The Canadian Journal of Statistics
Vol. 45, No. 1, 2017, Pages 29–43
La revue canadienne de statistique

Cure rate quantile regression accommodating both finite and infinite survival times

Yuanshan WU[1,2]* and Guosheng YIN[3]

[1]*School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China*
[2]*Computational Science Hubei Key Laboratory, Wuhan University, Wuhan, Hubei 430072, China*
[3]*Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam Road, Hong Kong*

Key words and phrases: Censored data; cure rate model; long-term survivor; regression quantile; survival analysis; survival fraction; susceptibility.

MSC 2010: Primary 62N01; secondary 62N02

Abstract: In survival analysis a proportion of patients may be cured by the treatment, and thus they become risk-free of the event of interest and their survival times change to infinity. The existence of such a survival fraction often makes the underlying population more heterogeneous and heavily right-skewed. Compared with the traditional mean- or hazard-based regression methods, quantile regression is more suitable for such survival data as it is more robust against outliers or infinite survival times. Moreover, it offers a comprehensive assessment of the covariate effects on the survival times at different quantile levels. We propose a new cure rate quantile regression model for the entire population including both finite and infinite survival times. By invoking non-parametric functional estimation an iterative algorithm is developed to estimate the cure rate parameters. The scheme of redistribution-of-mass to the right for censored data is adopted to estimate the quantile regression parameters. The consistency and asymptotic normality of the proposed estimators are established. Extensive simulation studies are conducted to evaluate the finite-sample performance of the proposed method, which is further illustrated with a phase III melanoma clinical trial study. *The Canadian Journal of Statistics* 45: 29–43; 2017 © 2016 Statistical Society of Canada

Résumé: En analyse de survie, certains patients peuvent guérir, de sorte que leur durée de vie devient l'infini puisqu'ils ne sont plus à risque de vivre l'événement à l'étude. L'existence de cette fraction de survivants rend souvent la population hétérogène en plus d'infliger une forte asymétrie vers la droite. En comparaison des méthodes de régression traditionnelles basées sur la moyenne ou le risque, les méthodes de régression quantile sont mieux adaptées à de telles données puisqu'elles sont plus robustes aux valeurs aberrantes et aux temps de survie infinis. De plus, elles offrent une évaluation détaillée de l'effet des covariables sur le temps de survie à différents quantiles. Les auteurs proposent un nouveau modèle de régression quantile pour la population entière, incluant les temps de survie finis et infinis. Ils développent un algorithme itératif invoquant l'estimation fonctionnelle non paramétrique afin d'estimer les paramètres de taux de guérison. Ils adoptent un schéma de redistribution de la masse des données censurées vers la droite pour estimer les paramètres de la régression quantile. Les auteurs établissent la convergence et la normalité asymptotique de leurs estimateurs. Ils présentent des simulations approfondies pour évaluer la performance de leur méthode sur des échantillons finis et ils l'illustrent avec les données d'une étude clinique de phase III sur le mélanome. *La revue canadienne de statistique* 45: 29–43; 2017 © 2016 Société statistique du Canada

Additional supporting information may be found in the online version of this article at the publisher's website.

* *Author to whom correspondence may be addressed.*
E-mail: shan@whu.edu.cn

1. INTRODUCTION

In many cancer studies a proportion of subjects may either be cured following treatment or be immune to the event of interest from the study origin. Such patients become risk-free and thus cannot be counted in the usual risk set in survival analysis. To explicitly incorporate such a survival fraction, cure rate models have been investigated extensively, among which a commonly used approach is the two-component mixture cure rate model. The mixture model assumes the underlying population is composed of susceptible (uncured) and immune (cured) subjects. As the susceptible subjects would eventually experience the event of interest if follow-up is sufficiently long, one can apply the usual survival models to their survival times, and logistic regression is often used to model the susceptibility indicator. For parametric models Berkson & Gage (1952) proposed the exponential–logistic mixture cure model, and Farewell (1982, 1986) considered the Weibull–logistic mixture cure model for censored data with a survival fraction. For semiparametric models Kuk & 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 cure indicator. Lu & Ying (2004) and Mao & Wang (2010) proposed transformation mixture cure models. In an extension of the accelerated hazards model (Chen & Wang, 2000), Zhang & Peng (2009) studied an accelerated hazards mixture cure model. Lu (2010) further developed the accelerated failure time mixture cure model using kernel-smoothed non-parametric maximum likelihood estimation. In the quantile regression framework Wu & Yin (2013) proposed a mixture cure rate model which separates the entire population into two subgroups and fits censored quantile regression to the susceptible subgroup. More recently Choi et al. (2014) proposed a semiparametric inverse-Gaussian cure rate model by taking into account the patients' health status prior to their failure.

