Biometrics 66, 105–114
March 2010
DOI: 10.1111/j.1541-0420.2009.01269.x

Bayesian Quantile Regression for Longitudinal Studies with Nonignorable Missing Data

Ying Yuan* **and Guosheng Yin**

Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.
**email:* yyuan@mdanderson.org

Summary. We study quantile regression (QR) for longitudinal measurements with nonignorable intermittent missing data and dropout. Compared to conventional mean regression, quantile regression can characterize the entire conditional distribution of the outcome variable, and is more robust to outliers and misspecification of the error distribution. We account for the within-subject correlation by introducing a ℓ_2 penalty in the usual QR check function to shrink the subject-specific intercepts and slopes toward the common population values. The informative missing data are assumed to be related to the longitudinal outcome process through the shared latent random effects. We assess the performance of the proposed method using simulation studies, and illustrate it with data from a pediatric AIDS clinical trial.

Key words: Bayesian inference; Informative missing data; Nonignorable dropout; Penalized function; Random effects; Repeated measures; Shared-parameter model.

1. Introduction

Longitudinal studies often suffer from attrition, which may lead to biased estimates of the model parameters if the missing data are nonignorable or informative. Modeling longitudinal data with nonignorable missing data has drawn substantial attention (Wu and Carroll, 1988; Wu and Bailey, 1989; Little, 1993; Diggle and Kenward, 1994; among others). Recent reviews on nonignorable dropout in longitudinal data are given in Little (1995, 2008); Verbeke and Molenberghs (2000); Hogan, Roy, and Korkontzelou (2004); and Molenberghs and Kenward (2007). The majority of these methods focuses on mean regression. In contrast, very limited research has been conducted on quantile regression (QR) for longitudinal studies with nonignorable missing data. In this article, we propose a shared-parameter QR model to deal with nonignorable missing data, in which the quantile regression model for the longitudinal process is linked to the missing data model via sharing common random effects.

QR models have become increasingly popular since the seminal work of Koenker and Bassett (1978). In contrast to the mean regression model, QR belongs to a robust model family, which can give an overall assessment of the covariate effects at different quantiles of the outcome (Koenker, 2005). In particular, we can model the lower or higher quantiles of the outcome to provide a natural assessment of covariate effects specific for those regression quantiles. Unlike conventional models, which address solely the conditional mean or the central effects of the covariates, QR models quantify the entire conditional distribution of the outcome variable. In addition, QR does not impose any distributional assumption on the error, except requiring that the error has a zero conditional quantile.

By inversely weighting the estimating equation with the probability of dropout, Lipsitz et al. (1997) studied QR for longitudinal data with ignorable dropouts. Noting that the classical random-effects model can be reformulated as a penalized least-square estimation, Koenker (2004) developed a ℓ_1-regularization QR method to shrink individual effects toward a common value. Geraci and Bottai (2007) proposed a random-effects QR model for longitudinal data based on the asymmetric Laplace distribution (ALD; Yu and Moyeed, 2001), in which the within-subject correlation was modeled by random intercepts. Other applications of quantile regression in correlated data include the work of Cole and Green (1992), Jung (1996), and Heagerty and Pepe (1999), among others.

Our motivating example is a double-blinded randomized pediatric AIDS trial (AIDS Clinical Trials Group 128, Brady et al., 1996), designed to compare the efficacy of a lower dosage (90 mg/m^2/dose) of zidovudine with a higher dosage (180 mg/m^2/dose) to treat HIV-infected children (3 months to 12 years of age) with mild to moderate symptoms. A total of 424 subjects were enrolled with 216 subjects randomized to a low-dose group and 208 to a high-dose group. The CD4 cell count was collected for the participants at the study entry and every 12 weeks up to 200 weeks. For HIV-infected patients, the CD4 cell count is often used for monitoring the progression of HIV infection, and a slower decline of the CD4 cell count represents a better treatment effect. It was of scientific interest to study the longitudinal trajectories of the CD4 cell counts in the two dosage groups. As these two dosages of zidovudine may have different efficacy effects on patients with different CD4 cell counts, the QR method can be naturally used to examine the treatment effects at various quantiles of the conditional distribution of the CD4 cell counts over time.

A prominent feature of the pediatric AIDS study that complicated the analysis was that there was a substantial amount of intermittent missing data and a large number of dropouts. Only 52% of the subjects completed 3 years of follow-up for

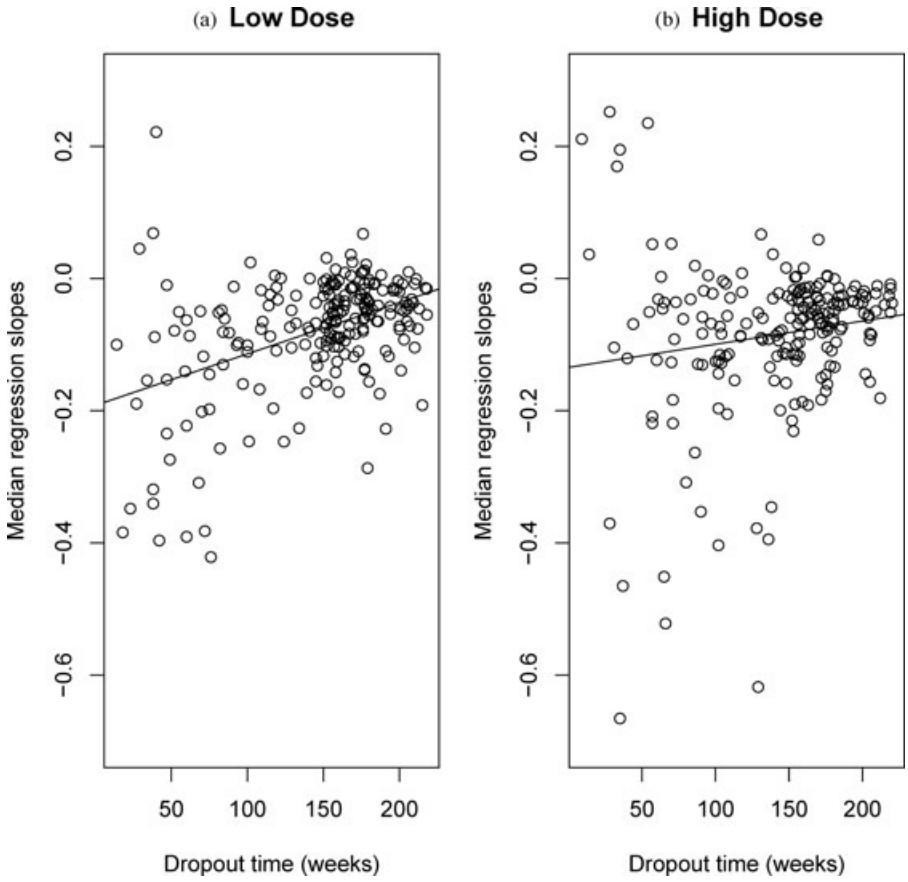

Figure 1. Individual median regression slopes for the square root of the CD4 cell count versus dropout times.

