JOINT MODELS FOR TIME-TO-EVENT AND MULTIVARIATE LONGITUDINAL DATA: A LIKELIHOOD APPROACH

Marcella Mazzoleni[1]
Department of Statistics and Quantitative Methods, University of Milano-Bicocca Via Bicocca degli Arcimboldi, 8, 20126 Milano, Italy

Abstract. *Joint models analyse the effects of longitudinal covariates on the risk of one or more events. The models are composed of two sub-models: a longitudinal model and a survival model. The longitudinal sub-model is typically a multivariate mixed model that considers fixed and random effects. The survival sub-model is usually a Cox proportional-hazards model that jointly considers the influence of more than one longitudinal covariate on the risk of the event. This study extends the estimation method based on a joint-likelihood formulation used in the univariate case to a multivariate longitudinal sub-model. The parameters are estimated by maximising the likelihood function using an expectation-maximisation algorithm. Here, the M-step employs a one-step Newton–Raphson update because it is not possible to obtain a closed-form expression for some of the parameter estimators. In addition, a Gauss–Hermite approximation is applied for some of the integrals.*

Keywords: *Joint Model; Multivariate Mixed Model; EM Algorithm; Joint Likelihood*

1. INTRODUCTION

Joint models analyse the effects of longitudinal covariates on the risk of one or more events. A longitudinal model and a survival sub-model compose the joint models. The longitudinal sub-model is commonly a multivariate mixed model that considers fixed and random effects. The survival sub-model is usually a Cox proportional-hazards model that jointly considers the influence of more than one longitudinal covariate on the risk of the event. This study extends the estimation method based on a joint-likelihood formulation used in the univariate case Rizopoulos (2012) to include a multivariate longitudinal sub-model. The parameters are estimated by maximising the likelihood function using an expectation-maximisation (EM) algorithm. In the E-step a Gauss–Hermite approximation is applied for some of the integrals, while the M-step employs a one-step Newton–Raphson update because it is not possible to obtain a closed-form expression for

[1] Marcella Mazzoleni, email: marcella.mazzoleni@unimib.it

some of the parameter estimators.
The first study to extend the joint model to include a multivariate longitudinal sub-model was that of Xu and Zeger (2001a). They propose a latent-variable model to jointly analyse the time to an event and repeated measures of multiple surrogate marker processes. The authors use two complementary approaches in which they compare the lengths of the predictive intervals to determine whether using multiple surrogate processes is better than using only one. They use a Markov chain Monte Carlo (MCMC) algorithm to estimate the parameters of the model, extending the univariate estimation method introduced by Xu and Zeger (2001b) and Faucett and Thomas (1996). The authors apply the model and the estimation method to schizophrenia trial data of risperidone, comparing the models with one and three biomarkers. They find a gain in precision, but that this gain is not sufficient to warrant the additional risk of bias from the more complex model.
Another work to conduct a joint analysis of time-to-event and multiple longitudinal variables was that of Lin et al. (2001), who extend the univariate estimation method of Wulfsohn and Tsiatis (1997). Their model allows for the simultaneous direct dependence of the event process on multiple longitudinal covariates, in addition to accommodating correlations between the covariates. They use a one-step-late EM algorithm to handle the direct dependence of the event process on the modelled longitudinal variables, along with the presence of other fixed covariates in both processes. The authors apply this new method to a data set of a beta-carotene trial, showing the benefits of the joint modelling of the longitudinal and time-to-event variables.
Subsequently, Song et al. (2002) extended the univariate estimation method presented by Tsiatis and Davidian (2001) to a semi-parametric conditional score estimation. They assume a proportional-hazards regression model for the survival sub-model, where the relationship between the hazard and the covariates is defined by a function that allows for flexibility. They apply the model to a data set on AIDS clinical trials, analysing the time trajectories of CD4 and CD8. Their results show that this estimation method works well, which they confirm using a simulation study.
Brown et al. (2005) propose a joint longitudinal and survival model with a non-parametric model for longitudinal markers. They use cubic B-splines to specify the longitudinal model, and a proportional-hazard model to link the longitudinal measures to the hazard. After posing several priors and using various rules for the approximations, the authors implement the Gibbs sampling method to estimate the parameters of the model. Then, they conduct a simulation study and apply the

model to an AIDS data set to determine the efficiency of the model. They show that the cubic B-spline model provides a better fit to the longitudinal data than can be obtained using simple parametric models.

Fieuws et al. (2008) also propose a multivariate mixed model, where they specify a joint distribution for the random effects. They combine the univariate mixed models into a multivariate mixed model by specifying a joint distribution for all of the random effects. In order to obtain estimations, they use a pairwise modelling strategy that fits all possible pairs of bivariate mixed models. Here, they use a pattern-mixture approach, pseudo-likelihood theory, and a Monte Carlo integration. The authors applied this model to analyse renal graft failure using several biomarkers.

Albert and Shih (2010) propose a regression calibration approach that appropriately accounts for informative drop-out to jointly model multiple longitudinal measurements and discrete time-to-event data. The authors argue that numerical integration techniques and the Monte Carlo method do not perform well, even for moderately high-dimensional random effects. Therefore, they propose a two-stage regression calibration approach. In the first stage, multivariate linear mixed models are used to model the longitudinal data. In the second stage, the time-to-event model is estimated by replacing the random effects with corresponding empirical Bayes estimates. Here, the discrete event-time distribution is modelled as a linear function of previous true values of the biomarkers, without measurement error, on a probit scale. The benefits of the models are shown in a simulation study in which they examine the effects of multiple longitudinal biomarkers on the short-term prognosis of patients with primary biliary cirrhosis (PBC).

Rizopoulos and Ghosh (2011) propose a new semiparametric multivariate joint model that relates multiple longitudinal outcomes to a time-to-event model. In particular, they use a spline-based approach for the subject-specific longitudinal evolutions, assume that the baseline risk function is piecewise constant, and model the distribution of the latent terms using a Dirichlet process prior formulation. To allow for flexible shapes of the subject-specific evolutions for each outcome, the authors suggest using a spline-based approach. They propose three parametrisations of the function that links the longitudinal covariates and the risk of the event. Then, they compare these in a simulation study and by analysing renal graft failure, using a Bayesian formulation for the joint semi-parametric multivariate joint model.

