Generalized linear mixed joint model for longitudinal and survival outcomes

Sara Viviani · Marco Alfó · Dimitris Rizopoulos

Abstract Longitudinal studies often entail categorical outcomes as primary responses. When dropout occurs, non-ignorability is frequently accounted for through shared parameter models (SPMs). In this context, several extensions from Gaussian to non-Gaussian longitudinal processes have been proposed. In this paper, we formulate an approach for non-Gaussian longitudinal outcomes in the framework of joint models. As an extension of SPMs, based on shared latent effects, we assume that the history of the response up to current time may have an influence on the risk of dropout. This history is represented by the current, expected, value of the response. Since the time a subject spends in the study is continuous, we parametrize the dropout process through a proportional hazard model. The resulting model is referred to as Generalized Linear Mixed Joint Model (GLMJM). To estimate model parameters, we adopt a maximum likelihood approach via the EM algorithm. In this context, the maximization of the observed data log-likelihood requires numerical integration over the random effect posterior distribution, which is usually not straightforward; under the assumption of Gaussian random effects, we compare Gauss-Hermite and Pseudo-Adaptive Gaussian quadrature rules. We investigate in a simulation study the behaviour of parameter estimates in the case of Poisson and Binomial longitudinal responses, and apply the GLMJM to a benchmark dataset.

1 Introduction

Several studies in different disciplines collect longitudinal non-Gaussian responses, such as binary outcomes or counts; an interesting review in this context is given by Molenberghs and Verbeke (2005). As in studies with Gaussian responses, it is common for some subjects to drop out prematurely; the occurrence of such a dropout leads to missing data and poses additional challenges to draw correct statistical inferences based on the observed data. One appealing approach to treat this issue is through the shared parameter model of Wu and Carrol (1988), where the longitudinal outcome and the time-to-event share a latent Gaussian random effect. Some proposals extend shared parameter models to discrete outcomes, with participation in the study described by a discrete time process. For instance, Albert and Follmann (2000) extend the shared random effect model of Wu and Carrol (1988) to count data. Albert et al. (2002) present a model for longitudinal binary data, where the binary response and the missing data process share a Gaussian autoregressive latent process. In both cases, the time spent in the study is considered as discrete and described by a multinomial model.

We start from a more general approach, the Joint Model (JM) of Wulfsohn and Tsiatis (1997), that takes into account a continuous time-to-dropout, jointly modeled with longitudinal measurements of a response of interest. This approach is based on the assumption that the error-free longitudinal response, together with dropout-specific covariates, influence the risk of dropout. Rizopoulos (2010) pointed out that

S. Viviani (✉) · M. Alfó
Department of Statistics, Sapienza University of Rome,
P.zza Aldo Moro 15, 00185 Rome, Italy
e-mail: sara.viviani@uniroma1.it

D. Rizopoulos
Department of Biostatistics, Erasmus University Medical Center,
3000 CA Rotterdam, The Netherlands

this model is based on the hypothesis that the primary outcome expected value, expressed as a function of fixed and random coefficients, is shared by the time-to-event process; Tsiatis and Davidian (2004) give a nice overview of joint models by proposing a unifying approach for all the previous existing work in the field. A recent contribution to handle JM with non-Gaussian longitudinal responses within a Bayesian framework is due to Rizopoulos and Ghosh (2011), who define a joint model for a multivariate (discrete and/or continuous) longitudinal outcome and a continuous time-to-event. The longitudinal response is assumed to follow a generalized linear mixed effect model, while the association between the longitudinal outcome and the survival time is captured via a relative risk model that includes, as time dependent covariate, a spline-based function of fixed and random effects in the longitudinal response model.

In this paper, we describe an extension of JMs to non-Gaussian data, the Generalized Linear Mixed Joint Models (GLMJMs), and show how parameter estimation can be performed in a likelihood-based framework.

Parameter estimates are obtained via the expectation-maximization (EM) algorithm, and corresponding standard errors are based on the observed information matrix. The optimization algorithm maximizes the observed data likelihood and requires integration over the random effect distribution via numerical approximation; for this purpose, we adopt two alternative quadrature rules: the standard Gaussian quadrature rule, which uses the current longitudinal and the time-to-event distributions, and the Pseudo-Adaptive Gaussian rule, see Rizopoulos (2012), which is based on the approximation of the random effect posterior distribution by a multivariate Gaussian distribution, according to the Bayesian Central Limit theorem, see e.g. Carlin and Louis (2009). This procedure has the advantage of an adaptive rule, i.e. it requires few quadrature points, and a reduced computational burden. We extend the pseudo-adaptive rule, that was originally proposed for continuous longitudinal outcomes, to the setting of GLMJM.

The paper is organized as follows. Section 2 describes the Generalized Linear Mixed Joint Model from an analytical point of view, while estimation issues are dealt with in Sects. 3 and 4. Section 5 specifies the GLMJM in the cases of Poisson and Binomial longitudinal outcomes. To investigate the behaviour of parameter estimates, a simulation study is described in Sect. 6. The application to the AIDS dataset of Goldman et al. (1996), see also Carlin and Louis (2009), is discussed in Sect. 7. Section 8 gives concluding remarks.

2 The generalized linear mixed joint model

In this section, we introduce the class of generalized linear mixed joint models (GLMJMs), which represents an ex-

tension of the linear mixed joint model, see Wulfsohn and Tsiatis (1997), to non-Gaussian responses in longitudinal studies with attrition. The proposed parametrization is based on joint modeling a longitudinal response, with distribution in the exponential family, and a time-to-event process, assuming that the *expected* value of the longitudinal outcome, measured at the dropout time, may influence the risk of dropout.

Let us denote by $T_i = \min(T_i^*, C_i)$ the observed failure time for the ith individual, $i = 1, \ldots, n$, taken as the minimum between the true event time T_i^* and the censoring time C_i, which may correspond to the end of the follow-up. Further, let δ_i be the event indicator defined by $\delta_i = I(T_i^* \leq C_i)$, where $I(\cdot)$ is the indicator function. The outcome $Y_i(t)$ is repeatedly observed before T_i at $t = 1, \ldots, n_i$ occasions, and is missing for $t \geq T_i$. We assume that the longitudinal process is associated with T_i^*, i.e the *true* event time, but it is independent of the censoring time C_i. Let $\mathbf{Y}$ be a random variable with distribution in the exponential family and natural parameter $\mu_i(t)$, that is $\mathbf{Y} \sim \mathrm{EF}(\mu_i(t))$.

