Published in final edited form as:

Biometrics. 2013 June ; 69(2): 375–385. doi:10.1111/biom.12010.

Quantile Regression for Recurrent Gap Time Data

Xianghua Luo^{1,*}, Chiung-Yu Huang², and Lan Wang³

¹Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55455, U.S.A.

²Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, U.S.A.

³School of Statistics, University of Minnesota, Minneapolis, Minnesota 55455, U.S.A.

Summary

Evaluating covariate effects on gap times between successive recurrent events is of interest in many medical and public health studies. While most existing methods for recurrent gap time analysis focus on modeling the hazard function of gap times, a direct interpretation of the covariate effects on the gap times is not available through these methods. In this article, we consider quantile regression that can provide direct assessment of covariate effects on the quantiles of the gap time distribution. Following the spirit of the weighted risk-set method by Luo and Huang (2011, *Statistics in Medicine* **30**, 301–311), we extend the martingale-based estimating equation method considered by Peng and Huang (2008, *Journal of the American Statistical Association* **103**, 637–649) for univariate survival data to analyze recurrent gap time data. The proposed estimation procedure can be easily implemented in existing software for univariate censored quantile regression. Uniform consistency and weak convergence of the proposed estimators are established. Monte Carlo studies demonstrate the effectiveness of the proposed method. An application to data from the Danish Psychiatric Central Register is presented to illustrate the methods developed in this article.

Keywords

Clustered survival data; Data perturbation; Gap times; Quantile regression; Recurrent events; Within-cluster resampling

1. Introduction

Recurrent event data frequently arise in medical and epidemiology studies, where each subject may experience a number of “failures” over the course of follow-up. Examples of recurrent failure events include tumor recurrences, recurrent infections, and repeated hospitalization. The times between successive recurrent events (gap times) are the natural

outcome of interest in these studies. Most existing methods for recurrent gap time analysis focus on modeling the hazard functions of gap times; for example, Huang and Chen (2003) studied proportional hazards (PH) models and Sun, Park, and Sun (2006) considered additive hazards models. Other researchers considered linear models for the covariate effects on transformed gap times. For example, Chang (2004) and Strawderman (2005) considered accelerated failure time (AFT) models for logarithm transformed gap times, and Lu (2005) studied semi-parametric linear transformation models by assuming linear covariate effects on an unspecified monotone increasing function of the gap times. However, these models only consider the effects of the covariates on the mean of the (transformed) gap time, thus important forms of heterogeneity might be overlooked. Furthermore, the associated theory generally requires that the data are homoskedastic in the sense that the error terms in the regression models are independent of the covariates. This assumption is often found to be too restrictive in practice.

In this article, we consider quantile regression for recurrent gap time data. Formulation of quantile regression is based on the conditional quantiles of the gap times given the covariates of interest. The covariate effects in quantile regression models are easier to interpret than those of hazard-based models. In addition, contrary to the PH model, quantile regression does not need the PH assumption and hence is more flexible. More importantly, quantile regression allows the covariates to have different effects at the tails or at different segments of the conditional distribution of the gap times and thus allows for heteroskedasticity in the data. Quantile regression provides a complete picture on the covariate effects on the gap time. This modeling framework is also robust for heavy-tailed gap times.

Quantile regression was originally proposed by Koenker and Bassett (1978) for analyzing uncensored continuous data. For univariate survival data, quantile regression was first studied by Powell (1984, 1986) for data with fixed censoring where the censoring times are known for all subjects. The random censoring case was studied by Ying, Jung, and Wei (1995) which assumed that the event time and the censoring time are unconditionally independent, and by Yang (1999) under a more relaxed conditional independence assumption but with independent and identically distributed (i.i.d.) error distributions. Portnoy (2003) considered a novel recursively reweighted estimator of the censored regression quantile process, which is a generalization of the Kaplan-Meier estimator. Recently, Peng and Huang (2008) proposed a martingale-based estimating equation approach that can be viewed as an extension of Nelson-Aalen type estimation and displays similar performance as Portnoy's estimator; Wang and Wang (2009) developed a locally weighted estimating procedure that relaxes the global linearity assumption.

In this article, we extend the martingale-based estimating equation approach by Peng and Huang (2008) for univariate survival data to analyze recurrent gap time data following the spirit of the weighted risk-set (WRS) method discussed by Luo and Huang (2011). The proposed estimation procedure can be easily implemented in existing software for univariate censored quantile regression. Applying the theory of empirical processes, we establish uniform consistency and weak convergence for the proposed estimator of the regression quantiles.

The rest of the article is organized as follows. In Section 2, we first present the estimation procedure for the first gap time only data by using Peng and Huang's method. We then discuss how this method can be extended to recurrent gap time data. We use the bootstrap resampling and data perturbation methods for estimating the variance of the proposed estimator. In Section 3, we report the results from simulation studies for both homoskedastic and heteroskedastic recurrent event data. In Section 4, we analyze the hospitalization data from the Danish Psychiatric Central Register. Some concluding remarks are provided in Section 5.

2. Method

2.1 Model Setup

We begin by introducing notation to describe the structure of the recurrent gap time data. Suppose n subjects are recruited into a study after experiencing an initial event. Let $i = 1, 2, \dots, n$ index subjects and $j = 0, 1, 2, \dots$ index the sequence of the recurrent events for a subject, where $j = 0$ indexes the initial event. For the i th subject, T_{ij} denotes the gap time between the $(j - 1)$ th and j th events and C_i denotes the time between enrollment and the end of follow-up. Let m_i be the index of the last censored gap time, so that m_i satisfies the

constraint $\sum_{j=1}^{m_i-1} T_{ij} \leq C_i$ and $\sum_{j=1}^{m_i} T_{ij} > C_i$, where $\sum_1^0 \equiv 0$ (see Figure 1 for an illustration). Define $\delta_i = I(m_i > 1)$, where $I(\cdot)$ is the indicator function. Thus $\delta_i = 0$ if the subject is free of events during follow-up, and $\delta_i = 1$ if the subject experienced recurrent events. As discussed in Wang and Chang (1999), the unique structure of recurrent events generates many difficulties in modeling gap time data. Firstly, the second and later gap times are subject to "induced" dependent censoring. Specifically, while the first gap time, T_{i1} may be subject to independent censoring C_i , the second and later gap times, T_{ij} ($j \geq 2$), are subject to dependent censoring $C_i - T_{i1} - \dots - T_{i,j-1}$, as gap times of the same subject are usually correlated. Second, due to intercept sampling, the last censored gap time, T_{imi} , tends to be longer than the uncensored gap times. Finally, the number of gap times is informative about the underlying distribution, as subjects at a higher risk of experiencing recurrent events are likely to have shorter, and hence more, gap times. Because of these reasons, naively treating recurrent gap time data as clustered survival data (e.g., family data) would not be appropriate.

Let $\mathbf{Z}_i$ be a $p \times 1$ vector of time-independent covariates, including the intercept. The τ th conditional quantile of T_{ij} given $\mathbf{Z}_i$ is defined as $Q_{ij}(\tau|\mathbf{Z}_i) = \inf_t \{\Pr(T_{ij} \leq t|\mathbf{Z}_i) \geq \tau\}$ for $\tau \in (0, 1)$. To model the effect of $\mathbf{Z}_i$ on the quantiles of T_{ij} , we impose the following model assumptions:

M1: There exists a nonnegative subject-specific frailty variable (or vector) γ_i such that, conditioning on γ_i and $\mathbf{Z}_i$, the individual recurrent event process is a renewal process.

M2: Given $\mathbf{Z}_i$, the conditional quantiles of T_{ij} satisfy the quantile regression model

$$Q_{ij}(\tau|\mathbf{Z}_i) = \exp \left\{ \mathbf{Z}_i^\top \beta(\tau) \right\} \text{ for } \tau \in (0, 1), \quad (1)$$

where $\beta(\tau)$ is a p -dimensional vector of quantile coefficients and may change with τ .

M3: Given $\mathbf{Z}_i$, the censoring time C_i is independent of γ_i and $(T_{i1}, T_{i2}, \dots)$.