the low-dose and 45% for the high-dose group. In addition, approximately 50% of the subjects experienced one or more intermittent missing CD4 cell counts. In AIDS studies, missing data are often associated with the level or underlying rate of the change in the CD4 cell counts, e.g., sicker patients often experience a steeper decline of the CD4 cell counts and consequently are more likely to drop out from the study (Hogan and Laird, 1997). These types of missing data are typically nonignorable (Wu and Carroll, 1988; De Gruttola and Tu, 1994; and Hogan and Laird, 1997). In a preliminary analysis, we applied the median regression model separately for each subject and plotted the estimated individual slopes against the dropout time. As shown in Figure 1, subjects with lower slopes tended to drop out earlier, especially in the low-dose arm, indicating that missing data might be informative. To deal with the nonignorable missing data, we propose a shared-parameter QR model, in which individual-level QR parameters are shrunk toward a population value by penalizing the standard check function of QR. Observing the link between the usual check function and the asymmetric Laplace distribution, we transform the penalized check function to a random-effects model in the likelihood framework. We assume that the missing data process is associated with the longitudinal outcome process via the shared latent subject-specific random effects. In the Bayesian paradigm, the estimation and inference based on the proposed model can be easily implemented using the Markov chain Monte Carlo (MCMC) procedure.

In Section 2, we present the shared-parameter random-effects QR model for longitudinal data with nonignorable intermittent missing data and dropout. We also outline the Bayesian MCMC estimation procedure. In Section 3, we carry out simulation studies to examine the performance of the proposed model, and in Section 4, we illustrate our method using the pediatric AIDS data. We conclude with a brief discussion in Section 5.

2. Methods

2.1 *Quantile Regression*

Let y_i denote the outcome of interest, and let $\boldsymbol{x}_i$ denote the corresponding covariate vector of subject i, for $i = 1, \ldots, n$. The τth QR model takes the form of

$$Q_{y_i}(\tau \mid \boldsymbol{x}_i) = \boldsymbol{x}_i^T \boldsymbol{\beta},$$

where $Q_{y_i}(\tau \mid \boldsymbol{x}_i)$ is the inverse cumulative distribution function of y_i given $\boldsymbol{x}_i$ evaluated at τ, and $0 < \tau < 1$. The regression coefficient vector $\boldsymbol{\beta}$ is estimated by minimizing

$$\sum_{i=1}^{n} \rho_\tau \left(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta} \right), \tag{1}$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$.

The check function (1) is closely related to the ALD; see Koenker and Machado (1999), Yu and Moyeed (2001), and

Yu and Stander (2007). The density function of an ALD with a scale parameter of 1 is

$$f(y \mid \mu, \tau) = \tau(1-\tau)\exp\{-\rho_\tau(y-\mu)\}, \tag{2}$$

where τ determines the quantile level, and μ is the location parameter. Minimizing equation (1) is equivalent to maximizing the likelihood function of y_i by assuming y_i from an ALD with $\mu = \boldsymbol{x}_i^T\boldsymbol{\beta}$.

The relationship between the check function and ALD can be used to reformulate the QR method in the likelihood framework. By utilizing this property, Koenker and Machado (1999) proposed a likelihood-based goodness-of-fit test for quantile regression. Yu and Moyeed (2001) developed Bayesian quantile regression, and Yu and Stander (2007) studied the Bayesian estimation procedure for the Tobit QR model with censored data. More recently, Geraci and Bottai (2007) proposed a random-intercept QR model for longitudinal data.

2.2 *Modeling Longitudinal Data*

Considering a longitudinal study in which n subjects are repeatedly measured at J prespecified time points, let y_{ij} denote the outcome for the ith subject measured at the jth time point, for $i = 1, \ldots, n$, $j = 1, \ldots, J$. Due to intermittent missing data or dropout, only a portion of the outcome measurements can be observed, and we use $\mathcal{J}_{i,\text{obs}}$ to denote the set of time points at which y_{ij} is observed. For the τth regression quantile, we propose the following ℓ_2-penalized check function,

$$\sum_{i=1}^{n}\sum_{j\in\mathcal{J}_{i,\text{obs}}}\rho_\tau\left(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{\beta} - \boldsymbol{z}_{ij}^T\boldsymbol{b}_i\right) + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{b}_i^T\boldsymbol{\Lambda}^{-1}\boldsymbol{b}_i, \tag{3}$$

where $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are vectors of covariates that may share common components, $\boldsymbol{b}_i$ is a vector of unknown subject-specific effects, and $\boldsymbol{\Lambda}$ is a symmetric nonsingular matrix. By introducing the penalty term $\sum_{i=1}^{n}\boldsymbol{b}_i^T\boldsymbol{\Lambda}^{-1}\boldsymbol{b}_i/2$ in (3), we shrink the individual effects $\boldsymbol{b}_i$ toward 0, and thereby borrow strength across subjects. The amount of shrinkage is controlled by the tuning parameter $\boldsymbol{\Lambda}$.

The penalized check function (3) can be cast into the likelihood framework of a random-effects model as follows:

$$\begin{aligned} y_{ij} \mid \boldsymbol{b}_i &\sim \text{ALD}\left(\tau, \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i\right) \\ \boldsymbol{b}_i &\sim N(0, \boldsymbol{\Lambda}). \end{aligned} \tag{4}$$

Reformulation of the ℓ_2 penalty $\sum_{i=1}^{n}\boldsymbol{b}_i^T\boldsymbol{\Lambda}^{-1}\boldsymbol{b}_i/2$ as random effects is analogous to representing a cubic smoothing spline as a linear mixed model, in which the roughness penalty is expressed as normal random effects (Ruppert, Wand, and Carroll, 2003; and Welham, 2008).

It is worth noting that the random-effects model (4) is merely a *working* model, in which the ALD and normal distribution assumptions imposed on y_{ij} and $\boldsymbol{b}_i$ are essentially artificial. These distributional assumptions are used solely to ensure that, conditional on the tuning parameter $\boldsymbol{\Lambda}$, the likelihood of model (4) matches the penalized check function in (3) so that minimizing the penalized check function can be achieved by maximizing the likelihood of the random-effects model. Such reformulation allows us to work with the usual likelihood function. Furthermore, the tuning parameter $\boldsymbol{\Lambda}$ can be directly determined, which is analogous to using the random-effects model representation of the smoothing spline to automatically determine the smoothing parameter through the restricted maximum likelihood or the Bayesian method. The tuning parameter $\boldsymbol{\Lambda}$ is often treated as nuisance, which controls the shrinkage of individual effects toward the population effects.

The working model (4) provides a generalization of the random-intercept model proposed by Geraci and Bottai (2007), as we allow for subject-specific curves via both random intercepts and slopes. Moreover, instead of using the expectation–maximization algorithm and bootstrap methods to estimate the parameters and associated variances, we proceed with our estimation using the Bayesian MCMC method, which automatically yields the posterior variance estimates based on the posterior samples of the model parameters.