The GLMJM is defined by the following equations:

$$
\begin{cases}
g(m_i(t)) = \boldsymbol{\beta}^\mathsf{T} \mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T} \mathbf{Z}_i(t) \\
h(t \mid M_i(t), \mathbf{W}_i) \\
\quad = h_0(t) \exp\{\boldsymbol{\gamma}^\mathsf{T} \mathbf{W}_i + \alpha m_i(t)\}.
\end{cases}
\tag{1}
$$

For what concerns the longitudinal process, $\mathbf{X}_i(t)$ is a vector of p predictors with fixed effects $\boldsymbol{\beta}$ and $\mathbf{Z}_i(t)$ is a vector of q predictors with random coefficients $\mathbf{b}_i$. Generally speaking, $\mathbf{X}_i(t)$ and $\mathbf{Z}_i(t)$ contain the time at which the response is measured, as well as the interaction between time and other covariates. $m_i(t) = m(\mu_i(t)) = g^{-1}(\boldsymbol{\beta}^\mathsf{T} \mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T} \mathbf{Z}_i(t))$ represents the expected value of the longitudinal response at time t, $g(\cdot)$ is a link function and $g(m_i(t)) = \mu_i(t)$. Assuming that the hazard of dropout depends on the expected value of the response at the event time is twice convenient: first, it allows to take into account the (expected) value of the response at that time, introducing a time-dependent covariate in the survival model. Further, it provides a practically more appealing non-ignorable approach when compared to the one where the actual response influences the drop-out process. Since it is extremely rare to have the response measured at the drop-out time, the choice of using $m_i(T_i)$ instead seems reasonable.

In the survival process, $M_i(t) = \{m_i(u) : 0 \leq u \leq t\}$ denotes the history of the true, but unobserved, longitudinal process up to t and $\mathbf{W}_i$ is a row vector of additional (time constant) covariates, with fixed parameter vector $\boldsymbol{\gamma}$. The term $h_0(t)$ denotes the baseline risk function and is typically left unspecified, see among others Cox (1972) and Andersen and Gill (1982). As Hsieh et al. (2006) have pointed out, leaving this function unspecified in the JM framework may lead to underestimate the parameter standard errors; there-

fore we consider a Weibull baseline function, i.e. $h_0(t) = \xi t^{\xi-1}$.

An interesting computational aspect arises when the survival function is considered. In the GLMJMs, it is given by

$$S\big(T_i \mid M_i(T_i), \mathbf{W}_i; \boldsymbol{\Phi}\big)$$
$$= \exp\left\{-\int_0^{T_i} h_i(s)\,ds\right\}$$
$$= \exp\left\{-\int_0^{T_i} \xi s^{\xi-1} h_0(s) \exp\{\boldsymbol{\gamma}^\mathsf{T} \mathbf{W}_i + \alpha m_i(s)\}\,ds\right\}, \quad (2)$$

where $\boldsymbol{\Phi} = (\boldsymbol{\gamma}, \alpha, \xi)$ is the survival parameter vector. Hence, the survival function involves an integral over the time which is not analytically tractable. We adopt a Gauss-Kronrod numerical integration method, see Kronrod (1964), as it is standard in the joint modeling framework, see for instance Rizopoulos (2012).

3 Likelihood and score

Following Rizopoulos et al. (2009), and assuming conditional independence, the observed data log-likelihood for model (1) is

$$\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta} \mid T_i, \delta_i, \mathbf{y}_i)$$
$$= \sum_i \log \int_{\mathbf{b}_i} p(T_i, \delta_i \mid \mathbf{b}_i; \boldsymbol{\Phi}) p(\mathbf{y}_i \mid \mathbf{b}_i; \boldsymbol{\beta})$$
$$\times p(\mathbf{b}_i; \mathbf{D})\,d\mathbf{b}_i, \quad (3)$$

where $\boldsymbol{\theta} = (\boldsymbol{\Phi}, \boldsymbol{\beta}, \mathbf{D})$ is the complete parameter vector for the longitudinal, time-to-event, and the latent process, respectively. The random effects $\mathbf{b}_i$ are assumed to follow a multivariate Gaussian distribution with zero mean and covariance matrix $\mathbf{D}$; they account for dependence over time among the repeated measurements corresponding to the same individual; moreover, through the term $m_i(t)$, they help account for the association between the longitudinal and the dropout process, see Molenberghs et al. (2011).

In the case of GLMJM, the longitudinal data density belongs to the exponential family of distributions and can be expressed by

$$p(\mathbf{y}_i \mid \mathbf{b}_i; \boldsymbol{\beta})$$
$$= \exp\{B(\boldsymbol{\mu}_i)\mathbf{y}_i - A(\boldsymbol{\mu}_i) + C(\mathbf{y}_i)\}$$
$$= \prod_{t=1}^{n_i} \exp\{B[\mu_i(t)]y_i(t) - A[\mu_i(t)] + C[y_i(t)]\},$$

where $A(\cdot)$, $B(\cdot)$ and $C(\cdot)$ are suitable functions. On the other hand, the time-to-event process is defined by

$$p(T_i, \delta_i \mid \mathbf{b}_i; \boldsymbol{\Phi})$$
$$= h\big(T_i \mid M_i(T_i), \mathbf{W}_i; \boldsymbol{\Phi}\big)^{\delta_i} S\big(T_i \mid M_i(T_i), \mathbf{W}_i; \boldsymbol{\Phi}\big). \quad (4)$$

The observed data score vector for a joint model can be expressed as the expected value of the complete data score vector with respect to the posterior distribution of the random effects:

$$\mathcal{S}(\boldsymbol{\theta}) = \sum_i \frac{\partial}{\partial \boldsymbol{\theta}^\mathsf{T}} \log \int p(T_i, \delta_i \mid \mathbf{b}_i; \boldsymbol{\Phi}) p(\mathbf{y}_i \mid \mathbf{b}_i; \boldsymbol{\beta})$$
$$\times p(\mathbf{b}_i; \mathbf{D})\,d\mathbf{b}_i$$
$$= \sum_i \int \omega(\boldsymbol{\theta}, \mathbf{b}_i) p(\mathbf{b}_i \mid T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\theta})\,d\mathbf{b}_i, \quad (5)$$

where $\omega(\boldsymbol{\theta}, \mathbf{b}_i) = \partial\{\log p(T_i, \delta_i \mid \mathbf{b}_i; \boldsymbol{\Phi}) + \log p(\mathbf{y}_i \mid \mathbf{b}_i; \boldsymbol{\beta}) + \log p(\mathbf{b}_i; \mathbf{D})\}/\partial\theta^\top$ denotes the complete data score vector. Equation (5) is central for parameter estimation in GLMJM, see Sect. 4 for further details.

4 Parameter estimation

Literature on JMs has so far presented different parameter estimation approaches, such as the Bayesian strategy, see Rizopoulos and Ghosh (2011), and multiple imputation, see Rizopoulos et al. (2010); other estimation procedures can be used as well, such as Monte Carlo ML algorithms, see Geyer and Thompson (1992). We obtain the maximum likelihood estimates for the complete parameter vector using the EM algorithm, as it is typical in mixed effect and joint modeling frameworks, since it is known to be robust to anomalous individual behaviour and to model misspecification. Sections 4.1 and 4.2 describe the expectation and the maximization steps in detail. Generally speaking, the estimation algorithm is based on the computation of the score vector (5) and the Hessian matrix; in both cases, numerical integration with respect to the posterior random effect distribution is required.

