Biometrics 73, 94–103
March 2017

DOI: 10.1111/biom.12574

Multiple Imputation for Cure Rate Quantile Regression with Censored Data

Yuanshan Wu[1,*] and Guosheng Yin[2,]**

[1]School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China
[2]Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong
[*]*email:* shan@whu.edu.cn
[**]*email:* gyin@hku.hk

Summary. The main challenge in the context of cure rate analysis is that one never knows whether censored subjects are cured or uncured, or whether they are susceptible or insusceptible to the event of interest. Considering the susceptible indicator as missing data, we propose a multiple imputation approach to cure rate quantile regression for censored data with a survival fraction. We develop an iterative algorithm to estimate the conditionally uncured probability for each subject. By utilizing this estimated probability and Bernoulli sample imputation, we can classify each subject as cured or uncured, and then employ the locally weighted method to estimate the quantile regression coefficients with only the uncured subjects. Repeating the imputation procedure multiple times and taking an average over the resultant estimators, we obtain consistent estimators for the quantile regression coefficients. Our approach relaxes the usual global linearity assumption, so that we can apply quantile regression to any particular quantile of interest. We establish asymptotic properties for the proposed estimators, including both consistency and asymptotic normality. We conduct simulation studies to assess the finite-sample performance of the proposed multiple imputation method and apply it to a lung cancer study as an illustration.

Key words: Censored data; Censored quantile regression; Cure rate model; Missing data; Multiple imputation; Survival fraction.

1. Introduction

It is commonly observed in oncology studies that a substantial proportion of subjects are either cured following treatment or are insusceptible to the event of interest, and thus they are risk-free of disease relapse or cancer recurrence. To explicitly incorporate the survival fraction for such data, cure rate models have been studied extensively, among which a widely used approach is the two-component mixture cure rate model, with the assumption that the underlying population is a mixture of susceptible and insusceptible subjects. For those susceptible subjects who would eventually experience the event of interest if the follow-up is sufficiently long, one can apply the usual survival models. However, when an observation is censored, we do not know whether the subject is susceptible or insusceptible to the event. Typically, the logistic regression is used to model the susceptibility indicator. In the parametric modeling framework, Berkson and Gage (1952) proposed the exponential–logistic mixture model, and Farewell (1982, 1986) considered the Weibull-logistic mixture model for censored data with a survival fraction. For the semiparametric models, Kuk and Chen (1992) proposed to use the Cox proportional hazards (PH) model (Cox, 1972) for the survival times of susceptible subjects and a logistic regression model for the susceptibility indicators. Extensive research has been conducted on other semiparametric cure rate models. For example, Lu and Ying (2004) and Mao and Wang (2010) proposed transformation cure rate models based on transformed linear regression. Zeng, Yin, and Ibrahim (2006), on the other hand, developed nonparametric maximum likelihood estimators for a general class of transformation cure rate models that include proportional hazards and proportional odds structures. Zhang and Peng (2009) studied an accelerated hazards cure rate model, and Lu (2010) further developed an accelerated failure time (AFT) model under the mixture cure rate structure through kernel-smoothed nonparametric maximum likelihood estimation. In the Bayesian paradigm, Yin (2005) incorporated frailty to cure rate modeling to account for correlations due to clustering or grouping, and Yin and Ibrahim (2005) proposed a unified class of cure rate models based on the Box–Cox type transformations, which can accommodate both mixture and non-mixture cure rate frameworks.

Censored quantile regression (CQR) has become a valuable complement to the traditional hazards-based survival models, as it allows for modeling heterogeneity and offers more complete assessment based on different quantiles of the survival data. In the field of econometrics, Powell (1984, 1986) studied CQR with fixed censoring, where the censoring times are known for all subjects. In survival analysis with random censoring, CQR has gained much popularity. Under the independence assumption between the survival and censoring times, Ying, Jung, and Wei (1995) proposed a semiparametric estimation procedure for CQR. Nevertheless, a more realistic assumption is the conditional independence of survival and censoring times given the covariates. Yang (1999) proposed a median regression model based on the weighted empirical survival function, which, however, requires homogeneous errors

or that the error distributions converge to a common distribution at a certain rate. Portnoy (2003) developed CQR based on the scheme of redistributing the censored data to the right. By utilizing the martingale structure of right-censored survival data, Peng and Huang (2008) proposed a martingale-based estimating equation for CQR. Both methods of Portnoy (2003) and Peng and Huang (2008) rely on a global linearity assumption; that is, in order to estimate the τth conditional quantile of the survival times, it requires estimation of all the conditional quantiles lower than τ in advance by assuming the same linear CQR structure. To relax such a global linearity assumption, Wang and Wang (2009) proposed a locally weighted quantile regression approach by adopting the weighting scheme of redistribution-of-mass to the right, which is applicable for any particular quantile of interest.

Due to the existence of a survival fraction, the underlying population could be heterogeneous and heavily right-skewed. As a result, CQR is more attractive than the traditional mean- or hazards-based mixture cure rate regression models. For survival data with a cure possibility, Wu and Yin (2013) proposed cure rate quantile regression, which, however, also relies on the global linearity assumption. Motivated by the fact that one never knows whether a censored subject is cured or not, we cast the cure rate indicator in a missing data framework and propose a new estimation procedure based on multiple imputation. To relax the global linearity assumption, we apply the locally weighted CQR of Wang and Wang (2009) to model any particular quantile of the survival times of the susceptible subjects after filling in the cure indicator. Specifically, we first consider an iterative method to estimate the coefficients in the logistic regression. We then evaluate the conditional uncured probability for each censored observation, via which we can utilize Bernoulli sampling to mark each subject either cured or uncured. Based on those subjects labeled with uncured, we employ the locally weighted CQR method (Wang and Wang, 2009) to estimate the coefficients in CQR. Repeating the imputation procedure for multiple times and taking an average over the resultant estimators, we can obtain the estimators for the quantile regression coefficients.

The rest of this article is organized as follows. In Section 2, we propose the multiple imputation method for cure rate quantile regression. We establish the asymptotic properties of the proposed method in Section 3, and conduct simulation studies to evaluate its finite-sample performance in Section 4. In Section 5, we illustrate our method with application to a real data example. Some remarks are concluded in Section 6, and technical details are provided in the supplementary material.

2. CQR with Multiple Imputation

Let T ($T < \infty$) denote the possibly transformed failure time of a susceptible (uncured) subject in the population under a known monotone transformation, for example, the log transformation. In the mixture cure rate model framework, we can decompose the failure time of a randomly selected subject from the entire population (including both susceptible and insusceptible subjects) as

$$Y = \eta T + (1-\eta)\infty,$$

where the indicator η takes a value of 1 if a subject is susceptible to the event of interest, and 0 otherwise. Let C be the censoring time, under the same transformation as T, let $\mathbf{Z}$ be a $(p+1)$-vector of covariates related to T, and let $\mathbf{W}$ be a $(q+1)$-vector of covariates associated with η. Both $\mathbf{Z}$ and $\mathbf{W}$ include 1 as an intercept, and they may share common components. The observed time is $X = Y \wedge C$, the minimum value of Y and C, and let $\Delta = I(Y \le C)$ be the censoring indicator. For $i = 1, \ldots, n$, $(X_i, \Delta_i, \mathbf{Z}_i, \mathbf{W}_i)$ are independent and identically distributed copies of $(X, \Delta, \mathbf{Z}, \mathbf{W})$, and (T_i, η_i) and C_i are assumed to be conditionally independent given covariates $\mathbf{Z}_i$ and $\mathbf{W}_i$.

