

Longitudinal quantile regression in presence of informative drop-out through longitudinal-survival joint modeling

Alessio Farcomeni

Department of Public Health and Infectious Diseases
Sapienza - University of Rome

Sara Viviani

Department of Statistics
Sapienza - University of Rome

Abstract

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 drop-out in quantile regression. A general 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, and parametric and non-parametric assumptions for the baseline hazard. Model properties are illustrated through a simulation study and an application to an original data set about dilated cardiomyopathies.

Key Words: Quantile Regression; Longitudinal Regression; Joint Models; Shared-parameter models.

1 Introduction

In longitudinal studies subjects may be lost to follow-up due to events, like death, which are associated with the outcome of interest. Failure to model drop-out 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 Follmann and Wu (1995). Participation to a study can in general be described by a survival model of time to drop-out. The simpler

way of taking into account informative drop-out is through pattern-mixture models (e.g., Wu and Bailey (1988, 1989); Little and Wang (1996)), where the outcome distribution is specified conditionally on the time to drop-out. Selection models on the other hand condition the drop-out mechanism to unobserved responses directly (Diggle and Kenward, 1994) or indirectly. A simple and effective indirect link between drop-out and unobserved outcomes is by assuming that these are independent conditionally on unobserved shared random effects, as in Wu and Carroll (1988). One can similarly assume that the risk of event at time t is influenced by the *expected value* of the longitudinal response, as in Rizopoulos (2010) and Rizopoulos and Ghosh (2011). In the resulting Joint Model (JM), the hazard of drop-out is a function of the predicted longitudinal outcome, that is, of shared random and fixed effects, and related covariates. See also Little (1995), Henderson *et al.* (2000), Tsiatis and Davidian (2004), and references therein.

Our concern in this paper 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 these issues is given by modeling conditional quantiles of the longitudinal outcome. Quantile regression models (Koenker, 2005) are robust with respect to outliers, so that one can simply model the median rather than the mean. In many biomedical applications interest lies furthermore in at least one of the tails, and covariates may have different effects on different quantiles of a distribution. Examples include longitudinal fetal growth studies, which are usually focused on low and high quantiles of key anthropometric measurements. Finally, quantiles are invariant to transformations so it is never needed to transform the outcome. Longitudinal quantile regression models are proposed among others in Koenker (2004), who maximizes a penalized version of the likelihood; and Geraci and Bottai (2007), who introduce a random intercept and assume the outcome follows the asymmetric Laplace distribution (ALD). Liu and Bottai (2009) extend the random intercept model to a general linear mixed quantile regression model. See also Geraci and Bottai (2013). The ALD assumption is also used in Farcomeni (2012), where 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 drop-out. In Lipsitz *et al.* (1997) and Yi and He (2009) estimating equations are weighted proportionally to the inverse of the probability of drop-out. Yuan and Yin (2010), in a Bayesian framework, model missingness as a binary time series sharing a random effect with the quantile regression process. A common limit of these approaches is that drop-out 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 death occurs between two visits). 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 non-ignorability parameter(s) in JM as the one we propose. A longitudinal quantile regression model with ignorable missingness is outlined in Geraci (2013).

In this paper we propose a joint model for a right-censored time-to-event outcome

and the quantile of a continuous response repeatedly measured over time. Drop-out is formally defined as a monotone missing pattern, that is, when the outcome is not measured for a subject, no further measurements take place for that subject until the end of the study.

In our approach the quantile and survival processes are associated not only via shared latent variables or the predicted longitudinal outcome. It is in fact assumed that the time-to-event outcome depends on a function, which for simplicity we assume linear, of both the latent variables and the predicted quantile of the longitudinal outcome. Two non-ignorability parameters are introduced, one for the fixed and the other for the random part of the linear predictor. Our joint model is therefore a flexible approach to handle informative drop-out in longitudinal quantile regression. Maximum likelihood estimates are efficiently derived by setting up a Monte Carlo Expectation-Maximization (MCEM) algorithm based on importance sampling (Doucet *et al.*, 2001; Levine and Casella, 2001). There are two clear advantages of using importance sampling: 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. The use of two non-ignorability parameters allows us also increase the flexibility of both the shared parameter model of Wu and Carroll (1988), by conditioning the drop-out process also on the residuals between the predicted longitudinal outcome and the random effects, and the JM of Wulfsohn and Tsiatis (1997) and Rizopoulos (2010), by allowing a residual dependence on the random effects. In the first case, we can say that the two processes are not only linked by unmeasured heterogeneity, but that their dependence can also be in part explained through shared observed heterogeneity. In the second case, we can say that the two processes are linked not only via a quantile of the longitudinal outcome, but also by a residual unmeasured heterogeneity.