Although extensive research has been conducted on the mixture cure rate models, the Cox PH cure rate model proposed by Yakovlev & Tsodikov (1996) is a viable alternative to the mixture modelling counterparts. It also has been further studied by Tsodikov (1998); Chen, Ibrahim, & Sinha (1999); and Tsodikov, Ibrahim, & Yakovlev (2003). The PH cure rate model is also known as the bounded cumulative hazards cure rate model or the non-mixture cure rate model. Instead of dividing the underlying population into cured and uncured groups the PH cure rate model directly characterizes the covariate effects on the survival times of the entire population, including the fraction of infinite survival times. Zeng, Yin, & Ibrahim (2006) proposed a variant of the PH cure rate model which introduces a gamma frailty to accommodate the possible heterogeneity in survival data with a cure fraction. However it is difficult to check the gamma distributional assumption for the frailty, which can be misspecified in practice. As the underlying population contains a fraction of immune subjects with infinitely long survival times, the overall survival times are thus heavily right-skewed. It is well-known that quantile regression can handle outliers or extreme observations (infinity in our case) naturally, and thus there is no need to fit quantile regression to the subgroup of susceptible subjects as in Wu & Yin (2013). By considering the infinite survival times to be outliers but viable observations we propose to model the conditional quantiles of the survival times of the entire population directly. Using the reweighting scheme of Wang & Wang (2009) for redistributing probability mass we can achieve the goal of fitting quantile regression to the entire population. To characterize those subjects who are cured (with infinite survival times) and those who are susceptible (with finite survival times) a logistic regression is used to model the latent cure indicator. Compared with the familiar mean- or hazard-based regression methods our cure rate quantile regression model can directly evaluate covariate effects on the survival times of the entire population at different quantile levels, which would provide a more complete assessment of covariate effects. The robust feature of quantile regression makes it particularly appealing for such survival data as the infinite survival times would all fall on one side of the quantile regression lines. The proposed cure rate quantile regression model also naturally takes into account the heterogeneity without requiring any explicit distributional assumption. Our

method can be regarded as the counterpart of the quantile-based non-mixture cure rate analysis, which is a useful alternative to quantile-based mixture cure rate analysis (Wu & Yin, 2013).

The rest of this article is organized as follows. In Section 2 we propose to apply the cure rate quantile regression model to the entire population and introduce the estimation procedure through a cure rate weighting scheme. The asymptotic properties of the proposed estimators are established in Section 3. We conduct extensive simulation studies to examine the finite-sample performance of the proposed method in Section 4 and illustrate it with a real data example in Section 5. We conclude with some remarks in Section 6.

2. QUANTILE REGRESSION AND CURE RATE ESTIMATION

Let T denote the survival time of a subject in the underlying population. As T could be ∞ with positive probability we assume it has a decomposition as

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

where $T^* < \infty$ denotes the survival time of a susceptible subject and the indicator η takes a value of 1 if a subject is susceptible and 0 otherwise. Let $\mathbf{Z}$ be a $(p + 1)$-vector of covariates that includes 1 as an intercept. For a fixed $\tau \in (0, 1)$ the quantile regression model for T associated with covariate $\mathbf{Z}$ takes the form