Let $F_T(t|\mathbf{z}) = P(T \le t|\mathbf{z})$ be the conditional distribution function of T given $\mathbf{Z} = \mathbf{z}$. For any given $\tau \in (0,1)$, let $Q_T(\tau|\mathbf{Z}) = \inf\{t\colon F_T(t|\mathbf{Z}) \ge \tau\}$ denote the τth conditional quantile function given covariate $\mathbf{Z}$, and the quantile regression model is given by

$$Q_T(\tau|\mathbf{Z}) = \mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}(\tau), \tag{1}$$

where $\boldsymbol{\beta}(\tau)$ is an unknown $(p+1)$-vector of regression coefficients. For ease of exposition, the quantile level τ is omitted hereafter in $\boldsymbol{\beta}(\tau)$, while it is understood that $\boldsymbol{\beta}$ is quantile-specific and we do not assume the usual global linearity across all τ's in $(0,1)$. Based on the logistic regression (Farewell, 1982), we can model the susceptibility indicator η as

$$P(\eta = 1|\mathbf{W}) = \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}) = \frac{\exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})}{1+\exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})}, \tag{2}$$

where $\boldsymbol{\gamma}$ is an unknown $(q+1)$-vector of regression coefficients.

Define the counting process $N_i(t) = \Delta_i I(X_i \le t)$, and the cumulative hazard function is given by

$$\begin{aligned}\Lambda_{Y,\boldsymbol{\gamma}}(t|\mathbf{Z}_i, \mathbf{W}_i) &= -\log\{1 - P(Y_i \le t|\mathbf{Z}_i, \mathbf{W}_i)\}\\ &= -\log\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_T(t|\mathbf{Z}_i)\}.\end{aligned}$$

Clearly, $M_i(t; \boldsymbol{\gamma}, F_T) = N_i(t) - \Lambda_{Y,\boldsymbol{\gamma}}(t \wedge X_i|\mathbf{Z}_i, \mathbf{W}_i)$ is a martingale (Fleming and Harrington, 1991). To estimate the cure rate parameter $\boldsymbol{\gamma}$, we note two facts in the observed data: (1) If a patient experiences the event of interest, this subject must belong to the uncured group; (2) If a patient's survival time is censored at X, the conditional probability that this subject belongs to the uncured group is $\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\{1 - F_T(X|\mathbf{Z})\}/\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_T(X|\mathbf{Z})\}$. As a result, the probability of a susceptible indicator is given by

$$P(\eta = 1|\Delta, X, \mathbf{Z}, \mathbf{W}) = \Delta + (1-\Delta)\frac{\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\{1 - F_T(X|\mathbf{Z})\}}{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_T(X|\mathbf{Z})}.$$

Based on (2), we have

$$E\left[\mathbf{W}\left\{P(\eta = 1|\Delta, X, \mathbf{Z}, \mathbf{W}) - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\right\}\right] = \mathbf{0},$$

which leads to an estimating equation for $\boldsymbol{\gamma}$,

$$\mathbf{S}_n(\boldsymbol{\gamma}, F_T) = n^{-1}\sum_{i=1}^{n} \frac{\mathbf{W}_i\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_T(X_i|\mathbf{Z}_i)} \times \left\{\Delta_i - \pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_T(X_i|\mathbf{Z}_i)\right\} = \mathbf{0}.$$

This can be rewritten in the martingale form,

$$\mathbf{S}_n(\boldsymbol{\gamma}, F_T) = n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{L} \frac{\mathbf{W}_i\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_i)F_T(t|\mathbf{Z}_i)}\mathrm{d}M_i(t;\boldsymbol{\gamma}, F_T) = \mathbf{0}, \tag{3}$$

where L denotes the transformed study end time.

Obviously, solving (3) would yield a consistent estimator for $\boldsymbol{\gamma}$ if the function $F_T(t|\mathbf{z})$ were known. Following the locally weighted Kaplan–Meier estimator in Wang and Wang (2009), we take a local Nelson–Aalen type estimator for the cumulative hazard function $\Lambda_T(t|\mathbf{z})$ in the context of cure rate analysis (Sy and Taylor, 2000; and Lu, 2010). For simplicity, suppose that p components of $\mathbf{Z}$ (excluding the first component 1) are continuous. We adopt a multivariate product kernel $K_p(\mathbf{u}) = \prod_{i=1}^{p} K(u_i)$, $\mathbf{u} \in \mathcal{R}^p$, where $K(\cdot)$ is a univariate kernel function. Let $h_n > 0$ be a bandwidth, and $h_n \to 0$ as $n \to \infty$. We first define a sequence of Nadaraya–Watson type weights,

$$B_{ni}(\mathbf{z}) = \frac{K_p\{(\mathbf{z}-\mathbf{Z}_i)/h_n\}}{\sum_{k=1}^{n} K_p\{(\mathbf{z}-\mathbf{Z}_k)/h_n\}},$$

and then we can estimate $\Lambda_T(t|\mathbf{z})$ via a local Nelson–Aalen type estimator,

$$\widehat{\Lambda}_T(t|\mathbf{z}) = \int_{-\infty}^{t} \frac{\sum_{i=1}^{n} B_{ni}(\mathbf{z})\mathrm{d}N_i(u)}{\sum_{k=1}^{n} I(X_k \geq u)\omega_k(\widehat{\boldsymbol{\gamma}}, \widehat{\Lambda}_T)B_{nk}(\mathbf{z})}, \tag{4}$$

where

$$\omega_k(\boldsymbol{\gamma}, \Lambda_T) = \Delta_k + (1-\Delta_k)\frac{\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_k)\exp\{-\Lambda_T(X_k|\mathbf{z})\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_k)+\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W}_k)\exp\{-\Lambda_T(X_k|\mathbf{z})\}}. \tag{5}$$

In the numerical algorithm, we first obtain an initial value $\widehat{\boldsymbol{\gamma}}^{(0)}$ by performing the logistic regression of Δ on $\mathbf{W}$, and then obtain $\widehat{\Lambda}_T^{(0)}(t|\mathbf{z})$ from (4) by setting all ω_k's to be one. Plugging $\widehat{\boldsymbol{\gamma}}^{(0)}$ and $\widehat{\Lambda}_T^{(0)}(t|\mathbf{z})$ into (5) leads to $\omega_k^{(0)}$. At the mth iteration, our algorithm for estimating $\boldsymbol{\gamma}$ and $\Lambda_T(t|\mathbf{z})$ proceeds as follows.

(i) Plug $\omega_k^{(m)}$ into (4) and obtain $\widehat{\Lambda}_T^{(m+1)}(t|\mathbf{z})$.
(ii) Plug $\widehat{\Lambda}_T^{(m+1)}(t|\mathbf{z})$ into (3) and solve the resultant equation using the Newton–Raphson algorithm to obtain $\widehat{\boldsymbol{\gamma}}^{(m+1)}$.
(iii) Plug $\widehat{\boldsymbol{\gamma}}^{(m+1)}$ and $\widehat{\Lambda}_T^{(m+1)}(t|\mathbf{z})$ into (5) and obtain $\omega_k^{(m+1)}$.
(iv) Repeat steps (i), (ii), and (iii) until a predetermined convergence criterion is met.

The resultant estimators for $\boldsymbol{\gamma}$ and $\Lambda_T(t|\mathbf{z})$ are denoted by $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\Lambda}_T(t|\mathbf{z})$, respectively. For identifiability and computational stability, we set $\widehat{\Lambda}_T(t|\mathbf{z}) = \infty$ if t is greater than the largest uncensored observation. Correspondingly, the conditional cumulative distribution function can be estimated by $\widehat{F}_T(t|\mathbf{z}) = 1-\exp\{-\widehat{\Lambda}_T(t|\mathbf{z})\}$.