The rest of the paper is as follows: in the next section we describe the proposed model. In Section 3 we outline inference, we illustrate the approach through simulations in Section 4, and in Section 5 where we apply the method to 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 i th individual, $i = 1, \dots, n$, 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 $\Delta_i = I(T_i^* \leq C_i)$, where $I(\cdot)$ is the indicator function. The continuous outcome Y_{it} is repeatedly observed at times $t = 1, \dots, n_i$ before T_i , and is missing for $t \geq T_i$. The longitudinal outcome at observation times is collected in $\mathbf{y}_i = \{y_i(t) : t \leq T_i\}$. We assume that the longitudinal process is associated with T_i^* , i.e with the true event time, but, as customary in survival analysis,

is independent of the censoring time C_i .

We let \mathbf{X}_{it} denote a vector of predictors used to model only the longitudinal outcome, \mathbf{H}_{it} a vector of shared predictors, and \mathbf{W}_i a vector of (time fixed) predictors used to model only the survival process. These are associated with fixed effects $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$, respectively. We then have a vector \mathbf{Z}_{it} of predictors associated with q dimensional random effects \mathbf{u}_i , and two non-ignorability parameters α_1 (associated with fixed effects) and α_2 (associated with random effects). Our model can be expressed by a set of two equations, one for the longitudinal and the other one for the survival outcome:

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

where the first equation gives the longitudinal model and the second the time-to-event model, and $h_0(s)$ is a baseline function. Specifically, the model for the longitudinal outcome $\tilde{\tau}_{it}$ is formulated along the usual lines for mixed effects models (Verbeke and Molenberghs, 2000) and the model for the time-to-event outcome is based on the subject-specific hazard function $h(T_i)$ (Cox, 1972; Andersen and Gill, 1982). The risk of drop-out is conditional on $\mathcal{T}_{iT_i} = \{\tilde{\tau}_{iu} : 0 \leq u \leq T_i\}$, i.e. 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 $\mathbf{u}_i \sim f(\mathbf{u}_i)$. Few options are discussed in Section 2.3.

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

It is worth noting how our proposed model (1) generalizes previous work. If we fix $\alpha = \alpha_1 = \alpha_2$, $\boldsymbol{\beta} = 0$, and assume a Gaussian distribution for the error, we obtain a usual formulation for the JM. Otherwise, with an ALD error distribution (see below), we obtain a joint quantile regression model. Here, the link between the time-to-event and longitudinal processes is summarized by the $\alpha(\boldsymbol{\delta}'\mathbf{H}_{it} + \mathbf{u}_i'\mathbf{Z}_{it}) = \alpha\tilde{\tau}_{it}$ term in the model for the hazard function. If $\alpha_1 \neq \alpha_2$ and $\boldsymbol{\beta} = 0$,

$$\alpha_1\boldsymbol{\delta}'\mathbf{H}_{it} + \alpha_2\mathbf{u}_i'\mathbf{Z}_{it} = \alpha_1\tilde{\tau}_{it} + (\alpha_2 - \alpha_1)\mathbf{u}_i'\mathbf{Z}_{it},$$

that is, we generalize JM models by allowing for a residual dependence on unobserved heterogeneity as summarized by the shared random effects. When $\boldsymbol{\beta} \neq 0$, we further assume that some predictors may be related only to the longitudinal outcome but may not contribute to explain non-ignorability. Similarly, a shared parameter model is obtained if we fix $\alpha_1 = 0$. Specifically, if $\mathbf{Z}_{it} = (1 \ t)$ and we have Gaussian error terms and random effects, we exactly obtain the classical model in Wulfsohn and Tsiatis (1997). When $\alpha_1 > 0$ we condition the survival process also on the difference between the predicted

longitudinal outcome and the random effects. It is straightforward to fully generalize (1) by including also random effects that are not shared in each part of the model, and by letting W_i be time-dependent. In this way we would have, in each model, separate and shared covariates both for random and fixed effects. We do not pursue this explicitly to keep notation simple, but we note that our MCEM strategy can be directly adapted to this completely general case.