Assumption M1 implies that, conditioning on γ_i and $\mathbf{Z}_i$, the underlying recurrent gap times $T_{i1}, T_{i2}, \dots$ are i.i.d. As a result, by integrating out γ_i , the gap times $T_{i1}, T_{i2}, \dots$ are identically (but not independently) distributed. The conditional i.i.d. assumption is a weaker condition than the i.i.d. assumption imposed by other competing approaches, including Strawderman (2005). We note that both the distribution of the subject-specific frailty γ_i and the dependency between γ_i and $(T_{i1}, T_{i2}, \dots)$ are left unspecified under the model assumptions. This contrasts with the conventional shared-frailty models for correlated survival data such as Huang and Liu (2007), where a parametric assumption for the frailty distribution is usually required for statistical inference.

We further note that Assumption M2 specifies the marginal distribution of the gap times unconditional on γ_i , hence the regression parameters in (1) have a marginal interpretation. Moreover, the proposed model reduces to the AFT model for recurrent gap times studied by Chang (2004) when the regression quantile coefficients corresponding to the nonintercept covariates are invariant with τ .

Define the observed gap time $X_{ij} = T_{ij}$ for $j < m_i$ and $X_{im_i} = C_i - \sum_{j=1}^{m_i-1} T_{ij}$. Denote the observed data for subject i by $\{\mathbf{Z}_i, X_{i1}, \dots, X_{im_i}\}$. The observed data from the n subjects are assumed to be i.i.d. For convenience, we define $m_i^* = \max\{m_i - 1, 1\}$. Hence, for people with no observed events ($m_i = 0$), $m_i^* = 1$; and for people with observed events ($m_i = 1$), m_i^* is the number of uncensored gap times. Figure 1 depicts a few examples.

2.2 WRS Estimators

To estimate $\beta(\tau)$, a simple approach would be applying the estimation procedure for quantile regression for univariate survival data to the subset of the data that are comprised of only the first gap times, $\{\mathbf{Z}_i, X_{i1}\}$, $i = 1, 2, \dots, n$. Specifically, we would apply the martingale-based estimation equation considered by Peng and Huang (2008) and estimate the quantile regression model by solving $U(\beta, \tau) = \mathbf{0}$, where

$$U(\beta, \tau) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i \left(N_{i1} \left[\exp \left\{ \mathbf{Z}_i^\top \beta(\tau) \right\} \right] - \int_0^\tau R_{i1} \left[\exp \left\{ \mathbf{Z}_i^\top \beta(u) \right\} \right] dH(u) \right), \quad (2)$$

with $N_{ij}(t) = I(X_{ij} \leq t, m_i = 1)$, $R_{ij}(t) = I(X_{ij} > t)$, and $H(x) = -\log(1 - x)$ for $0 \leq x < 1$. The time to first event analysis, however, is expected to be inefficient because the second and later gap times are not used in the formulation of (2).

For recurrent gap time data, Luo and Huang (2011) introduced two weighted stochastic processes as important building blocks for estimation procedure, namely, the averaged

counting process $N_i^*(t) = (m_i^*)^{-1} \sum_{j=1}^{m_i^*} N_{ij}(t)$ and the averaged at-risk process

$R_i^*(t) = (m_i^*)^{-1} \sum_{j=1}^{m_i^*} R_{ij}(t)$. Note that $N_i^*(t)$ and $R_i^*(t)$ have a jump size $1/m_i^*$ and $-1/m_i^*$,

respectively, at each uncensored gap time. This is different from the univariate survival analysis where the jump size of the counting processes $N_{i1}(t)$ and the at-risk processes $R_{i1}(t)$ is always 1 and -1 . In the same spirit as the WRS method for the PH regression discussed in Luo and Huang (2011), we replace N_{i1} and R_{i1} in $U(\beta, \tau)$ with their counterparts N_i^* and R_i^* to formulate a new estimating equation

$$U^*(\beta, \tau) = n^{-1} \sum_{i=1}^n \mathbf{Z}_i \left[N_i^* \left(\exp \left\{ \mathbf{Z}_i^\top \beta(\tau) \right\} \right) - \int_0^\tau R_i^* \left(\exp \left\{ \mathbf{Z}_i^\top \beta(u) \right\} \right) dH(u) \right], \quad (3)$$

and estimate $\beta(\tau)$ by solving the WRS estimating equation $U^*(\beta, \tau) = \mathbf{0}$. Note that under Assumptions M1 and M3, the observed uncensored gap times $X_{i1}, \dots, X_{im_i^*}$ are identically distributed conditional on $(\gamma_i; \mathbf{Z}_i, m_i, C_i)$. Hence, by double expectation we can show that $E[N_i^*(t)|\mathbf{Z}_i] = E[N_{i1}(t)|\mathbf{Z}_i]$ and $E[R_i^*(t)|\mathbf{Z}_i] = E[R_{i1}(t)|\mathbf{Z}_i]$. As a result, we have $E[U^*(\beta, \tau)] = E[U(\beta, \tau)]$, which ensures the consistency of the proposed estimator.

Next, we show how the proposed method can be implemented using existing software packages for univariate quantile regression. Note that the WRS estimating equation $U^*(\beta, \tau)$ can be reexpressed as

$$U^*(\beta, \tau) = n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i^*} \frac{\mathbf{Z}_i}{m_i^*} \left[\Delta_i I \left\{ \frac{\log X_{ij}}{m_i^*} \leq \frac{\mathbf{Z}_i^\top \beta(\tau)}{m_i^*} \right\} - \int_0^\tau I \left\{ \frac{\log X_{ij}}{m_i^*} \geq \frac{\mathbf{Z}_i^\top \beta(u)}{m_i^*} \right\} dH(u) \right].$$

This suggests that we can assign a weight $w_{ij} = 1/m_i^*$ to the first m_i^* observations of subject i to obtain working data $\{(\Delta_i, \log(X_{ij}^{(w)}), \mathbf{Z}_i^{(w)}); i=1, \dots, n, j=1, \dots, m_i^*\}$, where

$$\log(X_{ij}^{(w)}) = w_{ij} \log(X_{ij}) \text{ and } \mathbf{Z}_i^{(w)} = w_{ij} \mathbf{Z}_i.$$

Note that the censoring indicator δ_i remains unchanged for each subject and the last censored gap time is discarded for those subjects who have at least one complete gap time after reconstructing the working data set. We apply the martingale-based estimating equation method to the working data as if they were i.i.d. observations with sample size $\sum_{i=1}^n m_i^*$. This yields an estimating function

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i^*} \mathbf{Z}_i^{(w)} \left(N_{ij}^{(w)} \left[\exp \left\{ \mathbf{Z}_i^{(w)\top} \beta(\tau) \right\} \right] - \int_0^\tau R_{ij}^{(w)} \left[\exp \left\{ \mathbf{Z}_i^{(w)\top} \beta(u) \right\} \right] dH(u) \right), \quad (4)$$

where $N_{ij}^{(w)}(t) = I(X_{ij}^{(w)} \leq t, \Delta_i = 1)$ and $R_{ij}^{(w)}(t) = I(X_{ij}^{(w)} \geq t)$. By simple algebra, we have $N_{ij}^{(w)} \left\{ \exp(\mathbf{Z}_i^{(w)\top} \mathbf{b}) \right\} = N_{ij} \left\{ \exp(\mathbf{Z}_i^\top \mathbf{b}) \right\}$ and $R_{ij}^{(w)} \left\{ \exp(\mathbf{Z}_i^{(w)\top} \mathbf{b}) \right\} = R_{ij} \left\{ \exp(\mathbf{Z}_i^\top \mathbf{b}) \right\}$, and hence the working estimating equation (4) is equivalent to U^* in (3).

