The Netherlands

Springer

The primary focus of the study was to determine whether or not disease progression in individual CHF patients can be assessed using longitudinal measurements of several blood biomarkers (van Boven et al. 2018). Previous work on this data has focused mainly on the association between each individual biomarker and a single composite event, but it is likely that the predictive value of the biomarkers will be more accurately determined when they are assessed in concert.

Extension to the multivariate case is mathematically straightforward and may be easily combined with other extensions: allowing for longitudinal outcomes of varying types, left, right and interval censoring and the inclusion of competing risks, among others. There are also now a number of excellent software packages available, which makes the implementation of the more complex models easier. There are, however, technical challenges which hamper the widespread use of these models. As the number of longitudinal outcomes increases, and thus the number of random effects, standard methods become computationally prohibitive: under a Bayesian approach, the number of parameters to sample becomes unreasonably large, and in the case of maximum likelihood, we are required to numerically approximate the integrals over the random effects, which is challenging in high dimensions. The practical solution most commonly used in such settings is that of the two-stage approach, wherein a multivariate mixed model is first used for the longitudinal outcomes, following which, the output of this model is used to fit a survival submodel. Unfortunately, substantial research on this topic indicates that this approach results in biased estimates (Tsiatis and Davidian 2004; Rizopoulos 2012; Ye et al. 2008). In this paper, we propose an adaptation of the simple two-stage approach, which eliminates the bias and substantially reduces computational time. We propose the use of a correction factor, based on importance sampling theory (Press et al. 2007, Section 7.9). This correction factor allows us to re-weight each realization of the MCMC sample obtained from the Bayesian estimation of the two-stage approach such that the resulting estimates more closely approximate those obtained via the full multivariate joint model. The weights are given by the target distribution (the full posterior distribution of the multivariate joint model), divided by the product of the posterior distributions for each of the two stages, evaluated for each iteration of the MCMC sample. The use of this correction factor alone is not enough to eliminate the bias, but, prior to its application, the two-stage approach is itself modified: where before, in the second stage, only the parameters of the survival submodel were updated, we now also update the random effects. These adaptations combined achieve unbiased estimates in a fraction of the time required to compute the full multivariate model.

The rest of the paper is organized as follows: Sect. 2 introduces the full multivariate joint model, and Sect. 3 discusses the estimation of the model under the Bayesian paradigm. Section 4 introduces the importance-sampling-corrected two-stage approach and presents the results of a simple simulation, and Sect. 5 introduces the importance-sampling-corrected two-stage approach with updated random effects. Section 6 presents the results of a more complex simulation, and finally in Sect. 7, we look at an analysis of the Bio-SHiFT data.

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 T_i^* denotes the true event time for the ith subject and T_i and T_i^U the observed event times. Then, $\delta_i \in \{0, 1, 2, 3\}$ 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 kth outcome ($k = 1, \ldots, K$) and the ith subject, with elements y_{kij} denoting the value of the kth longitudinal outcome for the ith subject, 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

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

where $g_k(\cdot)$ denotes a known one-to-one monotonic link function, $y_{ki}(t)$ denotes the value of the kth longitudinal outcome for the ith subject at time point t and $\boldsymbol{x}_{ki}(t)$ and $\boldsymbol{z}_{ki}(t)$ denote the design vectors for the fixed effects $\boldsymbol{\beta}_k$ and the random effects $\boldsymbol{b}_{ki}$, respectively. The dimensionality and composition of these design vectors are allowed to differ between the multiple outcomes, and they may also contain a combination 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 $\boldsymbol{b}_i = (\boldsymbol{b}_{1i}^\top, \boldsymbol{b}_{2i}^\top, \ldots, \boldsymbol{b}_{Ki}^\top)^\top$ 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

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

where $\mathcal{H}_{ki}(t) = \{\eta_{ki}(s), 0 \leq s < t\}$ denotes the history of the underlying longitudinal process up to t for outcome k and subject i, $h_0(\cdot)$ denotes the baseline hazard function and $\boldsymbol{w}_i(t)$ is a vector of exogenous, possibly time-varying, covariates with corresponding regression coefficients $\boldsymbol{\gamma}$. Functions $f_{kl}(\cdot)$, parameterized by vector $\boldsymbol{\alpha}_{kl}$, specify which components/features of each longitudinal outcome are included in the linear predictor of the relative risk model, (Brown (2009), Rizopoulos and Ghosh (2011b), 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):

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

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 qth basis function of a B-spline with knots $v_1, \ldots, v_Q$ and $\boldsymbol{\gamma}_{h_0}$ the vector of spline coefficients; typically, $Q = 15$ or 20.

3 Likelihood and priors

As explained in Sect. 1, we use a Bayesian approach for the estimation of the joint model's parameters. 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:

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

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

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

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

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

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 is 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. 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 the 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 three 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. More specifically, for the sth element of $\boldsymbol{\alpha}$, we assume:

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

The global smoothing parameter τ has sufficient mass near zero to ensure shrinkage, while the local smoothing parameter ψ_s allows individual coefficients to attain large values. The motivation for using this type of prior distribution in this case is that we expect the different terms behind the specification of $f(\cdot)$ to be correlated and many of the corresponding coefficients to be nonzero. Nonetheless, other options of shrinkage or variable-selection priors could also be used (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):

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

where τ_h is the smoothing parameter that takes a Gamma $(1, \tau_{h\delta})$ prior distribution, with a hyper-prior $\tau_{h\delta} \sim$ 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 rth difference penalty matrix and where $\rho(\boldsymbol{K})$ 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, using either Markov chain Monte Carlo (MCMC) or Hamiltonian Monte Carlo (HMC). However, this estimation becomes very challenging from a computational viewpoint, due to the high number of random effects involved and the requirement for numerical integration in the calculation of the density of the survival outcome (5). This limitation has hampered the use of multivariate joint models in practice.

The two-stage approach, which entails fitting the longitudinal and survival outcomes separately, is the solution most often used to overcome this computational deadlock. Using

this approach, under the Bayesian framework, we would have the following two stages:

S-I: We fit a multivariate mixed model for 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,

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

where $\boldsymbol{\theta}_y$ denotes the subset of the parameters that are included 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,

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

where $\boldsymbol{\theta}_t$ denotes the subset of the parameters that are included in the definition of the survival submodel, and $\tilde{T} = (T, T^U)$.

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 we do not need to numerically integrate the survival submodel density function in Stage I. Even though this approach greatly reduces the computational burden, there exists a substantial body of work demonstrating that it results in biased estimates, even in the simpler case of univariate joint models (see Tsiatis and Davidian 2004; Rizopoulos 2012, and references therein). This bias is a result of not working with the full joint distribution, which would produce estimates of $\boldsymbol{\theta}_y$ and $\boldsymbol{b}$ that are appropriately corrected for informative dropout relating to the occurrence of an event.

To overcome this issue, we propose the correction of 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 $\{\boldsymbol{\theta}_t^{(m)}, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)}; m = 1, \ldots, M\}$ that we have obtained using the two-stage approach can be considered a weighted sample from the full posterior of the multivariate joint model with weights given by:

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

(6)

The numerator in this expression is the posterior distribution of the multivariate joint model, and the denominator, the corresponding posterior distributions from each of the two stages. As previously stated, from (6) we observe that the 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

$$\frac{p(\boldsymbol{\theta}_y^{(m)}, \boldsymbol{b} \mid \boldsymbol{y}, \tilde{T}, \delta)}{p(\boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)} \mid \boldsymbol{y})}$$

$$\propto \frac{\left(\prod_i p(\boldsymbol{y}_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}) \, p(\tilde{T}_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}) \right)}{\prod_i p(\boldsymbol{y}_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}) \, p(\boldsymbol{b}_i^{(m)} \mid \boldsymbol{\theta}_y^{(m)}) \, p(\boldsymbol{\theta}_y^{(m)})}$$

$$= \prod_i p(\tilde{T}_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)})$$

$$= \prod_i \int p(\tilde{T}_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{\theta}_t) \, d\boldsymbol{\theta}_t. \tag{7}$$

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

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

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

where

$$\widehat{\boldsymbol{\theta}}_t^{(m)} = \arg\max_{\boldsymbol{\theta}_t} \left[\log\{p(\tilde{T}_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}, \widehat{\boldsymbol{\theta}}_t)\} \right],$$