2.3 *Modeling Nonignorable Missing Data*

To account for the nonignorable missing data, we model the intermittent missing data and dropout processes, and connect them with the longitudinal outcome process. To this end, we define the indicator for the missing data status

$$s_{ij} = \begin{cases} \mathcal{O} & \text{if measurement } j \text{ of subject } i \text{ is observed,} \\ \mathcal{I} & \text{if measurement } j \text{ of subject } i \text{ is} \\ & \quad \text{intermittent missing,} \\ \mathcal{D} & \text{if subject } i \text{ drops out at measurement } j. \end{cases}$$

We assume that the outcome measurement at $j = 1$ is observed for all subjects, and dropout is an absorbing state, i.e., once a subject drops out, we will not observe measurements of this subject any more. We also require $\Pr(s_{ij} = \mathcal{D} \mid s_{i(j-1)} = \mathcal{I}) = 0$ because the intermittent missingness, by definition, cannot be immediately followed by a dropout.

We link the missing data process with the longitudinal outcome process by assuming that they share the same random effects $\boldsymbol{b}_i$. Particularly, we model the missing data process using the transition probabilities as follows:

$$\begin{aligned} \pi_{ij}^{(\mathcal{O})} &= \Pr(s_{ij} = \mathcal{O} \mid s_{i(j-1)} \neq \mathcal{D}, \boldsymbol{b}_i) \\ &= \frac{1}{1 + \sum_{k\in(\mathcal{I},\mathcal{D})}\exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\alpha}^{(k)} + \boldsymbol{b}_i^T\boldsymbol{\gamma}^{(k)}\right)} \\ \pi_{ij}^{(\mathcal{I})} &= \Pr(s_{ij} = \mathcal{I} \mid s_{i(j-1)} \neq \mathcal{D}, \boldsymbol{b}_i) \\ &= \frac{\exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\alpha}^{(\mathcal{I})} + \boldsymbol{b}_i^T\boldsymbol{\gamma}^{(\mathcal{I})}\right)}{1 + \sum_{k\in(\mathcal{I},\mathcal{D})}\exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\alpha}^{(k)} + \boldsymbol{b}_i^T\boldsymbol{\gamma}^{(k)}\right)} \\ \pi_{ij}^{(\mathcal{D})} &= \Pr(s_{ij} = \mathcal{D} \mid s_{i(j-1)} = \mathcal{O}, \boldsymbol{b}_i) \\ &= \frac{\exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\alpha}^{(\mathcal{D})} + \boldsymbol{b}_i^T\boldsymbol{\gamma}^{(\mathcal{D})}\right)}{1 + \sum_{k\in(\mathcal{I},\mathcal{D})}\exp\left(\boldsymbol{w}_{ij}^T\boldsymbol{\alpha}^{(k)} + \boldsymbol{b}_i^T\boldsymbol{\gamma}^{(k)}\right)} \end{aligned}$$

where $\boldsymbol{w}_{ij}$ is a vector of covariates, $\boldsymbol{\alpha}^{(k)}$ is its associated regression parameter, and $\boldsymbol{\gamma}^{(k)}$ governs the relationship between the random effects $\boldsymbol{b}_i$ and the missing data process. The logarithm of the conditional likelihood for the missing data process of subject i is

$$\begin{aligned} \sum_{j=2}^{J}\log f(s_{ij} \mid \boldsymbol{b}_i) = \sum_{j=2}^{J}\big\{ & I(s_{ij} = \mathcal{O})\log\pi_{ij}^{(\mathcal{O})} + I(s_{ij} = \mathcal{I})\log\pi_{ij}^{(\mathcal{I})} \\ & + I(s_{ij} = \mathcal{D})\log\pi_{ij}^{(\mathcal{D})}\big\}, \end{aligned}$$

where after subject i drops out, the rest of s_{ij}'s are undefined.

By sharing the random effects $\boldsymbol{b}_i$ with the longitudinal outcome process, the missing data model accounts for a nonignorable missing data mechanism (Little, 1995). Our model belongs to a class of shared-parameter models that have been extensively studied in the context of mean regression for longitudinal data with nonignorable dropout. For examples, see the work of Wu and Carroll (1988); De Gruttola and Tu (1994); Follman and Wu (1995); Ten Have et al. (1998); and Rizopoulos, Verbeke, and Molenberghs (2008); among others. The justification for the shared-parameter models is that the association between the longitudinal outcome process and the missing data process is fully captured by the latent subject-specific trajectories. In other words, conditional on the latent trajectories, the missing data process is independent of the outcome process.

2.4 *Posterior Estimation*

Let $(\boldsymbol{y}, \boldsymbol{s})$ denote the observed data; the likelihood of the observed data is given by

$$L(\boldsymbol{y}, \boldsymbol{s} \,|\, \boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\gamma}^{(k)}) = \prod_{i=1}^{n} \int \prod_{j \in \mathcal{J}_{i,\,\text{obs}}} f(y_{ij} \,|\, \boldsymbol{b}_i) \prod_{j=2}^{J} f(s_{ij} \,|\, \boldsymbol{b}_i) f(\boldsymbol{b}_i)\, \mathrm{d}\boldsymbol{b}_i, \quad k = \mathcal{I}, \mathcal{D}.$$

In the Bayesian paradigm, let $p(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\gamma}^{(k)})$ denote the prior distribution of the unknown parameters; the joint posterior distribution of these parameters is given by

$$p(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\gamma}^{(k)} \,|\, \boldsymbol{y}, \boldsymbol{s}) \propto L(\boldsymbol{y}, \boldsymbol{s} \,|\, \boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\gamma}^{(k)}) p(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\gamma}^{(k)}).$$

We assign noninformative prior distributions to the unknown parameters as follows,

$$\boldsymbol{\beta}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\gamma}^{(k)} \propto 1, \quad k = \mathcal{I}, \mathcal{D}$$
$$\boldsymbol{\Lambda}^{-1} \sim \mathrm{WI}(q, c\mathbf{I})$$

where $\mathrm{WI}(q, c\,\mathbf{I})$ denotes a Wishart distribution with q degrees of freedom, and a scale matrix $c\,\mathbf{I}$ with c a small constant and $\mathbf{I}$ the identity matrix.

We use the Gibbs sampler to obtain posterior distributions of the unknown parameters. The full conditional distributions of the model parameters, except for $\boldsymbol{\Lambda}$, do not have closed forms. The adaptive rejection Metropolis sampling algorithm (Gilks, Best, and Tan, 1995) is used to sample from these distributions. We monitor the convergence of the Gibbs sampler using graphical inspection of the trace plots and the method of Gelman and Rubin (1992). The computer programs are written in `C++` and available upon request.