As in Peng and Huang (2008), we can use a grid-based estimation procedure to obtain an estimate of the true quantile coefficient $\beta_0(\cdot)$. Specifically, we define an estimator of $\beta_0(\tau)$, denoted by $\hat{\beta}^*(\tau)$, as a right-continuous piecewise-constant function that jumps only at τ_k , $k = 1, \dots, L$, where $0 < \tau_1 < \dots < \tau_L = \tau_U < 1$ and τ_U is a constant subject to certain identifiability constraints due to censoring. We can obtain $\hat{\beta}^*(\tau_k)$ by sequentially solving the following estimating equation, which approximates (3), for $\beta(\tau_k)$:

$$n^{-1/2} \sum_{i=1}^n \sum_{j=1}^{m_i^*} \mathbf{Z}_i^{(w)} \left[N_{ij}^{(w)} \left(\exp \left\{ \mathbf{Z}_i^{(w)\top} \beta(\tau_k) \right\} \right) - \sum_{l=0}^{k-1} R_{ij}^{(w)} \left(\exp \left\{ \mathbf{Z}_i^{(w)\top} \hat{\beta}^*(\tau_l) \right\} \right) \times \{H(\tau_{l+1}) - H(\tau_l)\} \right] = 0, k=1, \dots, L, \quad (5)$$

where $\exp \left\{ \mathbf{Z}_i^{(w)\top} \hat{\beta}^*(0) \right\}$ for $i = 1, \dots, n$ are set to be 0. Since equation (5) is not continuous, an exact root may not exist. Following Peng and Huang (2008), we define $\hat{\beta}^*(\tau_k)$ as the generalized solutions of (5), in which a slight perturbation of any components results in changing sign of the estimating function (Fygenson and Ritov, 1994). In fact, solving the estimating equation (5) for 0 is equivalent to locating the minimizer of the following L_1 -type convex objective function:

$$\begin{aligned} V_k(\mathbf{b}) = & \sum_{i=1}^n \sum_{j=1}^{m_i^*} \left| \Delta_i \log \left(X_{ij}^{(w)} \right) - \Delta_i \mathbf{b}^\top \mathbf{Z}_i^{(w)} \right| \\ & + \left| A - \mathbf{b}^\top \sum_{i=1}^n \sum_{j=1}^{m_i^*} \left(-\Delta_i \mathbf{Z}_i^{(w)} \right) \right| \\ & + \left| A - \mathbf{b}^\top \sum_{i=1}^n \sum_{j=1}^{m_i^*} \left(2 \mathbf{Z}_i^{(w)} \sum_{l=0}^{k-1} R_{ij}^{(w)} \left[\exp \left\{ \mathbf{Z}_i^{(w)\top} \hat{\beta}^*(\tau_l) \right\} \right] \times \{H(\tau_{l+1}) - H(\tau_l)\} \right) \right|, \end{aligned} \quad (6)$$

where A is a very large number. Alternatively, as argued in Peng and Huang (2008) and Koenker (2008), finding the solution to (5) can be formulated as a linear programming problem. Therefore, the estimation of $\beta_0(\tau)$ also can be obtained by using software for linear programming problems, such as the R package *quantreg* (Koenker, 2009).

We now establish the uniform consistency and weak convergence of the proposed estimator $\hat{\beta}^*(\tau)$. Let $S_L = \{0 = \tau_0 < \tau_1 < \dots < \tau_L = \tau_U < 1\}$ denote the grid in τ , and let $\|S_L\| = \max \{ \tau_k - \tau_{k-1}; k = 1, \dots, L \}$ denote the size of S_L , where S_L may depend on n . For a column vector $\mathbf{a}$, let $\|\mathbf{a}\|$ denote the Euclidean norm of $\mathbf{a}$. Under the regularity conditions (C1–C4) provided in Web Appendix A, we summarize the asymptotic properties of the proposed estimator in the following theorem.

Theorem 1. Assuming that conditions M1–M3 and C1–C4 hold, provided $\lim_{n \rightarrow \infty} n^{1/2} \|S_L\| = 0$, then $\sup_{\tau \in [\nu, \tau_U]} \|\hat{\beta}^*(\tau) - \beta_0(\tau)\| \rightarrow 0$ in probability and $n^{1/2} \{\hat{\beta}^*(\tau) - \beta_0(\tau)\}$ converges weakly to a Gaussian process for $\tau \in [\nu, \tau_U]$, where $0 < \nu < \tau_U$.

The proof of asymptotic properties follows closely the proof for the univariate survival data given in Peng and Huang (2008) and is provided in Web Appendices A and B. Note that identifiability is an inherent issue in the censored quantile regression. In principle, τ_U can be chosen according to the range of interest. In practice, τ_U may be obtained in an adaptive manner as the estimating equations in (5) are being solved sequentially. For example, when the minimization of the k th objective function $V_k(b)$ is not feasible, it may suggest that the identifiability condition is not satisfied for all $\tau \leq T_k$. Hence, τ_U may be set to some value below the first τ_k in the sequence $\{0 < \tau_1 < \tau_2 < \dots\}$ where the estimation of $\beta(\tau_k)$ is not successfully carried out.

2.3 Resampling Methods for Variance Estimation

Both data perturbation and bootstrap resampling methods for the variance estimation of the regression quantiles are available in the R package *quantreg* for univariate censored quantile regressions. We can extend both methods to estimate the variance of the proposed estimator $\hat{\beta}^*(\tau)$ for recurrent gap time data. To apply the data perturbation method proposed by Jin, Ying, and Wei (2001) to our recurrent gap time data, we first generate a simple random sample of size n , $(\zeta_1, \dots, \zeta_n)$ from a nonnegative distribution with unit mean and unit variance, for example, exponential (1). We then multiply ζ_i to the weighted data of subject i .

The resulting perturbed data are $\{(\zeta_i \log(X_{ij}^{(w)}), \zeta_i Z_i^{(w)}); i=1, \dots, n, j=1, \dots, m_i^*\}$. We then consider a perturbed objective function for each τ_k , $k = 1, \dots, L$,

$$\begin{aligned} \tilde{V}_k(b) = & \sum_{i=1}^n \sum_{j=1}^{m_i^*} \left| \Delta_i \zeta_i \log(X_{ij}^{(w)}) - \Delta_i \zeta_i b^\top Z_i^{(w)} \right| \\ & + \left| A - b^\top \sum_{i=1}^n \sum_{j=1}^{m_i^*} (-\Delta_i \zeta_i Z_i^{(w)}) \right| \\ & + \left| A - b^\top \sum_{i=1}^n \sum_{j=1}^{m_i^*} \left(2\zeta_i Z_i^{(w)} \sum_{l=0}^{k-1} I \left[\{X_{ij}^{(w)}\}^{\zeta_i} \geq \exp\{\zeta_i Z_i^{(w)\top} \tilde{\beta}^*(\tau_l)\} \right] \times \{H(\tau_{l+1}) - H(\tau_l)\} \right) \right|, \end{aligned} \quad (7)$$

where $\exp\{\zeta_i Z_i^{(w)\top} \tilde{\beta}^*(0)\}$, $i = 1, \dots, n$, are set to be 0. Denote the minimizers of (7) by $\tilde{\beta}^*(\tau)$. With the data fixed at the observed values, we repeatedly generate perturbed data a large number of times, say, B times and obtain B realizations, $(\tilde{\beta}_1^*(\tau), \dots, \tilde{\beta}_B^*(\tau))$. Then we can estimate the variance–covariance matrix for $\hat{\beta}^*(\tau)$ based on the sample statistics of $(\tilde{\beta}_1^*(\tau), \dots, \tilde{\beta}_B^*(\tau))$, based on which the pointwise confidence interval (CI) for $\beta(\tau)$ can be constructed.

To use the bootstrap resampling method, we randomly sample n subjects with replacement from the n subjects in the original data set B times. For each resampled data set, we minimize the target function in (6) with the resampled data to obtain a bootstrap estimate for $\hat{\beta}(\tau)$. The resulting B realizations of the estimates can then be used to obtain the bootstrap estimate of the variance–covariance matrix for $\hat{\beta}^*(\tau)$. Note that both resampling methods

take subjects as the perturbation or resampling unit, so that the correlation structure within a subject is not disturbed by the resampling procedures.

3. Simulations

In this section, we investigate the performance of the proposed methods via Monte Carlo studies, each with 1000 replicates and $n = 100$ within each replicate. We apply the proposed WRS method to the simulated recurrent gap time data, with two different variance estimation methods. For variance estimation, we set the data perturbation times or the bootstrap resampling times $B = 100$. For all the following examples, the censoring times C_i , $i = 1, \dots, n$ are generated from a uniform distribution on $(0, \pi)$, where π is chosen so that the proportion of subjects without any complete gap times is 25% or 40%.