$\det(A)$ denotes the determinant of matrix A,

$$\widehat{\Sigma}^{(m)} = -\partial^2 \log\{p(\tilde{T}_i, \delta_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}, \widehat{\boldsymbol{\theta}}_t)\} / \partial \boldsymbol{\theta}_t^\top \partial \boldsymbol{\theta}_t \big|_{\boldsymbol{\theta}_t = \hat{\boldsymbol{\theta}}_t^{(m)}},$$

and q denotes the dimensionality of the $\boldsymbol{\theta}_t$ vector. The extra computational burden of performing this Laplace approximation is minimal in practice, since 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 observed with the simple
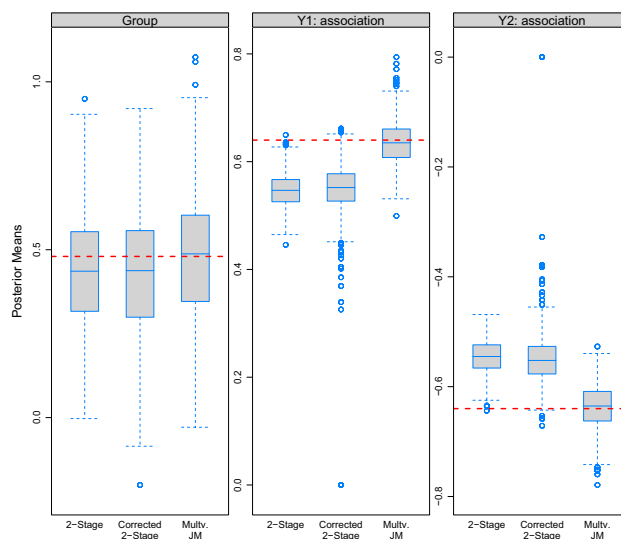

Fig. 1 Simulation results from 500 datasets comparing the two-stage approach and the importance-sampling-corrected two-stage approach with the full joint model for continuous longitudinal outcomes. The three panels show posterior means from the 500 datasets for the three coefficients in the survival submodel, namely the coefficient for the baseline group variable and the association parameters for the two longitudinal outcomes. The dashed horizontal line indicates the true value of the coefficients

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, as well as the full multivariate joint model in the case of two continuous longitudinal outcomes. The specific details of this simulation setting are given in "Appendix A.1." The results from 500 simulated datasets are presented in Fig. 1 and in the appendix, in Figs. 4 and 5. Figure 4 shows boxplots with the computing times required to fit the joint model under three approaches. Comparing the first two of these approaches, we see that the calculation of the importance sampling weights in the corrected two-stage approach had minimal computational cost, with the full multivariate joint model taking substantially more time to fit. Figure 5 shows boxplots of posterior means from the 500 datasets for the parameters of the two longitudinal submodels. We observe that all three approaches provide very similar results with minimal bias. Figure 1 shows the corresponding boxplots of posterior means for the parameters of the survival submodel. As expected, the full multivariate joint model returns unbiased results. Similarly, as has previously been reported in the literature, the simple two-stage approach exhibits considerable bias. We see that this bias persists for the corrected two-stage approach, although theoretically the use of the importance sampling weights should alleviate it (by adjusting the posterior means obtained via the simple two-stage approach such that they more closely resemble those from the full multivariate model).
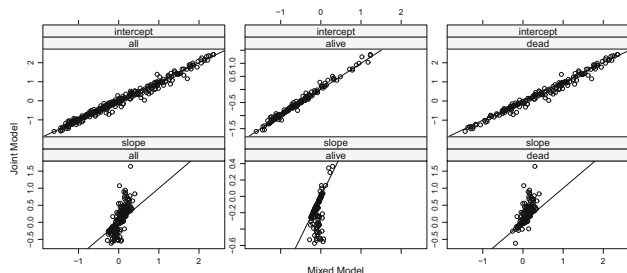
Fig. 2 Comparison of posterior mean estimates for the random intercepts and random slopes from one simulated dataset for the first longitudinal outcome between a linear mixed model and a joint model. The left column panels correspond to all subjects, the middle column to subjects without an event and the right column panel to subjects with an event

5 Corrected two-stage approach with random effects

5.1 Importance sampling correction with random effects

The above result is unexpected, since (as per Fig. 5), the corrected two-stage (and indeed the simple two-stage) approach unbiasedly estimates both the fixed effects and the variance components of the longitudinal submodels. However, further investigation shows that there is a considerable difference between the corrected two-stage approach and the multivariate joint model with regard to the posterior of the random effects. This is depicted in Fig. 2 for one of the longitudinal outcomes we have simulated. The data have been simulated such that higher values for longitudinal outcome y_1 are associated with a higher hazard of the event. From Fig. 2, we observe that the random effect estimates for the multivariate mixed model and, in particular, the random slope estimates for subjects with and without an event differ from those for the multivariate joint model. In particular, we observe that the random slope estimates from the joint model are larger for subjects with an event compared to the linear mixed model, and vice versa for subjects without an event. This observation suggests that we could improve the weights given in (6) by updating (in the second stage) not only the parameters of the survival submodel $\boldsymbol{\theta}_t$ but also the random effects $\boldsymbol{b}$. That is, we obtain a sample for the parameters of the survival submodel $\{\boldsymbol{\theta}_t^{(m)}, \boldsymbol{b}^{(m)}; m = 1, \dots, M\}$ from the corresponding joint posterior distribution,

$$
\begin{aligned}
&p(\boldsymbol{\theta}_t, \boldsymbol{b} \mid \tilde{T}, \delta, \boldsymbol{y}, \boldsymbol{\theta}_y^{(m)}) \\
&\propto \prod_{i=1}^{n} \prod_{k=1}^{K} \prod_{j=1}^{n_{ki}} p(y_{kij} \mid \boldsymbol{b}_{ki}, \boldsymbol{\theta}_y^{(m)}) \; p(\boldsymbol{b}_i \mid \boldsymbol{\theta}_y^{(m)}) \\
&\times \; p(\tilde{T}_i, \delta_i \mid \boldsymbol{\theta}_t, \boldsymbol{b}_i, \boldsymbol{\theta}_y^{(m)}) \; p(\boldsymbol{\theta}_t).
\end{aligned} \tag{8}
$$

Admittedly, simulating from $[\boldsymbol{\theta}_t, \boldsymbol{b} \mid \tilde{T}, \delta, \boldsymbol{y}, \boldsymbol{\theta}_y^{(m)}]$ is more computationally intensive than simulating from $[\boldsymbol{\theta}_t \mid \tilde{T}, \delta, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)}]$, the corresponding second stage presented in Sect. 4, since we now also need to calculate the densities of the mixed-effect models for the K longitudinal outcomes. Nonetheless, the computational gains compared to fitting the full joint model remain significant.

Under this second stage (8), the importance sampling weights now take the form:

$$
w^{(m)} = \frac{p(\boldsymbol{\theta}_t^{(m)}, \boldsymbol{b}^{(m)} \mid \tilde{T}, \delta, \boldsymbol{y}, \boldsymbol{\theta}_y^{(m)}) \; p(\boldsymbol{\theta}_y^{(m)} \mid \boldsymbol{y}, \tilde{T}, \delta)}{p(\boldsymbol{\theta}_t^{(m)}, \boldsymbol{b}^{(m)} \mid \tilde{T}, \delta, \boldsymbol{y}, \boldsymbol{\theta}_y^{(m)}) \; p(\boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)} \mid \boldsymbol{y})}. \tag{9}
$$