The Hessian matrix at convergence is also considered, as standard errors for parameter estimates are based on inverting the observed information matrix:

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\theta}}) = \big[\mathcal{I}(\hat{\boldsymbol{\theta}})\big]^{-1},$$

where $\mathcal{I}(\hat{\boldsymbol{\theta}}) = -\sum_{i=1}^n \frac{\partial \mathcal{S}_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$.

Several methods for the computation of standard errors when working with the EM algorithm have been proposed, see for instance Louis (1982) and Oakes (1999); these are used since the observed data log-likelihood is not tractable while the complete data log-likelihood can be easily computed. In our setting, the observed data log-likelihood have a tractable form and we use the EM algorithm primarily for its numerical stability. Calculation of the observed data log-likelihood and its derivatives under joint model requires, in fact, similar computations to the E-step of the EM algorithm, where the same integrals with respect to the random effects

have to be calculated. For this reason we opt to estimate standard errors by directly computing the observed information matrix, relying on Louis' method.

4.1 The E step

The E-step requires the calculation of the score vector (5); at the qth iteration, given the observed data and the parameter estimates calculated at the previous iteration, $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}^{(q-1)}$, the posterior random effect distribution $p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \hat{\boldsymbol{\theta}}^{(q-1)})$ can be written as follows:

$$
p\big(\mathbf{b}_i \mid T_i, \delta_i, \mathbf{y}_i; \hat{\boldsymbol{\theta}}^{(q-1)}\big)
$$
$$
= \frac{p(\mathbf{y}_i, T_i, \delta_i, \mathbf{b}_i; \hat{\boldsymbol{\theta}}^{(q-1)})}{\int p(\mathbf{y}_i, T_i, \delta_i, \mathbf{b}_i; \hat{\boldsymbol{\theta}}^{(q-1)}) d\mathbf{b}_i}. \tag{6}
$$

To compute the score vector as well as the joint distribution in the denominator of (6), we need numerical integration methods, e.g. Gaussian quadrature. In this framework, we consider Gauss-Hermite (GH) and Pseudo-Adaptive Gaussian (PA) rules, Rizopoulos (2012). The GH rule approximates the denominator in (6) through a weighted sum of the integrand calculated at pre-specified abscissas, taking into account the distribution of each process involved, i.e. $p(\mathbf{y}_i \mid \mathbf{b}_i; \boldsymbol{\beta})$, $p(T_i, \delta_i \mid \mathbf{b}_i; \boldsymbol{\Phi})$ and $p(\mathbf{b}_i; \mathbf{D})$. However, some limitations of this approach are well known, and basically concern the poor approximation of the integral when the random effect distribution is not well approximated by a Gaussian distribution, the need for a high number of quadrature points and the corresponding high computational burden, especially when $\dim(\mathbf{b}_i) > 1$. The PA method, on the other hand, is an appealing alternative to handle these problems. Effectively, it computes the integrand by scaling and centering it only once, at the beginning of the optimization algorithm, to approximate the random effect posterior distribution for faster fitting. See Appendix B for further details.

4.2 The M step

At the $q + 1$-th iteration, the maximization step provides maximum likelihood estimates for the random effect covariance matrix $\mathbf{D}$, the longitudinal parameters $\boldsymbol{\beta}$ and the set of survival parameters $\boldsymbol{\Phi} = (\boldsymbol{\gamma}, \xi, \alpha)$; the estimates are:

$$
\hat{\mathbf{D}}^{(q+1)} = n^{-1} \sum_i \widehat{\mathrm{Cov}}\big(\mathbf{b}_i \mid \mathbf{y}_i, T_i, \delta_i; \hat{\boldsymbol{\theta}}^{(q)}\big)
$$
$$
= n^{-1} \sum_i \mathbf{b}_i^\top \mathbf{b}_i \, p\big(\mathbf{b}_i \mid \mathbf{y}_i, T_i, \delta_i; \hat{\boldsymbol{\theta}}^{(q)}\big)
$$

$$
\hat{\boldsymbol{\beta}}^{(q+1)} = \hat{\boldsymbol{\beta}}^{(q)} - \left\{ \frac{\partial}{\partial \boldsymbol{\beta}^{(q)\top}} \mathcal{S}\big(\hat{\boldsymbol{\beta}}^{(q)}\big) \right\}^{-1} \mathcal{S}\big(\hat{\boldsymbol{\beta}}^{(q)}\big)
$$

$$
\hat{\boldsymbol{\Phi}}^{(q+1)} = \hat{\boldsymbol{\Phi}}^{(q)} - \left\{ \frac{\partial}{\partial \boldsymbol{\Phi}^{(q)\top}} \mathcal{S}\big(\hat{\boldsymbol{\Phi}}^{(q)}\big) \right\}^{-1} \mathcal{S}\big(\hat{\boldsymbol{\Phi}}^{(q)}\big).
$$

For faster fitting, after a fixed number of iterations of the EM algorithm, we allow an optional Quasi-Newton step, that maximizes the observed data log-likelihood:

$$
\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}).
$$

In this case, the score equations are solved with respect to $\boldsymbol{\theta}$ by considering $p(\mathbf{b}_i \mid T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$, leading to a direct maximization of the observed data log-likelihood $\ell(\boldsymbol{\theta})$.

5 Special cases

The GLMJM considers a general longitudinal response whose distribution is a member of the exponential family. In this section, we discuss two specific cases for the longitudinal response, by considering (conditionally) Poisson and Binomial outcomes. As the score vector and the Hessian matrix are required for parameter estimation and standard error calculation, their distribution-specific expressions are reported in Appendix A.

5.1 The Poisson case

Let us assume that $y_i(t)$, $i = 1, \ldots, n$ and $t = 1, \ldots, n_i$, are the observed values for a Poisson random variable $Y_i(t)$, conditional on the random effects $\mathbf{b}_i$. The corresponding conditional density function is

$$
p\big(y_i(t) \mid \mathbf{b}_i; \lambda_i(t)\big) = \frac{\exp(-\lambda_i(t))\lambda_i(t)^{y_i(t)}}{y_i(t)!},
$$

where $\lambda_i(t) = \mathbb{E}(Y_i(t)) = \mathbb{V}(Y_i(t)) = \exp(\boldsymbol{\beta}^\top \mathbf{X}_i(t) + \mathbf{b}_i^\top \mathbf{Z}_i(t))$.

The log-density can be written as $\log p(y_i(t) \mid \mathbf{b}_i; \boldsymbol{\theta}) = -\exp(\boldsymbol{\beta}^\top \mathbf{X}_i(t) + \mathbf{b}_i^\top \mathbf{Z}_i(t)) + y_i(t)(\boldsymbol{\beta}^\top \mathbf{X}_i(t) + \mathbf{b}_i^\top \mathbf{Z}_i(t)) - \log[y_i(t)!]$.