Choi et al. (2014) implement a joint model for mixed multivariate longitudinal measurements. Specifically, the authors formulate a unified Bayesian joint model

for mixed longitudinal responses and time-to-event outcomes. They build the log-likelihood for the observed data and implement a Bayesian approach for the parameter inferences using the Gibbs sampling algorithm. Then, they apply their estimation method in a simulation study and to mortality in a study on idiopathic pulmonary fibrosis outcomes.
He and Luo (2013) develop a joint model that consists of a multilevel item response theory model for the multiple longitudinal outcomes and a Cox proportional-hazard model with piecewise constant baseline hazards for the event-time data. Shared random effects are used to link the two models. Model inferences are conducted using a Bayesian framework in an MCMC simulation, implemented in the BUGS language. The model is applied to analysing Parkinson's disease.
Hickey et al. (2016) conduct an interesting review of joint models of time-to-event and multivariate longitudinal outcomes. The authors analyse longitudinal data, as well as the distribution and model assumptions, association structures, estimation approaches, software tools used in the implementation, and clinical applications of the methodologies. They highlight that despite developments in this area, there is a lack of software for estimating the parameters, which has translated into limited uptake by medical researchers. For this reason, Hickey et al. (2017) implemented a new package in R, called the joineRML package. This package fits the joint model proposed by Henderson et al. (2000), extended for multiple continuous longitudinal measures. The time-to-event data are modelled using a Cox proportional-hazards regression model with time-varying covariates, and the multiple longitudinal outcomes are modelled using a multivariate version of the Laird and Ware (1982) linear mixed model. The association is captured by a multivariate latent Gaussian process, and the parameters are estimated using a Monte Carlo EM algorithm.
Recently, Mazzoleni (2018) implemented a two-stage approach for estimating joint models with multivariate longitudinal sub-models. The author presents a simulation study and applies the estimation method to analyse undergraduates' paths in an Italian university, analysing the effect of one or more longitudinal covariates on the graduation event.
The model here proposed shows a different formulation from the models used by Rizopoulos (2010) and by Hickey et al. (2017), in fact this work aims to extend the model proposed by Rizopoulos (2010) using a multidimensional longitudinal approach, analysing the relation between the hazard of the event and the true and unobserved value of the longitudinal covariates, not only the relation between hazard and random effects, as in Hickey et al. (2017). In addition, in the estimation

method a Gauss–Hermite quadrature rule is used for the EM algorithm.
The remainder of the paper proceeds as follows. The second section presents the model and the estimation method. The third and fourth sections provide a simulation study and an application to a well-known data set, respectively. The final section concludes the paper and proposes ideas for future research.

2. THE MODEL

The joint model comprises longitudinal and survival sub-models. A proportional-hazard model is used for the survival sub-model, defined as a function of $m_{iq}(t)$, denoting the true and unobserved value of the longitudinal covariate q for subject i:

$$h_i(t|M_i(t),\omega_i) = h_0(t)\exp\left[\gamma'\omega_i + \sum_q \alpha_q m_{iq}(t)\right]. \tag{1}$$

In (1):

- $M_i(t) = \{m_{iq}(s), 0 \le s < t, \forall q = 1,\dots,Q\}$ indicates the history of the true unobserved longitudinal processes up to time t,
- α_q quantifies the effect of the longitudinal outcome q on the risk of an event,
- $h_0(t)$ indicates the baseline hazard function, and
- ω_i are the covariates that affect the risk of the event with coefficient γ.

In addition, in the survival sub-model T_i is the observed event time for the subject i defined as the minimum of the potential censoring time and the true event time, and δ_i is equal to 1 if the event occurs, 0 otherwise.
With regard to the longitudinal sub-model, the following linear multivariate mixed model is proposed:

$$\begin{cases} y_{iq}(t) = m_{iq}(t) + \varepsilon_{iq}(t) \\ m_{iq}(t) = x'_{iq}(t)\beta_q + z'_{iq}(t)b_{iq} \\ \varepsilon_{iq}(t) \sim N(0,\sigma_q^2) \\ b'_i = (b'_{1q},\dots,b'_{iQ}) \sim N(0,D) \\ b_{1q},\dots,b_{nQ},\varepsilon_{1q},\dots,\varepsilon_{nQ} \quad independent, \end{cases} \tag{2}$$

where q is the longitudinal variable index, $y_{iq}(t)$ is composed of $m_{iq}(t)$ and a random error term $\varepsilon_{iq}(t)$, β_q are the fixed effects for $x_{iq}(t)$, and b_{iq} are the random

effects for $z_{iq}(t)$. σ_q^2 is the variance of the random error term and D denotes random-effects variance–covariance matrix. These elements will be used in the subsequent sub-section in matrix form: $y_i' = (y_{1i}', ...y_{iq}', ..., y_{iQ}')$ where each vector y_{iq} is composed stacking the single elements $y_{iq}(t)$, $\beta = [\beta_1, ..., \beta_q, ..., \beta_Q]$ and $\sigma^2 = [\sigma_1^2, , ..., \sigma_q^2, ..., \sigma_Q^2]$. In addition, X_{iq} and Z_{iq} are the design matrices (with corresponding row vectors $x_{iq}'(t)$ and $z_{iq}'(t)$).

As anticipated in the introduction, the model formulation proposed is different from the models used in Rizopoulos (2010) and Hickey et al. (2017). In fact, Rizopoulos (2010) proposes an univariate formulation that relates the hazard of the event with the true and unobserved value of a longitudinal covariate $m_i(t)$, as follows:

$$h_i(t|M_i(t), \omega_i) = h_0(t) \exp\left[\gamma' \omega_i + \alpha m_i(t)\right]$$

The main difference between this and the (1) lies in the parameter α_q: a parameter for each longitudinal covariate q is introduced ($q = 1, \ldots, Q$).