Similarly to (6), the new weights have been formulated such that the difference lies in the second term in both the numerator and denominator. By doing an expansion of these two terms similar to that used in the previous section, we obtain:

$$
\begin{aligned}
w^{(m)} &= \frac{p(\boldsymbol{\theta}_y^{(m)} \mid \boldsymbol{y}, \tilde{T}, \delta)}{p(\boldsymbol{\theta}_y^{(m)}, \boldsymbol{b}^{(m)} \mid \boldsymbol{y})} \\
&\propto \varpi^{(m)} = \frac{\prod_i p(y_i, \tilde{T}_i, \delta_i \mid \boldsymbol{\theta}_y^{(m)}) \; p(\boldsymbol{\theta}_y^{(m)})}{\prod_i p(y_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}) \; p(\boldsymbol{b}_i^{(m)} \mid \boldsymbol{\theta}_y^{(m)}) \; p(\boldsymbol{\theta}_y^{(m)})} \\
&= \frac{\left(\begin{array}{c} \prod_i \int\int p(y_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}_y^{(m)}) p(\tilde{T}_i, \delta_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{\theta}_t) \\ \dots \times p(\boldsymbol{b}_i \mid \boldsymbol{\theta}_y^{(m)}) \, p(\boldsymbol{\theta}_t) \, d\boldsymbol{b}_i d\boldsymbol{\theta}_t \end{array} \right)}{\prod_i p(y_i \mid \boldsymbol{b}_i^{(m)}, \boldsymbol{\theta}_y^{(m)}) \; p(\boldsymbol{b}_i^{(m)} \mid \boldsymbol{\theta}_y^{(m)})},
\end{aligned} \tag{10}
$$

and the self-normalized weights are

$$
\tilde{w}^{(m)} = \varpi^{(m)} \Big/ \sum_m \varpi^{(m)}.
$$

The integrals in the numerator are once again approximated using the Laplace method; namely, we let

$$
\begin{aligned}
\{ \widehat{\boldsymbol{\theta}}_t^{\top}, \widehat{\boldsymbol{b}}_i^{\top} \}^{\top} = \underset{\theta_t, b_i}{\arg\max} \Big\{ &\sum_j \log p(y_{ij} \mid \boldsymbol{b}_i, \boldsymbol{\theta}_y^{(m)}) \\
&+ \log p(\tilde{T}_i, \delta_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{\theta}_t) \\
&+ \log p(\boldsymbol{b}_i \mid \boldsymbol{\theta}_y^{(m)}) + \log p(\boldsymbol{\theta}_t) \Big\},
\end{aligned}
$$

and

$$
\Sigma_{b_i} = -\frac{\partial^2 \big\{ \log p(y_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}_y^{(m)}) \dots + \log p(\tilde{T}_i, \delta_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}_y^{(m)}, \widehat{\boldsymbol{\theta}}_t) \dots + \log p(\boldsymbol{b}_i \mid \boldsymbol{\theta}_y^{(m)}) \big\}}{\partial \boldsymbol{b}^{\top} \partial \boldsymbol{b}} \Big|_{b=\hat{b}_i},
$$

denote the Hessian matrix for the random effects, and analogously,

$$\Sigma_{\theta_t} = -\frac{\partial^2 \sum_i \left\{ \log p(\tilde{T}_i, \delta_i \mid \widehat{\boldsymbol{b}}_i, \boldsymbol{\theta}_y^{(m)}, \boldsymbol{\theta}_t) + \log p(\boldsymbol{\theta}_t) \right\}}{\partial \boldsymbol{\theta}_t^\top \partial \boldsymbol{\theta}_t} \Bigg|_{\theta_t = \hat{\theta}_t},$$

denote the Hessian matrix for the $\boldsymbol{\theta}_t$ parameters. Then, we approximate the inner integral by

$$
\begin{aligned}
& p(\boldsymbol{y}_i, \tilde{T}_i, \delta_i \mid \boldsymbol{\theta}_y^{(m)}, \widehat{\boldsymbol{\theta}}_t) \\
& \approx \exp\Bigg[\frac{\kappa \log(2\pi) - \log\{\det(\Sigma_{b_i})\}}{2} \\
& \quad + \log p(\boldsymbol{y}_i \mid \widehat{\boldsymbol{b}}_i, \boldsymbol{\theta}_y^{(m)}) + \log p(\tilde{T}_i, \delta_i \mid \widehat{\boldsymbol{b}}_i, \boldsymbol{\theta}_y^{(m)}, \widehat{\boldsymbol{\theta}}_t) \\
& \quad + \log p(\widehat{\boldsymbol{b}}_i \mid \boldsymbol{\theta}_y^{(m)}) \Bigg],
\end{aligned}
$$

where κ denotes the number of random effects for each subject i. Similarly, the outer integral is approximated as

$$
\begin{aligned}
& p(\boldsymbol{y}_i, \tilde{T}_i, \delta_i \mid \boldsymbol{\theta}_y^{(m)}) \\
& \approx \exp\Bigg[\frac{q \log(2\pi) - \log\{\det(\Sigma_\theta)\}}{2} \\
& \quad + \sum_i \log p(\boldsymbol{y}_i, \tilde{T}_i, \delta_i \mid \boldsymbol{\theta}_y^{(m)}, \widehat{\boldsymbol{\theta}}_t) \Bigg].
\end{aligned}
$$

Given the requirement for a double Laplace approximation, and the fact that the denominator does not simplify, the calculation of the $\varpi^{(m)}$ weights given by (10) is more computationally intensive than the ones presented in Sect. 4. Nevertheless, these required computations still remain many orders of magnitude faster than fitting the full joint model.

5.2 Performance

To assess whether updating the random effects in the importance sampling weights alleviates the bias we observed in Sect. 4.2, we have re-analyzed the same simulated datasets. The details are again given in "Appendix A.1.". The results from 500 simulated datasets are presented in Figs. 3, 4 and 6. As anticipated, the corrected two-stage approach with updated random effects added only a small computational cost, with the full multivariate joint model still taking considerably more time to fit than either of the corrected two-stage approaches (Fig. 4). The boxplots depicting the posterior means from the 500 datasets for the parameters of the longitudinal submodels once again demonstrate similar results for all three approaches (Fig. 6). Figure 3 shows the posterior means for the parameters of the survival submodel. We observe that the bias seen for the corrected two-stage approach is now eliminated, with the posterior means from
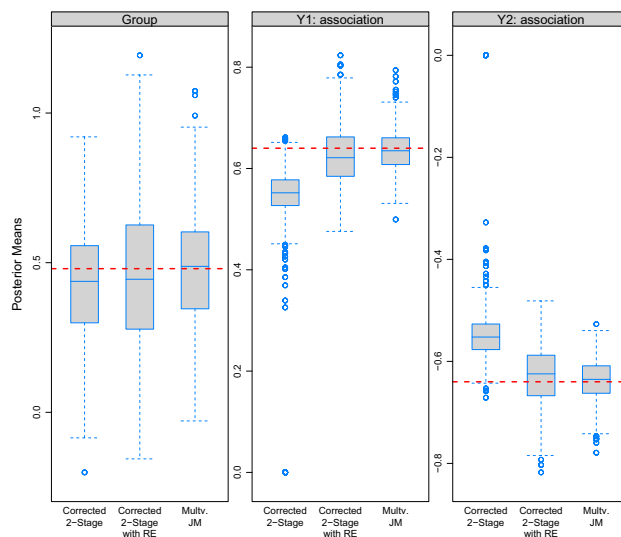

Fig. 3 Simulation results from 500 datasets comparing the importance-sampling-corrected two-stage approach and the corrected two-stage approach with random effects with the full joint model for continuous longitudinal outcomes. The three panels show posterior means from the 500 datasets for the three coefficients in the survival submodel, namely the coefficient for the baseline group variable and the association parameters for the two longitudinal outcomes. The dashed horizontal line indicates the true value of the coefficients

the approach with updated random effects closely approximating those from the full multivariate joint model.

6 Extra simulations

Further simulations were performed in order to assess the performance of the importance-sampling-corrected two-stage approach with the updated random effects, in different scenarios. Details of these simulations are given in Appendices A.2 and A.3.

6.1 Scenario II

Scenario II included six continuous longitudinal outcomes. Owing to the increased number of outcomes, the full multivariate joint model was not run. Table 1 shows the bias for the parameters of the survival submodel, together with the RMSE and coverage (based on the 2.5% and 97.5% credibility intervals for each parameter). Table 4 in the appendix shows the same information for the parameters of the six longitudinal outcomes.

6.2 Scenario III

Scenario III again included six longitudinal outcomes, now of varying types: three continuous and three binary. Table 2

Table 1 Simulation results for parameters of the survival submodel (Scenario II)

Parameter	True value	Bias	RMSE	Coverage
Group	-0.20	-0.02	0.19	0.96
α_1	1.09	-0.01	0.04	0.93
α_2	-1.09	0.01	0.04	0.93
α_3	-0.92	-0.01	0.16	0.95
α_4	0.92	0.03	0.17	0.95
α_5	0.43	0.03	0.05	0.83
α_6	-0.43	-0.02	0.04	0.88

Table 2 Simulation results for parameters of the survival submodel (Scenario III)

Parameter	True value	Bias	RMSE	Coverage
Group	-0.47	-0.01	0.91	0.97
α_1	0.17	-0.01	0.04	1.00
α_2	-0.17	0.01	0.04	1.00
α_3	0.17	0.00	0.43	0.90
α_4	-0.17	0.02	0.88	0.96
α_5	0.47	0.02	0.04	1.00
α_6	-0.47	-0.01	0.05	0.99

demonstrates yet again the alleviation of the bias achieved by updating the random effects.