We consider three different settings. In the first example, the regression quantile processes for covariates Z_1 is a constant; whereas in the second example, $\beta_{Z_1}(\tau)$ varies with τ . The third example differs from the second example in the functional form of $\beta_{Z_1}(\tau)$ and an additional covariate Z_2 .

Example 1. We generate recurrent event gap times from an AFT model

$$\log(T_{ij}) = b_0 + b_1 Z_{1i} + \gamma_i + \varepsilon_{ij}, i = 1, \dots, n, j = 1, 2, \dots,$$

where $b_0 = -1$, $b_1 = 1$, and $Z_{1i} \sim \text{uniform}(0, 1)$. The composite error term $\gamma_i + \varepsilon_{ij}$ is composed of the subject-specific random variable γ_i and the measurement error ε_{ij} , which are independent zero-mean normal variables with variance $\sigma^2 \in (0, 1)$ and $1 - \sigma^2$, respectively. Hence, the correlation between two observations $\log(T_{ij})$ and $\log(T_{ij'})$ from the same subject i is σ^2 if $j \neq j'$. We consider $\sigma^2 = 0.2, 0.4$, and 0.6 for different levels of within-subject correlation between gap times. The regression quantile for $(1, Z_1)^\top$ is given by $\beta(\tau) = (b_0 + Q_{\gamma + \varepsilon}(\tau), b_1)^\top$, where $Q_{\gamma + \varepsilon}(\tau)$, the quantile function of $\gamma_i + \varepsilon_{ij}$, is actually the probit function $\Phi(\tau)$ because $\gamma_i + \varepsilon_{ij}$ is a standard normal random variable. Note that in this example the regression quantile process for the covariate Z is a constant, $\beta_{Z_1}(\tau) = b_1$.

Example 2. For each subject i , $i = 1, \dots, n$, we generate recurrent event gap times T_{ij} based on a heteroskedastic model with the random errors dependent on the covariate,

$$\log(T_{ij}) = b_0 + b_1 Z_{1i} + (0.5 + Z_{1i})(\gamma_i + \varepsilon_{ij}), i = 1, \dots, n, j = 1, 2, \dots,$$

where $b_0 = -1$ and $b_1 = 1$. The covariate Z_{1i} and error terms γ_i and ε_{ij} are generated the same way as in Example 1. The regression quantile for $(1, Z_1)^\top$ is given by $\beta(\tau) = (b_0 + 0.5\Phi(\tau), b_1 + \Phi(\tau))^\top$. In this example, the regression quantile process for the covariate Z_1 varies with τ , $\beta_{Z_1}(\tau) = b_1 + \Phi(\tau)$. Note that, despite the heteroskedasticity in the error term $(0.5 + Z_i)(\gamma_i + \varepsilon_{ij})$, the correlation between $\log(T_{ij})$ and $\log(T_{ij'})$, $j \neq j'$, is still σ^2 as in Example 1.

Example 3. In this example, $\log(T_{ij})$ follows a multiple-covariate model

$$\log(T_{ij}) = b_0 + b_1 Z_{1i} + b_2 Z_{2i} + (0.5 + Z_{1i})(\gamma_i + \varepsilon_{ij}), i=1, \dots, n, j=1, 2, \dots,$$

where $b_0 = -1$, $b_1 = 0.5$, and $b_2 = 0.5$. The covariate $Z_{1i} \sim \text{uniform}(0, 1)$ and $Z_{2i} \sim \text{Bernoulli}(0.5)$. The frailty term γ_i is generated the same way as in the previous examples, while ε_{ij} is generated from an independent logistic distribution with zero mean and the scale parameter $s(s > 0)$. The scale parameter satisfies the condition $\sigma^2 + \pi^2 s^2/3 = 1$ so that the composite error term still has a unit variance, and hence the correlation between $\log(T_{ij})$ and $\log(T_{ij'}), j \neq j'$, is still σ^2 , as in the previous examples. The regression quantile process is $\beta(\tau) = (b_0 + 0.5Q_{\gamma+\varepsilon}(\tau), b_1 + Q_{\gamma+\varepsilon}(\tau), b_2)^\top$, where $Q_{\gamma+\varepsilon}(\tau)$ is the quantile function of the mixture of a normal and a logistic random variable, $\gamma_i + \varepsilon_{ij}$ and does not have a closed-form expression (Gupta and Nadarajah, 2008).

For comparison, we consider two naive approaches. The first approach is applying the univariate quantile regression model by Peng and Huang (2008) to the first gap time (censored or uncensored) only data (abbreviated as “first gap method”). The second approach is applying the AFT model by Chang (2004) to the recurrent gap time data by assuming a constant regression quantile process for each covariate. Note that, under the model assumption in Example 1, both the quantile regression models and the AFT model are valid methods to use. Tables 1–3 summarize the simulation results for Examples 1–3, respectively. For each scenario, we report the empirical biases (Bias), the empirical SD, the square root of the empirical mean of the estimated variances (SE), and the coverage rate (CR) of the 95% CI for the regression coefficient estimates, for $\tau = 0.2, 0.4$, and 0.6 .

The results in Tables 1–3 show that the point estimates of $\beta_z(\tau)$ from the proposed methods and the first gap method are approximately unbiased and that the CRs of the pointwise 95% CIs are close to the nominal level (0.95). Our method is clearly more efficient than the first gap method in all scenarios. In addition, identifiability is found to be less of an issue than in the first gap method in the sense that the range of τ satisfying the identifiability condition is wider for the former than the latter in the simulations (not shown). Tables 1–3 also show that the two variance estimation methods, data perturbation and bootstrap resampling methods, perform similarly well for the proposed WRS method. It is also shown that, for a fixed sample size, a smaller within-subject correlation results in a smaller variance of the estimated regression quantile of the covariate, as would be expected for correlated data. The AFT model estimates are approximately unbiased in Example 1 where the regression quantile process for the covariate is constant. In this case, the AFT model is more efficient than the proposed quantile regression method. For Example 2 with heteroskedastic data, the AFT model gives estimates with large biases. For Example 3 with two covariates, the AFT model estimate is biased for the coefficient of Z_1 in the conditional quantile function which changes with τ , while it is approximately unbiased for the coefficient of Z_2 which is independent of Z_1 and constant with respect to τ .

While it is natural to compare a quantile regression model to an AFT model due to their similar linear form in the conditional quantile function, it has been found less straight forward to compare a quantile regression model with the popular classical PH model, which

focuses on the hazard function. The PH model for recurrent gap times by Huang and Chen (2003) takes $\log \lambda$ to be linear in the covariates

$$\lambda(t|\mathbf{Z}_i) = \lambda_0(t) \exp(\mathbf{Z}_i^\top \boldsymbol{\theta}), \quad (8)$$

where $\boldsymbol{\theta}$ is the regression coefficient, $\lambda_0(t)$ is the baseline hazard function of the gap times for the subjects whose covariates are zero, and $\lambda(t|\mathbf{Z}_i)$ is the hazard function of T_{ij} given $\mathbf{Z}_i$,

$\lambda(t|\mathbf{Z}_i) = -\frac{d}{dt} \log S(t|\mathbf{Z}_i)$. Following the argument in Portnoy (2003), we can easily prove that under the PH model assumption in (8), the quantile for τ_{ij} , given $\mathbf{Z}_i = \mathbf{z}$, is

$$Q_{ph}(\tau|\mathbf{z}) = \Lambda_0^{-1}(-\log(1 - \tau) \exp(-\mathbf{z}^\top \boldsymbol{\theta})), \quad (9)$$

where $\Lambda_0(t) \equiv \int_0^t \lambda_0(u) du$ and $\Lambda_0^{-1}(\cdot)$ is the inverse function of $\Lambda_0(\cdot)$. Hence, under the PH model, the conditional quantile function is of a peculiar nonlinear form and it is not easy to make a direct comparison. However, Portnoy (2003) and Koenker and Geling (2001) suggested to use a local measure of the effect of the regression coefficient in the PH model on the conditional quantile at τ

$$b_{ph}(\tau; \boldsymbol{\theta}) = \frac{\partial}{\partial \mathbf{z}} Q_{ph}(\tau|\mathbf{z}) \Big|_{\mathbf{z}=\bar{\mathbf{z}}}$$

which is the derivative of the conditional quantile function in (9) with respect to the covariate $\mathbf{Z}$, evaluated at the average covariate value $\bar{\mathbf{z}}$. The PH regression coefficients, $\boldsymbol{\theta}$ can be estimated by $\hat{\boldsymbol{\theta}}$ using the method discussed in Huang and Chen (2003).

