Research Article

Statistics in Medicine

Received 4 April 2014, Accepted 27 November 2014 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6393

Longitudinal quantile regression in the presence of informative dropout through longitudinal–survival joint modeling

Alessio Farcomeni[a*†] **and Sara Viviani**[b]

We propose a joint model for a time-to-event outcome and a quantile of a continuous response repeatedly measured over time. The quantile and survival processes are associated via shared latent and manifest variables. Our joint model provides a flexible approach to handle informative dropout in quantile regression. A Monte Carlo expectation maximization strategy based on importance sampling is proposed, which is directly applicable under any distributional assumption for the longitudinal outcome and random effects. We consider both parametric and nonparametric assumptions for the baseline hazard. We illustrate through a simulation study and an application to an original data set about dilated cardiomyopathies. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: quantile regression; longitudinal regression; joint models; shared-parameter models

1. Introduction

During longitudinal studies, subjects may be lost at follow-up because of events that are associated with the outcome of interest. Failure to model dropout may lead to biased estimates in such cases. From the reverse perspective, the time trend of a longitudinal measurement may predict the risk of an event (e.g., a steadily decreasing CD4 count is predictive of adverse events in HIV patients). A general account of related longitudinal and survival processes can be found in [1]. Participation in a study can be described, in general, by a survival model of time to dropout. We treat terminating events (i.e., death) as dropout. This implies that the focus of the analysis is on describing the health status of the cohort [2], but other approaches are possible to target different aims.

The simplest way of taking into account informative dropout is through pattern-mixture models [3–5], where the outcome distribution is specified conditionally on the time to dropout. Selection models on the other hand condition the dropout mechanism to unobserved responses directly [6] or indirectly. A classical indirect link is given by the assumption that survival and longitudinal processes are independent, conditional on unobserved shared random effects [7]. One can similarly assume that the risk of event at time t is influenced by the *expected value* of the longitudinal response [8, 9]. In the resulting joint model (JM), the hazard of dropout is a function of the predicted longitudinal outcome, that is, of shared random and fixed effects, and related covariates. See also [10–12] and references therein.

Our main concern is that the expected value of the longitudinal outcome may not always be the summary of interest. Further, in some cases, it might be difficult to find a suitable transformation to normality for the outcome, or some resistance to outliers may be desired. An effective solution to all these issues is given by focusing on conditional quantiles [13] of the longitudinal outcome. Median regression models are informally robust (see [14, 15] for reviews of formally robust methods in medical research). Furthermore, in many biomedical applications, interest lies in at least one of the tails, and covariates may have different effects on different quantiles of a distribution. In HIV research, the effect of the treatment

[a] *Department of Public Health and Infectious Diseases, Sapienza University of Rome, Piazzale Aldo Moro 5, Rome 00186, Italy*
[b] *Department of Economics and Statistics, Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, Rome 00153, Italy*
**Correspondence to: Alessio Farcomeni, Department of Public Health and Infectious Diseases, Sapienza University of Rome, Piazzale Aldo Moro 5, Rome 00186, Italy.*
[†] *E-mail: alessio.farcomeni@uniroma1.it*

Copyright © 2014 John Wiley & Sons, Ltd.

is more important on the left tail of the distribution where subjects are at higher risk. Other examples include longitudinal fetal [16] and child [17] growth studies, which are usually focused on low and high quantiles of key anthropometric measurements. Longitudinal quantile regression models are proposed among others in [18], which maximizes a penalized version of the likelihood, and [19], which introduces a random intercept and assumes that the outcome follows the asymmetric Laplace distribution (ALD), conditional on the random effects. An extension of the random intercept model for general linear mixed quantile regression can be found in [20]. See also [21]. The ALD assumption is also used in [22], assuming that random effects are time varying and follow a discrete distribution. Informative missing data are ubiquitous in statistical applications, especially in longitudinal studies, but there are very few approaches to quantile regression with informative dropout. In [23] and [24], estimating equations are weighted proportionally to the inverse of the probability of dropout. In a Bayesian framework, Yuan and Yin [25] model missingness as a binary time series sharing a random effect with the quantile regression process. A common limit of these approaches is that dropout can occur only at one of the observation times of the longitudinal process. This does not hold in general (e.g., when measurements are scheduled according to a protocol and when death occurs between two visits, as can be expected in the CD4 example). Moreover, all approaches proposed so far do not directly model the strength of association between the longitudinal and time-to-event processes. The latter is summarized by nonignorability parameter(s) in JMs and shared-parameter models. A quantile regression model with ignorable missingness for potentially clustered data is outlined in [26].

In this paper, we propose a JM for a right-censored time-to-event outcome and a quantile of a continuous response repeatedly measured over time.

In our approach, the survival process depends on a linear combination of the latent variables and the predicted quantile of the longitudinal outcome, resulting in a hybrid between the shared-parameter model and JM. Maximum likelihood estimates are efficiently derived by setting up a Monte Carlo expectation maximization (MCEM) algorithm based on importance sampling (IS) [27,28]. There are two clear advantages of using IS: first, the resulting MCEM is completely general and straightforward to use with any distributional assumption for the longitudinal observations or the random effects. Secondly, it is computationally efficient given that we evaluate the posterior distribution only once for each sample. The MCEM approach, unlike commonly used quadrature methods, is amenable also to moderate dimensional random effects. This can be useful for instance when, in addition to a random intercept and slope, random effects are used to capture dependence arising from clustering of subjects (e.g., in multicenter studies).

The rest of the paper is as follows. In the next section, we describe the proposed model and derive inference in Section 3. The method is illustrated through simulations in Section 4. In Section 5, we analyze an original data set about patients with cardiomyopathy. Finally, in Section 6, we conclude with a brief discussion.

2. Joint longitudinal quantile and survival regressions

Let $T_i = \min(T_i^*, C_i)$ denote 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. Further, let Δ_i be the corresponding event indicator defined by $\Delta_i = I(T_i^* \leqslant C_i)$, where $I(\cdot)$ is the indicator function. In applications, we will therefore set $\Delta_i = 1$ when the informative dropout or terminating event has been recorded and $\Delta_i = 0$ when the loss at follow-up is deemed to be at random for the ith subject (e.g., we have complete records and no events, or we have missing values because the study has been closed before the subject reached the last observation time).

The continuous outcome Y_{it} is repeatedly observed at times $t = 1, \ldots, n_i$ before T_i and is missing for $t \geqslant T_i$. The longitudinal outcome is collected in $\boldsymbol{Y}_i = \{Y_i(t) : t \leqslant T_i\}$. We assume that the longitudinal process is associated with T_i^*, that is, with the true event time, but, as customary in survival analysis, is independent of the censoring time C_i.

We let $\boldsymbol{X}_{it}$ denote a vector of predictors used to model only the longitudinal outcome and $\boldsymbol{W}_i$ a vector of (time-fixed) predictors used to model only the survival process. Finally, $\boldsymbol{H}_{it}$ denotes a vector of predictors shared by the longitudinal and survival processes. Covariates are associated with fixed effects $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\delta}$. We then have a vector $\boldsymbol{Z}_{it}$ of predictors associated with q-dimensional random effects $\boldsymbol{u}_i$ and two nonignorability scalar parameters α_1 (associated with fixed effects) and α_2

Copyright © 2014 John Wiley & Sons, Ltd.

(associated with random effects). Our model can be expressed by a set of two equations, one for the longitudinal outcome and the other one for the survival outcome:

$$\begin{cases} Y_{it} = \beta' \boldsymbol{X}_{it} + \delta' \boldsymbol{H}_{it} + u_i' \boldsymbol{Z}_{it} + \epsilon_{it} = \tilde{\tau}_{it} + \epsilon_{it}, \\ h\left(T_i | \mathcal{T}_{iT_i}, \boldsymbol{W}_i; \gamma, \alpha_1, \alpha_2\right) = h_0(T_i) \exp\left\{\gamma' \boldsymbol{W}_i + \alpha_1 \delta' \boldsymbol{H}_{iT_i} + \alpha_2 u_i' \boldsymbol{Z}_{iT_i}\right\}, \end{cases} \tag{1}$$

where the first equation gives the longitudinal model and the second the time-to-event model and $h_0(\cdot)$ is a baseline function. Specifically, the model for the longitudinal outcome $\tilde{\tau}_{it}$ is formulated along the usual lines for mixed effects models [29], and the model for the time-to-event outcome is based on the subject-specific hazard function $h(T_i)$ [30,31]. The risk of dropout is conditional on $\mathcal{T}_{iT_i} = \{\tilde{\tau}_{is} : 0 \leqslant s \leqslant T_i\}$, that is, the error-free longitudinal process history up to time T_i. The model is completed by a distributional assumption for the shared latent distribution, that is, by specifying $\boldsymbol{u}_i \sim f(\boldsymbol{u}_i)$. Few options are discussed in Section 2.3.