7 Analysis of the Bio-SHiFT dataset

In this section, we present the analysis of data from the Bio-SHiFT cohort study. During a median follow-up period of 2.4 years (IQR: 2.32–2.45), estimated using the reverse Kaplan–Meier methodology (Shuster 1991), 66/254 (26%) patients experienced the primary event of interest (a composite event, consisting of hospitalization for heart failure, cardiac death, LVAD placement and heart transplantation). Biomarkers were measured at inclusion and subsequently every 3 months until the end of follow-up. We focus on six biomarkers: the glomerular marker cystatin C (CysC), two tubular markers: urinary N-acetyl-beta-D-glucosaminidase (NAG) and kidney injury molecule (KIM)-1, and the markers N-terminal propBNP (NT-proBNP), cardiac troponin T (HsTNT) and C-reactive protein (CRP). The latter three markers are known to be related to poor outcomes in CHF patients and measure various aspects of heart failure pathophysiology (wall stress, myocyte damage and inflammation, respectively). All biomarkers were logarithmically transformed for further analysis (log base 2) due to skewness.

For each of NT-probnp, HsTNT and CRP, we included natural cubic splines in both the fixed and random effects parts of their longitudinal models, with differing numbers of knots per outcome (Fig. 7). Simple linear models with random intercept and slope were used for CysC, NAG and KIM-1. Thus, for each of CysC, NAG and KIM-1 ($k = 1, 2, 3$), we fit:

$$E\{y_{ki}(t) \mid \mathbf{b}_{ki}\} = \beta_{k0} + b_{ki0} + (\beta_{k1} + b_{ki1}) \times \text{time}.$$

For the remaining outcomes ($k = 4, 5, 6$), we have:

$$E\{y_{ki}(t) \mid \mathbf{b}_{ki}\} = (\beta_{k0} + b_{ki0}) + \sum_{p_k}(\beta_{kp_k} + b_{kip_k})B_{kn}(t, \lambda_{p_k}),$$

where $B_{kn}(t; \lambda_{p_k})$ denotes the B-spline basis matrix for a natural cubic spline of time with two internal knots placed at the 25th and 75th percentiles of the follow-up times for NT-probnp ($p = 1, 2, 3$) and one internal knot placed at the 50th percentile of the follow-up times for each of HsTNT and CRP ($p = 1, 2$). Boundary knots were set at the fifth and 95th percentiles. We assume a multivariate normal distribution for the random effects, $\mathbf{b}_i = (\mathbf{b}_{1i}^T, \mathbf{b}_{2i}^T, \ldots, \mathbf{b}_{6i}^T)^T \sim MVN(\mathbf{0}, \mathbf{D})$, where $\mathbf{D}$ is a 16×16 unstructured variance covariance matrix. For the survival process, we included the baseline variables: (standardized) age, sex, NYHA class (class III / IV vs. class I / II), use of diuretics, presence or absence of ischemic heart disease (IHD), diabetes mellitus, (standardized) BMI and the estimated glomerular filtration rate (eGFR) value.

We fit three joint models, using the global–local ridge-type shrinkage prior previously described in each case. Model 1 included only the current underlying value of the longitudinal marker for each of the six markers. Model 2 included the current value and slope for each marker, and model 3 included the integrated longitudinal profile for each marker (AUC). We thus have:

$$\text{Model 1:} \quad h_i(t) = h_0(t)\exp\left[\boldsymbol{\gamma}^\top \boldsymbol{w}_i(t) + \sum_{k=1}^{K}\alpha_k\eta_{ki}(t)\right],$$

$$\text{Model 2:} \quad h_i(t) = h_0(t)\exp\left[\boldsymbol{\gamma}^\top \boldsymbol{w}_i(t) + \sum_{k=1}^{K}\alpha_{1k}\eta_{ki}(t) + \sum_{k=1}^{K}\alpha_{2k}\eta'_{ki}(t)\right],$$

$$\text{Model 3:} \quad h_i(t) = h_0(t)\exp\left[\boldsymbol{\gamma}^\top \boldsymbol{w}_i(t) + \sum_{k=1}^{K}\alpha_k\int_0^t \eta_{ki}(s)\,\mathrm{d}s\right].$$

The parameter estimates and 95% credibility intervals for the event process are presented in Table 3. Hazard ratios are presented per doubling of level, slope or AUC at any point in time. Following adjustment for covariates, the estimated association parameters in Model 1 indicate significant associations between the risk of the composite event and the current underlying values of NT-proBNP and CRP, such that there is a 1.86-fold increase in the risk of the composite event (95% CI: 1.45 to 2.37), per doubling of NT-probnp level, and a 1.44-fold increase in the risk of the composite event

Table 3 Parameter estimates and 95% credibility intervals under the joint modeling analysis for the Bio-SHiFT data

Event process

	HR (95% CI)	HR*	p value
Model 1: Current value			
Age	1.01 (0.99 to 1.03)	1.00	0.242
Sex (male vs. female)	1.00 (0.77 to 1.31)	1.06	0.970
NYHA Class (III/IV vs. I/II)	1.47 (0.97 to 2.64)	2.52	0.134
Diuretics (Yes vs. no)	1.12 (0.79 to 2.74)	0.96	0.770
IHD (Yes vs. no)	1.06 (0.88 to 1.55)	0.95	0.696
eGFR	1.01 (1.00 to 1.02)	1.01	0.048
BMI	1.04 (0.99 to 1.10)	1.03	0.140
Diabetes mellitus	1.08 (0.86 to 1.79)	0.93	0.638
α_{CRP}	1.44 (1.10 to 1.89)	1.25	0.008
α_{HsTNT}	1.32 (0.99 to 1.84)	1.35	0.078
$\alpha_{NT-proBNP}$	1.86 (1.45 to 2.37)	2.20	< 0.0001
α_{CysC}	1.04 (0.55 to 2.48)	0.90	0.966
α_{NAG}	0.89 (0.51 to 1.38)	0.55	0.660
α_{KIM-1}	0.96 (0.74 to 1.22)	1.21	0.768
Model 2: Current value and slope			
Age	1.01 (0.99 to 1.03)	1.00	0.162
Sex (male vs. female)	0.99 (0.66 to 1.34)	0.91	0.994
NYHA Class (III/IV vs. I/II)	1.61 (0.97 to 2.97)	1.53	0.104
Diuretics (Yes vs. no)	1.20 (0.79 to 4.64)	0.95	0.736
IHD (Yes vs. no)	1.06 (0.84 to 1.56)	1.36	0.750
eGFR	1.01 (1.00 to 1.02)	1.01	0.090
BMI	1.04 (0.98 to 1.10)	1.08	0.236
Diabetes mellitus	1.08 (0.84 to 1.73)	1.03	0.676
$\alpha_{CRP\ value}$	1.52 (1.19 to 1.99)	1.38	0.002
$\alpha_{HsTNT\ value}$	1.37 (1.00 to 1.89)	1.08	0.048
$\alpha_{NT-proBNP\ value}$	1.90 (1.48 to 2.44)	1.61	< 0.0001
$\alpha_{CysC\ value}$	1.05 (0.47 to 3.20)	0.92	0.996
$\alpha_{NAG\ value}$	0.74 (0.36 to 1.22)	1.04	0.292
$\alpha_{KIM-1\ value}$	0.92 (0.69 to 1.18)	1.02	0.534
$\alpha_{CRP\ slope}$	0.95 (0.55 to 1.50)	0.91	0.878
$\alpha_{HsTNT\ slope}$	0.84 (0.24 to 1.91)	1.05	0.734
$\alpha_{NT-proBNP\ slope}$	1.02 (0.70 to 1.48)	0.87	0.926
$\alpha_{CysC\ slope}$	1.75 (0.19 to 356.56)	1.35	0.826
$\alpha_{NAG\ slope}$	1.03 (0.08 to 6.63)	0.84	0.946
$\alpha_{KIM-1\ slope}$	2.37 (0.61 to 63.48)	0.93	0.460
Model 3: AUC			
Age	1.00 (0.98 to 1.03)	1.01	0.812
Sex	1.02 (0.83 to 1.38)	1.00	0.944
NYHA class (III/IV vs. I/II)	1.80 (0.99 to 3.25)	1.83	0.080
Diuretics (Yes vs. no)	1.18 (0.83 to 3.72)	1.22	0.680
IHD (Yes vs. no)	1.04 (0.85 to 1.40)	1.21	0.742
eGFR	1.01 (1.00 to 1.02)	1.00	0.166
BMI	1.03 (0.98 to 1.08)	1.02	0.286