Adopting a Weibull distribution, the survival model is given by (4), where the hazard function is

$$
h\big(T_i \mid M_i(T_i), \mathbf{W}_i; \boldsymbol{\Phi}\big)
$$
$$
= h_0(T_i) \exp\big\{\boldsymbol{\gamma}^\top \mathbf{W}_i + \alpha m_i(T_i)\big\}
$$
$$
= \xi T_i^{\xi-1} \exp\big\{\boldsymbol{\gamma}^\top \mathbf{W}_i + \alpha \exp(\boldsymbol{\beta}^\top \mathbf{X}_i(T_i) + \mathbf{b}_i^\top \mathbf{Z}_i(T_i))\big\}
$$

and the survival function is

$$
S\big(T_i \mid M_i(T_i), \mathbf{W}_i; \boldsymbol{\Phi}\big) = \exp\left\{ -\int_0^{T_i} h_i(s) ds \right\}. \tag{7}
$$

5.2 The Binomial case

When the primary outcome is the realization of a conditional Binomial random variable with size k, $y_i(t)$ denotes the observed number of successes; in this case, if we denote the success probability by $\pi_i(t)$, the conditional density function on the logarithmic scale is

$$\log\{p(y_i(t) \mid \mathbf{b}_i); \pi_i(t)\} = \log\binom{k}{y_i(t)} + y_i(t) \log \pi_i(t)$$
$$+ (k - y_i(t)) \log(1 - \pi_i(t)),$$

where $\pi_i(t) = \frac{\exp(\boldsymbol{\beta}^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t))}{1 + \exp(\boldsymbol{\beta}^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t))}$.

For a survival time following a Weibull distribution, the hazard function is defined by:

$$h(T_i \mid M_i(T_i), \mathbf{W}_i; \boldsymbol{\Phi})$$
$$= h_0(T_i) \exp\{\boldsymbol{\gamma}^\mathsf{T}\mathbf{W}_i + \alpha m_i(T_i)\}$$
$$= \xi T_i^{\xi-1} \exp\left\{\boldsymbol{\gamma}^\mathsf{T}\mathbf{W}_i + \alpha \frac{\exp(\boldsymbol{\beta}^\mathsf{T}\mathbf{X}_i(T_i) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(T_i))}{1 + \exp(\boldsymbol{\beta}^\mathsf{T}\mathbf{X}_i(T_i) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(T_i))}\right\},$$

while the survival function is given by formula (7).

Should a *probit* link be adopted for the longitudinal response, we would have

$$\begin{cases} m_i(t) = F[\boldsymbol{\beta}^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t)] \\ h(T_i \mid M_i(T_i), \mathbf{W}_i; \boldsymbol{\Phi}) \\ \quad = h_0(T_i) \exp\{\boldsymbol{\gamma}^\mathsf{T}\mathbf{W}_i + \alpha m_i(T_i)\}, \end{cases} \tag{8}$$

where $F[\cdot]$ is the CDF for a standard Gaussian random variable. However, in this case, we may notice that the inverse transformation for the logit link is directly interpretable as a log-odds, whereas $F^{-1}(\cdot)$ does not have this simple and direct interpretation; therefore we will mostly consider the logit link function in this paper.

6 Simulation study

To study the behaviour of the proposed model, we have considered the simulation study described below. The simulation results are discussed in Sect. 6.2.

6.1 Simulation design

The simulation study is aimed at analysing the proposed model behaviour under situations of ignorability and non-ignorability of the dropout mechanism. We consider the special cases of (conditionally) Poisson and Binomial responses, and compare the GH and the PA quadrature approaches with respect to parameter estimate bias. For each

setting, we have randomly drawn $N = 500$ samples with size $n = 550$ for the longitudinal response, observed at most for 20 (equally spaced) occasions. We induce attrition according to the survival mechanism in (10), and report the mean values, the MSEs and the standard errors for parameter estimates obtained by fitting the longitudinal model only (i.e. assuming an ignorable missing data mechanism) and the proposed joint model. To evaluate the standard errors performance, the estimated coverage of confidence intervals with nominal level 0.95 for α is reported. The censoring mechanism is induced by an exponential process with parameter 0.25.

For the Poisson case, data have been drawn by a conditional distribution with canonical parameter given by:

$$\eta_i(t) = \log(m_i(t)) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})x_{i1}(t) + \beta_2 x_{i2}$$
$$+ \beta_3(x_{i1}(t)x_{i2}). \tag{9}$$

Here, the covariate x_{i1} is a sequence from 0 to t.max$_i$, where t.max$_i$ is the maximum follow up time for subject i, x_{i2} is a time-constant binary random variable with parameter $p = 0.5$, representing a treatment variable, and $x_{i1}(t)x_{i2}$ is the corresponding interaction. Parameter values are fixed according to the application of the GLMJM to the AIDS data set, see Sect. 7. Hence, $\beta_0 = 1.74$, $\beta_1 = -0.08$, $\beta_2 = -0.04$ and $\beta_3 = 0.02$. The random effects are drawn from a multivariate Gaussian distribution with upper triangular covariance matrix $\mathbf{D} = (0.343, 0.007, 0.001)$.

The survival times are simulated according to a Weibull distribution with hazard function:

$$h(T_i) = \xi T_i^{\xi-1} \exp\{\gamma_0 + \gamma_1 x_{i2} + \alpha m_i(T_i)\}, \tag{10}$$

where $\xi = 1.8$, $\gamma_0 = -3.8$, $\gamma_1 = 0.32$, $\alpha = \{0, -0.3\}$ and $m_i(T_i)$ is the expected value of the longitudinal response at time T_i.

For the Binomial longitudinal outcome, we consider a size $k = 10$. The expected value at time t for the longitudinal process is given by

$$m_i(t) = \frac{\exp\{\eta_i(t)\}}{1 + \exp\{\eta_i(t)\}}, \tag{11}$$

where $\eta_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})x_{i1}(t) + \beta_2 x_{i2} + \beta_3(x_{i1}(t)x_{i2})$, $\beta_0 = 1.2$, $\beta_1 = -0.2$, $\beta_2 = -0.3$, $\beta_3 = 0.3$, and the covariate meaning is the same as in the Poisson simulation design.

On the other hand, the survival times are simulated from (10) with $\gamma_0 = -3.2$, $\gamma_1 = 0.5$, $\xi = 1.8$ and $\alpha = \{0, -1.2\}$; the random effect covariance matrix is in triangular form $\mathbf{D} = (0.343, 0.007, 0.001)$.