While, Hickey et al. (2017) propose a multivariate formulation analysing the relation between hazard and only the random effects b_{iq}, as follows:

$$h_i(t|M_i(t), \omega_i) = h_0(t) \exp\left\{\gamma' \omega_i + \sum_q \alpha_q \left[z_{iq}'(t) b_{iq}\right]\right\}$$

Whereas the model here proposed shows a multidimensional longitudinal extension of Rizopoulos (2010) model and relates the hazard of the event with the true and unobserved value of each of the Q longitudinal covariates, recalling (1):

$$h_i(t|M_i(t), \omega_i) = h_0(t) \exp\left[\gamma' \omega_i + \sum_q \alpha_q m_{iq}(t)\right].$$

There are two classes of estimation methods, namely, the two-stage approach and the joint-likelihood formulation. The two-stage approach is biased, but is less computationally demanding, while the joint-likelihood method is more efficient, but computationally slower.

The two-stage approach is based on two steps. In the first step, the random effects are estimated using a least-squares approach. In the second step, the estimates from the first step are used to impute appropriate values of $m_{iq}(t)$, which are then substituted into the classical partial likelihood of the Cox model.

The joint-likelihood method maximises the likelihood function using Bayesian or classical methods.

Rizopoulos (2012) proposes a new estimation method based on a joint-likelihood

formulation that maximises the log-likelihood function using the EM and Newton–Raphson algorithms, and a Gauss–Hermite quadrature rule. The author supposes that the vector of random effects b_i underlies both the longitudinal and the survival processes. Whereas, Hickey et al. (2016) maximise the joint-likelihood function using a Monte Carlo EM algorithm.

2.1. THE JOINT-LIKELIHOOD APPROACH

The aim of the study is to extend the estimation method of Rizopoulos (2012) to include a multivariate longitudinal sub-model. For each subject i, the classical log-likelihood equation is defined as:

$$\begin{aligned}
\log p(T_i, \delta_i, y_i; \theta) &= \log \int p(T_i, \delta_i, y_i, b_i; \theta) db_i \\
&= \log \int p(T_i, \delta_i, y_i | b_i; \theta, \beta) p(b_i, \theta) db_i \\
&= \log \int p(T_i, \delta_i | b_i; \theta_t, \beta) p(y_i | b_i; \theta_y) p(b_i, \theta_b) db_i \\
&= \log \int p(T_i, \delta_i | b_i; \theta_t, \beta) \left\{ \prod_q p(y_{iq} | b_{iq}; \theta_y) \right\} p(b_i, \theta_b) db_i
\end{aligned}$$

where $\theta = (\theta_t', \theta_y', \theta_b')'$ denotes the full parameter vector, where θ_t are the parameters for the event-time outcome, θ_y are the parameters for the longitudinal outcomes, and θ_b are the parameters of the random-effects variance–covariance matrix. In the formula, $\theta_y = [\beta', \sigma^2]$, $\theta_t = [\gamma', \alpha_1, ..., \alpha_q, ..., \alpha_Q, \theta_{h_0}]$, where θ_{h_0} is used when the baseline hazard is parametric; and $\theta_b = [vech(D)]$. The log-likelihood can be separated into three parts, each of which is related to part of the parameters vector.
The first part is related to the parameter vector θ_t (i.e. the parameters for the event-time outcome), and is defined as follows:

$$\begin{aligned}
p(T_i, \delta_i | b_i; \theta_t, \beta) &= \left\{ h_0(T_i) \exp\left[\gamma' \omega_i + \sum_{q=1}^{Q} \alpha_q m_{iq}(T_i) \right] \right\}^{\delta_i} \\
&\quad \exp\left\{ - \int_0^{T_i} h_0(s) \exp\left[\gamma' \omega_i + \sum_{q=1}^{Q} \alpha_q m_{iq}(s)) \right] ds \right\}.
\end{aligned} \tag{3}$$

The second part is related to the parameter vector θ_y (i.e. the parameters for the longitudinal outcomes), and is defined as follows:

$$p(y_i|b_i;\theta_y) = \prod_q p(y_{iq}|b_{iq};\theta_y) = \prod_q (2\pi\sigma_q^2)^{-n_{iq}/2} \exp\left[-\frac{\| y_{iq} - X_{iq}\beta_q - Z_{iq}b_{iq} \|^2}{2\sigma_q^2}\right], \tag{4}$$

where $\| \cdot \|$ denotes the Euclidean vector norm.
Lastly, the third part is related to the parameter vector θ_b (i.e. the parameters of the random-effects variance–covariance matrix), and is defined as follows:

$$p(b_i;\theta_b) = (2\pi)^{-R/2} det(D)^{-1/2} \exp\left[\frac{-b_i' D^{-1} b_i}{2}\right], \tag{5}$$

where $R = q_1 + ... + q_q + ... + q_Q$, where each q_q indicates the number of random effects considered for the longitudinal covariate q[2].
In order to maximise the likelihood function, the following score function must be considered (in Appendix the details):

$$\begin{aligned} S(\theta) &= \sum_{i=1}^{n} S_i(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log p(T_i, \delta_i, y_i; \theta) \\ &= \sum_i \int A(\theta, b_i) p(b_i|T_i, \delta_i, y_i; \theta) db_i, \end{aligned}$$

where $A(\theta, b_i) = \frac{\partial}{\partial\theta} [\log p(T_i, \delta_i|b_i;\theta_t) + \log p(y_i|b_i;\theta_y) + \log p(b_i;\theta_b)]$.