Table 3 continued

Event process

	HR (95% CI)	HR*	p value
Diabetes mellitus	1.11 (0.89 to 1.87)	0.99	0.564
α_{CRP}	1.18 (0.99 to 1.43)	1.10	0.064
α_{HsTNT}	1.20 (0.98 to 1.56)	1.06	0.082
$\alpha_{NT-proBNP}$	1.47 (1.21 to 1.79)	1.50	< 0.0001
α_{CysC}	1.13 (0.73 to 2.60)	1.38	0.704
α_{NAG}	0.96 (0.63 to 1.34)	1.12	0.846
α_{KIM-1}	1.00 (0.82 to 1.19)	1.01	0.960

Hazard ratios are presented per doubling of level, slope or AUC at any point in time. HR* is the estimate after importance sampling

(95% CI: 1.1 to 1.89), per doubling of CRP level. No significant associations were found for any of HsTNT, CysC, NAG or KIM-1. Similarly, no significant associations were found for the current underlying values of CysC, NAG or KIM-1 in Model 2, and nor were there any significant associations between the risk of the composite event and the slopes of the six continuous markers.

Model 3 indicates a significant association for NT-proBNP, with a 1.47-fold increase in the risk of the composite event (95% CI: 1.21 to 1.79) per doubling of the area under the NT-proBNP profile.

Since the parameter estimates for each of the longitudinal outcomes remained fairly constant across models, to avoid repetition, the estimates and 95% credibility intervals are presented for one model only (Table 7).

In previous analyses of these same six markers, the current underlying value, instantaneous slope and area under the curve of each marker were each assessed independently of one another. Van Boven et al., 2018 found significant associations in all cases for CRP, HsTNT and NT-proBNP, and Brankovic et al., 2018 found significant associations for the current underlying values and slopes of each of CysC, NAG and KIM-1, and the area under the curves for CysC, and NAG.

Van Boven et al., 2018, provided an additional multivariate analysis for CRP, HsTNT and NT-proBNP, wherein the predicted individual profiles for each marker were separately determined, and functions thereof were simultaneously included in a single extended Cox model as time-varying covariates. Models therefore included either the current underlying values, the instantaneous slopes or the area under the curves for all three markers simultaneously. In that analysis, only CRP and NT-proBNP were found to be independently predictive of the composite event, with significant associations for each of the current underlying values and slopes of these markers. In the model for the area under the curves, only NT-proBNP was significant.

8 Discussion

In this paper, we presented a novel approach for fitting joint models, which allows for the inclusion of multivariate longitudinal outcomes with realistic computing times. We demonstrated once again the bias of the estimated parameters for the survival process characteristic of the standard two-stage approach and proposed the use of an importance-sampling-corrected two-stage approach, with updated random effects, in its place. Our approach was shown to be successful, producing satisfactory results in a number of simulation scenarios: both survival and longitudinal estimates were unbiased, and computing times were reduced by several orders of magnitude, compared to the full multivariate joint model. We were easily able to incorporate multiple outcomes in the analysis of the Bio-SHiFT data, obtaining very similar results to those previously noted for the CHF-related biomarkers (CRP, HsTNT and NT-proBNP). We did not find any significant associations between any of the renal markers (CysC, NAG and KIM-1) and the risk of the composite event in the multivariate analysis, indicating that their predictive value may not be independent of the CHF-related markers. While the simulations included up to six multiple outcomes of varying types, it would be interesting to confirm our results in even more complex settings, (perhaps incorporating competing risks such as those present in the Bio-SHiFT study) and to try determining the limits of the methodology. A further topic for research would be methods for increasing the speed of computation involved in fitting the multivariate mixed model itself, so as to extend the number of outcomes even further.

The proposed importance-sampling-corrected two-stage estimation approach is implemented in function `mvJointModelBayes()` in the freely available package **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). An example of how these functions should be used can be found in the appendix.

Appendices

A simulation study design

A.1 Scenario I

Scenario I simulates 500 patients with a maximum of 15 repeated measurements per patient. We included $K = 2$ continuous longitudinal outcomes and one survival outcome. The k longitudinal outcomes each had form:

$$
\begin{aligned}
y_{ki}(t) &= \eta_{ki}(t) + \epsilon_{ki}(t) \\
&= \beta_{k0} + \beta_{k1} \times \text{time} + \beta_{k2} \times \text{group} \\
&\quad + \beta_{k3} \times \text{interaction} \\
&\quad + b_{ki0} + b_{ki1} \times \text{time} + \epsilon_{ki}(t),
\end{aligned}
$$

with $\epsilon_{ki}(t) \sim N(0, \sigma_k^2)$ and $\mathbf{b}_{ki} = (b_{ki0}, b_{ki1})^T$, with $\mathbf{b}_i = (\mathbf{b}_{1i}^T, \mathbf{b}_{2i}^T)^T \sim MVN(\mathbf{0}, \mathbf{D})$. The variance-covariance matrix $\mathbf{D}$ has general form:

$$
\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & \cdots & D_{\text{else}} \\ \cdots & \mathbf{D}_2 & \cdots \\ \cdots & \cdots & \ddots & \cdots \\ \cdots & & & \mathbf{D}_k \end{bmatrix}, \quad \mathbf{D}_k = \begin{bmatrix} D_{k11} & D_{k12} \\ D_{k21} & D_{k22} \end{bmatrix}
$$

and for Scenario I:

$$
\mathbf{D}_1 = \mathbf{D}_2 = \begin{bmatrix} 0.68 & -0.08 \\ -0.08 & 0.28 \end{bmatrix}, \text{ with } D_{\text{else}} = 0.10
$$

Time was simulated from a uniform distribution between 0 and 25. For the survival outcome, adjusting for group allocation, we used:

$$
\begin{aligned}
h_i(t) &= h_0(t) \exp\left[\gamma_0 + \gamma_1 \times \text{group} + \sum_{k=1}^{K} \alpha_k \eta_{ki}(t) \right] \\
&= h_0(t) \exp\left[\gamma_0 + \gamma_1 \times \text{group} + \alpha_1 \eta_{1i}(t) + \alpha_2 \eta_{2i}(t) \right].
\end{aligned}
$$

The baseline risk was simulated from a Weibull distribution $h_0(t) = \phi t^{\phi-1}$, with $\phi = 1.65$. For the simulation of the censoring times, an exponential censoring distribution was selected, with mean $\mu = 15$, such that the censoring rate was between 60% and 70%. More details are presented in Table 6.

A.2 Scenario II

Scenario II is an extension of Scenario I such that we now have $K = 6$ continuous longitudinal outcomes. We again simulate 500 patients with a maximum of 15 repeated measurements per patient. The k longitudinal outcomes each had form:

$$y_{ki}(t) = \eta_{ki}(t) + \epsilon_{ki}(t)$$
$$= \beta_{k0} + \beta_{k1} \times \text{time} + \beta_{k2} \times \text{group}$$
$$+ \beta_{k3} \times \text{interaction}$$
$$+ b_{ki0} + b_{ki1} \times \text{time} + \epsilon_{ki}(t),$$

with $\epsilon_{ki}(t) \sim N(0, \sigma_k^2)$ and $\mathbf{b}_{ki} = (b_{ki0}, b_{ki1})^T$, with $\mathbf{b}_i = (\mathbf{b}_{1i}^T, \mathbf{b}_{2i}^T, \ldots, \mathbf{b}_{6i}^T)^T \sim MVN(\mathbf{0}, \mathbf{D})$,

$$\mathbf{D}_1 = \mathbf{D}_2 = \begin{bmatrix} 0.97 & 0.78 \\ 0.07 & 0.032 \end{bmatrix}, \mathbf{D}_3 = \mathbf{D}_4 = \begin{bmatrix} 0.13 & -0.009 \\ -0.009 & 0.002 \end{bmatrix},$$

$$\mathbf{D}_5 = \mathbf{D}_6 = \begin{bmatrix} 0.56 & 0.05 \\ 0.05 & 0.03 \end{bmatrix}, \text{ and } \mathbf{D}_{\text{else}} = 0.00$$

Time was simulated from a uniform distribution between 0 and 25. For the survival outcome, adjusting for group allocation as in Scenario I, we used:

$$h_i(t) = h_0(t) \exp\left[\gamma_1 \times \text{group} + \sum_{k=1}^{K} \alpha_k \eta_{ki}(t)\right].$$

The baseline risk was simulated using B-splines with knots specified a priori. An exponential censoring distribution was used for the simulation of the censoring times, with mean $\mu = 15$, such that the censoring rate was between 60% and 70%. Further details are again available in Table 6.