3. Simulation Study

We conducted simulation studies to assess the performance of the proposed model. We mimicked the setting of the pediatric AIDS clinical trial by taking the sample size $n = 200$, and assuming that each subject had 14 scheduled longitudinal measurements. We simulated data from the model

$$y_{ij} \,|\, b_{0i}, b_{1i} = \beta_0 + \beta_1 x_{ij} + b_{0i} + b_{1i} x_{ij} + \epsilon_{ij}$$

$$\pi_{ij}^{(\mathcal{O})} = \frac{1}{1 + \sum_{k \in (\mathcal{I}, \mathcal{D})} \exp\left(\gamma_0^{(k)} + \gamma_1^{(k)} b_{1i}\right)}$$
$$\pi_{ij}^{(\mathcal{I})} = \frac{\exp\left(\gamma_0^{(\mathcal{I})} + \gamma_1^{(\mathcal{I})} b_{1i}\right)}{1 + \sum_{k \in (\mathcal{I}, \mathcal{D})} \exp\left(\gamma_0^{(k)} + \gamma_1^{(k)} b_{1i}\right)} \tag{5}$$
$$\pi_{ij}^{(\mathcal{D})} = \frac{\exp\left(\gamma_0^{(\mathcal{D})} + \gamma_1^{(\mathcal{D})} b_{1i}\right)}{1 + \sum_{k \in (\mathcal{I}, \mathcal{D})} \exp\left(\gamma_0^{(k)} + \gamma_1^{(k)} b_{1i}\right)}$$

where x_{ij} was the standardized measurement time, $x_{ij} = (j - 7.5)/4.18$ for $j = 1, \ldots, 14$. In this model, we assumed that the missing data process was associated with the outcome process via the random slope b_{1i}.

We simulated the error ϵ_{ij} from three different distributions: the standard normal distribution $N(0, 1)$, a $t_{(3)}$ distribution with three degrees of freedom, and a $\chi^2_{(3)}$ distribution with three degrees of freedom. We generated the random intercept $b_{0i} \sim N(0, 4)$, and the random slope b_{1i} from four different distributions: $N(0, 2)$, a $t_{(3)}$ distribution with a scale parameter of $\sqrt{2}$, a unimodal skewed mixture distribution $0.7N(1.3, 1.6^2) + 0.3N(-3.033, 1.6^2)$, and a bimodal mixture distribution $0.45N(-2, 1.5^2) + 0.55\ N(1.636, 1.5^2)$. In model (5), the parameter $\gamma_1^{(k)}$ determines the missing mechanism: if $\gamma_1^{(k)} = 0$, the missing data are missing at random (MAR); and if $\gamma_1^{(k)} \neq 0$, the missing data are nonignorable. We considered $(\gamma_0^{(\mathcal{I})}, \gamma_1^{(\mathcal{I})}, \gamma_0^{(\mathcal{D})}, \gamma_1^{(\mathcal{D})}) = (-2.4, 0, -2.8, 0)$, and $(-6,\ 1,\ -8.9,\ 1.5)$ to simulate the cases of MAR and nonignorable missing data, respectively. The values of $\gamma_0^{(\mathcal{I})}$ and $\gamma_0^{(\mathcal{D})}$ were chosen such that, on average, 50% of the subjects experienced one or more incidences of intermittent missing data or dropout. We set $\beta_0 = 2$, $\beta_1 = 4$, and simulated 500 replicated datasets under each configuration.

We compared our shared-parameter QR with the random-effects QR in (4) without adjusting for the missing data. As a benchmark, we also fitted the random-effects QR to the complete data before deletion. In Table 1, we present the simulation results for β_1, including the bias, the averaged standard error, and the coverage probability of the 95% credible interval (CI), when the missing data are nonignorable. Across various error and random-effects distributions, the random-effects QR yields biased estimates of β_1 and poor coverage probabilities. The bias ranges from -0.46 to -0.97, and most of the coverage probabilities are lower than 5%. In contrast, the shared-parameter QR took the missing data into account so that the estimation bias is effectively corrected and the coverage probabilities are reasonably accurate. As expected, the estimates based on the shared-parameter QR are less efficient than those under the ideal before-deletion QR, but the efficiency loss is quite small. For example, when the error follows a $t_{(3)}$ distribution and the random effects are normal, the standard error of $\hat{\beta}_1$ is 0.11 under the before-deletion QR and 0.13 under our shared-parameter QR. When the missing data are MAR (results are not shown), both the estimates based on the random-effects QR and the shared-parameter QR have negligible biases and reasonable coverage probabilities, but the shared-parameter QR is slightly less efficient due to its complex model structure. The results for β_0 are similar except that the bias of $\hat{\beta}_0$ under the random-effects QR

Table 1

Estimates of β_1 under different error and random effects distributions, including the bias, averaged standard error (SE), and coverage probability (CP) of the 95% CI, in the simulation study

			$\tau = 0.25$			$\tau = 0.5$		
ϵ_{ij}	b_{1i}	QR model	Bias	SE	CP(%)	Bias	SE	CP(%)
$N(0, 1)$	$N(0, 2)$	Before deletion	0.00	0.10	95.4	0.00	0.10	95.2
		Random effects	−0.47	0.11	0.8	−0.53	0.10	0.0
		Shared parameter	0.01	0.12	96.8	−0.04	0.12	94.0
	$t_{(3)}$	Before deletion	−0.01	0.17	93.8	−0.01	0.17	93.4
		Random effects	−0.70	0.15	0.4	−0.56	0.15	4.6
		Shared parameter	−0.06	0.16	94.2	−0.01	0.17	95.4
	Skewed	Before deletion	−0.01	0.16	94.8	−0.01	0.16	94.6
		Random effects	−0.86	0.16	0.2	−0.66	0.16	2.4
		Shared parameter	0.03	0.18	94.2	0.11	0.19	90.0
	Bimodal	Before deletion	0.00	0.18	95.4	0.00	0.18	95.4
		Random effects	−0.97	0.18	0.0	−0.72	0.19	2.0
		Shared parameter	0.03	0.21	95.0	0.12	0.21	92.4
$t_{(3)}$	$N(0, 2)$	Before deletion	0.00	0.11	94.6	0.00	0.10	94.8
		Random effects	−0.46	0.11	0.3	−0.53	0.11	0.2
		Shared parameter	0.02	0.13	95.0	−0.04	0.12	95.8
	$t_{(3)}$	Before deletion	0.00	0.17	95.0	0.00	0.17	94.6
		Random effects	−0.70	0.15	0.6	−0.56	0.15	2.4
		Shared parameter	−0.05	0.16	95.6	0.00	0.17	96.2
	Skewed	Before deletion	−0.01	0.16	94.4	0.00	0.16	94.2
		Random effects	−0.87	0.16	0.2	−0.67	0.17	0.8
		Shared parameter	0.04	0.18	93.4	0.12	0.19	90.0
	Bimodal	Before deletion	0.00	0.18	95.0	0.01	0.18	94.8
		Random effects	−0.97	0.19	0.0	−0.73	0.19	2.2
		Shared parameter	0.06	0.21	95.0	0.14	0.21	90.2
$\chi^2_{(3)}$	$N(0, 2)$	Before deletion	0.00	0.11	93.4	0.00	0.11	94.0
		Random effects	−0.54	0.11	0.0	−0.55	0.12	0.4
		Shared parameter	−0.03	0.13	94.2	−0.05	0.13	93.2
	$t_{(3)}$	Before deletion	−0.01	0.17	96.8	−0.01	0.17	97.2
		Random effects	−0.51	0.18	16.0	−0.58	0.17	8.0
		Shared parameter	0.08	0.19	95.6	0.01	0.18	96.6
	Skewed	Before deletion	0.00	0.18	94.0	−0.01	0.17	96.2
		Random effects	−0.64	0.21	16.2	−0.74	0.19	3.0
		Shared parameter	0.04	0.22	94.0	−0.05	0.19	96.0
	Bimodal	Before deletion	−0.01	0.17	96.2	−0.01	0.17	94.8
		Random effects	−0.74	0.19	3.0	−0.82	0.19	1.4
		Shared parameter	−0.05	0.19	96.0	−0.12	0.19	90.0

is substantially smaller than that of $\hat{\beta}_1$ because the dropout process only depends on the random slope b_{1i}. In other words, the observed data are a biased sample of the complete data with respect to β_1, but not β_0.