Using the heteroskedastic data generated in Example 2 ($\sigma^2 = 0.4$ and censoring rate = 25%), we compare the proposed regression quantile estimator, $\hat{\beta}^*(\tau)$ with the analogous PH model effect, $b_{ph}(\tau; \hat{\boldsymbol{\theta}})$. In Figure 2, the proposed estimator (left panel) and the PH model effect (right panel) are presented as dashed lines and compared with the true regression quantile coefficient $b_1 + \Phi(\tau)$ in solid lines. Clearly, the proposed estimator is virtually unbiased, whereas the PH model estimate differs significantly from the true regression quantile coefficient for a wide range of τ and also has a wider CI in the upper tail of τ .

4. An Example

To illustrate the proposed methods, we analyze a subset of the Danish Psychiatric Central Register data (Munk-Jørgensen and Mortensen, 1997), described in Luo and Huang (2011). The data set consists of 286 individuals who were first admitted to, or had contacts with Danish psychiatric services during the period from April 1, 1970 to December 31, 1970 with a diagnosis of schizophrenia. We set the maximum follow-up time for each person to be 3 years to avoid the potential change in the distributional pattern of recurrent gap times. Among the 286 individuals, 106 (37%) were females, 115 (40%) did not have another hospitalization after the first hospitalization or contact, 47 (20%) had only one

rehospitalization and the rest 115 (40%) had more than one rehospitalization. On average, each patient experienced 1.7 hospitalizations after the initial hospitalization or contact. Nine patients in this cohort died before the end of follow-up. Hence, the assumption of independent censoring is not expected to be seriously violated.

We are interested in assessing the effect of the onset age of schizophrenia on the length between two adjacent hospitalizations due to a schizophrenic episode. The onset age of this cohort ranges from 14 to 88, and the three quartiles are (21, 26, 39). A two-sample test comparing those who had schizophrenia onset as early as or before 20 years old (early onset) versus after 20 years old (late onset) and a PH regression analysis for the continuous onset age were reported in Luo and Huang (2011). Both analyses revealed that late onset ages of schizophrenia were significantly associated with longer gap times between hospitalizations. We apply the proposed WRS method to fit quantile regression models to evaluate the effect of age of onset (in years) on the quantiles of gap times, adjusting for gender. Figure 3 depicts the estimated regression quantile process for age of onset (left panel) and gender (right panel) based on different methods. Their associated 95% pointwise CIs are presented in Table 4. The estimated quantile process for onset age based on the WRS method is positive, ranging between 0.03 and 0.09, for τ between 0.1 and 0.6, suggesting that later onset ages are associated with longer gap times between two adjacent hospitalizations. We observe that the estimated quantile is virtually constant when τ ranges between 0.1 and 0.45, but after 0.45 it starts to increase steadily. The estimated regression quantile coefficient based on the first gap method (shown as dashed lines in Figure 3) indicates a similar effect of onset age as the WRS method, whereas the 95% pointwise CI is slightly wider than that of the WRS method. The estimate of the onset age effect based on the AFT model (Chang, 2004), depicted as the dot-dashed line in Figure 3, falls within the range of the maximum and minimum values of the estimated regression quantile coefficients of the WRS method or the first gap method, but it clearly differs from the estimated regression quantile coefficient for τ close to or larger than 0.5. The effect of gender, in the presence of onset age in the regression, is not significant in any of the models.

5. Remarks

Quantile regression has received much attention recently. Most existing censored quantile regression methods focus on univariate survival data. The limited work on multivariate survival data includes Yin and Cai (2005). They studied quantile regression for clustered survival data. In this article, we study quantile regression models for recurrent gap time data, which can serve as a useful alternative to hazard-based models and mean-based models. To estimate the regression quantile process for recurrent gap time data with the proposed method, the function *crq* in R package *quantreg* can be used directly on a properly constructed working data set. In the working data set, the last censored gap should be removed for those subjects where at least one complete gap time has been observed. A weight vector $((m_1^*)^{-1}1_{m_1^* \times 1}^\top, \dots, (m_n^*)^{-1}1_{m_n^* \times 1}^\top)^\top$, where 1 is a vector of one, should be provided to the R function *crq*. Modifications to the R functions *summary*, *crq* and *boot*, *crq* in the same R package are needed so that the perturbation or resampling schemes are performed on subjects rather than on individual gap times. The correlation structure of the

gap times within the same subject, hence, can be preserved when estimating the variance or CI of the regression quantile process.

The proposed WRS estimator for the regression quantile coefficients can be approximated by combining the modified within-cluster resampling (MWCR) scheme (Luo and Huang, 2011) with the univariate quantile regression by Peng and Huang (2008). Specifically, for each resampling, one gap time is randomly selected from the first m_i^* gap times for each person. Let $j(i) \in \{1, 2, \dots, m_i^*\}$ be the index for the selected gap time for subject i . The data $\{(X_{ijb(i)}, Z_i), i = 1, \dots, n\}$ from the b th resampling, $b = 1, \dots, M$, are i.i.d. Let $\tilde{\beta}_b(\tau)$ and $\tilde{\Sigma}_b(\tau)$, $\tau \in (0, 1)$ be the estimated quantile regression process and the corresponding variance-covariance process based on the b th resampled data by using Peng and Huang's method. The MWCR estimator for $\tilde{\beta}(\tau) = M^{-1} \sum_{b=1}^M \tilde{\beta}_b(\tau)$ and the variance of $\tilde{\beta}(\tau)$ can be estimated by $\tilde{\Sigma}(\tau) = \frac{1}{M} \sum_{b=1}^M \tilde{\Sigma}_b(\tau) - \frac{1}{M-1} \sum_{b=1}^M (\tilde{\beta}_b(\tau) - \tilde{\beta}(\tau))^{\otimes 2}$.

As suggested by a referee and also noted by Wang and Chang (1999), the gain in efficiency of the WRS-based methods relative to the first gap method may be diminished if the correlation among recurrent gap times is strong. More efficient estimators can be constructed by incorporating weight functions in the estimating equations, that is,

$$U^a(\beta, \tau) = n^{-1} \sum_{i=1}^n a_i Z_i \left[N_i^* \left(\exp \left\{ Z_i^\top \beta(\tau) \right\} \right) - \int_0^\tau R_i^* \left(\exp \left\{ Z_i^\top \beta(u) \right\} \right) dH(u) \right], \quad (10)$$

where a_i is a positive-valued and bounded weight function subject to the constraint $E(a_i^2) < \infty$. Alternatively, $U^a(\beta, \tau)$ can be expressed as

$$n^{-1} \sum_{i=1}^n Z_i \left[N_i^a \left(\exp \left\{ Z_i^\top \beta(\tau) \right\} \right) - \int_0^\tau R_i^a \left(\exp \left\{ Z_i^\top \beta(u) \right\} \right) dH(u) \right],$$

where $N_i^a(t) = (a_i/m_i^*) \sum_{j=1}^{m_i^*} N_{ij}(t)$ and $R_i^a(t) = (a_i/m_i^*) \sum_{j=1}^{m_i^*} R_{ij}(t)$. Similarly, $N_i^a(t)$ and $R_i^a(t)$ can be used in other WRS-based methods discussed in Luo and Huang (2011) to improve efficiency. By simple algebra, we can prove that $U^a(\beta, \tau)$ is an unbiased estimating function, which ensures the consistency of the estimator by solving $U^a(\beta, \tau) = 0$. Obviously, a convenient choice of $a_i = 1$ results in the proposed estimating function $U^*(\beta, \tau)$. Theoretically, the optimal weighted average can be achieved by using the weights proportional to the inverse of the variances of the summand,

$Z_i [N_i^* (\exp \{ Z_i^\top \beta(\tau) \}) - \int_0^\tau R_i^* (\exp \{ Z_i^\top \beta(u) \}) dH(u)]$ in (10). To this end, techniques in estimating the within-subject correlation structure among the gap times are needed and will be exploited in our future research.