It is recognized that the main challenge in cure rate analysis is that one never knows which censored observation is cured and which is uncured. Therefore, the susceptibility indicator can be viewed as missing data. By invoking the multiple imputation approach, we can estimate the parameter $\boldsymbol{\beta}$ in the quantile regression model (1). Let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^{\mathrm{T}}$, whose components are 1 for observed failure times and are missing for censored observations. Ideally, if both $\boldsymbol{\eta}$ and F_T were known, we can directly follow Wang and Wang (2009) to define the weight function by redistributing censored data to the right,

$$\upsilon_i(F_T) = \begin{cases} 1, & \text{if } \Delta_i = 1 \text{ or } F_T(C_i|\mathbf{Z}_i) > \tau, \\ \dfrac{\tau - F_T(C_i|\mathbf{Z}_i)}{1-F_T(C_i|\mathbf{Z}_i)}, & \text{if } \Delta_i = 0 \text{ and } F_T(C_i|\mathbf{Z}_i) < \tau. \end{cases}$$

The estimator of the quantile regression coefficient $\boldsymbol{\beta}$ can be obtained by minimizing the weighted objective function based on those susceptible subjects only,

$$Q_n(\boldsymbol{\beta}, \boldsymbol{\eta}, F_T) = n^{-1}\sum_{i=1}^{n}\eta_i\left[\upsilon_i(F_T)\rho_\tau(X_i - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}) + \{1-\upsilon_i(F_T)\}\rho_\tau(X^{\infty} - \mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta})\right], \tag{6}$$

where $\rho_\tau(u) = u\{\tau - I(u<0)\}$ is the check function and X^{∞} is any value sufficiently large to exceed all $\mathbf{Z}_i^{\mathrm{T}}\boldsymbol{\beta}$. Even though we obtain an estimator of F_T as a byproduct when estimating $\boldsymbol{\gamma}$, the components of $\boldsymbol{\eta}$ are unknown (missing) for censored subjects. To facilitate the minimization of $Q_n(\boldsymbol{\beta}, \boldsymbol{\eta}, F_T)$, it is vital to identify the value of each component of $\boldsymbol{\eta}$. Fortunately, once the estimators $\widehat{\boldsymbol{\gamma}}$ and $\widehat{F}_T$ are obtained, we can estimate the conditionally uncured probability for each subject given its observation,

$$\widehat{p}_i = \Delta_i + (1-\Delta_i)\frac{\pi(\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{W}_i)\{1-\widehat{F}_T(X_i|\mathbf{Z}_i)\}}{1-\pi(\widehat{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{W}_i)\widehat{F}_T(X_i|\mathbf{Z}_i)}, \tag{7}$$

from which multiple imputation of the missing values of η_i can be implemented.

Suppose that we conduct K times of imputation, and let $\widehat{\eta}_{i(k)}$ be the kth imputed value for η_i $(i = 1, \ldots, n)$, which is sampled from the Bernoulli distribution with success probability $\widehat{p}_i$. For the kth imputation, once $\widehat{\boldsymbol{\eta}}_{(k)} = (\widehat{\eta}_{1(k)}, \ldots, \widehat{\eta}_{n(k)})^{\mathrm{T}}$ are sampled, we employ the locally nonparametric Nelson–Aalen estimator to estimate $F_T(t|\mathbf{z})$ based on the filled-in data set consisting of $\{i\colon \widehat{\eta}_{i(k)} = 1, i = 1, \ldots, n\}$. The estimator of $F_T(t|\mathbf{z})$ is $\widehat{F}_{T(k)}(t|\mathbf{z}) = 1-\exp\{-\widehat{\Lambda}_{T(k)}(t|\mathbf{z})\}$, where

$$\widehat{\Lambda}_{T(k)}(t|\mathbf{z}) = \int_{-\infty}^{t} \frac{\sum_{i=1}^{n}\widehat{\eta}_{i(k)}B_{ni}(\mathbf{z})\mathrm{d}N_i(u)}{\sum_{i=1}^{n}\widehat{\eta}_{i(k)}I(X_i \geq u)B_{ni}(\mathbf{z})}.$$

As a result, a censored observation would be retained if the corresponding imputed value of η_i is 1. Thus, we can obtain the kth imputed estimator $\widehat{\boldsymbol{\beta}}_{(k)}$ by minimizing the objective function $Q_n(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}}_{(k)}, \widehat{F}_{T(k)})$. We repeat this imputation–estimation procedure K times and then take an average to obtain the multiple imputation estimator

$$\widehat{\boldsymbol{\beta}} = K^{-1} \sum_{k=1}^{K} \widehat{\boldsymbol{\beta}}_{(k)}.$$

In nonparametric functional estimation, it is critical to select the bandwidth for the kernel smoothing. We recommend a d-fold cross-validation method for choosing h_n as follows. We randomly divide the data into d nonoverlapping and equal-sized subgroups. For the jth subgroup, $\mathcal{D}_j$, we fit the model using $\mathcal{D}_{(-j)}$, the data excluding subgroup j, and calculate a loss function based on the martingale residuals,

$$\mathcal{L}_j^{\mathrm{CV}}(h) = \frac{1}{|\{i\colon \Delta_i = 1 \text{ and } i \in \mathcal{D}_j\}|} \sum_{k \in \mathcal{D}_j} \int_{-\infty}^{L} \{\mathcal{M}_{(-j)}^{\mathrm{CV}}(t, \mathbf{W}_k)\}^2 \mathrm{d}N_k(t),$$

where $|\mathcal{A}|$ denotes the cardinality of a set $\mathcal{A}$, and

$$\mathcal{M}_{(-j)}^{\mathrm{CV}}(t, \mathbf{w}) = \frac{1}{|\{i\colon i \in \mathcal{D}_{(-j)}\}|} \sum_{i \in \mathcal{D}_{(-j)}} \int_{-\infty}^{t} \frac{I(\mathbf{W}_i \le \mathbf{w})\{1 - \pi(\widehat{\boldsymbol{\gamma}}_{(-j)}^{\mathrm{T}} \mathbf{W}_i)\}}{1 - \pi(\widehat{\boldsymbol{\gamma}}_{(-j)}^{\mathrm{T}} \mathbf{W}_i) \widehat{F}_{T(-j)}(u|\mathbf{Z}_i)} \mathrm{d}M_i(u; \widehat{\boldsymbol{\gamma}}_{(-j)}, \widehat{F}_{T(-j)}).$$

Here, both $\widehat{\boldsymbol{\gamma}}_{(-j)}$ and $\widehat{F}_{T(-j)}$ are estimated using the data from $\mathcal{D}_{(-j)}$. Finally, we choose the bandwidth that minimizes the total loss, $\sum_{j=1}^{d} \mathcal{L}_j^{\mathrm{CV}}(h)$.

As a common practice in CQR, we adopt the resampling method to estimate the standard errors of the proposed estimators (Parzen, Wei, and Ying, 1994; Lin et al., 2000; Jin, Lin, and Ying, 2006), which proceeds as follows.

(i) Independently generate n variates, $G_1, \ldots, G_n$, from a distribution with mean 1 and variance 1, for example, Exp(1).