6.2 Simulation results

The simulation results are organized as follows. Tables 1–2 report the mean of parameter estimates for the GLMJM and

Table 1 Simulation study. Mean of longitudinal parameter estimates for the Poisson joint model, $\boldsymbol{\beta}_{\text{joint}}$, the MAR model, $\boldsymbol{\beta}_{\text{mar}}$, corresponding MSEs and standard errors (in brackets) under GH and PA Gaussian quadrature rules. For each setting, the mean rate of events, the true α value, the mean of the estimates $\hat{\alpha}$, the mean standard error (first bracket), MSE (second bracket) and coverage are reported. The nominal level is $1 - \alpha = 0.95$

		True	$\boldsymbol{\beta}_{\text{joint}}$	$\boldsymbol{\beta}_{\text{mar}}$	MSE ($\boldsymbol{\beta}_{\text{joint}}$)	MSE ($\boldsymbol{\beta}_{\text{mar}}$)
GH						
Events = 33 %	Intercept	1.74	1.731 (0.148)	1.751 (0.035)	0.022	0.001
$\alpha = 0$	Time	−0.08	−0.081 (0.032)	−0.018 (0.047)	0.001	0.006
$\hat{\alpha} = 0.002(0.001)(0.001)$	Group	−0.04	−0.038 (0.063)	−0.047 (0.049)	0.004	0.002
CILev(α) = 0.96	Group:Time	0.02	0.019 (0.063)	−0.031 (0.067)	0.006	0.007
Events = 25 %	Intercept	1.74	1.741 (0.155)	1.543 (0.026)	0.024	0.039
$\alpha = -0.3$	Time	−0.08	−0.082 (0.089)	−0.540 (0.040)	0.008	0.213
$\hat{\alpha} = -0.310(0.012)(0.001)$	Group	−0.04	−0.038 (0.194)	−0.046 (0.038)	0.038	0.001
CILev(α) = 0.95	Group:Time	0.02	0.021 (0.234)	0.013 (0.055)	0.055	0.004
PA						
Events = 33 %	Intercept	1.74	1.688 (0.055)	1.779 (0.026)	0.013	0.002
$\alpha = 0$	Time	−0.08	−0.072 (0.055)	−0.069 (0.046)	0.004	0.002
$\hat{\alpha} = 0.001(0.002)(0.001)$	Group	−0.04	−0.052 (0.071)	−0.033 (0.039)	0.005	0.002
CILev(α) = 0.95	Group:Time	0.02	0.010 (0.138)	−0.029 (0.066)	0.019	0.007
Events = 23 %	Intercept	1.74	1.640 (0.089)	1.537 (0.026)	0.018	0.042
$\alpha = -0.3$	Time	−0.08	−0.112 (0.054)	−0.545 (0.039)	0.004	0.218
$\hat{\alpha} = -0.27(0.12)(0.08)$	Group	−0.04	−0.046 (0.104)	−0.418 (0.037)	0.011	0.144
CILev(α) = 0.96	Group:Time	0.02	0.021 (0.063)	0.015 (0.055)	0.004	0.004

Table 2 Simulation study. Mean of longitudinal parameter estimates for the Binomial joint model, $\boldsymbol{\beta}_{\text{joint}}$, the MAR model, $\boldsymbol{\beta}_{\text{mar}}$, corresponding MSEs and standard errors (in brackets) under GH and PA Gaussian quadrature rules. For each setting, the mean rate of events, the true α value, the mean of its estimates $\hat{\alpha}$, its mean standard error (first bracket), MSE (second bracket) and coverage are reported. The nominal level is $1 - \alpha = 0.95$

		True	$\boldsymbol{\beta}_{\text{joint}}$	$\boldsymbol{\beta}_{\text{mar}}$	MSE ($\boldsymbol{\beta}_{\text{joint}}$)	MSE ($\boldsymbol{\beta}_{\text{mar}}$)
GH						
Events = 34 %	Intercept	1.2	1.219 (0.035)	1.201 (0.094)	0.016	0.009
$\alpha = 0$	Time	−0.2	−0.222 (0.027)	−0.209 (0.053)	0.001	0.003
$\hat{\alpha} = 0.014(0.009)(0.023)$	Group	−0.3	−0.301 (0.049)	−0.303 (0.133)	0.002	0.018
CILev(α) = 0.91	Group:Time	0.3	0.310 (0.032)	0.312 (0.075)	0.001	0.006
Events = 23 %	Intercept	1.2	1.198 (0.161)	0.448 (0.095)	0.026	0.574
$\alpha = -1.2$	Time	−0.2	−0.281 (0.089)	−0.475 (0.046)	0.014	0.078
$\hat{\alpha} = -1.01(0.548)(0.423)$	Group	−0.3	−0.372 (0.209)	−0.348 (0.145)	0.049	0.023
CILev(α) = 0.96	Group:Time	0.3	0.293 (0.122)	0.004 (0.071)	0.015	0.005
PA						
Events = 33 %	Intercept	1.2	1.199 (0.095)	1.201 (0.091)	0.009	0.008
$\alpha = 0$	Time	−0.2	−0.198 (0.055)	−0.200 (0.050)	0.003	0.003
$\hat{\alpha} = 0.01(0.014)(0.09)$	Group	−0.3	−0.312 (0.114)	−0.306 (0.133)	0.013	0.018
CILev(α) = 0.93	Group:Time	0.3	0.306 (0.077)	0.307 (0.075)	0.006	0.006
Events = 24 %	Intercept	1.2	1.022 (0.032)	0.497 (0.091)	0.032	0.503
$\alpha = -1.2$	Time	−0.2	−0.281 (0.055)	−0.426 (0.045)	0.009	0.053
$\hat{\alpha} = -1.170(0.058)(0.090)$	Group	−0.3	−0.331 (0.045)	−0.311 (0.137)	0.003	0.019
CILev(α) = 0.92	Group:Time	0.3	0.323 (0.032)	0.006 (0.068)	0.001	0.091

Table 3 Simulation study. Mean of longitudinal parameter estimates for the Gaussian joint model, the Gaussian MAR model and corresponding MSEs, when the response is Poisson distributed and PA Gaussian quadrature rule is adopted. The true α value, the mean of estimates $\hat{\alpha}$ is reported

	True	$\boldsymbol{\beta}_{\text{joint}}$	$\boldsymbol{\beta}_{\text{mar}}$	MSE ($\boldsymbol{\beta}_{\text{joint}}$)	MSE ($\boldsymbol{\beta}_{\text{mar}}$)	α	$\hat{\alpha}$
$n = 550$							
Intercept	1.766	2.457	2.456	4.565	4.566	−0.16	0.026
Time	−0.022	−0.022	−0.021	0.736	0.737		
Group	−0.014	−0.010	−0.009	0.797	0.796		
Group:Time	−0.001	−0.003	−0.003	0.719	0.718		
$n = 1000$							
Intercept	1.766	2.462	2.463	0.481	0.482	−0.16	0.012
Time	−0.022	−0.024	−0.023	0.000	0.002		
Group	−0.014	−0.022	−0.021	0.003	0.003		
Group:Time	−0.001	−0.004	−0.004	0.000	0.002		

the MAR model (calculated by fixing $\alpha = 0$), the corresponding MSEs and standard errors (in brackets), for the GH and the PA quadrature rules. Table 1 shows results for the Poisson case, while Table 2 shows the analogous results for the Binomial case.