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

2.1 The longitudinal model

The parametric assumption on the error distribution of the longitudinal outcome drives 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 pre-specified and fixed. The resulting density of Y_{it} , conditionally on covariates and random effects, is given by

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

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 (Geraci and Bottai, 2007). The ALD is justifiable since maximum likelihood is exactly equivalent to minimization of the quantile loss function, when no parametric assumptions are made on ε (Yu and Moyeed, 2001). Further, it can be seen in simulation studies to lead to good estimates even when the residuals are not ALD distributed (see for instance Liu and Bottai (2009), Farcomeni (2012)).

2.2 The survival model

The time-varying baseline risk function $h_0(T_i)$ 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 $h_0(t)$ (e.g., $h_0(t) = \eta t^{\eta-1}$, leading to a Weibull model), as in for instance Rizopoulos and Ghosh (2011), Rizopoulos (2012), Viviani *et al.* (2013) or we may leave it unspecified as in Follmann and Wu (1995). The inferential strategy for obtaining the MLE is slightly different in the two cases, as we discuss in the next section.

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

$$\begin{aligned} f(T_i, \Delta_i | \mathbf{u}_i) &= f(T_i | \mathcal{T}_{it}, \mathbf{W}_i)^{\Delta_i} S(T_i | \mathcal{T}_{it}, \mathbf{W}_i)^{1-\Delta_i} \\ &= h(T_i | \mathcal{T}_{it}, \mathbf{W}_i)^{\Delta_i} S(T_i | \mathcal{T}_{it}, \mathbf{W}_i), \end{aligned} \quad (3)$$

where $S(\cdot)$ denotes the survival function, i.e.

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

while $h(T_i | \mathcal{T}_{it}, \mathbf{W}_i)$ 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 (Rizopoulos *et al.* (2008), Hsieh *et al.* (2006)) and /or the dimensionality q of \mathbf{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**:

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

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 (Liu and Bottai, 2009):

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

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 Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC), see for instance Verbeke and Lesaffre (1996).

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 , Σ and any parameter associated with baseline hazard $h_0(s)$. The joint likelihood contribution of the longitudinal and survival processes for the i -th subject is obtained integrating the conditional distributions in (1) over the random effects space:

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

where $f(T_i, \Delta_i | \mathbf{u}_i)$ is given in equation (3) and

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

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

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

The integrals involved in (4) are usually tackled in the JM context through quadrature methods (see for instance Rizopoulos (2012), Viviani *et al.* (2013)). While effective in one or two dimensions, quadrature methods tend to become too slow or less precise as the dimensionality of the random effect distribution 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 next section we propose a Monte Carlo strategy which is completely general, and allows us to set up an algorithm which 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 a 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), (ii) maximizing the latter (M-step). The algorithm is guaranteed to converge to a local optimum. 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. A second remark regards how to obtain standard errors and confidence intervals. When performing quantile regression on the longitudinal outcome, we use a non-parametric bootstrap strategy (Buchinsky (1995), Andrews and Buchinsky (2000)). We preserve the dependency 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. The individual contribution to the complete data log-likelihood can be obtained as

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

The MCEM algorithm is completely general and can be simply adapted to any distributional assumption for the longitudinal error and random effects distribution. Assuming an ALD for the error distribution of the longitudinal measurements, the complete data

log-likelihood is as follows:

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

3.1 Monte Carlo E-step

The conditional expected value of (6) for the i -th subject at the j -th iteration of the algorithm is expressed as

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

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

$$\begin{aligned}
f(\mathbf{u}_i \mid T_i, \Delta_i, \mathbf{Y}_i; \boldsymbol{\theta}^{(j)}) &\propto f(\mathbf{Y}_i, T_i, \Delta_i, \mathbf{u}_i; \boldsymbol{\theta}^{(j)}) \\
&= f(T_i, \Delta_i | \mathbf{u}_i; \boldsymbol{\theta}^{(j)}) f(\mathbf{Y}_i | \mathbf{u}_i; \boldsymbol{\theta}^{(j)}) f(\mathbf{u}_i | \boldsymbol{\Sigma}^{(j)}).
\end{aligned} \tag{9}$$