(ii) Perturb the kernel weights

$$B_{ni}^*(\mathbf{z}) = \frac{K_p\{(\mathbf{z} - \mathbf{Z}_i)/h_n\}}{\sum_{k=1}^{n} K_p\{(\mathbf{z} - \mathbf{Z}_k)/h_n\} G_k}$$

and the local Nelson–Aalen type estimator

$$\widehat{\Lambda}_T^*(t|\mathbf{z}) = \int_{-\infty}^{t} \frac{\sum_{i=1}^{n} B_{ni}^*(\mathbf{z}) \mathrm{d}N_i(u) G_i}{\sum_{k=1}^{n} I(X_k \ge u) \omega_k(\widehat{\boldsymbol{\gamma}}, \widehat{\Lambda}_T^*) B_{nk}^*(\mathbf{z}) G_k}. \tag{8}$$

(iii) Perturb the estimating equation for $\boldsymbol{\gamma}$,

$$\mathbf{S}_n^*(\boldsymbol{\gamma}, F_T^*) = n^{-1} \sum_{i=1}^{n} \int_{-\infty}^{L} \frac{\mathbf{W}_i\{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}} \mathbf{W}_i)\}}{1 - \pi(\boldsymbol{\gamma}^{\mathrm{T}} \mathbf{W}_i) F_T^*(t|\mathbf{Z}_i)} \mathrm{d}M_i(t; \boldsymbol{\gamma}, F_T^*) G_i = \mathbf{0}, \tag{9}$$

and by employing the iterative procedure based on (8) and (9), we obtain the bootstrap estimates $\widehat{\boldsymbol{\gamma}}^*$ and $\widehat{F}_T^*(t|\mathbf{z})$.

(iv) Based on the uncured probability from the bootstrap estimates $\widehat{\boldsymbol{\gamma}}^*$ and $\widehat{F}_T^*(t|\mathbf{z})$,

$$\widehat{p}_i^* = \Delta_i + (1 - \Delta_i) \frac{\pi(\widehat{\boldsymbol{\gamma}}^{*\mathrm{T}} \mathbf{W}_i)\{1 - \widehat{F}_T^*(X_i|\mathbf{Z}_i)\}}{1 - \pi(\widehat{\boldsymbol{\gamma}}^{*\mathrm{T}} \mathbf{W}_i) \widehat{F}_T^*(X_i|\mathbf{Z}_i)},$$

we generate the kth set of imputation values $\widehat{\boldsymbol{\eta}}_{(k)}^* = (\widehat{\eta}_{1(k)}^*, \ldots, \widehat{\eta}_{n(k)}^*)^{\mathrm{T}}$, $k = 1, \ldots, K$.

Let

$$\widehat{\Lambda}_{T(k)}^*(t|\mathbf{z}) = \int_{-\infty}^{t} \frac{\sum_{i=1}^{n} \widehat{\eta}_{i(k)}^* B_{ni}^*(\mathbf{z}) \mathrm{d}N_i(u) G_i}{\sum_{i=1}^{n} \widehat{\eta}_{i(k)}^* I(X_i \ge u) B_{ni}^*(\mathbf{z}) G_i}.$$

Perturb the objective function for $\boldsymbol{\beta}$,

$$Q_n^*(\boldsymbol{\beta}, \widehat{\boldsymbol{\eta}}_{(k)}^*, \widehat{F}_{T(k)}^*) = n^{-1} \sum_{i=1}^{n} \widehat{\eta}_{i(k)}^* \left[\upsilon_i(\widehat{F}_{T(k)}^*) \rho_\tau(X_i - \mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\beta}) + \{1 - \upsilon_i(\widehat{F}_{T(k)}^*)\} \rho_\tau(X^\infty - \mathbf{Z}_i^{\mathrm{T}} \boldsymbol{\beta})\right] G_i. \tag{10}$$

Minimizing (10) results in the estimate $\widehat{\boldsymbol{\beta}}_{(k)}^*$. Thus, we obtain the bootstrap version of the multiple imputation estimator $\widehat{\boldsymbol{\beta}}^* = K^{-1} \sum_{k=1}^{K} \widehat{\boldsymbol{\beta}}_{(k)}^*$.

Repeating this perturbation procedure for a large number of times, we can obtain the bootstrap standard error estimates.

3. Asymptotic Properties

We denote $F_C(t|\mathbf{Z}, \mathbf{W}) = P(C \le t|\mathbf{Z}, \mathbf{W})$, $\bar{F}_C(t|\mathbf{Z}, \mathbf{W}) = 1 - F_C(t|\mathbf{Z}, \mathbf{W})$, and define $f_T(t|\mathbf{z}) = \mathrm{d}F_T(t|\mathbf{z})/\mathrm{d}t$ and $f_C(t|\mathbf{z}, \mathbf{w}) = \mathrm{d}F_C(t|\mathbf{z}, \mathbf{w})/\mathrm{d}t$. Let $\boldsymbol{\beta}_0$, $\boldsymbol{\gamma}_0$, F_{0T}, and f_{0T}, respectively, denote the true values of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, F_T, and f_T. Let $\mathcal{B} \subset \mathcal{R}^{p+1}$ and $\mathcal{K} \subset \mathcal{R}^{q+1}$ be compact subsets that contain $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ as the interior points, respectively. Let $\mathcal{H}$ denote an infinite dimensional distribution space whose support is the same as that of F_{0T}. Also, denote $\mathcal{Z}$ and $\mathcal{W}$ as the domains of covariates $\mathbf{Z}$ and $\mathbf{W}$, respectively. We impose the conditions as follows.

C1. With probability 1, both $\mathbf{Z}$ and $\mathbf{W}$ are bounded; $E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})$ and $E(\mathbf{W}\mathbf{W}^{\mathrm{T}})$ are positive definite matrices.

C2. The kernel function $K(\cdot)$ has a compact support in $\mathcal{R}$, and it is an ℓth order kernel satisfying that

Table 1
Simulation results for the proposed multiple imputation method under the cure rate quantile regression model when ϵ_τ follows a normal distribution

				$\gamma_0 = 1$ and $\beta_0 = 2$				$\gamma_1 = -1$ and $\beta_1 = 1$			
τ	n	Method	Coef.	Est.	SE	ESE	CP(%)	Est.	SE	ESE	CP(%)
0.5	200	Impute η	$\boldsymbol{\gamma}$	1.034	0.311	0.323	96.0	−1.027	0.521	0.544	95.2
			$\boldsymbol{\beta}$	2.010	0.247	0.280	95.2	0.972	0.548	0.622	93.8
		True η	$\boldsymbol{\beta}$	2.010	0.246	0.282	95.0	0.971	0.544	0.626	95.4
	400	Impute η	$\boldsymbol{\gamma}$	1.019	0.248	0.226	93.2	−1.035	0.433	0.380	92.0
			$\boldsymbol{\beta}$	1.999	0.180	0.194	95.8	0.995	0.395	0.429	95.4
		True η	$\boldsymbol{\beta}$	1.998	0.178	0.194	95.8	0.999	0.390	0.430	95.4
0.7	200	Impute η	$\boldsymbol{\gamma}$	1.030	0.318	0.319	95.0	−1.028	0.561	0.538	93.6
			$\boldsymbol{\beta}$	2.013	0.265	0.291	95.0	0.944	0.586	0.648	94.4
		True η	$\boldsymbol{\beta}$	2.007	0.265	0.292	95.0	0.959	0.582	0.652	94.8
	400	Impute η	$\boldsymbol{\gamma}$	1.009	0.226	0.224	95.2	−1.017	0.364	0.376	94.8
			$\boldsymbol{\beta}$	1.987	0.195	0.210	95.0	1.018	0.426	0.459	95.6
		True η	$\boldsymbol{\beta}$	1.986	0.191	0.211	95.0	1.020	0.414	0.459	96.8

Note: Coef. represents the regression coefficient, Est. is the average of the parameter estimates, SE is the sample standard error of the estimates, ESE is the average of bootstrap estimates of the standard error, and CP is the coverage probability of 95% confidence intervals. "Impute η" represents the proposed multiple imputation method, and "True η" refers to the ideal case where the true susceptibility indicators were known.