2.2. THE EM ALGORITHM

The EM algorithm is used to maximise the log-likelihood function, where the random effects are treated as 'missing data'.
Accordingly, in the E-step, the expected value of the complete data log-likelihood

[2] Specifically, if the model of the random effect for the longitudinal covariate q considers only the intercept, then $q_q = 1$. However, if the model of random effects considers both the intercept and the slope, then $q_q = 2$.

function that considers the random effects as missing data is:

$$\begin{aligned} Q(\theta|\theta^{(it)}) &= E[\log p(y;\theta)|y^0;\theta^{(it)}] = \int p(y^m,y^o;\theta)p(y^m|y^o;\theta^{(it)})dy^m \\ &= \sum_i \int \log p(T_i,\delta_i,y_i,b_i;\theta)p(b_i|T_i,\delta_i,y_i;\theta^{(it)})db_i \\ &= \sum_i \int \Big[\log p(T_i,\delta_i|b_i;\theta_t) + \log p(y_i|b_i;\theta_y) \\ &\quad + \log p(b_i;\theta_b)\Big] p(b_i|T_i,\delta_i,y_i;\theta^{(it)})db_i. \end{aligned}$$

A numerical integration procedure, such as the Gaussian quadrature rules, must be employed for the integral with respect to the random effects.

$$\begin{aligned} &E\{A(\theta,b_i)|T_i,\delta_i,y_i;\theta\} = \int A(\theta,b_i)p(b_i|T_i,\delta_i,y_i;\theta)db_i \\ &\approx 2^{R/2}\sum_{t_1=1}^{K}\ldots\sum_{t_R=1}^{K} w_{t_1}\ldots w_{t_R}A(\theta,b_t\sqrt{2})p(b_t\sqrt{2}|T_i,\delta_i,y_i;\theta)exp(\parallel b_t \parallel^2), \end{aligned} \tag{6}$$

where K denotes the quadrature points and $b_t = (b_{t_1},\ldots,b_{t_R})$ are the Hermite polynomials' roots with weights $w_{t_1},\ldots,w_{t_R}$.
In the M-step it is possible to obtain estimations for σ_q^2 and D in a closed-form solution.
Beginning with σ_q^2, for each $q = 1,\ldots,Q$, we have the following:

$$\begin{aligned} \hat{\sigma}_q^2 &= \sum_i \int \frac{\parallel y_{iq} - X_{iq}\beta_q - Z_{iq}b_{iq} \parallel^2}{N} p(b_i|T_i,\delta_i,y_i;\theta)db_i \\ &= \frac{1}{N}\sum_i (y_{iq} - X_{iq}\beta_q)'(y_{iq} - X_{iq}\beta_q - 2Z_{iq}\tilde{b}_{iq}) + tr(Z_{iq}'Z_{iq}\tilde{v}b_{iq}) + \tilde{b}_{iq}'Z_{iq}'Z_{iq}\tilde{b}_{iq}, \end{aligned}$$

where

- $\tilde{b}_{iq} = E(b_{iq}|T_i,\delta_i,y_i;\theta) = \int b_{iq}p(b_i|T_i,\delta_i,y_i;\theta)db_i$
- $\tilde{v}b_{iq} = var(b_{iq}|T_i,\delta_i,y_i;\theta) = \int (b_{iq} - \tilde{b}_{iq})^2 p(b_i|T_i,\delta_i,y_i;\theta)db_i.$

Considering D, we can obtain the following:

$$\hat{D} = \frac{1}{n}\sum_i \int b_i b_i' p(b_i|T_i,\delta_i,y_i;\theta)db_i = \frac{1}{n}\sum_i \tilde{v}b_i + \tilde{b}_i\tilde{b}_i'. \tag{7}$$

For the other parameters, there is no closed-form solution. Thus, it is necessary to use a one-step Newton–Raphson update:

$$\hat{\beta}^{it+1} = \hat{\beta}^{it} - \left\{ \frac{\partial}{\partial \beta} S(\hat{\beta}^{it}) \right\}^{-1} S(\hat{\beta}^{it}) \tag{8}$$

$$\hat{\theta}_t^{it+1} = \hat{\theta}_t^{it} - \left\{ \frac{\partial}{\partial \theta_t} S(\hat{\theta}_t^{it}) \right\}^{-1} S(\hat{\theta}_t^{it}), \tag{9}$$

where $\hat{\beta}^{it}$ and $\hat{\theta}_t^{it}$ denote the values of β and θ_t, respectively, at the current iteration it. In addition, $S(\hat{\beta}^{it})$ and $S(\hat{\theta}_t^{it})$ denote the corresponding blocks of the Hessian matrix, evaluated at $\hat{\beta}^{it}$ and $\hat{\theta}_t^{it}$, respectively. For the evaluation of the blocks of the Hessian matrix, the numerical derivative routine is used.
Then, starting with β_q, it is possible to obtain the score function as:

$$\begin{aligned} S(\beta_q) &= \sum_{i=1}^{n} S_i(\beta_q) = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_q} \log p(T_i, \delta_i, y_i; \theta) \\ &= \sum_i \int \left\{ \frac{X'_{iq}(y_{iq} - X_{iq}\beta_q - Z_{iq}b_{iq})}{\sigma_q^2} + \delta_i \alpha_q x_{iq}(T_i) + \right. \\ &\left. - \exp(\gamma' \omega_i) \int_0^{T_i} h_0(s) \alpha_q x_{iq}(s) \exp\left[\sum_{q=1}^{Q} \alpha_q m_{iq}(s) \right] p(b_i | T_i, \delta_i, y_i; \theta) ds \right\} db_i, \end{aligned}$$