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

where $Q_T(\tau|\mathbf{Z}) = \inf\{t\colon P(T \le t|\mathbf{Z}) \ge \tau\}$ is the τth conditional quantile function of survival time T given covariate $\mathbf{Z}$ and $\boldsymbol{\beta}(\tau)$ is an unknown $(p + 1)$-vector of regression coefficients. The key difference between Equation (1) and the work in Wu & Yin (2013) is that the latter models conditional quantiles of T^* instead of T directly. For ease of exposition hereafter we omit the quantile level τ in $\boldsymbol{\beta}(\tau)$ whenever there is no ambiguity. As there exists a constant $c_0 \in (0, 1)$ such that $P(T = \infty|\mathbf{Z}) > c_0$ we require $\tau \le 1 - c_0$ to ensure the identifiability of the model. Let C be the censoring time; then the observed time is denoted by $Y = T \wedge C$, the minimum of T and C. Let $\Delta = I(T \le C)$ be the censoring indicator. We assume that $(Y_i, \Delta_i, \mathbf{Z}_i)$, $i = 1, \ldots, n$, are independent and identically distributed copies of $(Y, \Delta, \mathbf{Z})$.

Let $F_T(t|\mathbf{z}) = P(T \le t|\mathbf{z})$ be the conditional cumulative distribution function (CDF) of the survival time T given $\mathbf{Z} = \mathbf{z}$. If $F_T(t|\mathbf{z})$ was known, we could define a weight function as in Wang & Wang (2009), that is,

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

Furthermore the quantile regression coefficient $\boldsymbol{\beta}$ could be estimated by minimizing the weighted objective function

$$Q_n(\boldsymbol{\beta}; F_T) = n^{-1} \sum_{i=1}^{n} \Big[\upsilon_i(F_T)\rho_\tau\{Y_i - \exp(\mathbf{Z}_i^\top \boldsymbol{\beta})\} + \{1 - \upsilon_i(F_T)\}\rho_\tau\{Y^\infty - \exp(\mathbf{Z}_i^\top \boldsymbol{\beta})\} \Big],$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the usual check function (Koenker, 2005) and Y^∞ is any sufficiently large value that exceeds all $\exp(\mathbf{Z}_i^\top \boldsymbol{\beta})$ for $i = 1, \ldots, n$. However both $F_T(t|\mathbf{Z}_i)$ and the weight function $\upsilon_i(F_T)$ are unknown. To carry out inference concerning $\boldsymbol{\beta}$ we first need to estimate F_T. As the survival times are a mixture of the susceptible and immune subjects, this places a point mass at infinity with positive probability. Thus the conditional CDF of T is improper and its estimation is nontrivial.

Following Farewell (1982) we use logistic regression to model the susceptibility indicator η, that is,

$$P(\eta = 1|\mathbf{Z}) = \pi(\boldsymbol{\gamma}^\top \mathbf{Z}) = \frac{\exp(\boldsymbol{\gamma}^\top \mathbf{Z})}{1 + \exp(\boldsymbol{\gamma}^\top \mathbf{Z})}. \tag{2}$$

Under the assumption that T^* and η are conditionally independent given $\mathbf{Z}$, it follows that $F_T(t|\mathbf{z}) = \pi(\boldsymbol{\gamma}^\top \mathbf{z}) F_{T^*}(t|\mathbf{z})$ for any $t < \infty$, where $F_{T^*}(t|\mathbf{z}) = P(T^* \le t|\mathbf{z})$ is the conditional CDF of T^* given $\mathbf{Z} = \mathbf{z}$. To obtain an estimator for $F_T(t|\mathbf{z})$ we need to estimate $\boldsymbol{\gamma}$ and $F_{T^*}(t|\mathbf{z})$ iteratively. Let $\boldsymbol{\gamma}_0$ be the true value of the parameter $\boldsymbol{\gamma}$ and F_{0T^*} be the true function F_{T^*}. We can extract the cure information to construct an estimating equation for $\boldsymbol{\gamma}$ (Lu & Ying, 2004). Subjects who have experienced the event of interest must belong to the susceptible subgroup, whereas censored observations may have been cured or may still remain susceptible; thus

$$P(\eta = 1|Y, \Delta, \mathbf{Z}) = \Delta + (1 - \Delta)\frac{\pi(\boldsymbol{\gamma}_0^\top \mathbf{Z})\{1 - F_{0T^*}(Y|\mathbf{Z})\}}{1 - \pi(\boldsymbol{\gamma}_0^\top \mathbf{Z}) + \pi(\boldsymbol{\gamma}_0^\top \mathbf{Z})\{1 - F_{0T^*}(Y|\mathbf{Z})\}}.$$