$\int_{\mathcal{R}} K(u)\mathrm{d}u = 1$, $\int_{\mathcal{R}} K^2(u)\mathrm{d}u < \infty$, $\int_{\mathcal{R}} u^j K(u)\mathrm{d}u = 0$ for $1 \le j < \ell$, and $\int_{\mathcal{R}} |u|^\ell K(u)\mathrm{d}u < \infty$. Moreover, it is Lipschitz continuous of order ℓ, with $\ell \ge 2$.

C3. The first ℓ partial derivatives of the density function of $\mathbf{Z}$, $f_{\mathbf{Z}}(\mathbf{z})$, with respect to $\mathbf{z}$ are uniformly bounded for $\mathbf{z} \in \mathcal{Z}$, and $f_{0T}(t|\mathbf{z})$ and $f_C(t|\mathbf{z}, \mathbf{w})$ are bounded (uniformly in t, $\mathbf{z}$, and $\mathbf{w}$) away from infinity, and the first ℓ partial derivatives of $f_{0T}(t|\mathbf{z})$ and $f_C(t|\mathbf{z}, \mathbf{w})$ with respect to $\mathbf{z}$ or $\mathbf{w}$ are uniformly bounded in $t \in (-\infty, L]$, $\mathbf{z} \in \mathcal{Z}$, and $\mathbf{w} \in \mathcal{W}$. In addition, $\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{w})$ is uniformly bounded away from 0 and 1 for $\mathbf{w} \in \mathcal{W}$ and $\boldsymbol{\gamma} \in \mathcal{K}$.

C4. For $\boldsymbol{\gamma}$ in the neighborhood of $\boldsymbol{\gamma}_0$, the "Hessian" matrix of $\boldsymbol{\gamma}$ (minus of the first derivative of (3) evaluated at

Table 2
Simulation results for the proposed multiple imputation method under the cure rate quantile regression model when ϵ_τ follows an extreme value distribution

				$\gamma_0 = 1$ and $\beta_0 = 2$				$\gamma_1 = -1$ and $\beta_1 = 1$			
τ	n	Method	Coef.	Est.	SE	ESE	CP(%)	Est.	SE	ESE	CP(%)
0.5	200	Impute η	$\boldsymbol{\gamma}$	1.033	0.324	0.323	94.4	−1.057	0.560	0.546	94.2
			$\boldsymbol{\beta}$	2.017	0.279	0.325	96.6	0.988	0.603	0.728	96.6
		True η	$\boldsymbol{\beta}$	2.010	0.277	0.326	97.0	0.995	0.590	0.729	97.6
	400	Impute η	$\boldsymbol{\gamma}$	1.025	0.237	0.227	94.0	−1.018	0.403	0.383	92.0
			$\boldsymbol{\beta}$	2.006	0.208	0.221	94.0	1.002	0.467	0.494	94.0
		True η	$\boldsymbol{\beta}$	2.005	0.205	0.221	94.4	0.999	0.459	0.494	94.4
0.7	200	Impute η	$\boldsymbol{\gamma}$	1.052	0.319	0.321	95.6	−1.069	0.536	0.540	94.6
			$\boldsymbol{\beta}$	2.031	0.369	0.419	95.2	0.960	0.806	0.928	95.6
		True η	$\boldsymbol{\beta}$	2.019	0.365	0.420	95.4	0.969	0.792	0.927	94.8
	400	Impute η	$\boldsymbol{\gamma}$	1.013	0.220	0.224	95.2	−1.015	0.371	0.376	95.0
			$\boldsymbol{\beta}$	2.007	0.260	0.285	95.6	0.965	0.568	0.638	96.2
		True η	$\boldsymbol{\beta}$	2.005	0.256	0.284	96.2	0.962	0.564	0.634	96.8

Note: See the footnote of Table 1.

Table 3
Simulation results for the proposed multiple imputation method under the cure rate quantile regression model when ϵ_τ follows a Cauchy distribution

				$\gamma_0 = 1$ and $\beta_0 = 2$				$\gamma_1 = -1$ and $\beta_1 = 1$			
τ	n	Method	Coef.	Est.	SE	ESE	CP(%)	Est.	SE	ESE	CP(%)
0.5	200	Impute η	$\boldsymbol{\gamma}$	0.996	0.332	0.330	96.6	−1.032	0.560	0.554	96.0
			$\boldsymbol{\beta}$	1.988	0.333	0.371	96.0	0.963	0.734	0.819	95.8
		True η	$\boldsymbol{\beta}$	1.995	0.333	0.373	95.6	1.005	0.739	0.828	95.2
	400	Impute η	$\boldsymbol{\gamma}$	1.002	0.252	0.235	95.2	−1.013	0.404	0.395	95.6
			$\boldsymbol{\beta}$	1.994	0.249	0.257	95.0	0.987	0.535	0.576	96.2
		True η	$\boldsymbol{\beta}$	1.995	0.237	0.254	95.0	1.007	0.508	0.564	96.6
0.7	200	Impute η	$\boldsymbol{\gamma}$	1.050	0.396	0.345	94.2	−1.078	0.658	0.584	94.4
			$\boldsymbol{\beta}$	1.996	0.364	0.380	96.8	0.951	0.806	0.849	95.6
		True η	$\boldsymbol{\beta}$	1.991	0.338	0.372	95.2	0.980	0.766	0.828	97.2
	400	Impute η	$\boldsymbol{\gamma}$	1.005	0.225	0.230	95.8	−1.034	0.373	0.384	96.4
			$\boldsymbol{\beta}$	2.024	0.329	0.385	95.8	0.957	1.340	1.236	95.8
		True η	$\boldsymbol{\beta}$	2.017	0.327	0.362	96.2	0.995	0.743	0.796	96.4

Note: See the footnote of Table 1.

$F_{0T}(t|\mathbf{z})$), $\Gamma(\boldsymbol{\gamma}, F_{0T})$, is positive definite, where

$$\begin{aligned}\Gamma(\boldsymbol{\gamma}, F_T) &= E\left[\int_{-\infty}^{L} \frac{\mathbf{W}^{\otimes 2}\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\}\{1-F_T(t|\mathbf{Z})\}}{\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_T(t|\mathbf{Z})\}^2}\mathrm{d}M(t;\boldsymbol{\gamma}, F_T)\right] \\ &\quad + E\left[\int_{-\infty}^{L}\left(\frac{\mathbf{W}\{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\}}{1-\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})F_T(t|\mathbf{Z})}\right)^{\otimes 2} I(X \geq t)\mathrm{d}\Lambda_{Y,\boldsymbol{\gamma}}(t|\mathbf{Z}, \mathbf{W})\right]\end{aligned}$$

with $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathrm{T}}$ for any column vector $\mathbf{a}$.

C5. For $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in the neighborhood of their true values,

$$\boldsymbol{\Phi}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = E[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\pi(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{W})\bar{F}_C(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}|\mathbf{Z}, \mathbf{W})f_{0T}(\mathbf{Z}^{\mathrm{T}}\boldsymbol{\beta}|\mathbf{Z})]$$

is positive definite.

C6. The bandwidth $h_n = O(n^{-v})$, where $0 < v < \min(1/p, 1/\ell)$.

C7. The bandwidth $h_n = O(n^{-v})$, where $1/(2\ell) < v < 1/(3p)$ and $\ell > 3p/2$.

Condition C2 requires $K(\cdot)$ to be an ℓth order kernel. Condition C3 guarantees the boundedness of the ℓth term in the Taylor expansion. The convergence rate of the bandwidth in condition C6 is required to obtain the consistency of the proposed estimators while a higher convergence rate is imposed in condition C7 for establishing their weak convergence properties. For a larger p, we need a higher order kernel function to control the bias.