A.3 Scenario III

In Scenario III, we simulate 500 patients with a maximum of 15 repeated measurements per patient, including three continuous and three binary longitudinal outcomes such that:

$$g_k\big[E\{y_{ki}(t) \mid \mathbf{b}_{ki}\}\big] = \eta_{ki}(t)$$
$$= \beta_{k0} + \beta_{k1} \times \text{time} + \beta_{k2} \times \text{group} + \beta_{k3} \times \text{interaction}$$
$$+ b_{ki0} + b_{ki1} \times \text{time},$$

where $g_k(\cdot)$ denotes the canonical link function appropriate to the response type (identity and logit for the Gaussian and binomial outcomes, respectively), and $\mathbf{b}_i = (\mathbf{b}_{1i}^T, \mathbf{b}_{2i}^T, \ldots, \mathbf{b}_{6i}^T)^T \sim MVN(\mathbf{0}, \mathbf{D})$, with

$$\mathbf{D}_1 = \mathbf{D}_2 = \begin{bmatrix} 0.97 & 0.78 \\ 0.07 & 0.032 \end{bmatrix}, \mathbf{D}_3 = \mathbf{D}_4 = \begin{bmatrix} 0.13 & -0.009 \\ -0.009 & 0.002 \end{bmatrix},$$

$$\mathbf{D}_5 = \mathbf{D}_6 = \begin{bmatrix} 0.56 & 0.05 \\ 0.05 & 0.03 \end{bmatrix}, \text{ and } \mathbf{D}_{\text{else}} = 0.00$$

For the survival outcome, adjusting for group allocation, we again used:

$$h_i(t) = h_0(t) \exp\left[\gamma_1 \times \text{group} + \sum_{k=1}^{K} \alpha_k \eta_{ki}(t)\right].$$

Scenario III maintains the use of the uniform distribution between 0 and 25 for time and the use of B-splines for the simulation of the baseline hazard. The censoring times were simulated using an exponential censoring distribution as before, with mean $\mu = 15$. Table 6 provides additional information.

B Example R Code

The below code fits a multivariate joint model for $K = 3$ longitudinal outcomes: y_1, y_2 and y_3, where y_1 is binary and both y_2 and y_3 are continuous. We fit a linear mixed model for y_1 with random intercept and slope (time is $futime$) and use natural cubic splines with two knots, (at $futime = 6$ and $futime = 15$ respectively) in both the fixed and random parts of the models for y_2 and y_3. The survival submodel adjusts for continuous baseline predictors x_1 and x_2.

```
# LIBRARIES: JMbayes, splines, survival

 # MULTIVARIATE MIXED MODEL

 MixedModel <-
mvglmer(list(y1 ~ futime + (futime | id),
  y2 ~ ns(futime, knots = c(6, 15),
    Boundary.knots = c(0, 27)) +
    (ns(futime, knots = c(6, 15),
    Boundary.knots = c(0, 27)) | id),
  y3 ~ ns(futime, knots = c(6, 15),
    Boundary.knots = c(0, 27)) +
    (ns(futime, knots = c(6, 15),
    Boundary.knots = c(0, 27)) | id)),
    data = longit,
    families =
    list(binomial, gaussian, gaussian))

 # SURVIVAL SUB MODEL

SurvFit <- coxph(Surv(months, pe) ~ x1 + x2,
    data = surv,
    model = TRUE)

 # MULTIVARIATE JOINT MODEL

JointFitAll <- mvJointModelBayes(MixedModel,
    SurvFit,
    timeVar = "futime")
 summary(JointFitAll, TRUE)
 # , TRUE is necessary to obtain the
 importance-sampling weighted estimates
```

C Tables

See Tables 4, 5, 6 and 7.

Table 4 Simulation results for parameters of the longitudinal submodel (Scenario II)

Parameter	True	Bias	RMSE	Coverage
Y1: Intercept	0.73	0.00001	0.06	0.96
Y1: Group	− 0.30	0.00000	0.09	0.95
Y1: Interaction	− 0.07	− 0.00001	0.02	0.92
Y1: Time	0.24	0.00001	0.01	0.95
Y1: Sigma	0.34	0.00000	0.00	0.95
Y2: Intercept	0.73	0.00001	0.06	0.94
Y2: Group	− 0.30	0.00000	0.09	0.93
Y2: Interaction	− 0.07	0.00000	0.02	0.97
Y2: Time	0.24	0.00001	0.01	0.96
Y2: Sigma	0.34	0.00001	0.00	0.96
Y3: Intercept	5.75	0.00000	0.02	0.96
Y3: Group	0.04	0.00000	0.03	0.95
Y3: Interaction	0.013	0.00000	0.00	0.96
Y3: Time	0.07	0.00000	0.00	0.96
Y3: Sigma	0.19	0.00000	0.00	0.96
Y4: Intercept	5.75	− 0.00001	0.02	0.96
Y4: Group	0.04	0.00001	0.03	0.95
Y4: Interaction	0.013	0.00001	0.00	0.95
Y4: Time	0.07	0.00001	0.00	0.96
Y4: Sigma	0.19	0.00000	0.00	0.94
Y5: Intercept	10.97	− 0.00001	0.06	0.95
Y5: Group	− 0.42	0.00000	0.09	0.95
Y5: Interaction	0.01	0.00000	0.02	0.95
Y5: Time	0.22	0.00000	0.01	0.96
Y5: Sigma	1.06	0.00001	0.01	0.93
Y6: Intercept	10.97	− 0.00001	0.07	0.94
Y6: Group	− 0.42	0.00001	0.09	0.94
Y6: Interaction	0.01	0.00001	0.02	0.95
Y6: Time	0.22	0.00001	0.01	0.95
Y6: Sigma	1.06	0.00000	0.01	0.94

Table 5 Simulation results for parameters of the longitudinal submodel (Scenario III)

Parameter	True	Bias	RMSE	Coverage
Y1: Intercept	10.98	0.00002	0.07	0.97
Y1: Group	− 0.45	0.00000	0.11	0.95
Y1: Interaction	0.05	− 0.00001	0.02	0.94
Y1: Time	0.21	0.00002	0.01	0.95
Y1: Sigma	1.10	− 0.00001	0.01	0.95
Y2: Intercept	10.98	0.00001	0.07	0.97
Y2: Group	− 0.45	− 0.00001	0.11	0.96
Y2: Interaction	0.05	− 0.00001	0.02	0.95
Y2: Time	0.21	0.00001	0.01	0.96
Y2: Sigma	1.10	0.00002	0.01	0.96
Y3: Intercept	10.98	− 0.00001	0.05	0.93
Y3: Group	− 0.45	0.00000	0.07	0.94
Y3: Interaction	0.05	0.00000	0.01	0.93
Y3: Time	0.21	0.00001	0.01	0.93
Y3: Sigma	1.10	0.00000	0.01	0.94
Y4: Intercept	1.11	− 0.00002	0.10	0.94
Y4: Group	− 1.09	0.00001	0.13	0.95
Y4: Interaction	0.01	0.00002	0.02	0.93
Y4: Time	0.14	0.00001	0.02	0.93
Y5: Intercept	1.11	0.00000	0.12	0.95
Y5: Group	− 1.09	0.00000	0.15	0.97
Y5: Interaction	0.01	0.00000	0.03	0.96
Y5: Time	0.14	− 0.00002	0.03	0.94
Y6: Intercept	1.11	− 0.00001	0.12	0.95
Y6: Group	− 1.09	− 0.00001	0.15	0.96
Y6: Interaction	0.01	0.00001	0.03	0.96
Y6: Time	0.14	0.00002	0.03	0.93