In most of the analyzed settings, GLMJM parameter estimates present smaller MSEs than the corresponding MAR model estimates. The α parameter seems to be accurately estimated regardless its true value, with an observed coverage proportion close to 0.95 (the nominal level is 0.95). The parameter estimates behaviour is convincing either when GH and PA rules are adopted. With respect to the computational time, however, the algorithm using PA Gaussian quadrature rule reaches convergence much faster than the one based on GH quadrature does. The average convergence time over simulated samples is 60.21 seconds for the PA, compared to 8.68 minutes for the GH rule.

Last, we have considered the following scenario: we have drawn the response from a conditional Poisson distribution and estimated model parameters by using a Gaussian joint model. The results are reported in Table 3; as it can be noted, the misspecification of the longitudinal response distribution may lead to biased parameter estimates, especially for α, although the MSE for longitudinal parameters decreases with larger sample sizes.

7 The AIDS data set

In this section, we discuss the application of the proposed GLMJM to the AIDS dataset of Goldman et al. (1996), see also Carlin and Louis (2009). The study is on 467 HIV-infected patients and has the aim at comparing the efficacy and safety of two randomly assigned antiretroviral drugs: didanosine (ddI) and zalcitabine (ddC). The longitudinal response is the CD4 cell count, recorded in five occasions, i.e.

at the randomization time and after 2, 6, 12 and 18 months. By the end of the study, 188 patients have died, corresponding to 60 % censoring. Standard procedures assume a conditional Gaussian distribution for the CD4 cell counts; we compare this approach to the one obtained through a joint model where the response is assumed to follow a conditional Poisson distribution.

Figure 1 shows the Kaplan-Meier and the GLMJM survival estimates.

The results are shown in Table 4. It can be noticed that for both the Gaussian and the Poisson joint model the non-ignorability parameter is significant. Also, the significance of the longitudinal parameter estimates agree in both the joint models, but for the treatment and the interaction between treatment and time. In this case, while the MAR Poisson model and the Gaussian Joint Model come up with non-significant, positive effects, the Poisson Joint Model gives significant (negative) estimates.

A further issue concerns the interpretation of the α parameter; in fact, it could be controversial to assess ignorability when it is significantly different from zero, see for instance Jansen et al. (2006). Effectively, it may happen that the hazard of dropout at a given time does not depend on the response expected value, but rather on the response measured at previous time occasions, or, more generally, that the joint model does not exactly describe the missing data mechanism behaviour. For this reason, an important issue is to assess the sensitivity of JMs to the assumptions on the dropout mechanism. In this context, Viviani et al. (2012) performed a sensitivity study on joint models through the index of local sensitivity to non-ignorability, see Troxel et al. (2004). The results showed a slight sensitivity of model parameter estimates when the departure from MAR is not too large, and a decreasing effect of the assumptions about ignorability for the missing data mechanism as the number of repeated measurements increase.

Fig. 1 Kaplan-Meier and
GLMJM survival estimate for
the AIDS dataset. The mean
curve is shown on the *left* side,
while the treatment effect is
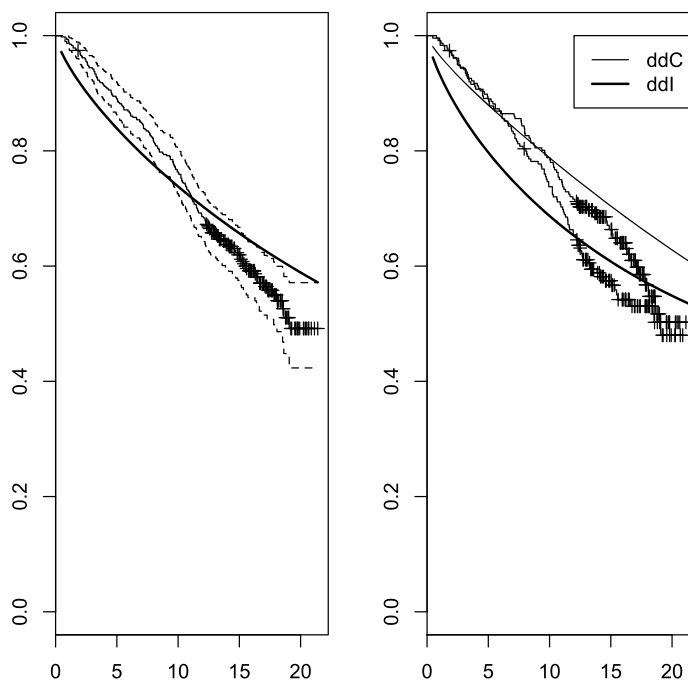distinctly shown on the *right*
side

Table 4 AIDS dataset.
Longitudinal parameter
estimates for the ignorable
model ($\hat{\boldsymbol{\beta}}_{mar}$) and the joint
model ($\hat{\boldsymbol{\beta}}_{jm}$) under Poisson and
Gaussian longitudinal
distribution. The p-values
correspond to $\hat{\boldsymbol{\beta}}_{jm}$, while the
MAR estimates are signed by
the symbol $*$ when they are
significant

	$\hat{\boldsymbol{\beta}}_{jm}$	$\hat{\boldsymbol{\beta}}_{mar}$	SE	z-value	p-value
Poisson JM, $\hat{\alpha} = -0.292$ ($p = 0.016$)					
Intercept	1.747*	1.666*	0.024	72.548	<0.001
Time	−0.079*	−0.033*	0.002	−17.362	<0.001
Treatment	−0.041*	0.072	0.033	−2.389	<0.001
Treatment:Time	0.018*	0.004	0.003	6.210	<0.001
Gaussian JM, $\hat{\alpha} = -1.023$ ($p < 0.001$)					
Intercept	2.510*	2.474*	0.060	41.756	<0.001
Time	−0.037*	−0.037*	0.005	−8.177	<0.001
Treatment	0.053	0.075	0.079	0.678	0.497
Treatment:Time	0.007	0.007	0.006	0.994	0.320

8 Discussion

In this paper, a Generalized Linear Mixed Joint Model
(GLMJM) has been proposed to model a longitudinal out-
come with distribution in the exponential family, and a
time-to-event process. The basic assumption underlying this
model is that the expected value of the response at a given
time influences the hazard of the dropout event at the same
time. This approach is particularly convenient in the context
of longitudinal studies, where either continuous or discrete
responses can be subject to non-ignorable attrition. In fact,
it allows to handle non random missingness by taking into
account the expected value of the response at time t as a
covariate in the dropout process.