In conjunction with Equation (2) we define an estimating function

$$\mathbf{S}_n(\boldsymbol{\gamma}; F_{T^*}) = n^{-1} \sum_{i=1}^{n} \int_0^L \frac{\mathbf{Z}_i\{1 - \pi(\boldsymbol{\gamma}^\top \mathbf{Z}_i)\}}{1 - \pi(\boldsymbol{\gamma}^\top \mathbf{Z}_i) F_{T^*}(t|\mathbf{Z}_i)} \mathrm{d}M_i(t; \boldsymbol{\gamma}, F_{T^*}) \tag{3}$$

for $\boldsymbol{\gamma}$, where L is the study ending time and

$$\begin{aligned} M_i(t; \boldsymbol{\gamma}, F_{T^*}) &= N_i(t) - \Lambda_{T,\boldsymbol{\gamma}}(t \wedge Y_i|\mathbf{Z}_i), \\ N_i(t) &= \Delta_i I(Y_i \le t), \\ \Lambda_{T,\boldsymbol{\gamma}}(t|\mathbf{Z}_i) &= -\log\{1 - \pi(\boldsymbol{\gamma}^\top \mathbf{Z}_i) F_{T^*}(t|\mathbf{Z}_i)\}. \end{aligned}$$

According to martingale theory (Fleming & Harrington, 1991), solving $\mathbf{S}_n(\boldsymbol{\gamma}; F_{T^*}) = \mathbf{0}$ leads to an asymptotically consistent estimator for $\boldsymbol{\gamma}$, provided F_{0T^*} is known. To estimate $F_{T^*}(t|\mathbf{z})$, or equivalently the cumulative hazard function $\Lambda_{T^*}(t|\mathbf{z})$, we construct a local Nelson–Aalen type estimator in the context of cure rate analysis. Let $K_p(\cdot)$ denote a p-variate kernel function and let $h_n > 0$ be a bandwidth that converges to zero as $n \to \infty$. For ease of exposition we assume that $\mathbf{Z}$ contains only continuous covariates and thus adopt a multivariate product kernel $K_p(\mathbf{u}) = \prod_{j=1}^{p} K(u_j)$, where $K(\cdot)$ is a univariate kernel function and $\mathbf{u} = (u_1, \ldots, u_p)^\top \in \mathbb{R}^p$. Then

$$\widehat{\Lambda}_{T^*}(t|\mathbf{z}) = \int_0^t \frac{\sum_{i=1}^{n} B_{ni}(\mathbf{z}) \mathrm{d}N_i(u)}{\sum_{k=1}^{n} I(Y_k \ge u)\omega_k(\widehat{\boldsymbol{\gamma}}, \widehat{\Lambda}_{T^*}) B_{nk}(\mathbf{z})}. \tag{4}$$

In (4), $\omega_k(\boldsymbol{\gamma}, \Lambda_{T^*})$ accommodates the cure information (Sy & Taylor, 2000; Lu, 2010) and is defined by

$$\omega_k(\boldsymbol{\gamma}, \Lambda_{T^*}) = \Delta_k + (1 - \Delta_k)\frac{\pi(\boldsymbol{\gamma}^\top \mathbf{Z}_k)\exp\{-\Lambda_{T^*}(Y_k|\mathbf{z})\}}{1 - \pi(\boldsymbol{\gamma}^\top \mathbf{Z}_k) + \pi(\boldsymbol{\gamma}^\top \mathbf{Z}_k)\exp\{-\Lambda_{T^*}(Y_k|\mathbf{z})\}}, \tag{5}$$

whereas $B_{ni}(\mathbf{z})$ represents a sequence of Nadaraya–Watson type weights equal to

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

A consistent estimator of $F_{0T^*}(t|\mathbf{z})$ is given by $\widehat{F}_{T^*}(t|\mathbf{z}) = 1 - \exp\{-\widehat{\Lambda}_{T^*}(t|\mathbf{z})\}$.