Table 6 Simulation scenarios

Scenario	Y	α_k	$\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{k1}, \beta_{k2}, \beta_{k3})$	σ_k	$\boldsymbol{\gamma}_k = (\gamma_{k0}, \gamma_{k1})$
I	1	0.64	(2.13, 0.24, −0.25, −0.05)	0.60	(−5.8, 0.48)
	2	−0.64			
II	1	1.09	(0.73, 0.24, −0.30, −0.07)	0.34	(na, −0.20)
	2	−1.09			
	3	−0.92	(5.75, 0.07, 0.04, 0.013)	0.19	(na, −0.20)
	4	0.92			
	5	0.43	(10.97, 0.22, −0.42, 0.01)	1.06	(na, −0.20)
	6	−0.43			
III	1	0.17	(10.98, 0.21, −0.45, 0.05)	1.10	(na, −0.47)
	2	−0.17			
	3	0.17			
	4	−0.17	(1.11, 0.14, −1.09, 0.01)	na	(na, −0.47)
	5	0.47			
	6	−0.47			

Table 7 Parameter estimates and 95% credibility intervals under the joint modeling analysis for the Bio-SHiFT data (Model 1)

Longitudinal process	Est. (95% CI)	Est.*	p value
Cystatin C			
Intercept	−0.36 (−0.42 to −0.31)	−0.38	<0.0001
Time (years since baseline)	−0.04 (−0.07 to −0.01)	−0.05	0.004
σ	0.43 (0.42 to 0.45)	0.43	<0.0001
NAG			
Intercept	2.46 (2.34 to 2.58)	2.42	<0.0001
Time (years since baseline)	−0.06 (−0.13 to 0.00)	−0.07	0.058
σ	0.93 (0.90 to 0.97)	0.93	<0.0001
KIM-1			
Intercept	8.93 (8.78 to 9.07)	8.99	<0.0001
Time (years since baseline)	0.01 (−0.05 to 0.08)	0.02	0.690
σ	0.85 (0.82 to 0.88)	0.81	<0.0001
CRP			
Intercept	1.06 (0.87 to 1.26)	1.05	<0.0001
$B_n(Time, \lambda_1)$	0.96 (0.65 to 1.26)	1.18	<0.0001
$B_n(Time, \lambda_2)$	0.50 (0.31 to 0.69)	0.35	<0.0001
σ	1.00 (0.96 to 1.04)	0.99	<0.0001
HsTNT			
Intercept	4.17 (4.02 to 4.32)	4.17	<0.0001
$B_n(Time, \lambda_1)$	0.17 (0.05 to 0.29)	0.26	0.010
$B_n(Time, \lambda_2)$	0.17 (0.10 to 0.26)	0.17	<0.0001
σ	0.28 (0.27 to 0.29)	0.29	<0.0001
NTS-proBNP			
Intercept	6.78 (6.54 to 7.01)	6.59	<0.0001
$B_n(Time, \lambda_1)$	0.00 (−0.17 to 0.18)	0.12	1.000
$B_n(Time, \lambda_2)$	−0.04 (−0.30 to 0.23)	0.05	0.750
$B_n(Time, \lambda_3)$	0.17 (0.03 to 0.34)	0.23	0.010
σ	0.50 (0.48 to 0.52)	0.50	<0.0001

Est.* are the estimates after importance sampling

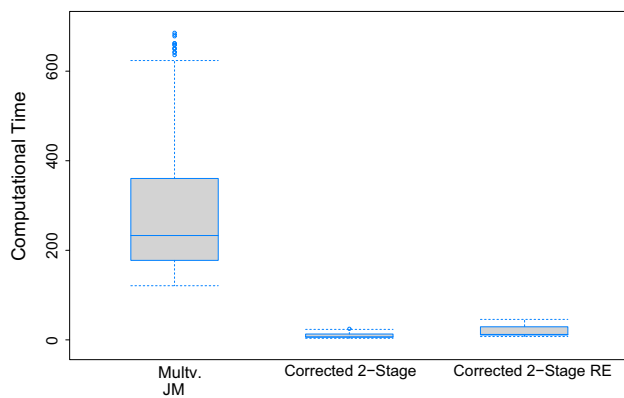# D Figures

See Figs. 4, 5, 6 and 7.

Fig. 4 Simulation results from 500 datasets comparing the importance-sampling-corrected two-stage approach with and without updated random effects with the full multivariate joint model. The boxplots show the mean computational time per approach (in minutes)
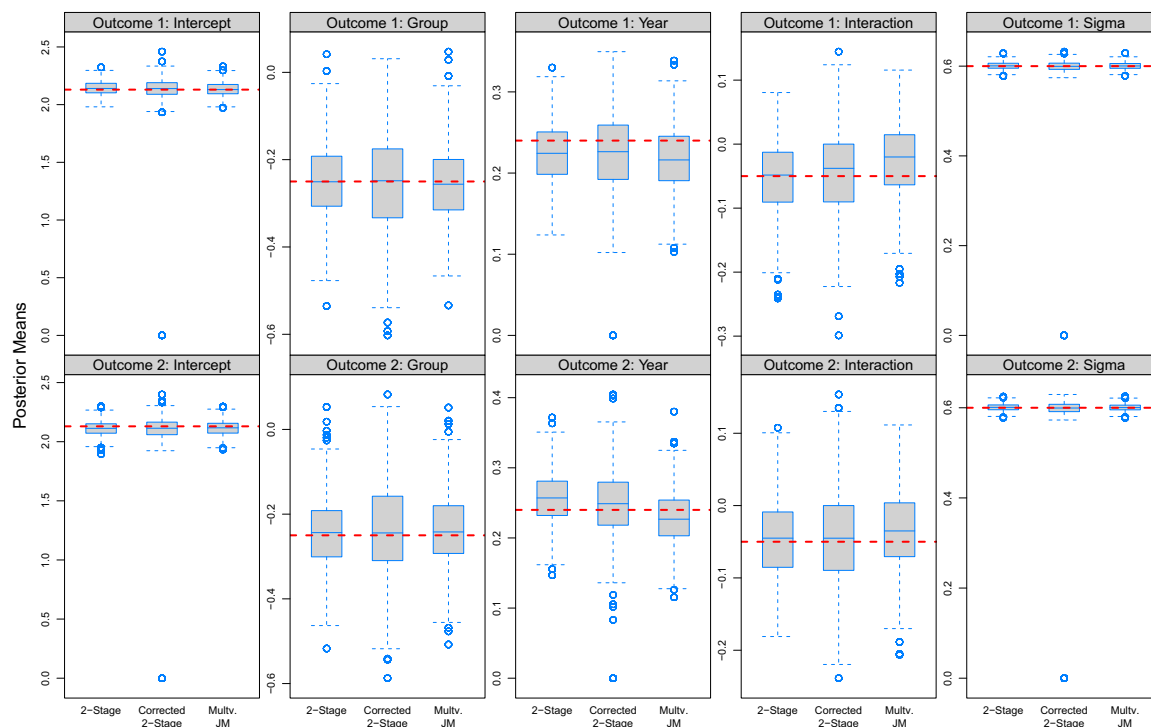