The degree of dependence between the longitudinal and survival processes is measured by the association parameters α_1 and α_2, which are introduced to assess potential nonignorability of the missing-data mechanism. In doing so, we admit two sources of nonignorability: a part that can be explained through observed heterogeneity in $\boldsymbol{H}_{it}$ (but not $\boldsymbol{X}_{it}$) and a part that is due to unobserved heterogeneity. The log-hazard ratios associated with $\boldsymbol{H}_{it}$ can be estimated as $\alpha_1 \boldsymbol{\delta}$, while those associated with $\boldsymbol{W}_i$ are directly estimated as $\boldsymbol{\gamma}$. All parameters are identifiable even if we multiply some of them in the survival model equation.

In the proposed JM, the individual heterogeneity terms $\boldsymbol{u}_i' \boldsymbol{Z}_{it}$ can be interpreted as the deviation of the ith subject from the population prediction in the longitudinal process and (after multiplication by α_2) as an additive frailty term in the survival process. The posterior estimates of these terms can be used to assess individual hazards and future values of the longitudinal process. Each dimension of $\boldsymbol{u}_i$ could also be interpreted separately in many cases (Section 5). All these elements can be used for conditional screening in order to identify subjects at higher risk of event and/or more needing intervention to control the longitudinal outcome.

It is important to underline that survival parameters γ corresponding to baseline covariates will also vary depending on τ. We believe different estimates should be treated differently according to the research aims. If the focus is on longitudinal modeling with dropout, then the survival part is simply used to remove bias, and the fact that there are different estimates can be ignored. If the focus is on survival with a longitudinal covariate measured with error, then one should fix a single τ. This can be by default set as $\tau = 0.5$, or different values of τ can be specified, and the best selected according to some measure of the goodness of fit. The latter strategy allows dealing with skewed measurement errors. Finally, if the focus is on the JM itself, the entire distribution of parameter estimates (as a function of τ) is of interest. In this case, stable estimates indicate that survival parameters are approximately orthogonal to longitudinal covariates, and varying estimates indicate the opposite. See Section 5 for an example.

We start describing each part of the model separately, and then we outline how they are linked in the model observed likelihood.

2.1. The longitudinal model

The parametric assumption on the error distribution of the longitudinal outcome identifies our target for inference. If we assume that ε_{it} follows a zero-centered Gaussian, we work with the conditional expectation of the outcome. On the other hand, if we assume an ALD, $\tilde{\tau}_{it}$ does not represent the conditional mean of Y_{it} anymore, but its conditional $0 < \tau < 1$ quantile. Note that τ is prespecified and fixed. The resulting density of Y_{it}, conditional on covariates and random effects, is given by

$$f\left(Y_{it} | \boldsymbol{X}_{it}, \boldsymbol{H}_{it}, \boldsymbol{Z}_{it}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{u}_i, \sigma\right) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho\left(\frac{Y_{it} - \boldsymbol{\beta}' \boldsymbol{X}_{it} - \boldsymbol{\delta}' \boldsymbol{H}_{it} - \boldsymbol{u}_i' \boldsymbol{Z}_{it}}{\sigma}\right)\right\}, \tag{2}$$

where $\rho(u) = u\{\tau - I(u < 0)\}$ is the quantile loss function and $\sigma > 0$ is a scale parameter. When $Z_{it} = 1$, we obtain a random intercept model [19]. The ALD is justifiable because the maximum likelihood is equivalent to minimization of the quantile loss function, which corresponds to the absence of parametric assumptions on ε [32]. Further, it can be seen in simulation studies to lead to precise estimates even when the residuals are not ALD distributed [20,22].

Copyright © 2014 John Wiley & Sons, Ltd.

2.2. The survival model

The time-varying baseline risk function $h_0(s)$ in (1) can be seen as the risk obtained when all covariates and random effects are exactly zero. We may wish to specify a flexible parametric form for $h_0(s)$; for example, $h_0(s) = \eta s^{\eta-1}$, leading to a Weibull model [9, 33, 34]. We may also leave it unspecified [1]. The inferential strategy for obtaining the maximum likelihood estimate under parametric or nonparametric assumptions is slightly different. We outline both in the next section.

The time-to-event distribution can in both cases be written as

$$\begin{aligned} f\left(T_i, \Delta_i \mid \boldsymbol{u}_i\right) &= f\left(T_i \mid \mathcal{T}_{iT_i}, \boldsymbol{W}_i\right)^{\Delta_i} S\left(T_i \mid \mathcal{T}_{iT_i}, \boldsymbol{W}_i\right)^{1-\Delta_i}, \\ &= h\left(T_i \mid \mathcal{T}_{iT_i}, \boldsymbol{W}_i\right)^{\Delta_i} S\left(T_i \mid \mathcal{T}_{iT_i}, \boldsymbol{W}_i\right), \end{aligned} \tag{3}$$

where $S(\cdot)$ denotes the survival function, that is,

$$S\left(T_i \mid \mathcal{T}_{it}, \boldsymbol{W}_i\right) = \exp\left\{ - \int_0^{T_i} h_0(s) \exp\left\{\boldsymbol{\gamma}'\boldsymbol{W}_i + \alpha_1 \boldsymbol{\delta}'\boldsymbol{H}_{is} + \alpha_2 \boldsymbol{u}_i'\boldsymbol{Z}_{is}\right\} ds \right\},$$

while $h(T_i \mid \mathcal{T}_{it}, \boldsymbol{W}_i)$ is given by the second equation in (1).

2.3. The random-effects model

A commonly used distribution for random effects is the multivariate normal. This is convenient to work with when a Gaussian assumption is formulated also for the longitudinal outcome. The multivariate normal may not be satisfactory anyway when the number of occasions is small [35, 36] and/or the dimensionality q of $\boldsymbol{Z}_{it}$ is large. Further, we may often expect a slower convergence of the posterior distribution of the random effects to a multivariate normal when modeling lower or upper quantiles. Two valid alternatives for the random-effects distribution are given by a multivariate T with k degrees of freedom:

$$f\left(\boldsymbol{u}_i|\boldsymbol{\Sigma}\right) \propto |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{1}{k}\boldsymbol{u}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_i\right)^{-\frac{k+q}{2}},$$

which can be used to capture fat tails of the random effects, and a multivariate ALD, which is often suggested in the quantile regression framework [20]:

$$f\left(u_i|\boldsymbol{\Sigma}\right) \propto |\boldsymbol{\Sigma}|^{-1/2} \left(\frac{\boldsymbol{u}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_i}{2}\right)^{v/2} K_v\left(\sqrt{\boldsymbol{u}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{u}_i}\right),$$

where $v = (2-q)/2$ and $K_v(\cdot)$ is the modified Bessel function of the third kind. The most appropriate random-effects distribution may be chosen for instance using the BIC or the AIC [37].

2.4. Observed likelihood

In what follows, $\boldsymbol{\theta}$ is a short-hand notation for model parameters, that is, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, α_1, α_2, $\boldsymbol{\gamma}$, σ, $\boldsymbol{\Sigma}$, and any parameter associated with baseline hazard $h_0(s)$. The joint likelihood contribution of the longitudinal and survival processes for the ith subject is obtained by integrating the conditional distributions in (1) over the random-effects space:

$$f\left(T_i, \Delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}\right) = \int f\left(\boldsymbol{Y}_i|\boldsymbol{u}_i\right) f\left(T_i, \Delta_i|\boldsymbol{u}_i\right) f\left(\boldsymbol{u}_i|\Sigma\right) d\boldsymbol{u}_i, \tag{4}$$

where $f(T_i, \Delta_i \mid \boldsymbol{u}_i)$ is given in Equation (3) and

$$f\left(\boldsymbol{Y}_i|\boldsymbol{u}_i\right) = \prod_{t=1}^{n_i} f\left(Y_{it}|\boldsymbol{u}_i\right),$$

while $f(Y_{it} \mid \boldsymbol{u}_i)$ is given in Equation (2). The observed data log-likelihood for the joint quantile regression model is then

$$\ell(\boldsymbol{\theta}) = \sum_i \log f\left(T_i, \Delta_i, y_i; \boldsymbol{\theta}\right). \tag{5}$$

Copyright © 2014 John Wiley & Sons, Ltd.

The integrals involved in (4) are usually tackled in the JM context through quadrature methods [33, 34]. Direct maximization of the likelihood can be achieved through SAS Proc NLMIXED [38, 39]. While effective in one or two dimensions, quadrature methods tend to become too slow or less precise as the dimensionality of the random effects grows. Furthermore, quadrature methods should have to be tailored to the random-effects distribution (e.g., a Gauss–Hermite quadrature would be best for Gaussian random effects, while a Gauss–Laguerre may be better under other assumptions). In the next section, we propose a Monte Carlo strategy that is completely general and allows us to set up an algorithm that is easily adapted to any assumption on the random effects and to any functional form for the two parts of the JM.