The simulation study shows that the estimates of the regression parameters based on the shared-parameter QR were robust to both the error and random-effects distributions. It is of interest to examine how the estimates of random effects are adjusted when their true distribution is not normal. Figure 2 depicts the posterior distribution of random slopes b_{1i} under various true distributions for $\tau = 0.25$ and 0.5. When the true distribution of b_{1i} is not normal, the posterior estimates of b_{1i} could be adaptively adjusted to the truth, providing a possible explanation for the robustness of estimating β_1 under various distributions for the random effects. The variance of b_{1i}, say λ_1, is the tuning parameter, and often treated as a nuisance parameter. We observe that the estimates of λ_1 based on the shared-parameter QR are generally much less biased than those based on the random-effects QR; see Table A1 in the Web Appendix.

4. Application

4.1 *Data Analysis*

We illustrate the proposed method with the pediatric AIDS data. Let t_{ij} be the jth measurement time for the ith subject, y_{ij} be the square root of the CD4 cell count measured at t_{ij}, and x_i be a binary treatment indicator with $x_i = 0$ denoting the high-dose arm. At the τth regression quantile, we considered

$$Q_{y_{ij}}(\tau \mid x_i, t_{ij}, b_{0i}, b_{1i}) = \beta_0 + \beta_1 x_i + \beta_2 t_{ij} + \beta_3 x_i t_{ij} + b_{0i} + b_{1i} t_{ij},$$

where the β's characterized the population-level trajectory, $b_{0i} \sim N(0, \lambda_0)$ and $b_{1i} \sim N(0, \lambda_1)$. The missing data model was given by

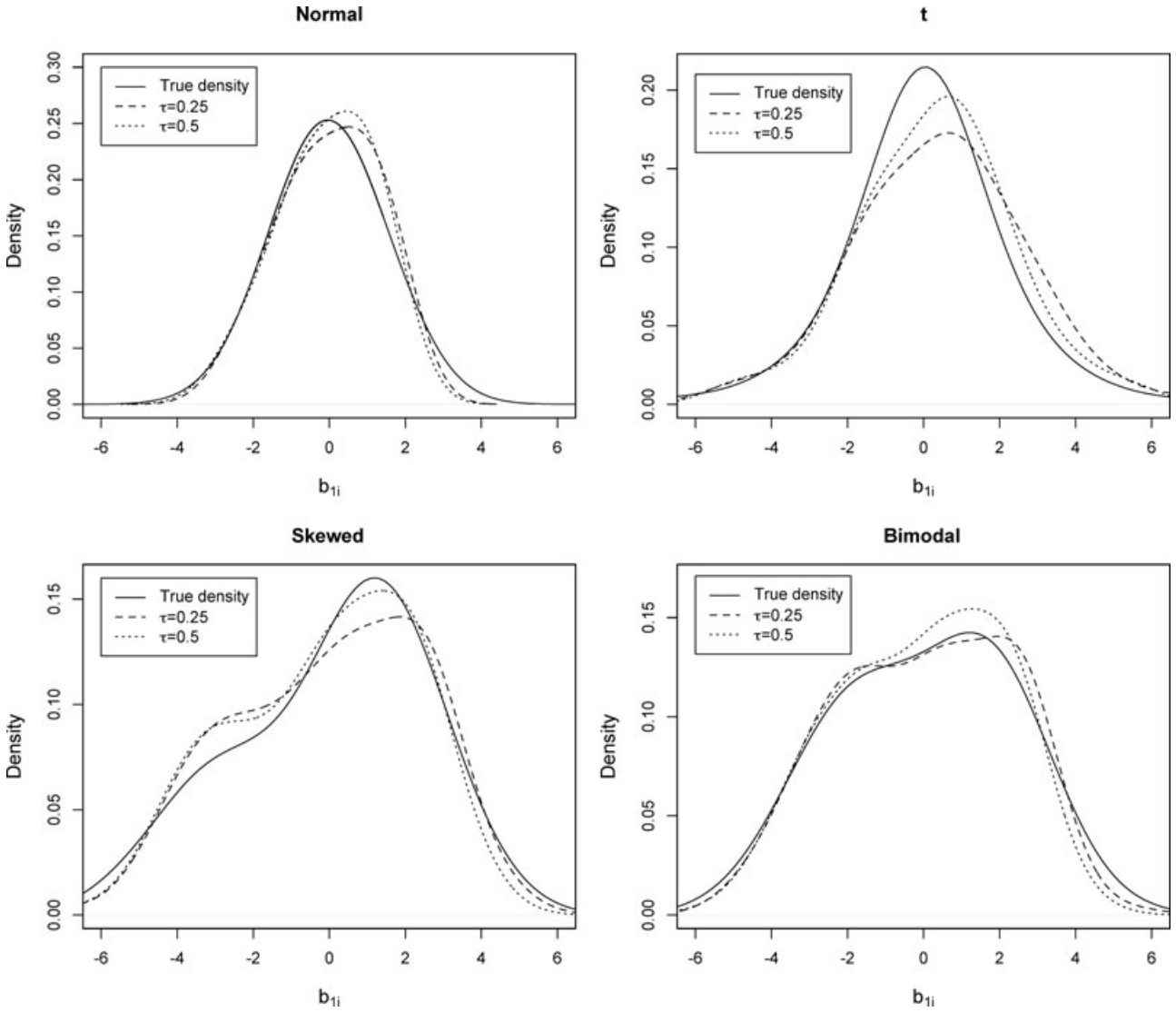

Figure 2. Posterior distributions of random effects b_{1i} under different true distributions for b_{1i}.