The proposed model in (1) assumes that the linearity between the covariates and the conditional quantiles hold for all $\tau \in (0, 1)$. This global linearity assumption may not hold in

practice. For univariate survival data, Peng and Huang (2008) considered a class of supremum test statistics based on the large sample distribution of a functional of the martingale residual processes,

$M_{i1}(\tau; \beta) = N_{i1}(\exp\{\mathbf{Z}_i^\top \beta(\tau)\}) - \int_0^\tau R_{i1}(\exp\{\mathbf{Z}_i^\top \beta(u)\}) dH(u)$. Along the lines of Huang, Luo, and Follmann (2011), we can construct similar test statistics for the quantile regression model in (1) for recurrent gap time by replacing $M_{i1}(\tau; \beta)$ with

$M_i^*(\tau; \beta) = N_i^*(\exp\{\mathbf{Z}_i^\top \beta(\tau)\}) - \int_0^\tau R_i^*(\exp\{\mathbf{Z}_i^\top \beta(u)\}) dH(u)$ and construct test statistics based on $M_i^*(\tau; \beta)$. In the case when the global linearity assumption in (1) is violated, a naive approach would be to apply the locally weighted quantile regression by Wang and Wang (2009) on the first gap times. Further research in extending quantile regression for recurrent gap time which only requires local linearity is certainly warranted.

The validity of the proposed estimation procedure relies on the exchangeability property implied by the conditionally i.i.d. assumption (M1) imposed in this article and many others, including Huang and Chen (2003), Chang (2004), and Lu (2005). The conditional frailty model is useful in describing heterogeneity due to omitted covariates—thus gap times from the same subject are more similar than those from different subjects. However, we acknowledge that the exchangeable correlation structure can be inadequate when the correlation among the gap times within a subject is complex. The readers are referred to Cook and Lawless (2007) for discussions on the limitations of the exchangeable correlation structure as well as for an alternative parametric log-normal model for recurrent gap times to allow for flexible correlation structures.

The method discussed in this article is only applicable for studying the effects of covariates when different episodes of uncensored recurrent gap times within a subject have the same distribution. Hence, when this condition fails to hold, the method may yield biased inferential results. We suggest that before using the proposed quantile regression, one should examine if there exists any change in the distributional pattern of the recurrent gap times by running the nonparametric trend test considered by Wang and Chen (2000). It will be interesting to exploit models that allow episode-specific regression quantile coefficients (i.e., $\beta_j(\tau)$) or covariates (i.e., $\mathbf{Z}_{ij}$).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Associate Editor and reviewer for their valuable comments, Drs Preben Bo Mortensen and William Eaton for providing schizophrenia data, Mr Todd DeFor for editing manuscript drafts, and University of Minnesota Supercomputing Institute for providing computing resources for this research. This research was supported by University of Minnesota Grant-in-Aid 22281 to XL and by grant NSF DMS-1007603 to LW.

References

Chang SH. Estimating marginal effects in accelerated failure time models for serial sojourn times among repeated events. *Lifetime Data Analysis*. 2004; 10:175–190. [PubMed: 15293631]

- Cook, RJ.; Lawless, JF. The Statistical Analysis of Recurrent Events. New York: Springer; 2007.
- Fygenson M, Ritov Y. Monotone estimating equations for censored data. The Annals of Statistics. 1994; 22:732–746.
- Gupta AK, Nadarajah S. Normal and logistic random variables: Distribution of the linear combination. Statistical Papers. 2008; 49:201–209.
- Huang CY, Luo X, Follmann DA. A model checking method for the proportional hazards model with recurrent gap time data. Biostatistics. 2011; 12:535–547. [PubMed: 21138876]
- Huang X, Liu L. A joint frailty model for survival and gap times between recurrent events. Biometrics. 2007; 63:389–397. [PubMed: 17688491]
- Huang YJ, Chen YQ. Marginal regression of gaps between recurrent events. Lifetime Data Analysis. 2003; 9:293–303. [PubMed: 14649847]
- Jin Z, Ying Z, Wei LJ. A simple resampling method by perturbing the minimand. Biometrika. 2001; 88:381–390.
- Koenker R. Censored quantile regression redux. Journal of Statistical Software. 2008; 27:1–25.
- Koenker, R. [accessed May 12, 2009] *quantreg*: quantile regression, R package version 4.30. 2009. Available at: <http://www.r-project.org>
- Koenker R, Bassett G. Regression quantiles. Econometrica. 1978; 46:33–50.
- Koenker R, Geling O. Reappraising medfly longevity: A quantile regression survival analysis. Journal of the American Statistical Association. 2001; 96:458–468.
- Lu W. Marginal regression of multivariate event times based on linear transformation models. Lifetime Data Analysis. 2005; 11:389–404. [PubMed: 16133886]
- Luo X, Huang CY. Analysis of recurrent gap time data using the weighted risk set method and the modified within-cluster resampling method. Statistics in Medicine. 2011; 30:301–311. [PubMed: 20963733]
- Munk-Jørgensen P, Mortensen PB. The Danish Psychiatric Central Register. Danish Medical Bulletin. 1997; 44:82–84. [PubMed: 9062767]
- Peng L, Huang Y. Survival analysis with quantile regression models. Journal of the American Statistical Association. 2008; 103:637–649.
- Portnoy S. Censored regression quantiles. Journal of the American Statistical Association. 2003; 98:1001–1012.
- Powell J. Least absolute deviations estimation for the censored regression model. Journal of Econometrics. 1984; 25:303–325.
- Powell J. Censored regression quantiles. Journal of Econometrics. 1986; 32:143–155.
- Strawderman RL. The accelerated gap times model. Biometrika. 2005; 92:647–666.
- Sun LQ, Park DH, Sun JG. The additive hazards model for recurrent gap times. Statistica Sinica. 2006; 16:919–932.
- Wang HJ, Wang L. Locally weighted censored quantile regression. Journal of the American Statistical Association. 2009; 104:1117–1128.
- Wang MC, Chang SH. Nonparametric estimation of a recurrent survival function. Journal of the American Statistical Association. 1999; 94:146–153. [PubMed: 24244058]
- Wang MC, Chen YQ. Nonparametric and semiparametric trend analysis for stratified recurrence time data. Biometrics. 2000; 56:789–794. [PubMed: 10985217]
- Yang S. Censoring median regression using weighted empirical survival and hazard functions. Journal of the American Statistical Association. 1999; 94:137–145.
- Yin G, Cai J. Quantile regression models with multivariate failure time data. Biometrics. 2005; 61:151–161. [PubMed: 15737088]
- Ying Z, Jung SH, Wei LJ. Survival analysis with median regression models. Journal of the American Statistical Association. 1995; 90:178–184.

Examples:

Figure 1.
Illustration of recurrent gap time data.

Figure 2.

Comparison of the proposed quantile regression coefficient estimator (panel a) and the proportional hazards (PH) model effect (panel b) for simulated data. The solid line in both panels is the true regression quantile coefficient for $Z_1, b_1 + \Phi(\tau)$; the dashed lines are for the proposed estimator (in panel a) and the PH model effect (in panel b), respectively; and the shaded areas are their corresponding empirical 95% pointwise confidence intervals.

Figure 3.

The estimated regression quantile for onset age of schizophrenia and gender for $0.1 \leq \tau \leq 0.6$. The solid line is the estimated regression quantile based on the weighted risk-set method; the dashed line is the first gap method estimate; and the dot-dashed line is the accelerated failure time model estimate.