Conditions C1–C4 and C7 entail the approximate representation (Wu and Yin, 2013),

$$n^{1/2}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = n^{-1/2}\sum_{i=1}^{n}\Gamma_0^{-1}\boldsymbol{\phi}(\mathcal{O}_i, \boldsymbol{\gamma}_0, F_{0T}) + o_P(1), \quad (11)$$

where $\Gamma_0 = \Gamma(\boldsymbol{\gamma}_0, F_{0T})$, $\mathcal{O}_i = (X_i, \Delta_i, \mathbf{Z}_i, \mathbf{W}_i)$, and $\boldsymbol{\phi}$ is a measurable function from $\mathcal{O}_i$ to $\mathcal{R}^{q+1}$ with mean zero and finite second moments. Thus, $n^{1/2}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ is asymptotically normal with mean zero and variance–covariance matrix $\Gamma_0^{-1}\Omega_0\Gamma_0^{-1}$, where $\Omega_0 = E\{\boldsymbol{\phi}(\mathcal{O}, \boldsymbol{\gamma}_0, F_{0T})^{\otimes 2}\}$.

Under conditions C1–C4 and C6 and following Theorem 2.3 in Liang, Una-Alvarez, and Iglesias-Perez (2012), we also have the Bahadur representation of $\widehat{F}_T(t|\mathbf{z})$,

$$\begin{aligned}\widehat{F}_T(t|\mathbf{z}) - F_{0T}(t|\mathbf{z}) = \frac{1}{nh_n^p f_{\mathbf{Z}}(\mathbf{z})}\sum_{i=1}^{n} K_p\left(\frac{\mathbf{z}-\mathbf{Z}_i}{h_n}\right) \\ \times\psi(X_i, \Delta_i, \mathbf{W}_i, t, \mathbf{z}; \boldsymbol{\gamma}_0, F_{0T}) \\ + O_P(\alpha_n), \end{aligned} \quad (12)$$

where $\psi(X, \Delta, \mathbf{W}, t, \mathbf{z}; \boldsymbol{\gamma}_0, F_{0T})$ is measurable function with mean zero and finite variance for any t and $\mathbf{z}$, and $\alpha_n = h_n^\ell + \{\log n/(nh_n^p)\}^{3/4}$.

THEOREM 1. *Under conditions C1–C6, $\widehat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$ as n goes to infinity.*

THEOREM 2. *Under conditions C1–C5 and C7, $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically normal with mean zero and variance–covariance matrix $\Phi_0^{-1}\Sigma_0\Phi_0^{-1}$, where $\boldsymbol{\Phi}_0 = \boldsymbol{\Phi}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ and Σ_0 is given in Web Appendix B.*

Proofs of Theorems 1 and 2 are relegated to Web Appendices A and B in the supplementary material, respectively. The derivation of (A.4) in Web Appendix A follows the similar arguments in Wang and Feng (2012).

Table 4
Comparison between the cure rate censored quantile regression (CQR) under global linearity and the proposed multiple imputation method with three covariates and a cure rate of 25%

			CQR under global linearity				Multiple imputation			
Error			Int.	Z_1	Z_2	Z_3	Int.	Z_1	Z_2	Z_3
Homo.	$\boldsymbol{\beta}(\tau)$	True	2.000	1.000	−1.000	1.000	2.000	1.000	−1.000	1.000
	0.1	Est.	0.686	0.948	−0.945	0.947	1.679	0.874	−0.806	0.804
		MSE	1.787	0.141	0.047	0.015	0.171	0.137	0.097	0.052
	0.2	Est.	1.121	0.917	−0.934	0.937	1.775	0.878	−0.852	0.836
		MSE	0.815	0.098	0.037	0.012	0.094	0.106	0.055	0.037
	0.3	Est.	1.428	0.914	−0.928	0.927	1.814	0.867	−0.845	0.857
		MSE	0.365	0.080	0.035	0.013	0.067	0.093	0.052	0.027
	0.4	Est.	1.695	0.897	−0.924	0.918	1.847	0.885	−0.876	0.863
		MSE	0.128	0.078	0.033	0.013	0.054	0.081	0.042	0.025
	0.5	Est.	1.947	0.877	−0.919	0.908	1.961	0.991	−0.971	0.965
		MSE	0.039	0.087	0.032	0.015	0.027	0.065	0.023	0.008
	0.6	Est.	2.192	0.870	−0.915	0.894	1.909	0.913	−0.914	0.898
		MSE	0.074	0.088	0.036	0.018	0.037	0.069	0.033	0.017
	$\boldsymbol{\gamma}$	True	1.000	−1.000	2.000	−1.000	1.000	−1.000	2.000	−1.000
		Est.	1.045	−1.146	2.232	−1.169	1.045	−1.146	2.232	−1.169
		MSE	0.183	0.422	0.226	0.215	0.183	0.422	0.226	0.215
Hete.	$\boldsymbol{\beta}(\tau)$	True	2.000	1.000	−1.000	1.000	2.000	1.000	−1.000	1.000
	0.1	Est.	0.652	0.263	−1.646	0.355	0.659	1.286	−0.686	1.092
		MSE	2.038	1.063	0.602	0.463	2.040	0.581	0.253	0.098
	0.2	Est.	1.330	0.496	−1.436	0.500	1.334	1.013	−0.859	1.542
		MSE	0.529	0.467	0.258	0.278	0.524	0.226	0.091	0.330
	0.3	Est.	1.665	0.630	−1.284	0.588	1.647	0.889	−0.944	1.172
		MSE	0.145	0.236	0.118	0.192	0.156	0.121	0.036	0.055
	0.4	Est.	1.857	0.764	−1.173	0.674	1.839	0.836	−1.046	0.888
		MSE	0.035	0.101	0.048	0.126	0.042	0.077	0.019	0.033
	0.5	Est.	1.987	0.845	−1.098	0.765	1.965	0.960	−1.028	0.898
		MSE	0.007	0.049	0.019	0.073	0.008	0.026	0.010	0.026
	0.6	Est.	2.123	0.884	−1.041	0.845	2.061	0.713	−1.195	0.468
		MSE	0.029	0.048	0.017	0.040	0.013	0.112	0.049	0.297
	$\boldsymbol{\gamma}$	True	1.000	−1.000	2.000	−1.000	1.000	−1.000	2.000	−1.000
		Est.	1.137	−1.106	1.868	−1.159	1.137	−1.106	1.868	−1.159
		MSE	0.189	0.381	0.141	0.045	0.189	0.381	0.141	0.045

Note: Homo. represents the homogeneous error case, Hete. represents the heterogeneous error case, Int. is the intercept, Est. is the average of the parameter estimates, and MSE is the mean squared error.

4. Simulation Studies

We conducted simulation studies to evaluate the finite-sample performance of the proposed multiple imputation method under the cure rate quantile regression model. We first generated survival times of susceptible subjects from the model with heterogeneous errors,

$$T \equiv \log \widetilde{T} = \beta_0 + \beta_1 Z + (1+Z)\epsilon_\tau,$$

where the true coefficients are $\beta_0 = 2$ and $\beta_1 = 1$, $Z \sim \text{Unif}(0,1)$, and the error ϵ_τ is normally distributed with the τth quantile being zero. The susceptibility indicator η was generated from the logistic regression model (2) with $\mathbf{W} = (1, Z)^{\mathrm{T}}$ and the true coefficients $\gamma_0 = 1$ and $\gamma_1 = -1$. We set the censoring time to be $C = \widetilde{C} \wedge L$, where $\widetilde{C}$ was generated from $\text{Unif}(c_0, L+2)$ if $Z < 0.5$, and from $\text{Unif}(c_0+1, L+2)$ otherwise. The constant c_0 and the study duration time L were chosen to yield a censoring rate of 40%, and the cure rate was approximately 37%. We adopted the biquadratic kernel function and selected the bandwidth h_n via the tenfold cross-validation procedure. We repeated the imputation procedure five times, that is, $K = 5$. To examine the effectiveness of the proposed imputation procedure, we also considered the ideal situation where the true η_i's were known, so the imputation method was not needed and one could directly estimate the model coefficients based on those truly known uncured subjects. For each configuration, we repeated 500 simulations, and for each replicated data set 200 bootstrap samples were generated for variance estimation based on the perturbation resampling procedure.