$$\pi_{ij}^{(\mathcal{O})} = \frac{1}{1+\sum_{k\in(\mathcal{I},\mathcal{D})}\exp\left(\gamma_0^{(k)}+\gamma_1^{(k)}b_{0i}+\gamma_2^{(k)}b_{1i}+\gamma_3^{(k)}x_i+\gamma_4^{(k)}x_ib_{0i}+\gamma_5^{(k)}x_ib_{1i}\right)}$$

$$\pi_{ij}^{(\mathcal{I})} = \frac{\exp\left(\gamma_0^{(\mathcal{I})}+\gamma_1^{(\mathcal{I})}b_{0i}+\gamma_2^{(\mathcal{I})}b_{1i}+\gamma_3^{(\mathcal{I})}x_i+\gamma_4^{(\mathcal{I})}x_ib_{0i}+\gamma_5^{(\mathcal{I})}x_ib_{1i}\right)}{1+\sum_{k\in(\mathcal{I},\mathcal{D})}\exp\left(\gamma_0^{(k)}+\gamma_1^{(k)}b_{0i}+\gamma_2^{(k)}b_{1i}+\gamma_3^{(k)}x_i+\gamma_4^{(k)}x_ib_{0i}+\gamma_5^{(k)}x_ib_{1i}\right)}$$

$$\pi_{ij}^{(\mathcal{D})} = \frac{\exp\left(\gamma_0^{(\mathcal{D})}+\gamma_1^{(\mathcal{D})}b_{0i}+\gamma_2^{(\mathcal{D})}b_{1i}+\gamma_3^{(\mathcal{D})}x_i+\gamma_4^{(\mathcal{D})}x_ib_{0i}+\gamma_5^{(\mathcal{D})}x_ib_{1i}\right)}{1+\sum_{k\in(\mathcal{I},\mathcal{D})}\exp\left(\gamma_0^{(k)}+\gamma_1^{(k)}b_{0i}+\gamma_2^{(k)}b_{1i}+\gamma_3^{(k)}x_i+\gamma_4^{(k)}x_ib_{0i}+\gamma_5^{(k)}x_ib_{1i}\right)}.$$

In the MCMC procedure, we recorded 10,000 draws after 1000 burn-in iterations. To assess the convergence of Markov chains, we calculated the Gelman–Rubin convergence statistic, the shrinkage factor, for the slopes based on three independent Markov chains with overly dispersed starting values. After 1000 burn-in iterations, the values of the shrinkage factors became very close to 1, suggesting the convergence of these chains.

Table 2 shows the estimates of the model parameters for regression quantiles of $\tau = 0.25,\ 0.5$, and 0.75, under the shared-parameter QR model and random-effects QR model, respectively. In both models, the estimates of β_1 are not significantly different from 0 at all three quantiles, indicating that the baseline CD4 count was well balanced between the two treatment arms by randomization.

Table 2
Estimates and 95% CIs of the model parameters for the pediatric AIDS data

		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
QR model		Est.	95% CI	Est.	95% CI	Est.	95% CI
Random effects	β_0	21.40	(20.40, 22.41)	23.62	(22.56, 24.68)	25.97	(24.86, 27.08)
	β_1	0.14	(−1.84, 2.11)	0.19	(−1.94, 2.31)	0.24	(−2.01, 2.47)
	β_2	−3.19	(−3.54, −2.86)	−3.54	(−3.89, −3.20)	−3.82	(−4.19, −3.46)
	β_3	0.78	(0.11, 1.45)	0.82	(0.13, 1.51)	0.83	(0.127, 1.55)
	λ_0	104.49	(90.94, 120.45)	117.98	(102.61, 135.76)	129.61	(112.56, 149.17)
	λ_1	9.49	(7.83, 11.38)	10.67	(8.86, 12.75)	11.09	(9.14, 13.36)
Shared parameter	β_0	21.09	(20.07, 22.08)	23.33	(22.24, 24.39)	25.58	(24.45, 26.70)
	β_1	0.17	(−1.86, 2.14)	0.21	(−1.92, 2.28)	0.27	(−1.98, 2.47)
	β_2	−3.40	(−3.76, −3.06)	−3.75	(−4.11, −3.40)	−4.09	(−4.47, −3.72)
	β_3	0.77	(0.06, 1.44)	0.81	(0.08, 1.52)	0.84	(0.11, 1.57)
	λ_0	105.87	(92.09, 121.59)	119.77	(104.40, 137.65)	132.02	(114.93, 151.51)
	λ_1	9.55	(7.92, 11.41)	10.95	(9.10, 13.12)	11.69	(9.68, 14.04)
	$\gamma_0^{(\mathcal{I})}$	−3.06	(−3.57, −2.59)	−3.00	(−3.49, −2.52)	−3.05	(−3.54, −2.59)
	$\gamma_1^{(\mathcal{I})}$	−0.03	(−0.08, 0.02)	−0.01	(−0.06, 0.04)	−0.03	(−0.07, 0.02)
	$\gamma_2^{(\mathcal{I})}$	0.02	(0.00, 0.03)	0.01	(0.00, 0.03)	0.01	(0.00, 0.03)
	$\gamma_3^{(\mathcal{I})}$	0.12	(−0.51, 0.79)	0.06	(−0.57, 0.70)	0.12	(−0.52, 0.77)
	$\gamma_4^{(\mathcal{I})}$	−0.03	(−0.11, 0.05)	−0.02	(−0.10, 0.05)	0.01	(−0.07, 0.08)
	$\gamma_5^{(\mathcal{I})}$	0.00	(−0.02, 0.02)	0.00	(−0.02, 0.02)	0.00	(−0.02, 0.03)
	$\gamma_0^{(\mathcal{D})}$	−2.05	(−2.53, −1.59)	−2.05	(−2.53, −1.58)	−2.07	(−2.59, −1.56)
	$\gamma_1^{(\mathcal{D})}$	−0.09	(−0.14, −0.03)	−0.09	(−0.14, −0.04)	−0.11	(−0.16, −0.05)
	$\gamma_2^{(\mathcal{D})}$	−0.07	(−0.09, −0.04)	−0.06	(−0.08, −0.04)	−0.06	(−0.08, −0.04)
	$\gamma_3^{(\mathcal{D})}$	−0.15	(−0.83, 0.54)	−0.25	(−0.99, 0.46)	−0.20	(−0.94, 0.56)
	$\gamma_4^{(\mathcal{D})}$	−0.06	(−0.15, 0.04)	−0.07	(−0.16, 0.02)	−0.06	(−0.15, 0.03)
	$\gamma_5^{(\mathcal{D})}$	0.00	(−0.03, 0.03)	0.00	(−0.03, 0.03)	0.00	(−0.03, 0.03)

The estimates of β_2 under the shared-parameter QR are smaller than those under the random-effects QR across all the three quantiles, as the shared-parameter QR takes into account the fact that early dropouts were associated with lower slopes (see Figure 1). For example, at the median, $\hat{\beta}_2$, i.e., the estimate of the slope corresponding to time in the high-dose arm, is −3.54 under the random-effects QR, and −3.75 under the shared-parameter QR. The random-effects QR did not adjust for the missing data, and thus led to an overestimated β_2. Under the shared-parameter QR, both the high-dose and low-dose regimens were more effective for sicker patients with lower CD4 counts, reflected by the slower decline of the CD4 count for these patients as $\hat{\beta}_2 = -3.40$ versus −4.09 for $\tau = 0.25$ and 0.75, respectively.