To solve the integral, a numerical method is needed. Here, the Gauss–Hermite quadrature is applied, as shown in equation (6). Accordingly, the former equation becomes:

$$\begin{aligned} S(\beta_q) \approx \sum_i 2^{R/2} \sum_{t_1=1}^{K} \dots \sum_{t_Q=1}^{K} w_{t_1} \dots w_{t_R} & \left\{ \frac{X'_{iq}(y_{iq} - X_{iq}\beta_q - Z_{iq}(b_t\sqrt{2}))}{\sigma_q^2} + \delta_i \alpha_q x_{iq}(T_i) \right. \\ & \left. - \exp(\gamma' \omega_i) \int_0^{T_i} h_0(s) \alpha_q x_{iq}(s) \exp\left[\sum_{q=1}^{Q} \alpha_q m^*_{iq}(s) \right] ds \right\} \\ & p(b_t\sqrt{2} | T_i, \delta_i, y_i; \theta) exp(\| b_t \|^2), \end{aligned}$$

posing $m^*_{iq}(s) = x'_{iq}(s)\beta_q + z'_{iq}(s)(b_t\sqrt{2})$.
Then, for the parameters that analyse the effect of the exogenous covariate on the

risk of the event γ, we have:

$$\begin{aligned} S(\gamma) &= \sum_{i=1}^{n} S_i(\gamma) = \sum_{i=1}^{n} \frac{\partial}{\partial \gamma} \log p(T_i, \delta_i, y_i; \theta) \\ &= \sum_i \omega_i \left\{ \delta_i - \exp(\gamma' \omega_i) \int \int_0^{T_i} h_0(s) \exp\left[\sum_{q=1}^{Q} \alpha_q m_{iq}(s) \right] p(b_i | T_i, \delta_i, y_i; \theta) ds db_i \right\}. \end{aligned}$$

Applying the Gauss–Hermite quadrature, it results:

$$\begin{aligned} S(\gamma) \approx \sum_i \omega_i \Bigg\{ \delta_i - \exp(\gamma' \omega_i) \int_0^{T_i} h_0(s) 2^{R/2} \sum_{t_1=1}^{K} \ldots \sum_{t_R=1}^{K} w_{t_1} \ldots w_{t_R} \\ \exp\left[\sum_{q=1}^{Q} \alpha_q m_{iq}^*(s) \right] ds \Bigg\}. \end{aligned}$$

In the next step, the parameter α_q is analysed:

$$\begin{aligned} S(\alpha_q) &= \sum_{i=1}^{n} S_i(\alpha_q) = \sum_{i=1}^{n} \frac{\partial}{\partial \alpha_q} \log p(T_i, \delta_i, y_i; \theta) \\ &= \sum_i \int \Bigg\{ \delta_i m_{iq}(T_i) - \exp(\gamma' \omega_i) \\ &\quad \int_0^{T_i} h_0(s) m_{iq}(s) \exp\left[\sum_{q=1}^{Q} \alpha_q m_{iq}(s) \right] p(b_i | T_i, \delta_i, y_i; \theta) ds \Bigg\} db_i. \end{aligned}$$

Applying the Gauss–Hermite quadrature, it is obtained:

$$\begin{aligned} S(\alpha_q) \approx \sum_i 2^{R/2} \sum_{t_1=1}^{K} \ldots \sum_{t_R=1}^{K} w_{t_1} \ldots w_{t_R} \Bigg\{ \delta_i m_{iq}^*(T_i) - \exp(\gamma' \omega_i) \\ \int_0^{T_i} h_0(s)\, m_{iq}^*(s) \exp\left[\sum_{q=1}^{Q} \alpha_q m_{iq}^*(s) \right] ds \Bigg\}. \end{aligned}$$

In each iteration, the baseline hazard is updated using a non-parametric estimation based on the Breslow (Cox, 1972) method:

$$\hat{h}_0(t) = \sum_{i=1}^{n} \frac{\delta_i I(T_i = t)}{\sum_{j=1}^{n} I(T_i \leq t) \int \exp(\gamma' \omega_j + \sum_{q=1}^{Q} \alpha_q m_{jq}(t)) p(b_i | T_i, \delta_i, y_i; \hat{\theta}) db_i},$$

where $I(\cdot)$ denotes an indicator function that takes the value one if an event occurs, and zero otherwise. In addition, the estimations of the random effects b_i are updated at each iteration using the conditional expected value:

$$\bar{b}_i = \int b_i p(b_i|T_i, \delta_i, y_i; \theta) db_i,$$

where

$$p(b_i|T_i, \delta_i, y_i; \theta) = \frac{p(T_i, \delta_i|b_i; \theta) p(y_i|b_i; \theta) p(b_i; \theta)}{p(T_i, \delta_i, y_i; \theta)}.$$

Convergence is achieved when the parameter estimates and/or the log-likelihood are stable. The standard errors are estimated at convergence, recalling that:

$$\hat{var}(\hat{\theta}) = [I(\hat{\theta})]^{-1} \quad \text{where} \quad I(\hat{\theta}) = -\sum_{i=1}^{n} \frac{\partial S_i(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}}.$$

The standard errors for the joint models obtained using this estimation method are underestimated. For this reason, the following empirical information matrix is used (Scott, 2002):

$$I_e(\theta) = \sum_{i=1}^{n} S_i(\theta) S_i'(\theta) - \frac{1}{n} \left(\sum_{i=1}^{n} S_i(\theta) \right) \left(\sum_{i=1}^{n} S_i(\theta) \right)', \tag{10}$$

where $S_i(\theta) = \frac{\partial}{\partial \theta} \log p(T_i, \delta_i, y_i; \theta)$.
The standard errors are obtained from the empirical information matrix in the usual way:

$$\hat{var}(\hat{\theta}) = [I_e(\hat{\theta})]^{-1}.$$

The algorithm is as follows:

1. The initial values are estimated using the two-stage approach (Mazzoleni, 2018).

2. In the E-step, the expected value of the complete data log-likelihood function is used, considering the random effects as the missing data, in addition to the Gauss–Hermite quadrature rule, as shown in formula (6).

3. In the M-step, for σ_q^2 and D, it is possible to obtain closed-form solutions. However, for the parameters γ, α_q, and β_q, a one-step Newton–Raphson update is implemented, as shown in formula (9). Then, in each step the random effects and the baseline hazard are updated.

4. Iterate between steps 2 and 3 until the algorithm converges, which occurs when the parameter estimates and/or the log-likelihood are stable.

5. At convergence, the standard error for each parameter is calculated using the empirical information matrix.

The algorithm was implemented in R using ad hoc code.
As anticipated in the introduction, the estimation method here proposed is different from that used in Hickey et al. (2017). Indeed, in addition to a different formulation of the model, in this EM algorithm a Gauss–Hermite quadrature rule is used, while Hickey et al. (2017) use a Monte Carlo EM algorithm.