3. Estimation of the proposed model

We propose an MCEM algorithm for fitting the proposed model. This involves alternating two steps until convergence: (i) sampling from the posterior distribution for the random effects, given the data and current values of the parameters (Monte Carlo step) and obtaining the conditional expected value of the complete data log-likelihood (E-step) and (ii) maximizing the latter (M-step). The algorithm is guaranteed to converge to a local optimum, or close to it. In order to increase the odds of obtaining the global optimum, we perform a multistart. The first run starts from estimates obtained from separate longitudinal and survival models, which are readily available. Other runs are initialized by randomly perturbing the deterministic initial solution. An issue with MCEM algorithms is how to obtain standard errors and confidence intervals. When performing quantile regression on the longitudinal outcome, we use nonparametric bootstrap [40, 41]. We preserve the dependence structure in the data by resampling subjects rather than separately resampling the outcomes (and related predictors). Under the usual regularity conditions, tests on the regression parameters may then be simply performed by using Wald statistics based on the standard errors.

The MCEM algorithm is based on the complete likelihood, that is, the likelihood we would have if we could observe the random effects. Recalling that T, Δ, and $\boldsymbol{Y}$ are conditionally independent, the individual contribution to the complete data log-likelihood can be obtained as

$$\log f\left(T_i, \Delta_i, \boldsymbol{Y}_i, \boldsymbol{u}_i; \theta\right) = \log f\left(T_i, \Delta_i | \boldsymbol{u}_i; \boldsymbol{\theta}\right) + \log f\left(\boldsymbol{Y}_i | \boldsymbol{u}_i; \boldsymbol{\theta}\right) + \log f\left(\boldsymbol{u}_i | \boldsymbol{\Sigma}\right). \tag{6}$$

Assuming an ALD for the error distribution of the longitudinal measurements, the complete data log-likelihood is as follows:

$$\begin{aligned}
\ell_c(\boldsymbol{\theta}) &= \sum_i \log f\left(T_i, \Delta_i, \boldsymbol{Y}_i, \boldsymbol{u}_i; \boldsymbol{\theta}\right), \\
&= \sum_i \log f\left(Y_i | \boldsymbol{u}_i; \boldsymbol{\theta}\right) + \sum_i \log f\left(T_i, \Delta_i | \boldsymbol{u}_i; \boldsymbol{\theta}\right) + \sum_i \log f\left(\boldsymbol{u}_i | \boldsymbol{\Sigma}\right), \\
&= -\log \sigma \sum_i n_i - \sum_i \sum_{t=1}^{n_i} \rho\left(\frac{Y_{it} - \boldsymbol{\beta}'\boldsymbol{X}_{it} - \boldsymbol{\delta}'\boldsymbol{H}_{it} - \boldsymbol{u}_i'\boldsymbol{Z}_{it}}{\sigma}\right) \\
&\quad + \sum_i \Delta_i \log h_0\left(T_i\right) + \sum_i \Delta_i \boldsymbol{\gamma}'\boldsymbol{W}_i + \alpha_1 \sum_i \Delta_i \boldsymbol{\delta}'\boldsymbol{H}_{iT_i} + \alpha_2 \sum_i \Delta_i \boldsymbol{u}_i'\boldsymbol{Z}_{iT_i} \\
&\quad - \sum_i \int_0^{T_i} h_0(s) \exp\left\{\boldsymbol{\gamma}'\boldsymbol{W}_i + \alpha_1 \boldsymbol{\delta}'\boldsymbol{H}_{is} + \alpha_2 \boldsymbol{u}_i'\boldsymbol{Z}_{is}\right\} ds \\
&\quad + \sum_i \log f\left(\boldsymbol{u}_i | \boldsymbol{\Sigma}\right).
\end{aligned} \tag{7}$$

3.1. Monte Carlo E-step

The conditional expected value of (6) for the ith subject at the jth iteration of the algorithm is expressed as

$$\begin{aligned}
\mathbb{E}\left[\ell_c\left(\boldsymbol{\theta} \mid T_i, \Delta_i, \boldsymbol{Y}_i, \boldsymbol{u}_i\right)\right] = \sum_i \int &\left[\log f\left(\boldsymbol{Y}_i | \boldsymbol{u}_i; \boldsymbol{\theta}\right) + \log f\left(T_i, \Delta_i | \boldsymbol{u}_i; \boldsymbol{\theta}\right)\right. \\
&\left. + \log f\left(\boldsymbol{u}_i; \boldsymbol{\theta}\right)\right] f\left(\boldsymbol{u}_i \mid T_i, \Delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(j)}\right) d\boldsymbol{u}_i,
\end{aligned} \tag{8}$$

where $\boldsymbol{\theta}^{(j)}$ denotes the current value of the parameters. The posterior distribution of the random effects is

Copyright © 2014 John Wiley & Sons, Ltd.

$$f\left(\boldsymbol{u}_i \mid T_i, \Delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(j)}\right) \propto f\left(\boldsymbol{Y}_i, T_i, \Delta_i, \boldsymbol{u}_i; \boldsymbol{\theta}^{(j)}\right), \\ = f\left(T_i, \Delta_i | \boldsymbol{u}_i; \boldsymbol{\theta}^{(j)}\right) f\left(\boldsymbol{Y}_i | \boldsymbol{u}_i; \boldsymbol{\theta}^{(j)}\right) f\left(\boldsymbol{u}_i | \boldsymbol{\Sigma}^{(j)}\right). \tag{9}$$

Straightforward algebra can be used to see that (9) simplifies to

$$f\left(\boldsymbol{u}_i \mid T_i, \Delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(j)}\right) \propto \exp\left\{-\rho\left(\frac{Y_{it} - \boldsymbol{\beta}^{(j)'}\boldsymbol{X}_{it} - \boldsymbol{\delta}^{(j)'}\boldsymbol{H}_{it} - \boldsymbol{u}_i'\boldsymbol{Z}_{it}}{\sigma^{(j)}}\right) + \Delta_i \alpha_2^{(j)} \boldsymbol{u}_i' \boldsymbol{Z}_{iT_i}\right\} \\ \exp\left\{-\int_0^{T_i} h_0^{(j)}(s) \exp\left\{\alpha_2^{(j)} \boldsymbol{u}_i' \boldsymbol{Z}_{is}\right\}\right\} f\left(u_i | \boldsymbol{\Sigma}^{(j)}\right).$$

In order to work with (8), we need to marginalize the joint distribution with respect to the multivariate random-effects posterior distribution. The resulting integral is conveniently approximated through IS. IS proceeds by obtaining a random sample $(v_{i1}, \ldots, v_{im_i^{(j)}})$ from a proposal distribution $g(\cdot)$. The IS identity

$$\int \ell_c(\boldsymbol{\theta}) f\left(\boldsymbol{u}_i \mid T_i, \Delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(j)}\right) d\boldsymbol{u}_i = \int \ell_c(\boldsymbol{\theta}) \frac{f\left(\boldsymbol{u}_i \mid T_i, \Delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(j)}\right)}{g\left(\boldsymbol{u}_i\right)} g\left(\boldsymbol{u}_i\right) d\boldsymbol{u}_i$$

is used to approximate (8). More formally, the expression in (8) is approximated as

$$\mathbb{E}\left[\ell_c(\boldsymbol{\theta})\right] \approx \sum_i \sum_{b=1}^{m_i^{(j)}} \left[\log f\left(\boldsymbol{Y}_i | \boldsymbol{v}_{ib}; \boldsymbol{\theta}^{(j)}\right) + \log f\left(T_i, \Delta_i | \boldsymbol{v}_{ib}; \boldsymbol{\theta}^{(j)}\right) \right. \\ \left. + \log f\left(\boldsymbol{v}_{ib} \mid \boldsymbol{\theta}^{(j)}\right)\right] w_{ib}, \tag{10}$$

where

$$\tilde{w}_{ib} = \frac{f\left(\boldsymbol{v}_{ib} \mid T_i, \Delta_i, \boldsymbol{Y}_i; \boldsymbol{\theta}^{(j)}\right)}{g\left(\boldsymbol{v}_{ib}\right)}$$

and $w_{ib} = \tilde{w}_{ib} / \sum_b \tilde{w}_{ib}$. In this work, we proceed as in [28], where we sample $(v_{i1}, \ldots, v_{im_i^{(j)}})$ from the posterior distribution at the initial parameter estimates using adaptive rejection Metropolis sampling [42] and then update the weights at each iteration. The observed likelihood (5) can be directly approximated as $\prod_i \sum_b \tilde{w}_{ib}$. The latter is used to check convergence of the MCEM and after convergence for testing and computation of information criteria. We refer the reader to [28] and [43] for additional details on checking convergence of the MCEM algorithm and dynamically selecting the Monte Carlo sample size $m_i^{(j)}$. In summary, at the E-step, we update $\tilde{w}_{ib}$ and w_{ib} and evaluate the likelihood.