Table 1 summarizes simulation results for two different quantile levels $\tau = 0.5$ and 0.7 under sample sizes $n = 200$ and 400, respectively. The cure rate parameter estimates via the iterative nonparametric method appear to be unbiased, and the proposed multiple imputation estimators for

Table 5

Analysis results of the lung cancer data under the proposed imputation cure rate censored quantile regression (CQR) model with three particular quantiles ($\tau = 0.3$, 0.5, and 0.7), the Cox proportional hazards cure rate model, and the accelerated failure time (AFT) cure rate model

Cure model	Coef.	Covariate	Est.	ESE	p-value
CQR	$\boldsymbol{\beta}(0.3)$	Intercept	1.391	0.500	0.005
		Histology	0.749	0.484	0.122
		Age	0.028	0.276	0.919
		Sex	−0.193	0.517	0.708
	$\boldsymbol{\beta}(0.5)$	Intercept	2.762	0.442	< 0.001
		Histology	0.884	0.474	0.062
		Age	−0.067	0.303	0.825
		Sex	−0.832	0.474	0.079
	$\boldsymbol{\beta}(0.7)$	Intercept	3.955	0.585	< 0.001
		Histology	0.719	0.725	0.322
		Age	−0.065	0.361	0.856
		Sex	−0.700	0.728	0.337
	$\boldsymbol{\gamma}$	Intercept	1.161	0.505	0.022
		Histology	−0.506	0.473	0.285
		Age	0.816	0.289	0.005
		Sex	−0.656	0.511	0.199
Cox	$\boldsymbol{\beta}$	Histology	−0.423	0.211	0.045
		Age	0.135	0.130	0.297
		Sex	0.160	0.229	0.485
	$\boldsymbol{\gamma}$	Intercept	1.602	0.620	0.010
		Histology	−0.246	0.506	0.626
		Age	0.860	0.345	0.013
		Sex	−0.675	0.590	0.252
AFT	$\boldsymbol{\beta}$	Histology	0.850	0.435	0.051
		Age	−0.015	0.317	0.962
		Sex	−0.156	0.540	0.772
	$\boldsymbol{\gamma}$	Intercept	0.924	1.064	0.385
		Histology	0.773	0.916	0.399
		Age	0.967	0.576	0.093
		Sex	−0.686	1.234	0.578

the quantile regression coefficients are also virtually unbiased. Furthermore, the estimated standard errors using the bootstrap method agree well with the sampling standard errors, and the coverage probabilities of 95% confidence intervals are around the nominal level. More remarkably, the multiple imputation estimator is comparable with the ideal estimator (when we know the true η_i under the hypothetical situation), which demonstrates that the multiple imputation method can on average "identify" the susceptible subjects in the censored group. As expected, more precise estimators can be obtained when the sample size is increased to $n = 400$.

We also explored different distributions for the model error ϵ_τ including the extreme value distribution and the heavy-tailed Cauchy distribution, while keeping the rest of data generation scheme the same as before. The corresponding simulation results are presented in Tables 2 and 3, from which we can draw similar conclusions.

To gain more insight, we further generated survival times $\widetilde{T}$ of susceptible subjects from a more complicated model with multiple covariates,

$$T \equiv \log \widetilde{T} = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \sigma(\mathbf{Z})\epsilon_\tau, \quad (13)$$

where $\mathbf{Z} = (1, Z_1, Z_2, Z_3)^{\mathrm{T}}$, Z_1 was simulated from Unif(0, 1), Z_2 from Bernoulli(0.5), and Z_3 from a truncated standard normal distribution between −3 and 3. The true quantile regression coefficients are $(\beta_0, \beta_1, \beta_2, \beta_3) = (2, 1, -1, 1)$. The error ϵ_τ is normally distributed with the τth quantile being zero, and for ease of exposition, we fixed τ at 0.5. We considered both the homogeneous and heteroscedastic cases with $\sigma(\mathbf{Z}) \equiv 1$ and $\sigma(\mathbf{Z}) = (Z_1^2 + Z_2)^{1/2} \sin(Z_3) + \sum_{j=1}^{3} Z_j$, respectively. For the homogeneous model error, the global linearity assumption under the cure rate CQR model is satisfied by appropriately shifting the intercept. However, for the heterogeneous case, the global linearity does not hold, while we can still fit the CQR for the particular quantile with $\tau = 0.5$ using the proposed imputation method. The censoring times were generated in the same manner as before by considering different censoring distributions for $Z_1 < 0.5$ and $Z_1 \geq 0.5$, which yielded a censoring rate of 40%. We took $\mathbf{W} = \mathbf{Z}$ and the true parameter values in the logistic model were $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (1, -1, 2, -1)$, which led to a cure rate of 25%. We used the product fourth-order kernel and adopted the cross-validation procedure to select the bandwidth.

Although the data were simulated from model (13) with τ being fixed at 0.5, we applied the proposed multiple imputation method at different quantile levels such as $\tau = 0.1, \ldots, 0.6$. Note that only at quantile level 0.5 the fitted cure rate quantile regression model is correct. We compare the proposed multiple imputation method and the cure rate CQR method under the global linearity assumption. Table 4 summarizes the simulation results with the sample size $n = 400$. It is evident that, at the quantile level of 0.5, the performance of the multiple imputation method is satisfactory with three covariates including one discrete and two continuous variates and a lower cure rate. However, at other quantile levels, the resulting estimates for the regression slope coefficients are slightly biased. The global linearity assumption holds under the homogeneous model error by appropriately shifting the intercept, and thus under this case the globally linear cure rate CQR method produces unbiased estimates for the regression slope coefficients across all of the considered quantile levels. For the heterogeneous model error, when $\tau \neq 0.5$, the estimates of the regression coefficients are biased by using either the multiple imputation method or the cure rate CQR method with global linearity assumption. Furthermore, the bias tends to be more serious when the quantile level moves farther away from the median, because the fitted regression model deviates farther from the truth for the lower quantiles. For the cure rate parameters, both methods produce the same (unbiased) estimates as they use the same estimation approaches.

5. Lung Cancer Data Example

As an illustration, we applied the proposed multiple imputation method under the cure rate CQR to a lung cancer study. This study involved 280 lung cancer patients with a censoring rate of 64.3%. In our analysis, the covariates of