3. SIMULATION STUDY

Here, we present three simulation studies, based on samples with 50, 100, and 200 units, respectively. Each simulated data set is composed of two longitudinal covariates $m_{i1}(t)$ and $m_{i2}(t)$, a continuous exogenous covariate *cont*, and a binary *bin* exogenous covariate. The following formula is used:

$$\begin{cases} y_{i1}(t) = \beta_{01} + \beta_{11}t + \beta_{21}cont + \beta_{31}bin + b_{i01} + b_{i11}t + \varepsilon_{i1}(t) \\ y_{i2}(t) = \beta_{02} + \beta_{12}t + \beta_{22}cont + \beta_{32}bin + b_{i01} + b_{i12}t + \varepsilon_{i2}(t) \\ h_i(t) = h_0(t)\exp[\gamma_1 cont + \gamma_2 bin + \alpha_1 m_{i1}(t) + \alpha_2 m_{i2}(t)] \end{cases} \tag{11}$$

The event times are simulated as in Austin (2013). In this simulation, the event times follow a Gompertz distribution. Then, we have:

$$T_i = \frac{1}{\psi_{ti} + \alpha}\log\left\{1 + \frac{(\psi_{ti} + \alpha)(-\log(u_i))}{\lambda\exp(\psi_i' x_i)}\right\}, \tag{12}$$

where $\lambda > 0$ and $-\infty < \alpha < \infty$ are the scale and shape parameters, respectively, of the Gompertz distribution and $u_i \sim U(0,1)$. In addition, for each subject i, ψ_{ti} is the sum of all time-dependent parameters and ψ_i is a vector of all parameters that are not time dependent, and are related to the vector x_i, which contains all covariates that are not time dependent[3]. Independent right-censoring is also considered, but the censoring percentage is always lower than 30% in order to avoid the influence of censoring on the parameter estimation and unstable results.
Table 1 contains the results of the simulation study with 50 units. The rate of convergence when each data set contains 50 units is 89.4%. When we have so few

[3] In the simulation studies (11), for each subject i, $\psi_{ti} = \alpha_1\beta_{11} + \alpha_1 b_{i11} + \alpha_2\beta_{12} + \alpha_2 b_{i12}$, and $\psi_i = [\gamma_1, \gamma_2, \alpha_1\beta_{01}, \alpha_1\beta_{21}, \alpha_1\beta_{31}, \alpha_1 b_{i01}, \alpha_2\beta_{02}, \alpha_2\beta_{22}, \alpha_2\beta_{32}, \alpha_2 b_{i02}]$.

Tab. 1: Simulation study for 50 units

Par.	true	mean	MSE	C.I. 95%
α_1	0.5	0.4268	0.0224	(0.1871 ; 0.7037)
α_2	-1	-0.8925	0.0491	(-1.2942 ; -0.5739)
β_{01}	1	1.0476	0.0884	(0.4781 ; 1.6051)
β_{11}	1	0.8286	0.0967	(0.3195 ; 1.3693)
β_{21}	1	0.9775	0.0517	(0.5485 ; 1.4398)
β_{31}	1	0.9466	0.1699	(0.2255 ; 1.7655)
β_{02}	1	0.8906	0.1097	(0.3261 ; 1.5312)
β_{12}	1	1.3854	0.2074	(0.8772 ; 1.8227)
β_{22}	1	1.0254	0.0512	(0.5679 ; 1.4669)
β_{32}	1	1.0232	0.1666	(0.1967 ; 1.8090)
γ_1	1	0.9238	0.1158	(0.3540 ; 1.6838)
γ_2	1	0.8933	0.2498	(0.0042 ; 1.9212)

units for each data set, the variance–covariance matrix D is not always invertible. Based on the results, we can argue that the mean value of each parameter is close to the true value, and that the 95% confidence interval contains the true value.

Table 2 contains the results for the simulation with 100 units. Here, we find that increasing the number of units for each data set resolves the problem related to the variance–covariance matrix D. In fact, the rate of convergence is now 100%. In addition, the mean value of the parameter estimates is closer to the true value, the mean squared error (MSE) has decreased, and the length of the 95% confidence interval is shorter.

Table 3 contains the results of the simulation with 200 units. Once again, increasing the number of units for each data set has moved the mean value of the parameter estimates closer to the true value. In addition, the MSE has decreased further and the length of the 95% confidence interval is shorter still.

In summary, increasing the number of units in each data set yields better results, but also increases the computation time. The difference between the mean and the true value is related to the limit on the computation time. Increasing the number of units in each data set decreases this difference. In conclusion, the implemented algorithm seems to perform well, as confirmed by the simulation results.

Tab. 2: Simulation study for 100 units

Par.	true	mean	MSE	C.I. 95%
α_1	0.5	0.4313	0.0139	(0.2591 ; 0.6268)
α_2	-1	-0.8994	0.0313	(-1.2065 ; -0.6420)
β_{01}	1	1.0531	0.0476	(0.6277 ; 1.4768)
β_{11}	1	0.8435	0.0745	(0.4385 ; 1.3024)
β_{21}	1	0.9779	0.0253	(0.6614 ; 1.2628)
β_{31}	1	0.9587	0.0780	(0.4526 ; 1.5069)
β_{02}	1	0.8978	0.0542	(0.4931 ; 1.3254)
β_{12}	1	1.3690	0.1786	(0.9487 ; 1.7831)
β_{22}	1	1.0308	0.0240	(0.7484 ; 1.3343)
β_{32}	1	1.0114	0.0710	(0.4679 ; 1.5064)
γ_1	1	0.9080	0.0586	(0.4812 ; 1.3690)
γ_2	1	0.8646	0.1285	(0.2547 ; 1.5778)