3.2. M-step

We outline here the M-step, which consists of maximizing the approximated conditional expected value of the complete likelihood with respect to $\boldsymbol{\theta}$.

When $h_0(s)$ is left completely unspecified, we obtain a Nelson–Aalen [44, 45] type estimator as

$$\hat{h}_0(s) = \sum_{i=1}^{n} \frac{\Delta_i I\left(T_i = s\right)}{\sum\limits_{i: T_i \geqslant s} \sum_{b=1}^{m_i^{(j)}} w_{ib} \exp\left\{\boldsymbol{\gamma}' \boldsymbol{W}_i + \alpha_1 \boldsymbol{\delta}' \boldsymbol{H}_{is} + \alpha_2 \boldsymbol{v}_{ib}' \boldsymbol{Z}_{is}\right\}}. \tag{11}$$

This corresponds to the solution of the score equation for $h_0(s)$ [46–48]. Expression (11) is not an explicit solution as it depends on other parameters. Nevertheless, it can be plugged in the conditional expected value of the complete likelihood, thus obtaining a profile complete likelihood (complete with respect to the random effects and profiled with respect to the nonparametric baseline along the lines of [30]). If instead we specify a parametric form for $h_0(s)$, for example, $h_0(s) = \eta s^{\eta - 1}$, we can plug in this expression and update any parameter involved in $h_0(s)$ within the rest of the M-step.

When a Gaussian distribution is assumed for the longitudinal measurements, regression coefficients and hazard ratios are updated through a one-step Newton–Raphson algorithm (which is easily adapted from [46]),

Copyright © 2014 John Wiley & Sons, Ltd.

while the variance of the random term ϵ_{it} in (1) is estimated through the usual closed-form expression.

When we assume an ALD for the longitudinal measurements, the M-step is complicated by the presence of the quantile loss function. An estimator of σ, dependent on the other parameters, can be explicitly obtained as

$$\hat{\sigma} = \frac{1}{\sum_i n_i} \sum_{i=1}^{n} \sum_{t=1}^{n_i} \sum_{b=1}^{m_i^{(j)}} w_{ib} \rho \left(Y_{it} - \boldsymbol{\beta}' \boldsymbol{X}_{it} - \boldsymbol{\delta}' \boldsymbol{H}_{it} - \boldsymbol{v}_{ib}' \boldsymbol{Z}_{it} \right). \tag{12}$$

The vector of regression parameters and dispersion parameter for the ALD is block updated using one step of the Nelder–Mead [49] numerical optimization algorithm, majorizing (10) after plug-in of (12) and (11) or the parametric formula of $h_0(s)$. The resulting expression to be maximized is

$$\begin{aligned} \sum_i \Delta_i \log \left(\hat{h}_0 \left(T_i \right) \right) + \sum_{i=1}^{n} \Delta_i \boldsymbol{\gamma}' \boldsymbol{W}_i + \alpha_1 \sum_{i=1}^{n} \Delta_i \boldsymbol{\delta}' \boldsymbol{H}_{iT_i} + \alpha_2 \sum_{i=1}^{n} \Delta_i \sum_{b=1}^{m_i^{(j)}} w_{ib} \boldsymbol{v}_{ib}' \boldsymbol{Z}_{iT_i} \\ - \sum_{i=1}^{n} \sum_{b=1}^{m_i^{(j)}} w_{ib} \int_0^{T_i} \hat{h}_0(s) \exp \left\{ \boldsymbol{\gamma}' \boldsymbol{W}_i + \alpha_1 \boldsymbol{\delta}' \boldsymbol{H}_{is} + \alpha_2 \boldsymbol{v}_{ib}' \boldsymbol{Z}_{is} \right\} ds \\ - \sum_{i=1}^{n} \sum_{t=1}^{n_i} \sum_{b=1}^{m_i^{(j)}} w_{ib} \rho \left(\frac{\boldsymbol{Y}_{it} - \beta' \boldsymbol{X}_{it} - \boldsymbol{\delta}' \boldsymbol{H}_{it} - \boldsymbol{v}_{ib}' \boldsymbol{Z}_{it}}{\hat{\sigma}} \right) - \log \left(\hat{\sigma} \right) \sum_{i=1}^{n} n_i. \end{aligned}$$

The integral involved in preceding the expression (and similarly in any expression where $f(T_i, \Delta_i | \boldsymbol{v}_{ib}; \boldsymbol{\theta})$ appears also at the E-step) reduces to a sum when a nonparametric baseline is used and can instead be approximated using a one-dimensional Gauss–Kronrod quadrature [50] when $h_0(s)$ is continuous.

The M-step is concluded by maximizing the approximated conditional expected value of the complete likelihood with respect to parameters involved in the distribution of the random effects. This is readily accomplished under any of the assumptions we have proposed in Section 2.3 using the method of moments. In all cases, in fact, $\boldsymbol{\Sigma}$ can be updated as the weighted empirical covariance matrix of the random effects sampled at the E-step.

4. Simulations

In this section, we illustrate our approach through a simulation study. We evaluate the performance of our proposed model for data missing not at random (MNAR) and compare it with a model that ignores informative dropout (missing at random, or MAR).

For $n = \{250, 500\}$, $\alpha_1 = \{0, 1\}$, $\alpha_2 = \{0, 1\}$, and $\tau = \{0.25, 0.5, 0.75\}$, we fix $\boldsymbol{\beta} = \boldsymbol{\delta} = \boldsymbol{\gamma} = (1\ 1)$ and $\sigma = 1$. We assume random effects arise from a centered bivariate normal distribution with standard deviations equal to 0.3 and correlation 0.16. We let $\boldsymbol{Z}_{it} = (1\ t)$, $\boldsymbol{H}_{it} = (h_{i1}\ h_{i2} * t)$, and $\boldsymbol{X}_i = (1\ x_i)$, with h_{i1}, h_{i2}, x_i, W_{i1}, and W_{i2} generated from independent standard normals. By also fixing $h_0(s) = 1$, it is possible to exactly obtain the survival distribution as

$$S\left(T_i | \boldsymbol{u}_i, \boldsymbol{H}_i, \boldsymbol{W}_i\right) = \exp \left\{ - \frac{e^{\alpha_1 \left(\delta_1 H_{i1} + \delta_2 H_{i2T_i} \right) + \alpha_2 \left(u_{i1} + u_{i2} T_i \right) + \boldsymbol{\gamma}' \boldsymbol{W}_i} - e^{\alpha_1 \delta_1 H_{i1} + \alpha_2 u_{i1} + \boldsymbol{\gamma}' \boldsymbol{W}_i}}{\alpha_2 u_{i2} + \alpha_1 \delta_2 h_{i1}} \right\}$$

when $\alpha_1 \neq 0$ or $\alpha_2 \neq 0$ and

$$S\left(T_i | U, H, W\right) = \exp \left\{ -T_i e^{\boldsymbol{\gamma}' \boldsymbol{W}_i} \right\}$$

when $\alpha_1 = \alpha_2 = 0$. The preceding expression can be inverted to obtain T_i after generating n random variates uniformly distributed on the unit interval. We then let the censoring time $C_i/5$ be distributed according to a beta with parameters 4 and 1, in order to obtain a censoring proportion around 25%. We allow for a maximum of six observation times for each subject, at $t = 0$, 1/4, 1/2, 3/4, 1, and 3. Longitudinal observations before dropout are independently generated from an ALD for the τth quantile, centered on $\boldsymbol{\beta}' \boldsymbol{X}_i + \boldsymbol{\delta}' \boldsymbol{H}_{it} + \boldsymbol{u}' \boldsymbol{Z}_{it}$ and with dispersion parameter σ. On each fabricated data set, we fit our JM

Copyright © 2014 John Wiley & Sons, Ltd.

Table I. Bias and standard deviation of the estimates of the proposed model on simulated data for different values of n, α_1, and α_2.