The estimate of β_3, corresponding to the difference in the slopes of time between the low-dose arm and high-dose arm, is similar between the two models. Across the three regression quantiles, the slope of time in the low-dose arm is significantly higher than that in the high-dose arm as $\hat{\beta}_3 > 0$, suggesting that the low dose of zidovudine was superior to the high dose because the decline in the CD4 cell count was less steep in the low-dose arm. However, the superiority of the low-dose arm is slightly less for lower quantiles, as $\hat{\beta}_3 = 0.77$ versus 0.84 for $\tau = 0.25$ and 0.75, respectively. QR allows us to examine the treatment effects at different quantiles of the CD4 count. In contrast, mean regression only models the average or central effects, and thus cannot detect such quantile differences. For comparison, we also implemented a shared-parameter mean regression model, which gave the estimates of β_1, β_2, and β_3 as 0.36, −3.77, and 0.94, with standard errors of 1.23, 0.50, and 0.41, respectively.

4.2 *Sensitivity Analysis*

In the shared-parameter QR model, $(\gamma_1^{(\mathcal{I})}, \gamma_2^{(\mathcal{I})}, \gamma_4^{(\mathcal{I})}, \gamma_5^{(\mathcal{I})})$ determine the intermittent missing data mechanism. For example, if $\gamma_1^{(\mathcal{I})} = \gamma_2^{(\mathcal{I})} = \gamma_4^{(\mathcal{I})} = \gamma_5^{(\mathcal{I})} = 0$, the intermittent missing data are ignorable because the missing data process is independent of the outcome process; otherwise the intermittent missing data are nonignorable. Similarly, $(\gamma_1^{(\mathcal{D})}, \gamma_2^{(\mathcal{D})}, \gamma_4^{(\mathcal{D})}, \gamma_5^{(\mathcal{D})})$ govern the missing mechanism associated with dropout. In Table 2, the 95% CIs of $\gamma_1^{(\mathcal{I})}, \gamma_2^{(\mathcal{I})}, \gamma_4^{(\mathcal{I})}$, and $\gamma_5^{(\mathcal{I})}$ all contain 0, suggesting that the intermittent missing data might be MAR in both treatment arms. However, the dropout process seems to be nonignorable, as the 95% CIs of $\gamma_1^{(\mathcal{D})}$ and $\gamma_2^{(\mathcal{D})}$ do not contain 0.

Unfortunately, the observed data often contain limited information to determine the missing data mechanism reliably. In the Bayesian paradigm, we conducted a sensitivity analysis by assigning a series of independent informative normal priors to each γ with a mean ranging from −10 to 10, but the same variance of 0.2. Figure 3 displays the estimates of β_2 and β_3 under these informative priors, indicating that our results were not particularly sensitive to the values of the γ's.

A key assumption of the shared-parameter model is the conditional independence between the outcome process and

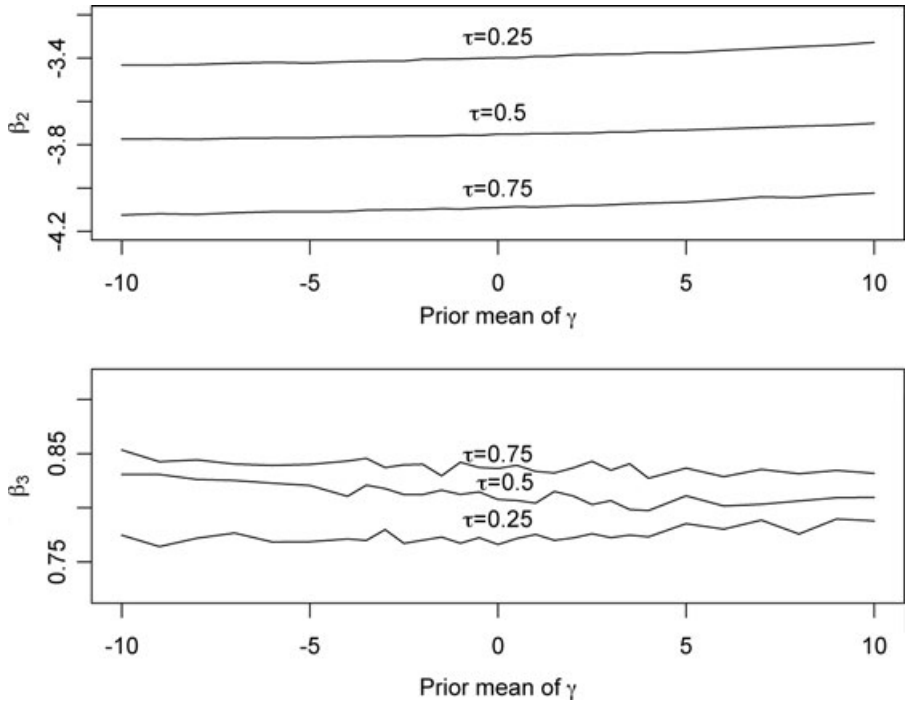

Figure 3. Sensitivity analysis by assigning a series of informative normal priors with different means to the γ's in the missing data model.

the missing data process given the random effects. To examine this assumption, Pulkstenis, Ten Have, and Landis (1998) proposed a sensitivity analysis that assesses the changes in the parameter estimates and standard errors when the observed outcome prior to the dropout is also included as a covariate in the dropout model. However, our case is more complicated, because we must consider both the intermittent missing data and dropout. At the jth measurement time, we use the last observed CD4 cell count before time j as the covariate if the intermittent missingness or dropout occurs; otherwise, the observed CD4 cell count at time j is used as the covariate. We denote the regression coefficients corresponding to this new covariate and its interaction with x_i by $\alpha_1^{(\mathcal{I})}$ and $\alpha_2^{(\mathcal{I})}$ for intermittent missingness, and $\alpha_1^{(\mathcal{D})}$ and $\alpha_2^{(\mathcal{D})}$ for dropout.

The parameter estimates with this additional augmented covariate are summarized in Table 3. In order to evaluate the validity of the missing data mechanism assumed in the shared-parameter QR, we compare the estimates of the $\gamma^{(k)}$'s ($k = \mathcal{I}, \mathcal{D}$) in Table 3 with their counterparts in Table 2. One of the most notable differences is that the estimate of $\gamma_2^{(\mathcal{I})}$ substantially increases, i.e., the dependency between the intermittent missing data and the random slopes is strengthened, when accounting for the dependency of data missingness on the last observed CD4 cell count. This provides empirical evidence of violation of the conditional independence assumption. However, the estimates for the parameters of interest, particularly the β's, are similar between Table 3 and Table 2, indicating robustness of our model to the conditional independence assumption.

5. Conclusion

We have studied QR for longitudinal data with nonignorable intermittent missing data and dropout. We used the ℓ_2 regularization to shrink the subject-specific regression lines toward the population line, thereby accounting for the within-subject correlations. We assumed that the missing data process is related to the longitudinal outcome process through sharing the common underlying random effects. By utilizing the relationship between the QR check function and the ALD, we cast the QR problem into the usual likelihood framework. We implemented a Bayesian MCMC approach, which naturally provides the posterior estimates of the model parameters and variances. Moreover, it automatically updates the tuning parameter for shrinkage in the Gibbs sampler. The simulation studies have demonstrated that our method can effectively remove the estimation bias caused by nonignorable intermittent missing data and dropout.

Our approach does not pose any distributional assumption on the outcome variable, and is thus more robust than conventional mean regression. However, we do make a model assumption on the dropout process. Because the nonignorable missing data mechanism cannot be directly verified based on the observed data (Molenberghs et al., 2008), some form of sensitivity analysis, such as the one described in Section 4, should be carried out.