Parameter estimation is performed through an EM algo-
rithm, requiring numerical integration over the posterior ran-
dom effect distribution in the expectation step. The integral
is approximated by Gauss-Hermite (GH) and Pseudo Adap-
tive (PA) Gaussian rules.

The main findings we may derive from a simulation study
and the application to a benchmark dataset are that param-
eter estimates obtained through the GLMJM are less biased
than the ones under ignorability. This phenomenon is also
evident when the PA quadrature rule is adopted. In light of
this, we can conclude that, for GLMJMs and in the pro-
posed settings, the PA quadrature rule performs satisfacto-
rily and it can be considered as a proper approximation pro-
cedure, as it substantially reduces the computational time
with very limited loss in terms of precision. Applications to

the AIDS dataset of Goldman et al. (1996) and Carlin and Louis (2009) has suggested that the GLMJM highlights two further significant effects, which are not found in the Gaussian JM. This points out, once again, the need for a sensitivity analysis.

Appendix A: Score vectors and Hessian matrices for the Poisson and the Binomial joint model

As highlighted in Sect. 2, for parameter estimation in the GLMJMs, the EM algorithm requires the computation of the observed data score vector and Hessian matrix for the complete joint model. The score can assume the form:

$$S(\theta) = \sum_i \int \omega(\theta, \mathbf{b}_i) p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta) d\mathbf{b}_i, \tag{12}$$

where $\omega(\theta, \mathbf{b}_i) = \partial \{\log p(T_i, \delta_i | \mathbf{b}_i; \theta) + \log p(\mathbf{y}_i | \mathbf{b}_i; \theta) + \log p(\mathbf{b}_i; \theta)\}/\partial \theta^\top$ denotes the complete data score vector. Therefore, it can be considered as the expected value of the complete data score vector with respect to the posterior distribution of the random effects. In the M-step, at the q-th iteration, equation (12) is solved with respect to θ, with $p(b_i | T_i, \delta_i, y_i; \theta)$ calculated by fixing $\theta = \theta^{(q-1)}$.

Since this work has focused on the longitudinal parameter estimation under non-ignorability, in this section we report the score vector and the Hessian matrices for the specific longitudinal processes considered (Poisson and Binomial distributions).

For the Poisson longitudinal distribution, the score vector for the fixed effects in the longitudinal model can be written as

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \int_{\mathbf{b}_i} -\mathbf{x}_i \exp\{\boldsymbol{\beta}^\top \mathbf{x} + \mathbf{b}_i^\top \mathbf{z}_i)\} + \mathbf{x}_i \mathbf{y}_i$$
$$- \int_0^t h_0(s)\alpha \mathbf{x}_i \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}$$
$$\times \exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \alpha \exp(\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i)\} ds d\mathbf{b}_i,$$

while the Hessian matrix assumes the form

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \int_{\mathbf{b}_i} -\mathbf{x}_i^\top \mathbf{x}_i \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}$$
$$+ \delta_i \alpha \mathbf{x}_i^\top \mathbf{x}_i \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}$$
$$- \int_0^t h_0(s)\alpha \mathbf{x}_i \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}$$
$$\times \exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \alpha \exp(\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i)\} ds$$
$$- \int_0^t h_0(s)\alpha^2 \mathbf{x}_i \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\} \mathbf{x}_i^\top$$
$$\times \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}^\top$$
$$\times \exp\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \alpha \exp(\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i)\} ds d\mathbf{b}_i.$$

For the Binomial case, on the other hand, the score vector and the Hessian matrix are given by, respectively:

$$S(\boldsymbol{\beta}) = y_i(t)\mathbf{x}_i - n\mathbf{x}_i^\top \frac{\exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}}{1 + \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}}$$
$$+ \delta_i \alpha n \mathbf{x}_i^\top \frac{\exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}}{[1 + \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}]^2}$$
$$- \int_0^t h_0(s)\mathbf{x}_i^\top \frac{\exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}}{[1 + \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}]^2} \alpha n$$
$$\times \exp\left\{\boldsymbol{\gamma}^\top \mathbf{w}_i + \frac{\exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}}{1 + \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}}\right\} ds d\mathbf{b}_i,$$

and

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -n\mathbf{x}_i^\top \mathbf{x}_i \mathbf{B}_i + \delta_i \alpha n \mathbf{x}_i \mathbf{B}_i$$
$$- \int_0^t h_0(s)\mathbf{x}_i \mathbf{A}_i \alpha^2 n \exp\left\{\boldsymbol{\gamma}^\top \mathbf{w}_i \right.$$
$$\left. + \frac{\exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}}{1 + \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}}\right\} \mathbf{x}_i^\top \mathbf{B}_i ds d\mathbf{b}_i,$$

where

$$\mathbf{A}_i = \frac{[2\mathbf{x}_i \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}][1 - 2\mathbf{x}_i \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\} - 1/2\mathbf{x}_i \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}^2]}{[1 + \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}]^4},$$

while

$$\mathbf{B}_i = \frac{\exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}}{[1 + \exp\{\boldsymbol{\beta}^\top \mathbf{x}_i + \mathbf{b}_i^\top \mathbf{z}_i\}]^2}.$$

Appendix B: The pseudo-adaptive Gaussian quadrature

As it has been pointed out in Sect. 3, the computation of the observed data log-likelihood requires the use of quadra-

ture rules, such as Gaussian quadrature. In this Appendix, we describe an alternative numerical integration approach based on the Pseudo adaptive Gauss-Hermite rule, proposed by Rizopoulos (2012) for standard JMs.

The standard Gauss-Hermite rule (GH) approximates the integral in (3) by a weighted sum of integrand evaluations at pre-specified abscissas, see e.g. Press et al. (2007). The quality of the approximation improves with the number of quadrature points. However, as the GH requires the evaluation of the integrand over the Cartesian product of the abscissas corresponding to each dimension in the random effect vector, the computational effort increases exponentially with the random effect dimension. Another limitation of the GH approach is that the corresponding approximation depends on the location of the quadrature points with respect to the main mass of the integrand, that may be far from each other. For instance, the random effect distribution may be far from Gaussianity.

To solve these problems, an adaptive quadrature approach may be used, that appropriately centres and scales the integrand at each iteration of the optimization algorithm. However, this kind of approach implies a substantial computational burden due to the location of the mode and the computation of the Hessian matrix of the random effect posterior distribution at each iteration.