			β		δ		γ		α_1		α_2	
n	α_1	α_2	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
						$\tau = 0.25$						
250	0	0	−0.004	0.107	0.001	0.124	0.020	0.085	−0.001	0.044	−0.002	0.131
250	0	1	−0.013	0.104	−0.002	0.119	−0.023	0.100	−0.001	0.066	−0.076	0.177
250	1	0	−0.003	0.109	0.009	0.094	0.020	0.088	0.021	0.101	−0.008	0.146
250	1	1	−0.012	0.105	−0.003	0.108	−0.021	0.106	−0.025	0.108	−0.068	0.189
500	0	0	−0.002	0.074	−0.000	0.082	0.006	0.058	−0.001	0.031	−0.003	0.091
500	0	1	−0.011	0.072	0.002	0.082	−0.034	0.072	0.001	0.046	−0.076	0.120
500	1	0	−0.003	0.074	0.004	0.061	0.008	0.061	0.008	0.067	−0.006	0.092
500	1	1	−0.011	0.074	−0.004	0.075	−0.033	0.073	−0.038	0.076	−0.063	0.118
						$\tau = 0.5$						
250	0	0	−0.002	0.096	−0.000	0.113	0.020	0.084	−0.001	0.044	−0.000	0.126
250	0	1	−0.014	0.092	−0.003	0.107	−0.020	0.100	−0.001	0.066	−0.064	0.164
250	1	0	−0.001	0.097	0.005	0.087	0.020	0.087	0.022	0.096	−0.006	0.138
250	1	1	−0.015	0.094	−0.005	0.101	−0.019	0.105	−0.030	0.107	−0.074	0.175
500	0	0	0.001	0.067	−0.000	0.076	0.005	0.057	−0.001	0.031	−0.004	0.083
500	0	1	−0.008	0.063	0.000	0.074	−0.028	0.071	−0.001	0.045	−0.060	0.106
500	1	0	−0.000	0.066	0.001	0.058	0.006	0.059	0.010	0.064	−0.003	0.088
500	1	1	−0.008	0.066	−0.006	0.069	−0.031	0.072	−0.035	0.072	−0.056	0.119
						$\tau = 0.75$						
250	0	0	−0.001	0.108	0.002	0.122	0.019	0.085	−0.003	0.044	−0.007	0.134
250	0	1	−0.013	0.104	−0.003	0.119	−0.023	0.100	−0.001	0.066	−0.076	0.169
250	1	0	0.003	0.111	0.004	0.091	0.021	0.088	0.024	0.099	−0.002	0.145
250	1	1	−0.012	0.107	−0.006	0.107	−0.020	0.106	−0.028	0.112	−0.079	0.178
500	0	0	0.004	0.075	−0.003	0.083	0.006	0.058	−0.001	0.031	−0.004	0.088
500	0	1	−0.006	0.070	−0.000	0.080	−0.030	0.072	−0.001	0.046	−0.066	0.109
500	1	0	0.004	0.075	−0.002	0.062	0.007	0.060	0.014	0.068	−0.004	0.092
500	1	1	−0.007	0.073	−0.009	0.075	−0.032	0.073	−0.033	0.076	−0.062	0.120

Results are based on $B = 1000$ replicates.

with parametric baseline distribution and another based on two separate models (one for the longitudinal process and one for the time-to-event process, therefore obtaining MAR estimates).

For each setting, we report the bias and standard deviation of the estimates averaged over $B = 1000$ replicates and further averaged over groups of parameters for $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and $\boldsymbol{\gamma}$. Results are shown in Table I, where it can be seen that the MNAR model has a very low bias and standard deviation of the estimates for all values of α_1 and α_2, with very few exceptions, which are likely due to random fluctuation. Results are consistent across all quantiles, with a slightly larger mean squared error (MSE) for quantiles distant from the median, as expected.

In Table II, we report the ratio of the bias and the variance of the estimates of the MAR over the MNAR model, separately for the longitudinal and survival parameters. The bias has been computed by taking the square root of the sum of the squared bias for each group of parameters. These ratios are close to the unity when $\alpha_1 = \alpha_2 = 0$, with our model generally performing slightly better given that the MNAR model assumes $\boldsymbol{\delta}$ parameters are equal in the longitudinal and survival parts. When α_1 or α_2 is nonzero, the ratios of the variances of the estimates are still close to unity, but the ratios of biases increase substantially. The bias of the MAR model may be up to 30 times the bias of our JM. The effect of α_1 is often stronger than the effect of α_2, but this is likely only because in all simulated settings, there is a larger heterogeneity due to shared covariates with respect to the unobserved heterogeneity due to random effects. As could be expected, the ratios are generally increasing with n, given that the MSE of the JM is infinitesimal.

In order to see the effect of a violation of parametric assumptions, we run two separate simulation studies with a misspecified random-effects distribution. In the first case, we generate random effects from a multivariate T with 3 degrees of freedom and the same covariance matrix as before, while still assuming multivariate Gaussian random effects. Simulation results are reported in Table III. It can be seen that in this case, estimates have a bias and standard error that is rather close to the ones reported in Table I,

Copyright © 2014 John Wiley & Sons, Ltd.

Table II. Ratios of bias and variance of the estimates obtained with the missing-at-random (MAR model) ($bias_{MAR}$ and var_{MAR}) and with our model ($bias$ and var) for the longitudinal (Y) and survival (T) parts of the model.

			$\frac{bias_{MAR}(Y)}{bias(Y)}$	$\frac{var_{MAR}(Y)}{var(Y)}$	$\frac{bias_{MAR}(T)}{bias(T)}$	$\frac{var_{MAR}(T)}{var(T)}$
n	α_1	α_2				
			$\tau = 0.25$			
250	0	0	1.548	1.203	0.800	1.115
250	0	1	6.616	1.241	1.266	1.315
250	1	0	2.876	1.573	9.990	0.929
250	1	1	5.868	1.385	8.051	0.814
500	0	0	1.264	1.252	0.802	1.197
500	0	1	4.214	1.270	1.954	1.287
500	1	0	6.966	1.654	5.954	0.996
500	1	1	7.227	1.412	24.926	0.834
			$\tau = 0.5$			
250	0	0	1.091	1.121	0.909	1.129
250	0	1	1.679	1.204	1.464	1.333
250	1	0	5.659	1.439	7.579	0.948
250	1	1	4.646	1.289	10.572	0.804
500	0	0	0.865	1.188	0.930	1.221
500	0	1	6.936	1.221	1.208	1.339
500	1	0	2.756	1.501	6.758	1.041
500	1	1	5.797	1.290	29.194	0.851
			$\tau = 0.75$			
250	0	0	1.308	1.223	1.010	1.151
250	0	1	4.972	1.280	1.283	1.318
250	1	0	3.895	1.586	7.153	0.934
250	1	1	4.191	1.367	10.346	0.783
500	0	0	1.866	1.255	1.304	1.201
500	0	1	6.733	1.276	1.135	1.303
500	1	0	4.370	1.672	5.414	1.014
500	1	1	7.216	1.347	27.891	0.836

Results are shown for different values of n, α_1, and α_2 and are based on $B = 1000$ replicates.

where the model is correctly specified. As far as the fixed-effects parameters are concerned, the MSE still decreases with n. It shall be noted anyway that there is a large bias for $\hat{\alpha}_2$ when the true $\alpha_2 \neq 0$. This is a consequence of the heavy tails of the true underlying distribution. This bias may be expected to decrease only as the follow-up time and number of repeated measurements are increased. In a second simulation study with misspecified models, we generate multivariate Gaussian random effects as before but assume a multivariate ALD when fitting the model. Results are reported in Table IV. It can be seen that bias and variance are approximately equal to those reported in Table I, with the only exception of $\hat{\alpha}_2$, which has the same bias but a slightly larger standard error in some cases. In conclusion, the ALD assumption seems to be rather robust with respect to model misspecification when the true underlying distribution is Gaussian.

5. Application to dilated cardiomyopathy data

In this section, we briefly illustrate the proposed approach on an original data set about patients with dilated cardiomyopathy. Data refer to $n = 659$ consecutive patients who begun treatment for dilated cardiomyopathy in the cardiovascular department of 'Ospedali Riuniti' in Trieste, Italy. Patients were enrolled at first treatment and scheduled for follow-up after 6 months and 1, 2, 3, 4, 6, and 10 years, with only 25% of the patients having complete records. Maximum follow-up time before cardiovascular death or censoring due to loss at follow-up or transplant is 25 years, with a total of 212 events (32%).

Dilation of the left ventricular is known to lead to heart failure, and in many cases, an etiological basis cannot be identified [51]. The goal of this study is to compare patients with mild dilation of the

Copyright © 2014 John Wiley & Sons, Ltd.

Table III. Bias and standard deviation of the estimates of the proposed model on simulated data for different values of n, α_1, and α_2.