Fig. 5 Simulation results from 500 datasets comparing the simple two-stage approach and the importance-sampling-corrected two-stage approach with the full multivariate joint model. The panels show the posterior means from the 500 datasets for the coefficients from the two longitudinal outcomes in Scenario I. The dashed horizontal line indicates the true value of the coefficients
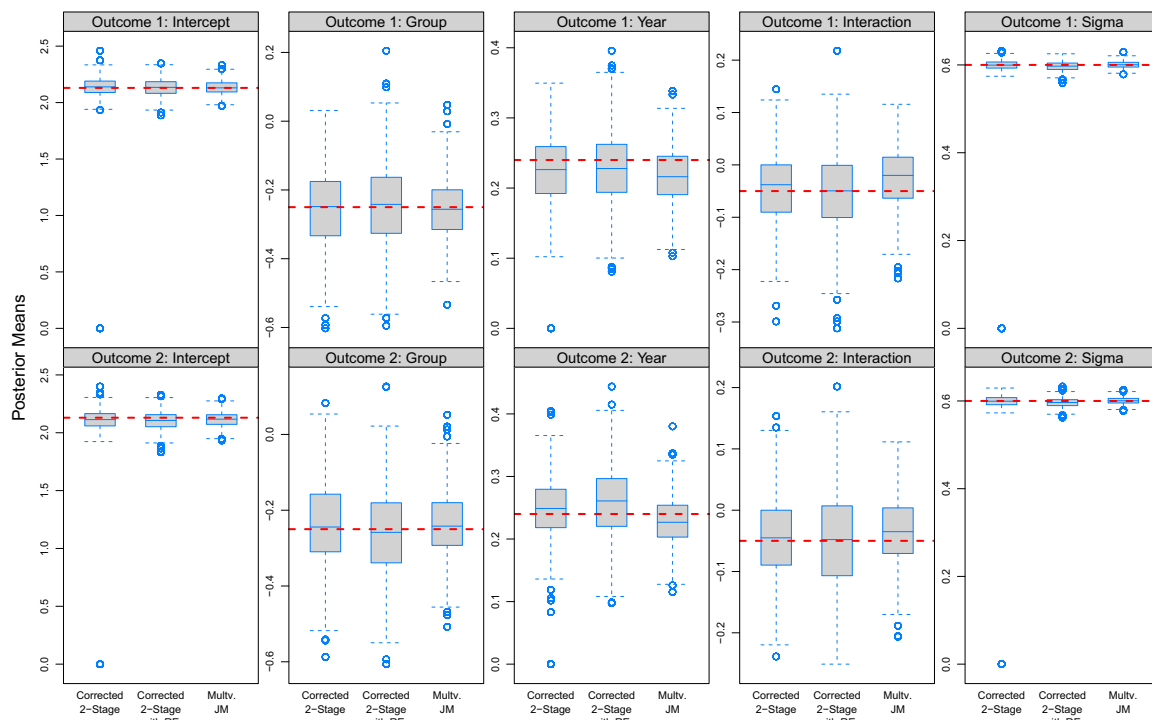
Fig. 6 Simulation results from 500 datasets comparing the importance-sampling-corrected two-stage approach with and without updated random effects with the full multivariate joint model. The panels show the posterior means from the 500 datasets for the coefficients from the two longitudinal outcomes in Scenario I. The dashed horizontal line indicates the true value of the coefficients

Fig. 7 Longitudinal profiles of continuous biomarkers (log base 2) for a randomly selected subset of individuals from the Bio-SHiFT cohort study that did/did not experience the primary event of interest
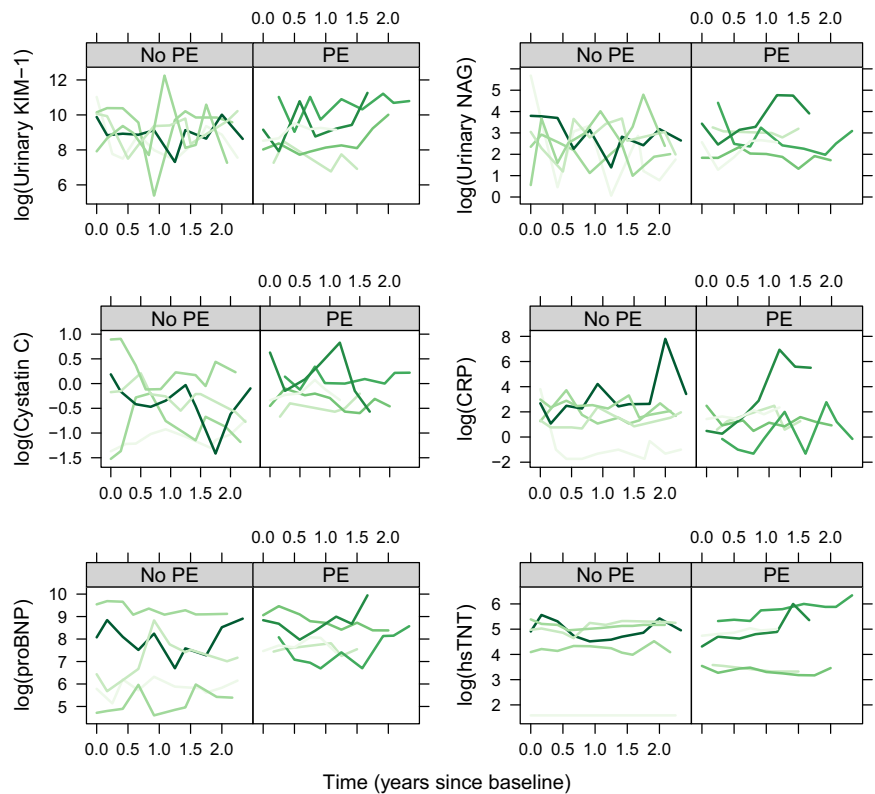
References

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

Andrinopoulou, E.R., Rizopoulos, D., Takkenberg, J., Lesaffre, E.: Joint modeling of two longitudinal outcomes and competing risk data. Stat. Med. **33**, 3167–3178 (2014)

Andrinopoulou, E.R., Eilers, P.H.C., Takkenberg, J.J.M., Rizopoulos, D.: Improved dynamic predictions from joint models of longitudinal and survival data with time-varying effects using P-splines. Biometrics (2018). https://doi.org/10.1111/biom.12814

Brown, E.R.: Assessing the association between trends in a biomarker and risk of event with an application in pediatric HIV/AIDS. Ann. Appl. Stat. **3**, 1163–1182 (2009)

Brown, E.R., Ibrahim, J.G., DeGruttola, V.: A flexible B-spline model for multiple longitudinal biomarkers and survival. Biometrics **61**, 64–73 (2005a)

Brown, E.R., Ibrahim, J.G., DeGruttola, V.: A flexible B-spline model for multiple longitudinal biomarkers and survival. Biometrics **61**(1), 64–73 (2005b)

Chi, Y.Y., Ibrahim, J.G.: Joint models for multivariate longitudinal and multivariate survival data. Biometrics **62**(2), 432–445 (2006)

Elashoff, R., Li, G., Li, N.: A joint model for longitudinal measurements and survival data in the presence of multiple failure types. Biometrics **64**, 762–771 (2008)

Faucett, C., Thomas, D.: Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Stat. Med. **15**, 1663–1685 (1996)

Ferrer, L., Rondeau, V., Dignam, J., Pickles, T., Jacqmin-Gadda, H., Proust-Lima, C.: Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. Stat. Med. **35**, 3933–3948 (2016)

Jullion, A., Lambert, P.: Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. Comput. Stat. Data Anal. **51**, 2542–2558 (2007)

Lang, S., Brezger, A.: Bayesian P-splines. J. Comput. Gr. Stat. **13**, 183–212 (2004)

Lewandowski, D., Kurowicka, D., Joe, H.: Generating random correlation matrices based on vines and extended onion method. J. Multivar. Anal. **100**, 1989–2001 (2009)

Lin, H., McCulloch, C.E., Mayne, S.T.: Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. Stat. Med. **21**(16), 2369–2382 (2002)

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

Proust-Lima, C., Taylor, J.M.G.: Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. Biostatistics **10**, 535–549 (2009)

Rizopoulos, D.: Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometrics **67**, 819–829 (2011)

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

Rizopoulos, D., Ghosh, P.: A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Stat. Med. **30**, 1366–1380 (2011a)

Rizopoulos, D., Ghosh, P.: A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. Stat. Med. **30**, 1366–1380 (2011b)

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

Rizopoulos, D., Molenberghs, G., Lesaffre, E.M.E.H.: Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. Biom. J. **59**, 1261–1276 (2017)

Shuster, J.J.: Median follow-up in clinical trials. J. Clin. Oncol. **9**(1), 191–192 (1991)

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

van Boven, N., Battes, L.C., Akkerhuis, K.M., Rizopoulos, D., Caliskan, K., Anroedh, S.S., et al.: Toward personalized risk assessment in patients with chronic heart failure: detailed temporal patterns of NT-proBNP, troponin T, and CRP in the Bio-SHIFT study. Am. Heart J. **196**, 36–48 (2018)

Wulfsohn, M., Tsiatis, A.: A joint model for survival and longitudinal data measured with error. Biometrics **53**, 330–339 (1997)

Ye, W., Lin, X., Taylor, J.: Semiparametric modeling of longitudinal measurements and time-to-event data—a two stage regression calibration approach. Biometrics **64**, 1238–1246 (2008)