Tab. 3: Simulation study for 200 units

Par.	true	mean	MSE	C.I. 95%
α_1	0.5	0.4353	0.0087	(0.3201 ; 0.5739)
α_2	-1	-0.9027	0.0203	(-1.1149 ; -0.7093)
β_{01}	1	1.0482	0.0238	(0.7494 ; 1.3200)
β_{11}	1	0.8535	0.0591	(0.5043 ; 1.2907)
β_{21}	1	0.9780	0.0109	(0.7723 ; 1.1750)
β_{31}	1	0.9701	0.0418	(0.5806 ; 1.3870)
β_{02}	1	0.9080	0.0354	(0.6008 ; 1.2426)
β_{12}	1	1.3456	0.1514	(0.9878 ; 1.7002)
β_{22}	1	1.0286	0.0109	(0.8333 ; 1.2377)
β_{32}	1	0.9933	0.0425	(0.6187 ; 1.3949)
γ_1	1	0.9081	0.0295	(0.6046 ; 1.1745)
γ_2	1	0.8665	0.0666	(0.4326 ; 1.3284)

4. APPLICATION TO PRIMARY BILIARY CIRRHOSIS DATA

In this section, the proposed model is applied to the well-known PBC data set.

The Mayo Clinic has established a database of 424 patients suffering from PBC (Dickson et al., 1989; Fleming and Harrington, 1991; Murtaugh et al., 1994; Therneau and Grambsch, 2000). These 424 units represent all PBC patients referred

to Mayo between January 1974 and May 1984 who met the standard eligibility criteria for a randomized, double-blinded, placebo-controlled clinical trial of the drug D-penicillamine (DPCA). Each patient and his/her treating physician agreed to randomization in 312 of the 424 cases. For each of the 312 clinical trial patients, data on clinical, biochemical, sérologie, and histologie parameters were collected. For this analysis, a complete follow up until July 1986 was attempted on all patients. By the end of the trial, 125 of the 312 had died. PBC is a fatal chronic liver disease of unknown cause. The primary pathologic event appears to be destruction of interlobular bile ducts, which may be mediated by immunological mechanisms.

The *PBCSEQ* data set is available from the package survival (Therneau and Lumley, 2015) in R, which contains multiple laboratory results collected for each patient during each follow-up visit, with different baseline and longitudinal covariates. Two longitudinal covariates are considered: the level of serum bilirubin in mg/dl (*serBilir*) and the level of albumin in mg/dl (*albumin*). The observation time is expressed in days. In the survival sub-model, we analyse the exogenous covariate patient's age at registration, given in years (*age*).

Accordingly, the longitudinal and the survival sub-models are as follows:

$$\begin{cases} y_{i1}(t) = \beta_{01} + \beta_{11}t + b_{i01} + b_{i11}t + \varepsilon_{i1}(t) \\ y_{i2}(t) = \beta_{02} + \beta_{12}t + b_{i02} + b_{i12}t + \varepsilon_{i2}(t) \\ h_i(t) = h_0(t)\exp[\alpha_1 m_{i1}(t) + \alpha_2 m_{i2}(t) + \gamma_1 age] \end{cases}, \tag{13}$$

where $y_{i1}(t)$ is *log(serBilir)* , the logarithm of the level of serum bilirubin, and $y_{i2}(t)$ is *albumin*, the level of albumin.

The results obtained using the proposed algorithm are shown in Table 4, where all parameter results are statistically significant. In particular, *log(serBilir)* has a positive effect on the risk of death, with a one-point increase in *log(serBilir)* associated with a 3.0038 $(=\exp(1.0999))$-fold increase in the risk of death. Then, *albumin* has a negative effect on the risk of death, with a one-point increase in *albumin* resulting in a 0.1582 $(=\exp(-1.8434))$-fold decrease in the risk of death. Moreover, the exogenous variable *age* has a positive effect on the risk of death, with a one-point increase in *age* resulting in a 1.0481 $(=\exp(0.0472))$-fold increase in the risk of death.

Analysing the estimates for the longitudinal sub-model, the slope indicates the change in the longitudinal covariates from an increase of one day. Accordingly, the observation time has a positive effect of $(\beta_{11} = 0.0005)$ on the level of *log(serBilir)*, but has a negative effect of $(\beta_{12} = -0.0003)$ on the level of *albumin*.

Tab. 4: PBC Resuts

Parameter	Est.	SE	p-value
α_1 (log(*serBilir*))	1.0999	0.1036	< 0.0001
α_2 (*albumin*)	-1.8434	0.1695	< 0.0001
γ_1 (*age*)	0.0472	0.0084	< 0.0001
β_{01} (*Intercept*)	0.6004	0.0262	< 0.0001
β_{11} (*Time*)	0.0005	$1.3543 * 10^{-05}$	< 0.0001
β_{02} (*Intercept*)	3.5488	0.0204	< 0.0001
β_{12} (*Time*)	-0.0003	$1.2989 * 10^{-05}$	< 0.0001
Log-lik	-2934.6710		

Tab. 5: PBC Resuts using joineRML

Parameter	Est.	SE	p-value
α_1 (log(*serBilir*))	1.0402	0.1221	< 0.0001
α_2 (*albumin*)	-2.3425	0.3247	< 0.0001
γ_1 (*age*)	0.0478	0.0086	< 0.0001
β_{01} (*Intercept*)	0.4848	0.0496	< 0.0001
β_{11} (*Time*)	0.0005	0.0000	< 0.0001
β_{02} (*Intercept*)	3.5512	0.0223	< 0.0001
β_{12} (*Time*)	-0.0003	0.0000	< 0.0001
Log-lik	-3076.646		

For comparative purposes, Table 5 reports the results of the parameter estimations for model (13) using the package joineRML (Hickey et al., 2017). The estimates based on the proposed algorithm are coherent with those obtained using joineRML.

5. DISCUSSION

The proposed algorithm that extends the maximum-likelihood estimation method to the case of a multivariate longitudinal sub-model shows encouraging results, which are confirmed by the simulation studies. In fact, the mean value of the estimates is close to the true value. Increasing the number of units in the data set yields better results, but also increases the computation time.