			β		δ		γ		α_1		α_2	
n	α_1	α_2	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
						$\tau = 0.25$						
250	0	0	0.000	0.103	−0.001	0.118	0.012	0.084	−0.001	0.047	0.001	0.133
250	0	1	−0.006	0.105	0.000	0.122	−0.026	0.105	−0.003	0.063	−0.254	0.199
250	1	0	−0.002	0.105	0.002	0.092	0.015	0.088	0.026	0.104	−0.009	0.139
250	1	1	−0.007	0.104	−0.012	0.110	−0.020	0.108	−0.019	0.111	−0.204	0.228
500	0	0	0.000	0.070	−0.002	0.081	0.002	0.058	−0.003	0.030	0.002	0.087
500	0	1	−0.003	0.072	0.001	0.086	−0.039	0.073	−0.002	0.042	−0.261	0.148
500	1	0	−0.004	0.073	0.006	0.060	0.006	0.059	0.006	0.064	−0.005	0.095
500	1	1	−0.004	0.071	−0.016	0.078	−0.030	0.075	−0.030	0.076	−0.175	0.166
						$\tau = 0.5$						
250	0	0	−0.000	0.090	0.007	0.103	0.019	0.084	−0.001	0.048	−0.003	0.128
250	0	1	−0.004	0.088	0.006	0.104	−0.027	0.103	−0.001	0.065	−0.247	0.185
250	1	0	−0.001	0.092	0.010	0.085	0.014	0.087	0.015	0.100	−0.004	0.133
250	1	1	−0.005	0.089	−0.000	0.099	−0.025	0.106	−0.029	0.108	−0.186	0.203
500	0	0	0.002	0.063	−0.000	0.074	0.002	0.057	−0.003	0.030	0.000	0.085
500	0	1	−0.003	0.064	−0.000	0.073	−0.035	0.075	−0.003	0.043	−0.236	0.138
500	1	0	−0.000	0.065	0.001	0.056	0.008	0.060	0.010	0.061	−0.001	0.089
500	1	1	−0.003	0.064	−0.015	0.077	−0.027	0.075	−0.033	0.076	−0.171	0.155
						$\tau = 0.75$						
250	0	0	−0.002	0.103	0.001	0.124	0.018	0.087	0.001	0.046	−0.002	0.129
250	0	1	−0.005	0.098	0.006	0.115	−0.031	0.105	0.000	0.063	−0.246	0.189
250	1	0	−0.001	0.104	0.002	0.089	0.020	0.090	0.026	0.102	0.007	0.140
250	1	1	−0.004	0.104	−0.008	0.105	−0.026	0.109	−0.027	0.111	−0.191	0.210
500	0	0	0.002	0.069	0.002	0.078	0.003	0.058	−0.003	0.030	−0.001	0.092
500	0	1	−0.003	0.069	0.002	0.080	−0.030	0.077	0.001	0.042	−0.239	0.144
500	1	0	−0.001	0.072	0.002	0.061	0.009	0.059	0.011	0.066	0.002	0.090
500	1	1	−0.004	0.072	−0.012	0.079	−0.033	0.078	−0.036	0.084	−0.172	0.151

The model is misspecified as true random effects are multivariate T with 3 degrees of freedom, while the model assumes multivariate Gaussian random effects. Results are based on $B = 1000$ replicates.

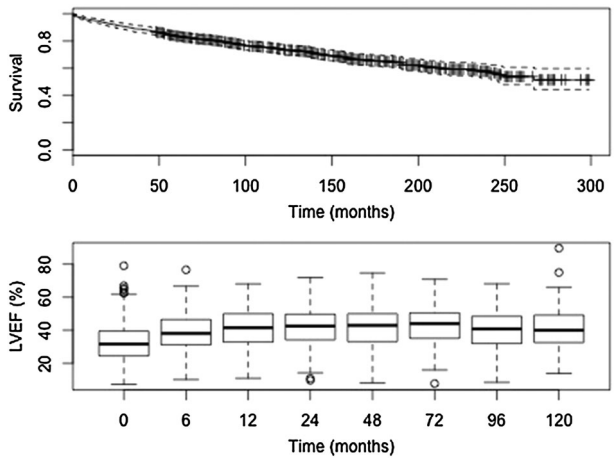

Figure 1. Mildly dilated cardiomyopathy data: Kaplan–Meier estimate for time to cardiovascular death (upper panel) and occasion-specific box plots of measured left ventricular ejection fraction (LVEF; lower panel).

left ventricular (mildly dilated cardiomyopathy, or MDCM) with respect to patients with a dilation of unrecognized etiology (idiopathic dilated cardiomyopathy). MDCM patients are generally believed to be at a slightly lower risk [52], but the physiological reasons are still unrecognized.

Copyright © 2014 John Wiley & Sons, Ltd.

Table IV. Bias and standard deviation of the estimates of the proposed model on simulated data for different values of n, α_1, and α_2.

			β		δ		γ		α_1		α_2	
n	α_1	α_2	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
						$\tau = 0.25$						
250	0	0	−0.005	0.106	−0.002	0.124	0.017	0.087	−0.001	0.046	−0.008	0.131
250	0	1	−0.019	0.105	0.000	0.125	−0.032	0.104	0.000	0.065	−0.030	0.190
250	1	0	−0.001	0.111	0.008	0.095	0.016	0.091	0.021	0.104	−0.008	0.140
250	1	1	−0.020	0.106	−0.004	0.109	−0.031	0.108	−0.041	0.115	−0.052	0.209
500	0	0	−0.007	0.072	0.001	0.085	0.005	0.059	−0.000	0.030	0.003	0.088
500	0	1	−0.019	0.070	0.001	0.081	−0.040	0.073	−0.002	0.046	−0.017	0.130
500	1	0	−0.004	0.074	0.011	0.062	0.005	0.059	−0.000	0.066	0.001	0.092
500	1	1	−0.018	0.073	−0.002	0.075	−0.043	0.073	−0.055	0.069	−0.042	0.140
						$\tau = 0.5$						
250	0	0	0.000	0.095	0.000	0.114	0.017	0.087	−0.001	0.046	−0.002	0.128
250	0	1	−0.015	0.092	0.001	0.111	−0.029	0.104	0.001	0.065	−0.015	0.189
250	1	0	−0.000	0.097	0.004	0.087	0.017	0.088	0.024	0.099	−0.006	0.134
250	1	1	−0.016	0.095	−0.006	0.102	−0.033	0.108	−0.039	0.109	−0.041	0.195
500	0	0	0.001	0.066	−0.002	0.077	0.008	0.056	−0.001	0.033	0.003	0.087
500	0	1	−0.016	0.064	−0.001	0.072	−0.039	0.073	−0.002	0.046	−0.008	0.126
500	1	0	−0.002	0.068	0.004	0.058	0.005	0.059	0.005	0.066	−0.000	0.086
500	1	1	−0.013	0.066	−0.005	0.072	−0.038	0.071	−0.049	0.071	−0.022	0.141
						$\tau = 0.75$						
250	0	0	0.008	0.109	0.001	0.123	0.014	0.087	0.001	0.047	0.003	0.135
250	0	1	−0.011	0.105	0.001	0.123	−0.031	0.106	0.001	0.064	−0.027	0.194
250	1	0	0.004	0.111	−0.003	0.096	0.019	0.089	0.031	0.105	−0.001	0.137
250	1	1	−0.010	0.105	−0.013	0.110	−0.033	0.106	−0.031	0.110	−0.049	0.195
500	0	0	0.004	0.074	0.001	0.086	0.005	0.059	−0.000	0.030	−0.003	0.082
500	0	1	−0.010	0.072	−0.002	0.080	−0.041	0.073	−0.002	0.045	−0.018	0.129
500	1	0	0.005	0.076	−0.000	0.064	0.005	0.060	0.010	0.069	−0.001	0.089
500	1	1	−0.012	0.071	−0.016	0.073	−0.035	0.070	−0.041	0.073	−0.034	0.131

The model is misspecified as true random effects are multivariate Gaussian, while the model assumes multivariate Laplace random effects. Results are based on $B = 1000$ replicates.

Our longitudinal outcome of interest is the left ventricular ejection fraction (LVEF), that is, the volumetric fraction of blood pumped out of the ventricle with each heartbeat. Note that LVEF is a bounded outcome, but all measurements are far from the minimum and maximum allowed values, so there are no boundary issues. In Figure 1, we report the Kaplan–Meier estimate for time to cardiovascular death (upper panel) and box plots of the LVEF of subjects surviving to each time point (lower panel). No extreme LVEF measurements seem to have been recorded.

Patients with a lower ejection fraction are at higher risk of death, which links the two processes. A univariate Cox model for the baseline LVEF gives a hazard ratio of 0.95 for each percentage point, with $p < 1e^{-16}$. Furthermore, LVEF is skewed, and its skewness seems to change over time, with a right tail at the beginning of the observation period and a left tail later. This leads to two issues: first, using a classical JM after transformation of the LVEF would be awkward, as the optimal transformation is different at each time point. Secondly, modeling the mean of the transformed LVEF would not be as meaningful from a clinical perspective than directly modeling quantiles. We mostly are interested in low quantiles (like the 10th or the 15th), in the terziles or quartiles for the outcome [53–56]. Consequently, we explore the ejection fraction distribution by evaluating $\tau = 0.1, 0.15, 0.25, 0.33, 0.5, 0.66$, and 0.75.