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

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

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

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

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

$$\begin{aligned} \mathbb{E}[\ell_c(\boldsymbol{\theta})] \approx & \sum_i \sum_{b=1}^{m_i^{(j)}} [\log f(\mathbf{Y}_i | \mathbf{v}_{ib}; \boldsymbol{\theta}^{(j)}) + \log f(T_i, \Delta_i | \mathbf{v}_{ib}; \boldsymbol{\theta}^{(j)}) \\ & + \log f(\mathbf{v}_{ib} | \boldsymbol{\theta}^{(j)})] w_{ib}, \end{aligned} \quad (10)$$

where

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

and $w_{ib} = \frac{\tilde{w}_{ib}}{\sum_b \tilde{w}_{ib}}$. In this work we proceed as in Levine and Casella (2001), where we sample $(v_{i1}, \dots, v_{im_i^{(j)}})$ from the posterior distribution at the initial parameter estimates using adaptive rejection Metropolis sampling (Gilks *et al.*, 1995), and then update the weights at each iteration. The MC sample size $m_i^{(j)}$ is increased to control the MC error (Levine and Casella, 2001; Eickoff *et al.*, 2004). Finally, 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. In summary, the E-step is given by updating of \tilde{w}_{ib} , w_{ib} , and evaluation of the likelihood.

3.2 M-step

We outline here the M-step, which consists in 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 (Nelson, 1972; Aalen, 1978) type estimator as

$$\hat{h}_0(s) = \sum_{i=1}^n \frac{\Delta_i I(T_i = s)}{\sum_{i: T_i \geq s} \sum_{b=1}^{m_i^{(j)}} w_{ib} \exp\{\boldsymbol{\gamma}' \mathbf{W}_i + \alpha_1 \boldsymbol{\delta}' \mathbf{H}_{is} + \alpha_2 \mathbf{v}_{ib}' \mathbf{Z}_{is}\}}, \quad (11)$$

by setting to zero the score equations for $h_0(s)$ (Wulfsohn and Tsiatis, 1997). See also Nielsen *et al.* (1992) and Gill (1992). The 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 non-parametric baseline along the lines of Cox (1972)). If instead we specify a parametric form for $h_0(s)$, e.g, $h_0(s) = \eta s^{\eta-1}$, 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 Wulfsohn and Tsiatis (1997)), 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 check function. An estimator of σ , dependent on the other parameters, can be explicitly obtained as:

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

The vector of regression parameters and dispersion parameter for the ALD is block updated using one step of the Nelder and Mead (1965) numerical optimization algorithm, majorizing (10) after plug-in of (12) and (11) or the parametric formula of $h_0(s)$. The resulting expected complete likelihood is

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

The integral involved in the expression above (and similarly in any expression where $f(T_i, \Delta_i | \mathbf{v}_{ib}; \boldsymbol{\theta})$ appears also at the E-step) reduce to sums when a non-parametric baseline is used, and can instead be approximated using one-dimensional Gauss-Kronrod quadrature (e.g., Kahaner *et al.* (1989)) when a parametric assumption is formulated for $h_0(s)$.

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 bias and standard deviation of the estimates for our proposed model for data Missing Not At Random (MNAR), and compare with a model which ignores informative drop-out (Missing At Random - MAR model).

For $n = \{250, 500\}$, $\alpha_1 = \{0, 1\}$, $\alpha_2 = \{0, 1\}$, $\tau = \{0.25, 0.5, 0.75\}$ we fix $\beta = \delta = \gamma = (1 \ 1)$ 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

$\mathbf{Z}_{it} = (1 \ t)$, $\mathbf{H}_{it} = (h_{i1} \ h_{i2} * t)$, $\mathbf{X}_i = (1 \ x_i)$; 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

$$S(t|\mathbf{u}_i, \mathbf{H}_i, \mathbf{W}_i) = \exp \left\{ - \frac{e^{\alpha_1(\delta_1 H_{i1} + \delta_2 H_{i2}t) + \alpha_2(u_{i1} + u_{i2}t) + \boldsymbol{\gamma}' \mathbf{W}_i} - e^{\alpha_1 \delta_1 H_{i1} + \alpha_2 u_{i1} + \boldsymbol{\gamma}' \mathbf{W}_i}}{\alpha_2 u_{i2} + \alpha_1 \delta_2 h_i} \right\}$$