As already argued, the model here proposed is different from the models used in Rizopoulos (2010) and Hickey et al. (2017). Indeed, this work uses a multivariate formulation of the Rizopoulos (2010) model and analyses the relation between the

hazard of the event and the true and unobserved value of the longitudinal covariates, not only the relation between hazard and random effects, as in Hickey et al. (2017). In addition, in the estimation method, for the EM algorithm, a Gauss–Hermite quadrature rule is used.

The results of applying the method to PBC data quantify the influence of the two longitudinal covariates on the event. Here, we find that *log(serBilir)* has a positive effect on the risk of death, whereas *albumin* has a negative effect on the risk of death. The results are coherent with those obtained from the other package joineRML (Hickey et al., 2017).

The results are encouraging and lead to several possibilities for future work. For instance, it will be worthwhile developing diagnostic analyses and dynamic predictions. Moreover, we would like to extend the survival sub-model by studying the joint effect of more than one longitudinal covariate on more than one terminal event.

REFERENCES

Albert, P. and Shih, J. (2010). An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. In *The Annals of Applied Statistics*, 4: 1517–1532.

Austin, P.C. (2013). Generating survival times to simulate Cox proportional hazards models with time varying covariates. In *Statistics in Medicine*, 31: 3946– 3958.

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

Choi, J., Anderson, S., Richards, T. and Thompson, W. (2014). Prediction of transplant-free survival in idiopathic pulmonary fibrosis patients using joint models for event times and mixed multivariate longitudinal data. In *Journal of Applied Statistics*, 41: 2192–2205.

Cox, D. (1972). Regression models and life-tables (with discussion). In *Journal of the Royal Statistical Society*, 34: 187–220.

Dickson, E., Grambsch, P., Fleming, T., Fisher, L. and Langworthy, A. (1989). Prognosis in primary biliary cirrhosis: Model for decision making. In *Hepatology*, 10: 1–7.

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

Fieuws, S., Verbeke, G., Maes, B. and Vanrenterghem, Y. (2008). Predicting renal graft failure using multivariate longitudinal profiles. In *Biostatistics*, 9: 419–431.

Fleming, T. and Harrington, D. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

He, B. and Luo, S. (2013). Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. In *Statistical Methods in Medical Research*, 0: 1–13.

Henderson, A., De Gruttola, V. and Wulfsohn, M. (2000). Joint modelling of longitudinal measurements and event time data. In *Biostatistics*, 1: 465–480.

Hickey, G.L., Philipson, P., Jorgensen, A. and Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. In *BMC Medical Research Methodology*, 16: 117– 131.

Hickey, G., Philipson, P., Jorgensen, A., Kolamunnage-Dona, R., Williamson, P. and Rizopoulos, D. (2017). *joineRML: Joint Modelling of Multivariate Longitudinal Data and Time-to-Event Outcomes*. URL https://cran. r-project.org/web/packages/joineRML/index.html. Version 0.4.1.

Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. In *Biometrics*, 38: 963–974.

Lin, H., McCulloch, C. and Mayne, S. (2001). Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. In *Statistics in Medicine*, 21: 2369–2382.

Mazzoleni, M. (2018). The analysis of student paths at the university using the multivariate joint models. In *Communications in Statistics - Theory and Methods*, 1–15.

Murtaugh, P., Dickson, E., Van Dam, G., Malinchoc, M., Grambsch, P., Langworthy, A. and Gips, C. (1994). Primary biliary cirrhosis: Prediction of short-term survival based on repeated patient visits. In *Hepatology*, 20: 126–134.

Rizopoulos, D. (2010). Jm: An r package for the joint modelling of longitudinal and time-to-event data. In *Journal of Statistical Software*, 35.

Rizopoulos, D. (2012). *Joint model for Longitudinal and Time-to-Event Data with applications in R*. CRC Press, Boca Raton.

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

Scott, A. (2002). Maximum likelihood estimation using the empirical fisher information matrix. In *Journal of Statistical Computation and Simulation*, 72 (8): 599–611.

Song, X., Davidian, M. and Tsiatis, A. (2002). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. In *Biostatistics*, 3: 511–528.

Therneau, T. and Grambsch, P. (2000). *Modeling Survival Data: extending the Cox Model*. Springer-Verlang, New York.

Therneau, T. and Lumley, T. (2015). *survival: Survival Analysis*. URL https: //cran.r-project.org/web/packages/survival/index.html. Version 2.41-3.

Tsiatis, A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. In *Biometrika*, 88: 447–458.

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

Xu, J. and Zeger, S. (2001a). The evaluation of multiple surrogate endpoints. In *Biometrics*, 57: 81–87.

Xu, J. and Zeger, S. (2001b). Joint analysis of longitudinal data comprising repeated measures and times to events. In *Applies Statistics*, 50: 375–387.

APPENDIX

The score function is defined as:

$$
\begin{aligned}
S(\theta) &= \sum_{i=1}^{n} S_i(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log p(T_i, \delta_i, y_i; \theta) \\
&= \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \frac{\partial}{\partial \theta'} \int p(T_i, \delta_i | b_i; \theta_t) p(y_i | b_i; \theta_y) p(b_i; \theta_b) db_i \\
&= \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \int \frac{\partial}{\partial \theta'} \left[p(T_i, \delta_i | b_i; \theta_t) p(y_i | b_i; \theta_y) p(b_i; \theta_b) \right] db_i \\
&= \sum_i \int \left\{ \frac{\partial}{\partial \theta'} \log \left[p(T_i, \delta_i | b_i; \theta_t) p(y_i | b_i; \theta_y) p(b_i; \theta_b) \right] \right\} \\
&\quad \frac{p(T_i, \delta_i | b_i; \theta_t) p(y_i | b_i; \theta_y) p(b_i; \theta_b)}{p(T_i, \delta_i, y_i; \theta)} db_i = \\
&= \sum_i \int A(\theta, b_i) p(b_i | T_i, \delta_i, y_i; \theta) db_i
\end{aligned}
$$