We model the longitudinal outcome conditionally on an intercept and age at baseline ($\boldsymbol{X}$ matrix), on the indicator of MDCM and its interaction with time ($\boldsymbol{H}$ matrix). Besides covariates in the $\boldsymbol{H}$ matrix, we let the hazard of death depend on gender (1 for male) and indicator of New York Heart Association (NYHA) functional classes I or II at baseline ($\boldsymbol{W}$ matrix). NYHA classes are related to limitation of physical activities and experience of symptoms like fatigue, palpitation, and shortness of breath. Patients in class I have no limitations and symptoms, and those in class II only slight limitation and/or symptoms during moderate activity. Patients in class III on the other hand have marked limitation of physical activity and/or

Copyright © 2014 John Wiley & Sons, Ltd.

Table V. Log-likelihood at convergence, AIC, and BIC for our model fit with a parametric and nonparametric baseline, at different quantiles of interest, on the dilated cardiomyopathy data.

	Nonparametric baseline			Parametric baseline		
τ	$\ell(\theta)$	BIC	AIC	$\ell(\theta)$	BIC	AIC
0.1	−10,515.214	21,064.255	21,054.428	−10,602.111	21,240.867	21,228.222
0.15	−10,449.858	20,933.543	20,923.716	−10,461.581	20,959.807	20,947.162
0.25	−10,289.721	20,613.269	20,603.442	−10,322.538	20,681.721	20,669.076
0.33	−10,197.313	20,428.453	20,418.626	−10,235.107	20,506.859	20,494.214
0.5	−10,154.227	20,342.281	20,332.454	−10,175.175	20,386.995	20,374.350
0.66	−10,223.996	20,481.819	20,471.992	−10,229.257	20,495.159	20,482.514
0.75	−10,300.913	20,635.653	20,625.826	−10,308.303	20,653.251	20,640.606

Table VI. Estimates for the MNAR and MAR models for the MDCM data.

	Longitudinal outcome				Survival outcome			
τ	Int*	Age*	MDCM*	MDCM:time*	Males*	NYHA*	α_1*	α_2*
				MNAR model				
0.1	22.535	−0.038	7.754	−0.038	0.538	−0.912	−0.044	−0.042
0.15	25.134	−0.048	7.768	−0.028	0.527	−0.895	−0.037	−0.057
0.25	29.255	−0.068	7.752	−0.042	0.593	−0.990	−0.017	−0.068
0.33	32.813	−0.095	7.752	−0.018	0.592	−0.988	−0.018	−0.087
0.5	39.086	−0.124	7.749	−0.093	0.542	−0.918	−0.031	−0.063
0.66	44.437	−0.148	7.734	−0.136	0.566	−0.952	−0.011	−0.097
0.75	47.355	−0.156	7.744	−0.044	0.591	−0.987	−0.021	−0.094
				MAR model				
0.1	29.700	−0.141	7.580	0.041	0.600	−1.096	0.000	0.000
0.15	36.266	−0.251	8.027	0.076	0.600	−1.096	0.000	0.000
0.25	35.175	−0.130	5.262	0.011	0.600	−1.096	0.000	0.000
0.33	35.246	−0.109	6.516	0.068	0.600	−1.096	0.000	0.000
0.5	40.048	−0.137	8.289	0.028	0.600	−1.096	0.000	0.000
0.66	44.541	−0.123	5.711	0.022	0.600	−1.096	0.000	0.000
0.75	44.284	−0.088	5.558	0.017	0.600	−1.096	0.000	0.000

The models are based on a nonparametric baseline for the survival process and two dimensional shared random effects.
MAR, missing-at-random; MDCM, mildly dilated cardiomyopathy; MNAR, missing-not-at-random; NYHA, New York Heart Association.
*Estimates in the column are significant at the 5% level for all quantiles τ.

experience the aforementioned symptoms during less than ordinary physical activity. Patients in class IV are unable to perform any physical activity without discomfort and/or experience symptoms of cardiac insufficiency even at rest. For more details, see for instance [51] and references therein. We also include a shared subject-specific random intercept and a random slope, that is, $\boldsymbol{Z} = (1\ t)$. Given that the number of follow-up times is slightly large, we use a normality assumption for $\boldsymbol{u}_i$. We also have compared with a multivariate T and multivariate Laplace, with analogous results that we do not report for reasons of space. The multivariate normal distribution is also chosen using AIC and BIC. We estimate our proposed model both with a parametric Weibull baseline and with a nonparametric baseline. Likelihood, AIC, and BIC at each quantile are reported in Table V. Based on these results, we select the nonparametric baseline for all quantiles.

We report the estimates and compare with those obtained under a MAR model in Table VI. The effect of age on LVEF becomes stronger as τ increases, while the effect of MDCM is approximately constant as a function of τ, with a negative interaction with time. Men are at a slightly higher risk of death, and patients in lower functional NYHA classifications are at a lower risk. After adjusting for these covariates, the significant and negative estimates for $\hat{\alpha}_1$ lead us to conclude that MDCM is an *independent* predictor of a slightly lower risk of death, even after considering its effect on LVEF. We could expect negative estimates for α_1 and α_2 as longitudinal measurements and survival time are positively dependent.

Copyright © 2014 John Wiley & Sons, Ltd.

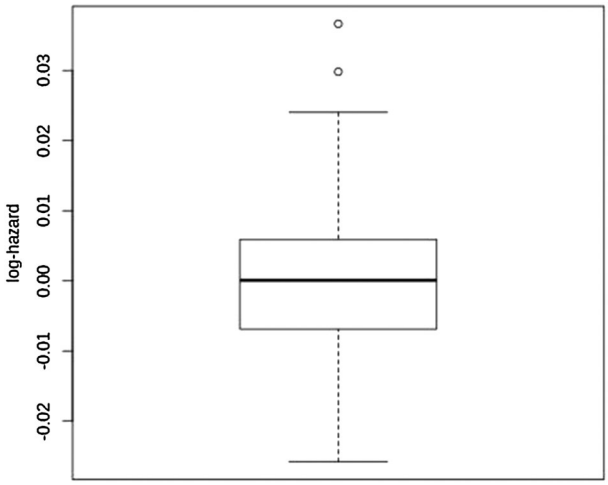

Figure 2. Mildly dilated cardiomyopathy data: box plot of the estimated subject-specific frailties at time zero.

Ignoring dropout may lead to an important bias. First of all, the intercepts estimated with the MAR models are slightly larger than those obtained with the MNAR models for all τ, except $\tau = 75\%$. This happens as dropout is more likely in the low quantiles of the longitudinal outcome distribution. Secondly, under the MAR model, a significant *positive* interaction between MDCM and time is estimated. This may be due to the fact that subjects with higher LVEF tend to drop out later and to be in the MDCM class more often, resulting in a positive bias when ignoring the informative dropout. Given the results in Table VI, we can conclude there is some sensitivity to dropout for the data at hand within the proposed class of models. As clearly noted in [57], one can never test the MAR assumption. From Table VI, we can see that parameter estimates for the survival model are rather stable as a function of τ. This indicates that gender and NYHA score are approximately orthogonal of MDCM and age after adjusting for common unobserved heterogeneity. We conclude with a brief discussion on the random-effects estimates $\boldsymbol{u}_i$. As stated in Section 2, these can be used for conditional screening. If we focus on the survival outcome and fix $\tau = 0.5$ to control for LVEF measured with error, a box plot of $\alpha_2 u_{i1}$ (Figure 2) indicates that at time zero, there are two subjects with slightly outlying frailty (higher risk of dropout) after adjusting for all covariates. The first of these subjects has indeed experienced an early event at time 6, which therefore can be probably attributed to unmeasured covariates. Similarly, the predicted subject-specific LVEF should be adjusted by $\boldsymbol{u}_i' Z_{it}$. This adjustment is negative at time zero for all τ for the two aforementioned subjects, indicating that unobserved heterogeneity slightly lowered their baseline LVEF measurements.

6. Conclusions

Informative dropout may bias parameter estimates both in mean and quantile regression if ignored. As our data example suggests, the problem may be stronger for quantiles corresponding to a higher rate of events. In our example, we have checked that sensitivity to dropout is milder in high quantiles than in low quantiles, for instance. Moreover, the brief simulation study confirms that MAR estimates might be substantially biased under MNAR mechanisms. This bias may not decrease as the sample size is increased.