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

$$S(t|U, H, W) = \exp\{-te^{\boldsymbol{\gamma}' \mathbf{W}_i}\}$$

when $\alpha_1 = \alpha_2 = 0$. The expression above 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 $t = 0, 1/4, 1/2, 3/4, 1, 3$. Longitudinal observations before drop-out are independently generated from an ALD for the τ -th quantile, centered on

$$\boldsymbol{\beta}' \mathbf{X}_i + \boldsymbol{\delta}' \mathbf{H}_{it} + \mathbf{u}' \mathbf{Z}_{it},$$

and with dispersion parameter σ . We fit our joint model with parametric baseline distribution and two separate models (one for the longitudinal process and one for the time-to-event, therefore obtaining MAR estimates), based on each generated data set.

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 1, 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 MSE for quantiles distant from the median, as expected.

In Table 2 we report the ratio of the bias and the variance of the estimates of the MAR over the MNAR model, based on the average bias for all parameters, separately for the longitudinal and survival 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 are non-zero, the ratios of the variances of the estimates are still close to the unity, but the ratios of biases increase substantially. The bias of the MAR model may be up to 30 times the bias of our joint model. The effect of α_1 is often stronger than the effect of α_2 , but this is likely only due to the fact that in all simulated settings there is a larger heterogeneity due to the 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 joint model is infinitesimal.

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 refers 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, 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 hearth failure, and in many cases an ethiological basis cannot be identified (Merlo *et al.*, 2011). The goal of this study is to compare patients with mild dilation of the left ventricular (Mildly Dilated CardioMyopathy or MDCM) with respect to patients with a general dilation of unrecognized ethiology (Idiopathic Dilated CardioMyopathy or IDCM). MDCM patients are generally believed to be at a slightly lower risk (e.g., Keren *et al.* (1990)), but the physiological reasons are still unrecognized.

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 heart-beat. Note that LVEF is a bounded outcome, but all measurements are far from the boundaries so that we unlike Bottai *et al.* (2010) we can avoid any transformations. Dropout occurs due to cardiovascular death, and it can be easily expected that the two processes are related as patients with a lower ejection fraction are at higher risk of death. For example, a univariate Cox model for the baseline LVEF gives an hazard ratio (HR) of 0.95 for each percentage point, with $p < 1e - 16$. Furthermore, LVEF is skewed and its skewness seems to change over time. This leads to two issues: first, using a classical joint model 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, which is also straightforward to interpret. We mostly are interested in low quantiles (like the 10th or the 15th), in the terziles or quartiles for the outcome. See for instance Sandri *et al.* (2003), Clements (2005), Ndrepepa *et al.* (2007), Cowie *et al.* (2012). Consequently, we explore the ejection fraction distribution by evaluating $\tau = 0.1, 0.15, 0.25, 0.33, 0.5, 0.66, 0.75$.

We model the longitudinal outcome conditionally on an intercept and age at baseline (\mathbf{X} matrix), on the indicator of MDCM and its interaction with time (\mathbf{H} matrix). Besides covariates in the \mathbf{H} matrix, we let the hazard of death depend on gender (1 for males) and indicator of New York Hearth Association (NYHA) functional classes I or II at baseline (\mathbf{W} matrix). For more details on NYHA functional classes see for instance Merlo *et al.* (2011) and references therein. We also include a shared subject-specific random intercept and a random slope, that is, $\mathbf{Z} = (1 \ t)$. Given that the number of follow-up times is slightly large, we use a normality assumption for \mathbf{u}_i . We also have compared with a multivariate T and multivariate Laplace, with analogous results which we do not report for reasons of space. We only mention that the multivariate normal

distribution is chosen using AIC and BIC criteria. We estimate our proposed model both with a parametric Weibull baseline and with a non-parametric baseline. Likelihood, AIC, and BIC at each quantile are reported in Table 3.

Based on those results we select the non-parametric baseline for all quantiles. We compare the estimates with those obtained under a MAR model in Table 4.

There is a stronger and stronger effect of age on LVEF as τ increases, while the effect of MDCM is slightly constant with τ , with a negative interaction with time. Males are at 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.