6. Supplementary Materials

The Web Appendix referenced in Section 3 is available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Table 3

Sensitivity analysis of the pediatric AIDS data by augmenting the covariate with the last observed CD4 cell count in the model

	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
	Est.	95% CI	Est.	95% CI	Est.	95% CI
β_0	21.05	(20.04, 22.06)	23.28	(22.20, 24.35)	25.54	(24.42, 26.67)
β_1	0.20	(−1.81, −2.18)	0.25	(−1.89, 2.33)	0.29	(−1.94, 2.50)
β_2	−3.43	(−3.78, −3.09)	−3.78	(−4.14, −3.42)	−4.12	(−4.50, −3.74)
β_3	0.79	(0.09, 1.48)	0.83	(0.12, 1.55)	0.85	(0.09, 1.60)
λ_0	106.4	(92.6, 122.0)	120.2	(104.6, 137.8)	132.4	(115.3, 152.2)
λ_1	9.69	(8.06, 11.53)	11.10	(9.25, 13.25)	11.81	(9.79, 14.11)
$\gamma_0^{(\mathcal{I})}$	−2.99	(−3.49, −2.52)	−2.30	(−3.50, −2.50)	−3.06	(−3.59, −2.58)
$\gamma_1^{(\mathcal{I})}$	−0.03	(−0.08, 0.03)	−0.02	(−0.07, 0.03)	−0.02	(−0.07, 0.03)
$\gamma_2^{(\mathcal{I})}$	0.05	(0.03, 0.07)	0.05	(0.03, 0.07)	0.05	(0.03, 0.07)
$\gamma_3^{(\mathcal{I})}$	0.10	(−0.57, 0.78)	0.09	(−0.57, 0.73)	0.15	(−0.52, 0.84)
$\gamma_4^{(\mathcal{I})}$	−0.03	(−0.12, 0.05)	−0.02	(−0.10, 0.06)	0.00	(−0.08, 0.08)
$\gamma_5^{(\mathcal{I})}$	0.00	(−0.03, 0.03)	0.00	(−0.03, 0.03)	0.00	(−0.02, 0.03)
$\gamma_0^{(\mathcal{D})}$	−1.98	(−2.45, −1.53)	−2.01	(−2.49, −1.53)	−2.05	(−2.55, −1.55)
$\gamma_1^{(\mathcal{D})}$	−0.09	(−0.15, −0.04)	−0.10	(−0.15, −0.04)	−0.11	(−0.16, −0.06)
$\gamma_2^{(\mathcal{D})}$	−0.05	(−0.07, −0.02)	−0.04	(−0.07, −0.02)	−0.04	(−0.07, −0.02)
$\gamma_3^{(\mathcal{D})}$	−0.18	(−0.87, 0.52)	−0.25	(−0.95, 0.43)	−0.20	(−0.97, 0.53)
$\gamma_4^{(\mathcal{D})}$	−0.06	(−0.17, 0.04)	−0.08	(−0.17, 0.02)	−0.07	(−0.16, 0.02)
$\gamma_5^{(\mathcal{D})}$	0.00	(−0.03, 0.04)	0.00	(−0.04, 0.04)	0.00	(−0.03, 0.04)
$\alpha_1^{(\mathcal{I})}$	−0.04	(−0.05, −0.03)	−0.04	(−0.05, −0.03)	−0.04	(−0.05, −0.03)
$\alpha_2^{(\mathcal{I})}$	0.00	(−0.02, 0.02)	0.00	(−0.02, 0.02)	0.00	(−0.02, 0.02)
$\alpha_1^{(\mathcal{D})}$	−0.02	(−0.04, 0.00)	−0.03	(−0.04, −0.01)	−0.02	(−0.04, 0.00)
$\alpha_2^{(\mathcal{D})}$	0.00	(−0.03, 0.02)	0.00	(−0.03, 0.02)	−0.01	(−0.03, 0.02)

Acknowledgements

We thank the editor, associate editor, and two referees for insightful and constructive comments that substantially improved the article.

References

Brady, M. T., McGrath, N., Brouwers, P., Gelber, R., Fowler, M. G., Yogev, R., Hutton, N., Bryson, Y. J., Mitchell, C. D., Fikrig, S., Borkowsky, W., Jimenez, E., McSherry, G., Rubenstein, A., Wilfert, C. M., McIntosh, K., Elkins, M. M., Weintrub, P. S., and the Pediatric AIDS Clinical Trials Group. (1996). Randomized study of the tolerance and efficacy of high- versus low-dose zidovudine in human immunodeficiency virus-infected children with mild to moderate symptoms (ACTG 128). *Journal of Infectious Disease* **173,** 1097–1106.

Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine* **11,** 1305–1319.

De Gruttola, V. and Tu, X. M. (1994). Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* **50,** 1003–1014.

Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics* **43,** 49–73.

Follman, D. and Wu, M. C. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51,** 151–168.

Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* **8,** 140–154.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7,** 457–472.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). Adaptive rejection metropolis sampling. *Applied Statistics* **44,** 455–472.

Heagerty, P. J. and Pepe, M. S. (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. *Journal of the Royal Statistical Society, Series C* **48,** 533–551.

Hogan, J. W. and Laird, N. M. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* **16,** 259–272.

Hogan, J. W., Roy J., and Korkontzelou C. (2004). Biostatistics tutorial: Handling drop-out in longitudinal studies. *Statistics in Medicine* **23,** 1455–1497.

Jung, S. (1996). Quasi-likelihood for median regression models. *Journal of the American Statistical Association* **91,** 251–257.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis* **91,** 74–89.

Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46,** 33–50.

Koenker, R. and Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association* **94,** 1296–1310.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G., and Zhao, L. P. (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.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88,** 125–134.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90,** 1112–1121.

Little, R. J. A. (2008). Selection and pattern-mixture models. In *Advances in Longitudinal Data Analysis*, G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs (eds). London: CRC Press.

Molenberghs, G. and Kenward, M. G. (2007). *Missing Data in Clinical Studies*. New York: Wiley.

Molenberghs, G., Beunckens, C., Sotto C., and Kenward M. G. (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.

Pulkstenis, E., Ten Have, T. R., and Landis, J. R. (1998). Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association* **93,** 438–450.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika* **95,** 63–74.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

Ten Have, T. R., Kunselman, A., Pulkstenis, E., and Landis, J. R. (1998). Mixed effects logistic regression models for longitudinal binary response data with informative dropout. *Biometrics* **54,** 367–383.

Verbeke G. and Molenberghs G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

Welham S. J. (2008). Smoothing spline models for longitudinal data. In *Longitudinal Data Analysis*, G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs (eds). London: CRC Press.

Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics* **45,** 939–955.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44,** 175–188.

Yu, K. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters* **54,** 437–447.

Yu, K. and Stander, J. (2007). Bayesian analysis of a Tobit quantile regression model. *Journal of Econometrics* **137,** 260–276.

Received July 2008. *Revised February* 2009.
Accepted February 2009.