Table 1

Summary of the simulation results for Example 1

Proposed method																		
First gap method										AFT model				Perturb			Boot strap	
σ^2	m^a	τ	Bias ^b	SD ^c	SE ^d	CR ^e	Bias	SD	SE	CR	Bias	SD	SE	CR	SE	CR		
Censoring rate ^f = 25%																		
0.2	3.42	0.2	0.006	0.520	0.525	0.925	0.011	0.362	0.366	0.925	0.022	0.363	0.397	0.932	0.386	0.937		
		0.4	-0.004	0.467	0.481	0.942	0.011	0.362	0.366	0.925	0.004	0.371	0.379	0.936	0.381	0.935		
		0.6	0.004	0.479	0.512	0.937	0.011	0.362	0.366	0.925	0.007	0.406	0.431	0.933	0.426	0.937		
0.4	3.78	0.2	0.005	0.516	0.534	0.926	0.025	0.378	0.377	0.925	0.012	0.380	0.383	0.928	0.392	0.934		
		0.4	0.002	0.463	0.478	0.937	0.025	0.378	0.377	0.925	0.014	0.372	0.385	0.944	0.390	0.947		
		0.6	0.027	0.477	0.513	0.944	0.025	0.378	0.377	0.925	0.014	0.418	0.435	0.946	0.441	0.942		
0.6	4.28	0.2	0.001	0.501	0.528	0.937	0.018	0.394	0.384	0.923	0.015	0.403	0.396	0.923	0.402	0.931		
		0.4	-0.006	0.465	0.470	0.933	0.018	0.394	0.384	0.923	0.010	0.395	0.392	0.936	0.394	0.934		
		0.6	0.007	0.495	0.508	0.934	0.018	0.394	0.384	0.923	0.013	0.442	0.444	0.930	0.447	0.941		
Censoring rate = 40%																		
0.2	2.27	0.2	0.039	0.506	0.553	0.919	0.012	0.411	0.406	0.931	0.026	0.434	0.462	0.913	0.466	0.918		
		0.4	0.026	0.488	0.533	0.918	0.012	0.411	0.406	0.931	0.023	0.448	0.473	0.917	0.478	0.932		
		0.6	0.060	0.721	0.680	0.923	0.012	0.411	0.406	0.931	0.039	0.592	0.601	0.921	0.624	0.932		
0.4	2.44	0.2	0.038	0.518	0.560	0.914	0.012	0.402	0.411	0.929	0.023	0.437	0.460	0.913	0.468	0.928		
		0.4	0.010	0.485	0.525	0.922	0.012	0.402	0.411	0.929	0.009	0.438	0.463	0.921	0.472	0.927		
		0.6	0.025	0.566	0.649	0.919	0.012	0.402	0.411	0.929	0.024	0.535	0.580	0.913	0.593	0.920		
0.6	2.69	0.2	0.023	0.493	0.565	0.925	0.017	0.407	0.418	0.922	0.014	0.434	0.467	0.909	0.478	0.917		
		0.4	0.014	0.482	0.525	0.927	0.017	0.407	0.418	0.922	0.022	0.447	0.475	0.919	0.481	0.921		
		0.6	0.053	0.567	0.638	0.939	0.017	0.407	0.418	0.922	0.047	0.552	0.620	0.931	0.614	0.944		

^a Average number of censored and uncensored gap times per subject.

^b Empirical bias.

^c Empirical standard deviation.

^d Square root of the empirical mean of the estimated variances.

e Coverage rate of the nominal 95% confidence intervals.

f Proportion of subjects without any complete gap times.

Table 2

Summary of the simulation results for Example 8

σ^2		First gap method						AFT model						Proposed method																							
		m^{σ^2}			τ			Bias ^b			SD ^c			SE ^d			CR ^e			AFT model						Proposed method											
																				Bias						SD						SE					
		Censoring rate ^f = 25%																																			
0.2	3.52	0.2	-0.003	0.408	0.411	0.937	1.161	0.360	0.351	0.095	0.022	0.317	0.309	0.921	0.311	0.933																					
		0.4	0.024	0.369	0.378	0.933	0.573	0.360	0.351	0.592	0.026	0.311	0.306	0.925	0.308	0.931																					
		0.6	0.033	0.394	0.400	0.926	0.066	0.360	0.351	0.916	0.024	0.338	0.340	0.933	0.343	0.934																					
0.4	3.82	0.2	-0.006	0.410	0.406	0.920	1.147	0.349	0.360	0.112	0.017	0.314	0.306	0.923	0.310	0.926																					
		0.4	0.019	0.373	0.374	0.927	0.559	0.349	0.360	0.640	0.012	0.306	0.308	0.941	0.310	0.936																					
		0.6	0.024	0.374	0.402	0.937	0.052	0.349	0.360	0.929	0.017	0.329	0.352	0.943	0.355	0.945																					
0.6	4.29	0.2	-0.010	0.394	0.414	0.935	1.155	0.352	0.362	0.120	0.015	0.313	0.314	0.925	0.320	0.929																					
		0.4	-0.001	0.357	0.371	0.939	0.567	0.352	0.362	0.616	0.006	0.302	0.313	0.944	0.317	0.937																					
		0.6	0.011	0.375	0.397	0.939	0.060	0.352	0.362	0.926	0.014	0.345	0.354	0.941	0.354	0.937																					
		Censoring rate = 40%																																			
0.2	2.24	0.2	0.041	0.490	0.548	0.916	0.942	0.364	0.363	0.291	0.044	0.424	460	0.911	0.469	0.912																					
		0.4	0.030	0.487	0.521	0.925	0.354	0.364	0.363	0.801	0.056	0.461	0.477	0.905	0.486	0.919																					
		0.6	0.086	0.602	0.663	0.934	-0.153	0.364	0.363	0.902	0.081	0.569	0.595	0.907	0.623	0.928																					
0.4	2.40	0.2	0.036	0.508	0.550	0.925	0.933	0.363	0.369	0.297	0.049	0.441	0.453	0.920	0.467	0.921																					
		0.4	0.018	0.476	0.519	0.918	0.344	0.363	0.369	0.813	0.045	0.440	0.470	0.911	0.478	0.916																					
		0.6	0.071	0.576	0.675	0.919	-0.162	0.363	0.369	0.897	0.073	0.546	0.623	0.910	0.628	0.927																					
0.6	2.61	0.2	0.037	0.501	0.549	0.921	0.940	0.370	0.374	0.284	0.047	0.439	0.462	0.920	0.472	0.920																					
		0.4	0.023	0.479	0.519	0.915	0.351	0.370	0.374	0.802	0.045	0.445	0.481	0.919	0.484	0.921																					
		0.6	0.103	0.614	0.660	0.922	-0.155	0.370	0.374	0.914	0.095	0.565	0.620	0.928	0.640	0.929																					

^a Average number of censored and uncensored gap times per subject.

^b Empirical bias.

^c Empirical standard deviation.

^d Square root of the empirical mean of the estimated variances.

e Coverage rate of the nominal 95% confidence intervals.

f Proportion of subjects without any complete gap times.