Ignoring drop-out 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 is in line with the expected consequences of drop-out in the low quantiles of the longitudinal outcome distribution. Secondly, under the MAR models 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 drop-out. We conclude by noting that given the results in Table 4 we can conclude there is some sensitivity to drop-out for the data at hand within the proposed class of models. As clearly noted in Molenberghs *et al.* (2008) one can never test the MAR assumption.

6 Conclusions

Informative drop-out 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 drop-out is milder in high quantiles than in low quantiles, for instance. Moreover, the brief simulation study reported confirms that, as long as the informative drop-out process is ignored, bias of the parameter estimates may be substantial and, more importantly, may not decrease with the sample size.

The proposed approach allows to simultaneously model the quantile of a longitudinal outcome and the hazard of drop-out, 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 drop-out 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 joint-models in different directions: first of all, we have proposed an alternative parametric assumption for the longitudinal error, the ALD, which allows 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.

In our example we have specified different values for the quantile of interest τ for illustration. It can be argued that 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 allow dependence (e.g., over the random effects at each quantile) or τ -homogeneity (e.g., for the variance of the random effects). Model estimation under these assumptions is at the moment grounds for further work.

Acknowledgements

The authors are grateful to the Cardiovascular department of “Ospedali riuniti”, Trento, Italy, and in particular to Giulia Barbati for permission to use the dilated cardiomyopathy data.

References

- O. AALEN (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics*, **6**, 701–726.
- P. ANDERSEN AND R. GILL (1982). Cox’s regression model for counting processes: A large sample study. *Annals of Statistics*, **10**, 1100–1120.
- D. W. K. ANDREWS AND M. BUCHINSKY (2000). A three-step method for choosing the number of bootstrap repetitions. *Econometrica*, **68**, 23–52.
- M. BOTTAI, B. CAI, AND E.R. MCKEOWN (2010). Logistic quantile regression for bounded outcomes. *Statistics in Medicine*, **29**, 309–317.
- M. BUCHINSKY (1995). Estimating the asymptotic covariance matrix for quantile regression models: a Monte Carlo study. *Journal of Econometrics*, **68**, 303–338.
- I. P. CLEMENTS (2005). Combined systolic and diastolic dysfunction in the presence of preserved left ventricular ejection fraction. *European Journal of Heart Failure*, **7**, 490–497.
- B. COWIE, R. KLUGER, AND M. KALPOKAS (2012). Left ventricular volume and ejection fraction assessment with transoesophageal echocardiography: 2D vs 3D imaging. *British Journal of Anaesthesia*.
- D. COX (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.
- P. DIGGLE AND M.G. KENWARD (1994). Informative drop-out in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–93.

- A. DOUCET, N. DE FREITAS, AND N. GORDON (2001). *Sequential Monte Carlo Methods in Practice*. Springer.
- J. C. EICKOFF, J. ZHU, AND Y. AMAMIYA (2004). On the simulation size and convergence of the Monte Carlo EM algorithm via likelihood-based distances. *Statistics and Probability letters*, **67**, 161–171.
- A. FARCOMENI (2012). Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Statistics and Computing*, **22**, 141–152.
- D. FOLLMANN AND M. WU (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics*, **51**, 151–168.
- M. GERACI (2013). Estimation of regression quantiles in complex surveys with data missing at random: An application to birthweight determinants. *Statistical Methods in Medical Research*, available online.
- M. GERACI AND M. BOTTAI (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8**, 140–154.
- M. GERACI AND M. BOTTAI (2013). Linear quantile mixed models. *Statistics and Computing*, available online.
- W. R. GILKS, N. BEST, AND K. TAN (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics*, **44**, 455–472.
- R.D. GILL (1992). Marginal partial likelihood. *Scandinavian Journal of Statistics*, **19**, 133–137.
- R. HENDERSON, P. DIGGLE, AND A. DOBSON (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, **1**, 465–480.
- F. HSIEH, Y.-K. TSENG, AND J.-K. WANG (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, **62**, 1037–1043.
- D. KAHANER, C. MOLER, AND S. NASH (1989). *Numerical Methods and Software*. PrenticeHall.
- A. KEREN, S. GOTTLIEB, D. TZIVONI, S. STERN, R. YAROM, M.E. BILLINGHAM, AND R.L. POPP (1990). Mildly dilated congestive cardiomyopathy. Use of prospective diagnostic criteria and description of the clinical course without heart transplantation. *Circulation*, **81**, 506–517.
- R. KOENKER (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, **91**, 74–89.
- R. KOENKER (2005). *Quantile Regression*. Cambridge University Press, New York.