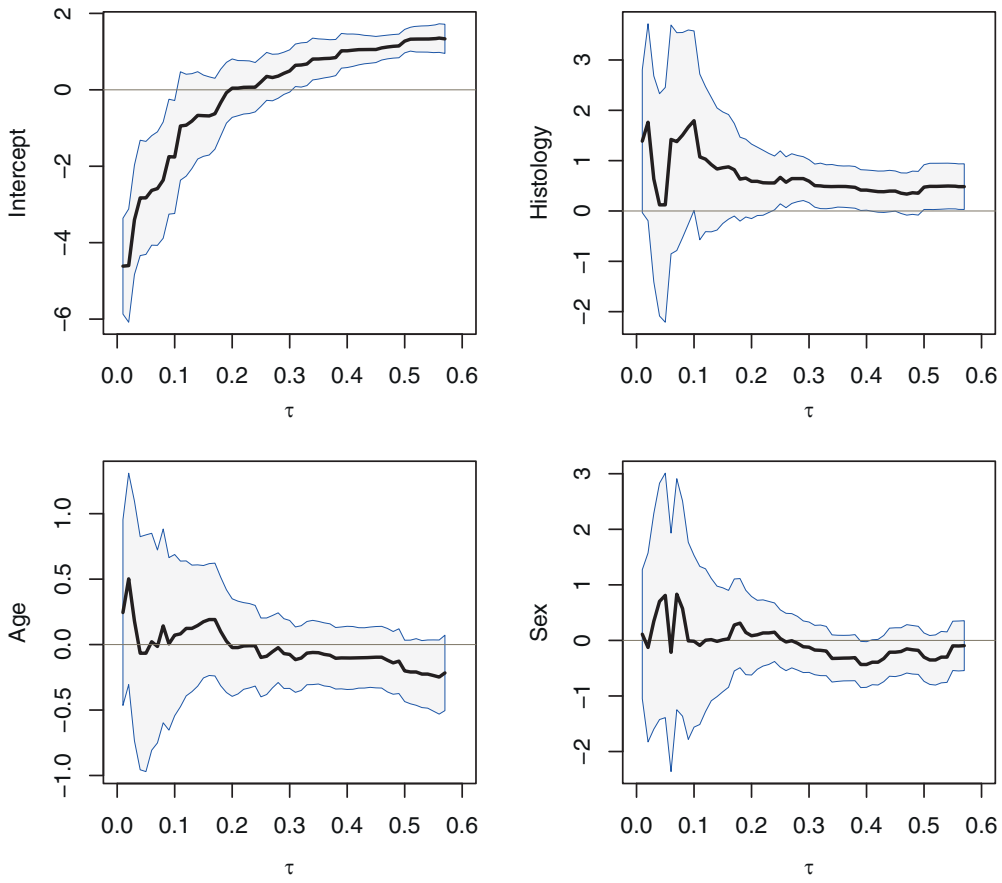

Figure 1. Estimated quantile covariate effects for the lung cancer data and the corresponding 95% pointwise confidence intervals under the cure rate censored quantile regression assuming global linearity.

interest included tumor histology (61% adenocarcinoma = 1; 39% squamous cell carcinoma = 0), age (ranging from 34 to 90 years with a mean of 66 years), and sex (52% female = 1; 48% male = 0). The covariate age was standardized to have mean 0 and variance 1. The effect of tumor histology on patient survival was of particular interest in this study. Based on the usual Kaplan–Meier survival curves (shown in Web Appendix C), after approximately seven years of follow-up, a stable plateau at the tail of the survival curves can be observed, which indicates the existence of a possible cure fraction. We applied the proposed models (1) and (2) by taking $\mathbf{Z} = \mathbf{W}$ with both involving (Histology, Age, Sex) to fit the lung cancer survival data, and the analysis results are summarized in Table 5. We are particularly interested in the conditional quantiles of survival times at $\tau = 0.3$, 0.5, and 0.7, representing the lower, median, and higher quantiles, respectively. For the considered regression quantiles, none of the covariates showed an significant effect on the survival time, while there was a trend that patients with adenocarcinoma tended to have a longer median survival time, which nevertheless requires further confirmation. For the estimation of cure rate parameters, patient's age was significantly associated with the cure rate in the lung cancer and, in particular, younger patients were more likely to be insusceptible to the event of interest.

For comparison, we also fitted the lung cancer data using the cure rate CQR method under the global linearity assumption (Wu and Yin, 2013). Figure 1 displays the quantile regression estimates of covariate effects and the corresponding 95% pointwise confidence intervals for $\tau \in [0.01, 0.58]$ with a step size of 0.01. We did not find any significant covariate effects for patients' age and sex on the survival times across the range of considered regression quantiles. Nevertheless, the tumor histology appeared to have an significant influence on patients' survival, for regression quantiles near 0.3. Furthermore, we analyzed the lung cancer data using the Cox PH cure rate model (Sy and Taylor, 2000) and the AFT cure rate model (Lu, 2010). Both delivered the same conclusion that histology appeared to be an important factor on patient survival. In terms of the cure rate estimation, the Cox PH cure rate model coincides with the proposed cure rate quantile regression with multiple imputation; both showed a significant age effect in the logistic regression. However, the AFT cure rate model did not show the age effect to be statistically significant, although there was such a trend. The estimated histology effects using the AFT cure rate model and the quantile cure rate model were both insignificant but with different signs. This is possibly due to the misspecification of the AFT model structure upon which the cure rate estimation relies. The proposed cure rate quantile regression with multiple imputation does not impose any model assumption for the survival times of the susceptible subjects, and thus is more robust for practical use.

6. Conclusions

The multiple imputation method is prevalent in handling missing data problems. We recast the uncertainty of whether a censored subject is susceptible or not in the missing data framework and invoke the multiple imputation approach to the cure rate analysis. We impose the usual censoring assumption that survival times and censoring times are conditionally independent given covariates. Global linearity is often assumed by existing CQR methods (Portnoy, 2003; Peng and Huang, 2008; Wu and Yin, 2013), while the proposed method relaxes such a stringent model assumption for the cure rate quantile regression. Numerical results indicate that the proposed method performs well with finite samples. The proposed multiple imputation method opens a new door to the traditional mean- or hazards-based mixture cure rate regression models.

7. Supplementary Materials

Web Appendices referenced in Sections 3 and 5, and the R script to obtain the proposed estimators and generate the simulative data sets are available with this article at the *Biometrics* website on Wiley Online Library.

Acknowledgements

The authors thank the editor, the associate editor, and the referee for very constructive comments that improved the presentation of the article. The research was supported in part by the National Natural Science Foundation of China and the Research Grants Council of Hong Kong.

References

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* **47**, 501–515.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**, 1041–1046.

Farewell, V. T. (1986). Mixture models in survival analysis: Are they worth the risk? *The Canadian Journal of Statistics* **14**, 257–262.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.

Jin, Z., Lin, D. Y., and Ying, Z. (2006). On least-squares regression with censored data. *Biometrika* **93**, 147–161.

Kuk, A. Y. C. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531–541.

Liang, H.-Y., Una-Alvarez, J., and Iglesias-Perez, M. C. (2012). Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data. *Test* **21**, 790–810.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society, Series B* **62**, 711–730.

Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure fraction. *Statistica Sinica* **20**, 661–674.

Lu, W. and Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika* **91**, 331–343.

Mao, M. and Wang, J. L. (2010). Semiparametric efficient estimation for a class of generalized proportional odds cure models. *Journal of the American Statistical Association* **105**, 302–311.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56**, 227–236.

Parzen, M. I., Wei, L. J., and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association* **103**, 637–649.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association* **98**, 1001–1012.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression. *Journal of Econometrics* **25**, 303–325.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics* **32**, 143–155.

Wang, H. and Feng, X. (2012). Multiple imputation for *M*-regression with censored covariates. *Journal of the American Statistical Association* **107**, 194–204.

Wang, H. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association* **104**, 1117–1128.

Wu, Y. and Yin, G. (2013). Cure rate quantile regression for censored data with a survival fraction. *Journal of the American Statistical Association* **108**, 1517–1531.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard function. *Journal of the American Statistical Association* **94**, 137–145.

Yin, G. (2005). Bayesian cure rate frailty models with application to a root canal therapy study. *Biometrics* **61**, 552–558.

Yin, G. and Ibrahim, J. (2005). Cure rate models: A unified approach. *Canadian Journal of Statistics* **33**, 559–570.

Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90**, 178–184.

Zeng, D., Yin, G., and Ibrahim, J. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association* **101**, 670–684.

Zhang, J. and Peng, Y. (2009). Accelerated hazards mixture cure model. *Lifetime Data Analysis* **15**, 455–467.

Received September 2014. *Revised June* 2016.
Accepted June 2016.