To decrease the computational effort, we propose to use a pseudo-adaptive quadrature rule. This rule is based on the Bayesian Central Limit (BCL) theorem, see Cox and Hinkley (1974); the posterior distribution of the random effects can be approximated by a multivariate Gaussian distribution as the number of repeated measurements increases. The logarithm of this distribution can in fact be written as

$$\log p(\mathbf{b}_i | T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\theta})$$

$$\propto \sum_{j=1}^{n_i} \log p\big(y_i(t_{ij}) | \mathbf{b}_i; \boldsymbol{\beta}\big) + \log p(T_i, \delta_i | \mathbf{b}_i; \boldsymbol{\Phi})$$

$$+ \log p(\mathbf{b}_i; \mathbf{D}). \tag{13}$$

It is clear from (13) that as n_i increases, the predominant term becomes the longitudinal density $\log p(\mathbf{y}_i | \mathbf{b}_i; \boldsymbol{\beta})$. According to the BCL and under general regularity conditions, as $n_i \to \infty$, we have

$$p(\mathbf{b}_i; T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\beta}) \xrightarrow{p} \mathcal{N}\big(\tilde{\mathbf{b}}_i, \tilde{\mathbf{H}}_i^{-1}\big),$$

where $\tilde{\mathbf{b}}_i = \arg\max_b \{\log p(\mathbf{y}_i | \mathbf{b}_i; \boldsymbol{\beta})\}$ and $\tilde{\mathbf{H}}_i = -\partial^2 \times \log p(\mathbf{y}_i | \mathbf{b}_i; \boldsymbol{\beta}) / \partial \mathbf{b} \partial \mathbf{b}^\top |_{\mathbf{b}=\tilde{\mathbf{b}}}$. Practically, as n_i increases, it is sufficient to re-center and re-scale the integrand for each subject using only the information coming from the longitudinal outcome model.

Applying this procedure to the joint model, we propose to first fit a generalized linear mixed model, extract the empirical Bayes estimates $\tilde{\mathbf{b}}_i = \arg\max_b \{\log p(\mathbf{y}_i | \mathbf{b}_i; \tilde{\theta}_y)\}$ and

the corresponding covariance matrix $\tilde{\mathbf{H}}_i^{-1} = (\mathbf{B}_i \mathbf{B}_i^T)^{-1}$; successively, we apply the following transformation to compute the score vector:

$$\mathbb{E}\big\{\mathcal{S}(\boldsymbol{\theta}, \mathbf{b}_i) | T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\theta}\big\}$$

$$\approx 2^{q/2} |\tilde{\mathbf{B}}_i|^{-1} \sum_{t_1, \dots, t_q} \pi_t \mathcal{S}(\boldsymbol{\theta}, \tilde{r}_t) p(\tilde{r}_t | T_i, \delta_i, \mathbf{y}_i; \boldsymbol{\theta})$$

$$\times \exp\big(\|\mathbf{b}_t\|^2\big),$$

where q is the number of random effects, $\sum_{t_1, \dots, t_q}$ is a shorthand for $\sum_{t_1=1}^{K}, \dots, \sum_{t_q=1}^{K}$, K denoting the common number of quadrature points, $\mathbf{b}_t = (b_{t1}, \dots, b_{tq})$ are the abscissas with corresponding weights π_t, $\tilde{r}_t = \tilde{\mathbf{b}}_i + \sqrt{2}\mathbf{B}_i^{-1}\mathbf{b}_t$ are the re-scaled abscissas and $\tilde{\theta}_i$ denotes the maximum likelihood estimates for the longitudinal model.

This procedure is very similar to the GH rule, but it is implemented only once, at the start of the EM algorithm, leading to faster estimation.

References

Albert, P.S., Follmann, D.A.: Modeling repeated count data subject to informative dropout. Biometrics **56**, 667–677 (2000)

Albert, P.S., Follmann, D.A., Wang, S.A., Suh, E.B.: A latent autoregressive model for longitudinal binary data subject to informative missingness. Biometrics **58**, 631–642 (2002)

Andersen, P., Gill, R.: Cox's regression model for counting processes: a large sample study. Ann. Stat. **10**, 1100–1120 (1982)

Carlin, B.P., Louis, T.A.: Bayesian Methods for Data Analysis. Chapman & Hall/CRC Press, Boca Raton (2009)

Cox, D.R.: Regression models and life tables (with discussion). J. R. Stat. Soc., Ser. B, Stat. Methodol. **34**, 187–220 (1972)

Cox, D.R., Hinkley, D.: Theoretical Statistics. Chapman & Hall, London (1974)

Geyer, J.C., Thompson, E.A.: Constrained Monte Carlo maximum likelihood for dependent data. J. R. Stat. Soc., Ser. B, Stat. Methodol. **54**, 657–699 (1992)

Goldman, A., Carlin, B., Crane, L., Launer, C., Korvick, J., Deyton, L., Abrams, D.: Response of CD4+ and clinical consequences to treatment using ddI or ddC in patients with advanced HIV infection. J. Acquir. Immune Defic. Syndr. Human Retrovirol. **11**, 161–169 (1996)

Hsieh, F., Tseng, Y.-K., Wang, J.-K.: Joint modeling of survival and longitudinal data: likelihood approach revisited. Biometrics **62**, 1037–1043 (2006)

Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G., Kenward, M.G.: The nature of sensitivity in missing not at random models. Comput. Stat. Data Anal. **50**, 830–854 (2006)

Kronrod, A.S.: On the theory of quadratic pencils of selfadjoint operators. Dokl. Akad. Nauk SSSR **154**, 283–286 (1964)

Louis, T.A.: Finding the observed information matrix when using the EM algorithm. J. R. Stat. Soc. B **44**, 226–233 (1982)

Molenberghs, G., Verbeke, G.: Models for Discrete Longitudinal Data. Springer, New York (2005)

Molenberghs, G., Verbeke, G., Demetrio, C.G.B., Vieira, M.C.: A family of generalized linear models for repeated measures with normal and conjugate random effects. Stat. Sci. **25**, 325–347 (2011)

Oakes, D.: Direct calculation of the information matrix via the EM algorithm. J. R. Stat. Soc., Ser. B, Stat. Methodol. **61**, 479–482 (1999)

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

Rizopoulos, D.: Jm: an R package for the joint modelling of longitudinal and time-to-event data. J. Stat. Softw. **35**, 9 (2010)

Rizopoulos, D.: Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. Comput. Stat. Data Anal. **56**, 491–501 (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 (2011)

Rizopoulos, D., Verbeke, G., Lesaffre, E.: Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. J. R. Stat. Soc., Ser. B, Stat. Methodol. **71**, 637–654 (2009)

Rizopoulos, D., Verbeke, G., Molenberghs, G.: Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. Biometrics **66**, 20–29 (2010)

Troxel, A.B., Ma, G., Heitjan, F.: Shared parameter models for the joint analysis of longitudinal data and event times. Stat. Sin. **14**, 1221–1237 (2004)

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

Viviani, S., Rizopoulos, D., Alfó, M.: Local Sensitivity to non-ignorability in Joint Models. Stat. Model. (2012, major revision)

Wu, M., Carrol, R.: Estimation and comparison of changes in presence of informative right censoring by modelling the censoring process. Biometrics **45**, 939–955 (1988)

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