- R. A. LEVINE AND G. CASELLA (2001). Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics*, **10**, 422–439.
- S. R. LIPSITZ, G.M. FITZMAURICE, G. MOLENBERGHS, AND L.P. ZHAO (1997). 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*, **46**, 463–476.
- R. J. A. LITTLE (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.
- R. J. A. LITTLE AND Y. WANG (1996). Pattern-mixing models for multivariate incomplete data with covariates. *Biometrics*, **52**, 98–111.
- Y. LIU AND M. BOTTAI (2009). Mixed-effects models for conditional quantiles with longitudinal data. *The International Journal of Biostatistics*, **5**.
- M. MERLO, S.A. PYXARAS, B. PINAMONTI, G. BARBATI, A. DI LENARDA, AND G. SINAGRA (2011). Prevalence and prognostic significance of left ventricular reverse remodeling in dilated cardiomyopathy receiving tailored medical treatment. *Journal of the American College of Cardiology*, **57**, 1468–1476.
- G. MOLENBERGHS, C. BEUNCKENS, C. SOTTO, AND M. G. KENWARD (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society (Series B)*, **70**, 371–388.
- G. NDREPEPA, J. MEHILLI, S. MARTINOFF, M. SCHWAIGER, A. SCHÖMIG, AND A. KASTRATI (2007). Evolution of left ventricular ejection fraction and its relationship to infarct size after acute myocardial infarction. *Journal of the American College of Cardiology*, **50**, 149–156.
- J. A. NELDER AND R. MEAD (1965). A simplex algorithm for function minimization. *Computer Journal*, **7**, 308–313.
- W. NELSON (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, **14**, 945–965.
- G.G. NIELSEN, R.D. GILL, P.K. ANDERSEN, AND T.I.A. SØRENSEN (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, **19**, 25–43.
- D. RIZOPOULOS (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, **35**, 9.
- D. RIZOPOULOS (2012). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics and Data Analysis*, **56**, 491–501.

- D. RIZOPOULOS AND P. GHOSH (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, **30**, 1366–1380.
- D. RIZOPOULOS, G. VERBEKE, AND G. MOLENBERGHS (2008). Shared parameter models under random effects misspecification. *Biometrika*, **95**, 63–74.
- M.T. SANDRI, D. CARDINALE, L. ZORZINO, R. PASSERINI, P. LENTATI, A. MARTINONI, G. MARTINELLI, AND C.M. CIPOLLA (2003). Minor increases in plasma troponin I predict decreased left ventricular ejection fraction after high-dose chemotherapy. *Clinical Chemistry*, **49**, 248–252.
- A. TSIATIS AND M. DAVIDIAN (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica*, **14**, 793–818.
- G. VERBEKE AND E. LESAFFRE (1996). A linear mixed model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**, 217–221.
- G. VERBEKE AND G. MOLENBERGHS (2000). *Linear mixed models for longitudinal data*. Springer.
- S. VIVIANI, M. ALFÓ, AND D. RIZOPOULOS (2013). Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing*, available online.
- M. WU AND R. CARROL (1988). Estimation and comparison of changes in presence of informative right censoring by modelling the censoring process. *Biometrics*, **45**, 939–955.
- M. C. WU AND K. R. BAILEY (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, **7**, 337–346.
- M. C. WU AND K. R. BAILEY (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, **45**, 939–955.
- M. WULFSOHN AND A. TSIATIS (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330–339.
- G. Y. YI AND W. HE (2009). Median regression models for longitudinal data with dropouts. *Biometrics*, **65**, 618–625.
- K. YU AND R.A. MOYEED (2001). Bayesian quantile regression. *Statistics and Probability Letters*, **54**, 437–447.
- Y. YUAN AND G. YIN (2010). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, **66**, 105–114.

Table 1: Bias and standard deviation of the estimates of the proposed model on simulated data for different values of n , α_1 and α_2 . Results are based on $B = 1000$ replicates.