The proposed approach allows us to simultaneously model the quantile of a longitudinal outcome and the hazard of dropout, allowing them to share part of the observed and unobserved heterogeneity. Our model can be applied with right-censored event times occurring between two scheduled visits, when dropout times coincide with observation times for the longitudinal process and also when the observation times are not scheduled in advance. We have generalized shared parameters and JMs in different directions: first of all, we have proposed an alternative parametric assumption for the longitudinal error, the ALD, which allows us to fit quantile regression models. Secondly, we have proposed two alternative random-effects distributions. A general efficient MCEM strategy has been used to fit our model under any of those assumptions. It shall be noted that MCEM strategies are inherently slow. Our non-optimized implementation based on `R` code takes around 2 h to obtain the estimates and run a nonparametric bootstrap for the standard errors when $n = 250$ in the simulated setting, but times range up to around 4 days in the real example where $n = 659$. These figures are based on median times observed when running

Copyright © 2014 John Wiley & Sons, Ltd.

the algorithm on a single CPU. In our implementation, we have actually used parallel computing for the bootstrap, which substantially reduced the runtimes. Additionally, optimization and implementation in C++ of some routines can be expected to speed up the algorithm about by 30 times.

In our example, we have specified different values for the quantile of interest τ. Under conditional independence assumptions, the total likelihood is the sum of the likelihood based on each quantile; hence, this approach is equivalent to simultaneously fitting the model for different values of τ and τ-specific parameters. When more than one quantile is of interest in applications, one could also use dependence (e.g., over the random effects at each quantile) or τ-homogeneity (e.g., for the variance of the random effects) assumptions. Model estimation under these assumptions is at the moment grounds for further work.

Acknowledgements

The authors are grateful to an associate editor and two referees for their kind suggestions. Acknowledgements go also to the Cardiovascular Department of 'Ospedali Riuniti', Trento, Italy, and in particular to Giulia Barbati, for permission to use the dilated cardiomyopathy data.

References

1. Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. *Biometrics* 1995; **51**:151–168.
2. Kurland BF, Johnson LL, Egleston BL, Diehr PH. Longitudinal data with follow-up truncated by death: match the analysis method to research aims. *Statistical Science* 2009; **24**:211–222.
3. Wu MC, Bailey KR. Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine* 1988; **7**:337–346.
4. Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* 1989; **45**:939–955.
5. Little R JA, Wang Y. Pattern-mixing models for multivariate incomplete data with covariates. *Biometrics* 1996; **52**:98–111.
6. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics* 1994; **43**:49–93.
7. Wu M, Carrol R. Estimation and comparison of changes in presence of informative right censoring by modelling the censoring process. *Biometrics* 1988; **45**:939–955.
8. Rizopoulos D. JM: an R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* 2010; **35**(9):1–33.
9. Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine* 2011; **30**:1366–1380.
10. Little R JA. Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* 1995; **90**:1112–1121.
11. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**:465–480.
12. Tsiatis A, Davidian M. An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica* 2004; **14**:793–818.
13. Koenker R. *Quantile Regression*. Cambridge University Press: New York, 2005.
14. Heritier S, Cantoni E, Copt S, Victoria-Feser MP. *Robust Methods in Biostatistics*. Wiley: Chichester, U.K., 2009.
15. Farcomeni A, Ventura L. An overview of robust methods in medical research. *Statistical Methods in Medical Research* 2012; **21**:111–133.
16. Daniel-Spiegel E, Weiner E, Yarom I, Doveh E, Friedman P, Cohen A, Shalev E. Establishment of fetal biometric charts using quantile regression analysis. *Journal of Ultrasound in Medicine* 2013; **32**:23–33.
17. Wei Y, Pere A, Koenker R, He X. Quantile regression methods for reference growth charts. *Statistics in Medicine* 2006; **25**:1369–1382.
18. Koenker R. Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 2004; **91**:74–89.
19. Geraci M, Bottai M. Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* 2007; **8**:140–154.
20. Liu Y, Bottai M. Mixed-effects models for conditional quantiles with longitudinal data. *The International Journal of Biostatistics* 2009; **5**:1–24.
21. Geraci M, Bottai M. Linear quantile mixed models. *Statistics and Computing* 2014; **24**:461–479.
22. Farcomeni A. Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Statistics and Computing* 2012; **22**:141–152.
23. Lipsitz S R, Fitzmaurice GM, Molenberghs G, Zhao LP. Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *Journal of the Royal Statistical Society, Series C* 1997; **46**:463–476.
24. Yi GY, He W. Median regression models for longitudinal data with dropouts. *Biometrics* 2009; **65**:618–625.
25. Yuan Y, Yin G. Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* 2010; **66**:105–114.

Copyright © 2014 John Wiley & Sons, Ltd.

26. Geraci M. Estimation of regression quantiles in complex surveys with data missing at random: an application to birthweight determinants. *Statistical Methods in Medical Research* 2014. DOI: 10.1177/0962280213484401.
27. Doucet A, de Freitas N, Gordon N. *Sequential Monte Carlo Methods in Practice*. Springer: New York, 2001.
28. Levine R A, Casella G. Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics* 2001; **10**:422–439.
29. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer: New York, 2000.
30. Cox D. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* 1972; **34**: 187–220.
31. Andersen P, Gill R. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 1982; **10**:1100–1120.
32. Yu K, Moyeed RA. Bayesian quantile regression. *Statistics and Probability Letters* 2001; **54**:437–447.
33. Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics and Data Analysis* 2012; **56**:491–501.
34. Viviani S, Alfó M, Rizopoulos D. Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing* 2014; **24**:417–427.
35. Rizopoulos D, Verbeke G, Molenberghs G. Shared parameter models under random effects misspecification. *Biometrika* 2008; **95**:63–74.
36. Hsieh F, Tseng YK, Wang JK. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* 2006; **62**:1037–1043.
37. Verbeke G, Lesaffre E. A linear mixed model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* 1996; **91**:217–221.
38. Vonesh Edward F, Greene Tom, Schluchter Mark D. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine* 2006; **25**:143–163.
39. Liu L, Huang X, O'Quigley J. Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* 2008; **64**:950–958.
40. Buchinsky M. Estimating the asymptotic covariance matrix for quantile regression models: a Monte Carlo study. *Journal of Econometrics* 1995; **68**:303–338.
41. Andrews DWK, Buchinsky M. A three-step method for choosing the number of bootstrap repetitions. *Econometrica* 2000; **68**:23–52.
42. Gilks WR, Best N, Tan K. Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics* 1995; **44**: 455–472.
43. Eickoff JC, Zhu J, Amamiya Y. On the simulation size and convergence of the Monte Carlo EM algorithm via likelihood-based distances. *Statistics and Probability letters* 2004; **67**:161–171.
44. Nelson W. Theory and applications of hazard plotting for censored failure data. *Technometrics* 1972; **14**:945–965.
45. Aalen O. Nonparametric inference for a family of counting processes. *Annals of Statistics* 1978; **6**:701–726.
46. Wulfsohn M, Tsiatis A. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**: 330–339.
47. Nielsen GG, Gill RD, Andersen PK, Sørensen TIA. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* 1992; **19**:25–43.
48. Gill RD. Marginal partial likelihood. *Scandinavian Journal of Statistics* 1992; **19**:133–137.
49. Nelder JA, Mead R. A simplex algorithm for function minimization. *Computer Journal* 1965; **7**:308–313.
50. Kahaner D, Moler C, Nash S. *Numerical Methods and Software*. Prentice Hall: Englewood Cliffs, N.J., 1989.
51. Merlo M, Pyxaras SA, Pinamonti B, Barbati G, Di Lenarda A, Sinagra G. Prevalence and prognostic significance of left ventricular reverse remodeling in dilated cardiomyopathy receiving tailored medical treatment. *Journal of the American College of Cardiology* 2011; **57**:1468–1476.
52. Keren A, Gottlieb S, Tzivoni D, Stern S, Yarom R, Billingham ME, Popp RL. Mildly dilated congestive cardiomyopathy. Use of prospective diagnostic criteria and description of the clinical course without heart transplantation. *Circulation* 1990; **81**:506–517.
53. Sandri MT, Cardinale D, Zorzino L, Passerini R, Lentati P, Martinoni A, Martinelli G, Cipolla CM. Minor increases in plasma troponin I predict decreased left ventricular ejection fraction after high-dose chemotherapy. *Clinical Chemistry* 2003; **49**:248–252.
54. Clements IP. Combined systolic and diastolic dysfunction in the presence of preserved left ventricular ejection fraction. *European Journal of Heart Failure* 2005; **7**:490–497.
55. Ndrepepa G, Mehilli J, Martinoff S, Schwaiger M, Schömig A, Kastrati A. Evolution of left ventricular ejection fraction and its relationship to infarct size after acute myocardial infarction. *Journal of the American College of Cardiology* 2007; **50**:149–156.
56. Cowie B, Kluger R, Kalpokas M. Left ventricular volume and ejection fraction assessment with transoesophageal echocardiography: 2D vs 3D imaging. *British Journal of Anaesthesia* 2013; **110**:201–206.
57. Molenberghs G, Beunckens C, Sotto C, Kenward MG. Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society (Series B)* 2008; **70**:371–388.

Copyright © 2014 John Wiley & Sons, Ltd.