Table 3

Summary of the simulation results for Example 3

Proposed method																			
First gap method							AFT model					Perturb				Bootstrap			
σ^2	m^{α}	τ	Bias ^b	SD ^c	SE ^d	CR ^e	Bias	SD	SE	CR	Bias	SD	SE	CR	SE	CR			
Censoring rate ^f = 25%							$\beta_1(\tau) = b_1 + Q_{\gamma,\varepsilon}(\tau)$												
0.2	3.44	0.2	0.031	0.488	0.520	0.913	1.111	0.349	0.352	0.116	0.049	0.366	0.382	0.927	0.397	0.939			
		0.4	0.028	0.409	0.465	0.939	0.566	0.349	0.352	0.648	0.046	0.331	0.367	0.931	0.378	0.942			
		0.6	0.042	0.425	0.484	0.940	0.080	0.349	0.352	0.931	0.051	0.380	0.408	0.932	0.418	0.941			
0.4	3.81	0.2	0.017	0.509	0.534	0.919	1.141	0.358	0.366	0.130	0.037	0.391	0.398	0.912	0.410	0.920			
		0.4	0.027	0.438	0.474	0.922	0.578	0.358	0.366	0.612	0.040	0.374	0.383	0.921	0.395	0.925			
		0.6	0.010	0.457	0.493	0.927	0.083	0.358	0.366	0.939	0.035	0.401	0.426	0.920	0.439	0.944			
0.6	4.33	0.2	0.052	0.498	0.527	0.911	1.177	0.376	0.370	0.122	0.059	0.398	0.408	0.914	0.419	0.926			
		0.4	0.061	0.450	0.488	0.917	0.605	0.376	0.370	0.607	0.069	0.385	0.401	0.931	0.411	0.928			
		0.6	0.051	0.458	0.502	0.927	0.107	0.376	0.370	0.916	0.068	0.414	0.440	0.924	0.448	0.926			
Censoring rate = 40%																			
0.2	2.26	0.2	0.036	0.521	0.546	0.922	0.926	0.370	0.369	0.322	0.046	0.439	0.459	0.920	0.470	0.915			
		0.4	0.045	0.444	0.492	0.926	0.381	0.370	0.369	0.798	0.054	0.408	0.440	0.938	0.447	0.931			
		0.6	0.057	0.518	0.570	0.921	-0.104	0.370	0.369	0.897	0.062	0.492	0.524	0.915	0.543	0.925			
0.4	2.43	0.2	0.025	0.515	0.549	0.916	0.932	0.368	0.379	0.330	0.032	0.430	0.456	0.915	0.474	0.924			
		0.4	0.016	0.470	0.505	0.925	0.370	0.368	0.379	0.815	0.039	0.434	0.446	0.914	0.456	0.933			
		0.6	0.041	0.544	0.587	0.926	-0.125	0.368	0.379	0.912	0.055	0.520	0.550	0.931	0.561	0.941			
0.6	2.68	0.2	0.046	0.513	0.550	0.917	0.098	0.371	0.386	0.297	0.063	0.446	0.454	0.915	0.469	0.921			
		0.4	0.045	0.465	0.516	0.937	0.405	0.371	0.386	0.795	0.063	0.429	0.463	0.924	0.475	0.928			
		0.6	0.084	0.547	0.595	0.936	-0.092	0.371	0.386	0.918	0.088	0.528	0.556	0.926	0.575	0.940			
$\beta_2(\tau) = b_2$																			
Censoring rate = 25%																			
0.2	3.44	0.2	0.006	0.266	0.287	0.922	-0.022	0.200	0.198	0.929	0.002	0.191	0.206	0.948	0.211	0.945			
		0.4	0.000	0.236	0.250	0.944	-0.022	0.200	0.198	0.929	-0.005	0.185	0.194	0.939	0.197	0.940			

Proposed method																			
First gap method						AFT model						Perturb						Bootstrap	
σ^2	m^a	τ	Bias ^b	SD ^c	SE ^d	CR ^e	Bias	SD	SE	CR	Bias	SD	SE	CR	SE	CR			
0.4	3.81	0.2	-0.010	0.242	0.261	0.940	-0.022	0.200	0.198	0.929	-0.003	0.200	0.214	0.936	0.216	0.943			
			-0.004	0.272	0.290	0.932	-0.014	0.202	0.204	0.939	0.007	0.202	0.211	0.945	0.219	0.947			
			-0.006	0.240	0.256	0.937	-0.014	0.202	0.204	0.939	-0.001	0.196	0.204	0.951	0.207	0.942			
0.6	4.33	0.2	-0.005	0.247	0.263	0.931	-0.014	0.202	0.204	0.939	0.000	0.214	0.226	0.937	0.228	0.944			
			-0.014	0.272	0.291	0.937	-0.028	0.209	0.208	0.920	-0.012	0.211	0.219	0.937	0.225	0.942			
			-0.009	0.248	0.262	0.925	-0.028	0.209	0.208	0.920	-0.009	0.208	0.214	0.935	0.216	0.938			
Censoring rate = 40%	0.2	2.26	-0.002	0.253	0.267	0.944	-0.028	0.209	0.208	0.920	-0.007	0.226	0.231	0.935	0.235	0.938			
			0.4	2.43	0.2	-0.014	0.271	0.293	0.934	-0.036	0.216	0.209	0.914	-0.006	0.229	0.245	0.943	0.252	0.955
						-0.008	0.247	0.267	0.936	-0.036	0.216	0.209	0.914	-0.006	0.222	0.231	0.936	0.237	0.934
0.6	2.68	0.2				0.002	0.269	0.280	0.918	-0.036	0.216	0.209	0.914	-0.001	0.256	0.257	0.922	0.262	0.923
			-0.007	0.274	0.299	0.935	-0.020	0.206	0.215	0.936	0.007	0.231	0.245	0.943	0.250	0.947			
			-0.001	0.250	0.274	0.952	-0.020	0.206	0.215	0.936	0.003	0.227	0.242	0.938	0.246	0.950			
0.6	2.68	0.2	0.005	0.258	0.284	0.923	-0.020	0.206	0.215	0.936	0.005	0.253	0.265	0.922	0.273	0.940			
			-0.004	0.273	0.301	0.935	-0.027	0.212	0.220	0.932	-0.001	0.235	0.247	0.934	0.253	0.939			
			0.4	0.002	0.248	0.273	0.941	-0.027	0.212	0.220	0.932	0.004	0.230	0.247	0.941	0.251	0.945		
0.6	2.68	0.2	0.002	0.268	0.296	0.938	-0.027	0.212	0.220	0.932	0.000	0.258	0.277	0.945	0.280	0.943			

^a Average number of censored and uncensored gap times per subject.

^b Empirical bias.

^c Empirical standard deviation.

^d Square root of the empirical mean of the estimated variances.

^e Coverage rate of the nominal 95% confidence intervals.

^f Proportion of subjects without any complete gap times.

Table 4
Summary of estimated regression quantile for schizophrenia data

τ	First gap method			AFT model			Proposed method		
	$\beta \hat{\sigma}^a$	s.e. ^b	95% CI ^c	$\beta \hat{\sigma}$	s.e.	95% CI	$\beta \hat{\sigma}$	s.e.	95% CI
Age of onset									
0.1	0.040	0.011	(0.019, 0.062)	0.042	0.008	(0.027, 0.057)	0.035	0.010	(0.016, 0.053)
0.2	0.033	0.012	(0.009, 0.057)	0.042	0.008	(0.027, 0.057)	0.030	0.008	(0.015, 0.045)
0.3	0.032	0.013	(0.007, 0.060)	0.042	0.008	(0.027, 0.057)	0.032	0.012	(0.009, 0.055)
0.4	0.042	0.009	(0.023, 0.060)	0.042	0.008	(0.027, 0.057)	0.036	0.009	(0.018, 0.053)
0.5	0.105	0.044	(0.019, 0.191)	0.042	0.008	(0.027, 0.057)	0.075	0.036	(0.005, 0.145)
0.6	0.117	0.038	(0.042, 0.192)	0.042	0.008	0.027, 0.057	0.093	0.030	(0.034, 0.153)
Gender (female)									
0.1	-0.211	0.539	(-1.267, 0.845)	0.070	0.234	(-0.455, 0.469)	-0.059	0.346	(-0.737, 0.619)
0.2	-0.163	0.360	(-0.868, 0.541)	0.070	0.234	(-0.455, 0.469)	0.054	0.349	(-0.630, 0.739)
0.3	0.116	0.288	(-0.449, 0.681)	0.070	0.234	(-0.455, 0.469)	0.192	0.251	(-0.300, 0.683)
0.4	-0.042	0.255	(-0.541, 0.457)	0.070	0.234	(-0.455, 0.469)	0.148	0.175	(-0.195, 0.490)
0.5	0.150	0.413	(-0.660, 0.960)	0.070	0.234	(-0.455, 0.469)	0.128	0.394	(-0.645, 0.900)
0.6	0.173	0.352	(-0.517, 0.864)	0.070	0.234	(-0.455, 0.469)	0.332	0.237	(-0.133, 0.797)

^aPoint estimate based on different methods.

^bStandard error (the data perturbation method was used for the first gap method and the proposed method.

^c95% pointwise confidence interval.