$\tau = 0.25$												
n	α_1	α_2	β		δ		γ		α_1		α_2	
			bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.
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
$\tau = 0.5$												
n	α_1	α_2	β		δ		γ		α_1		α_2	
			bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.
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
$\tau = 0.75$												
n	α_1	α_2	β		δ		γ		α_1		α_2	
			bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.
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

Table 2: Ratios of bias and variance of the estimates obtained with the MAR model ($bias_{MAR}, var_{MAR}$) and with our model ($bias, var$) for the longitudinal (Y) and survival (T) part of the model. Results are shown for different values of n , α_1 and α_2 and are based on $B = 1000$ replicates.

$\tau = 0.25$						
n	α_1	α_2	$\frac{bias_{MAR}(Y)}{bias(Y)}$	$\frac{var_{MAR}(Y)}{var(Y)}$	$\frac{bias_{MAR}(T)}{bias(T)}$	$\frac{var_{MAR}(T)}{var(T)}$
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
$\tau = 0.5$						
n	α_1	α_2	$\frac{bias_{MAR}(Y)}{bias(Y)}$	$\frac{var_{MAR}(Y)}{var(Y)}$	$\frac{bias_{MAR}(T)}{bias(T)}$	$\frac{var_{MAR}(T)}{var(T)}$
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
$\tau = 0.75$						
n	α_1	α_2	$\frac{bias_{MAR}(Y)}{bias(Y)}$	$\frac{var_{MAR}(Y)}{var(Y)}$	$\frac{bias_{MAR}(T)}{bias(T)}$	$\frac{var_{MAR}(T)}{var(T)}$
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

Table 3: Log-likelihood at convergence, AIC and BIC for our model fit with a parametric and non-parametric baseline, at different quantiles of interest, on the dilated cardiomyopathy data.

τ	non-parametric baseline			parametric baseline		
	$\ell(\boldsymbol{\theta})$	BIC	AIC	$\ell(\boldsymbol{\theta})$	BIC	AIC
.1	-10515.214	21064.255	21054.428	-10602.111	21240.867	21228.222
.15	-10449.858	20933.543	20923.716	-10461.581	20959.807	20947.162
.25	-10289.721	20613.269	20603.442	-10322.538	20681.721	20669.076
.33	-10197.313	20428.453	20418.626	-10235.107	20506.859	20494.214
.5	-10154.227	20342.281	20332.454	-10175.175	20386.995	20374.350
.66	-10223.996	20481.819	20471.992	-10229.257	20495.159	20482.514
.75	-10300.913	20635.653	20625.826	-10308.303	20653.251	20640.606

Table 4: Estimates for the MNAR and MAR models for the MDCM data. An asterisk indicates that estimates in the column are significant at the 5% level for all quantiles τ . The models are based on a non-parametric baseline for the survival process and two dimensional shared random effects.

MNAR model								
τ	Longitudinal outcome				Survival outcome			
	Int*	Age*	MDCM*	MDCM:time*	Gender*	NYHA*	α_1^*	α_2^*
.1	22.535	-0.038	7.754	-0.038	0.538	-0.912	-0.044	-0.042
.15	25.134	-0.048	7.768	-0.028	0.527	-0.895	-0.037	-0.057
.25	29.255	-0.068	7.752	-0.042	0.593	-0.990	-0.017	-0.068
.33	32.813	-0.095	7.752	-0.018	0.592	-0.988	-0.018	-0.087
.5	39.086	-0.124	7.749	-0.093	0.542	-0.918	-0.031	-0.063
.66	44.437	-0.148	7.734	-0.136	0.566	-0.952	-0.011	-0.097
.75	47.355	-0.156	7.744	-0.044	0.591	-0.987	-0.021	-0.094
MAR model								
	Longitudinal outcome				Survival outcome			
	Int*	Age*	MDCM*	MDCM:time*	Gender*	NYHA*	α_1	α_2
.1	29.700	-0.141	7.580	0.041	0.600	-1.096	0.000	0.000
.15	36.266	-0.251	8.027	0.076	0.600	-1.096	0.000	0.000
.25	35.175	-0.130	5.262	0.011	0.600	-1.096	0.000	0.000
.33	35.246	-0.109	6.516	0.068	0.600	-1.096	0.000	0.000
.5	40.048	-0.137	8.289	0.028	0.600	-1.096	0.000	0.000
.66	44.541	-0.123	5.711	0.022	0.600	-1.096	0.000	0.000
.75	44.284	-0.088	5.558	0.017	0.600	-1.096	0.000	0.000