In our numerical algorithm the initial value $\widehat{\boldsymbol{\gamma}}^{(0)}$ is set by evaluating a logistic regression of Δ on $\mathbf{Z}$, and $\widehat{\Lambda}^{(0)}_{T^*}(t|\mathbf{z})$ is obtained from (4) by setting all the ω_k's to be one, which then leads to $\omega_k^{(0)}$ in (5). At the mth iteration our algorithm for estimating $\boldsymbol{\gamma}_0$ proceeds as follows:

(i) Plug $\omega_k^{(m)}$ into (4) and obtain $\widehat{\Lambda}^{(m+1)}_{T^*}(t|\mathbf{z})$;
(ii) Substitute $\widehat{\Lambda}^{(m+1)}_{T^*}(t|\mathbf{z})$ into (3) and solve the resulting equation using the Newton–Raphson algorithm to obtain $\widehat{\boldsymbol{\gamma}}^{(m+1)}$;
(iii) Obtain $\omega_k^{(m+1)}$ by plugging $\widehat{\boldsymbol{\gamma}}^{(m+1)}$ and $\widehat{\Lambda}^{(m+1)}_{T^*}(t|\mathbf{z})$ into (5).

Repeat these three steps until a predetermined convergence criterion has been met; the resulting estimate is $\widehat{\boldsymbol{\gamma}}$. Both $\widehat{\boldsymbol{\gamma}}$ and $\widehat{F}_{T^*}(t|\mathbf{z})$ have been shown to be consistent estimators (Wu & Yin, 2013). It follows that $\widehat{F}_T(t|\mathbf{z}) = \pi(\widehat{\boldsymbol{\gamma}}^\top \mathbf{z})\widehat{F}_{T^*}(t|\mathbf{z})$, and hence $\widehat{\boldsymbol{\beta}}$ is the minimizer of $Q_n(\boldsymbol{\beta}; \widehat{F}_T)$. For identifiability and computational stability we set $\widehat{F}_{T^*}(t|\mathbf{z}) = 1$ if t is greater than the largest uncensored observation. This is a standard approach in the context of cure rate analysis.

Bandwidth selection plays an important role in non-parametric functional estimation, for which we recommend using a d-fold cross-validation method to select h_n. We randomly divide the data into d non-overlapping and roughly equal-sized subgroups. For the jth subgroup, $\mathcal{D}_j$, we fit the model using $\mathcal{D}_{(-j)}$, the data excluding subgroup j. Let $\widehat{\boldsymbol{\gamma}}_{(-j)}$ and $\widehat{F}_{T^*(-j)}$ denote the estimators of $\boldsymbol{\gamma}$ and F_{T^*} based on $\mathcal{D}_{(-j)}$. For each $\mathcal{D}_j$ we define

$$\mathcal{M}_j^{\mathrm{CV}}(h) = \frac{1}{|\{i\colon \Delta_i = 1 \text{ and } i \in \mathcal{D}_j\}|} \sum_{k \in \mathcal{D}_j} \int_0^L \{\mathcal{M}^{\mathrm{CV}}_{(-j)}(t, \mathbf{Z}_k)\}^2 \mathrm{d}N_k(t),$$

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

$$\mathcal{M}^{\mathrm{CV}}_{(-j)}(t, \mathbf{z}) = \frac{1}{|\{i\colon i \in \mathcal{D}_{(-j)}\}|} \times \sum_{i \in \mathcal{D}_{(-j)}} \int_0^t \frac{I(\mathbf{Z}_i \le \mathbf{z})\{1 - \pi(\widehat{\boldsymbol{\gamma}}^\top_{(-j)}\mathbf{Z}_i)\}}{1 - \pi(\widehat{\boldsymbol{\gamma}}^\top_{(-j)}\mathbf{Z}_i)\widehat{F}_{T^*(-j)}(u|\mathbf{Z}_i)} \mathrm{d}M_i(u; \widehat{\boldsymbol{\gamma}}_{(-j)}, \widehat{F}_{T^*(-j)}).$$

Finally we choose the optimal bandwidth by minimizing $\sum_{j=1}^d \mathcal{M}_j^{\mathrm{CV}}(h)$.

We adopt the usual resampling method for variance estimation (Lin et al., 2000). Thus we generate n independent variables $G_1, \ldots, G_n$ with unit mean and unit variance from the Exp(1) distribution, for example. We then multiply each individual term i in the estimating function $\mathbf{S}_n(\boldsymbol{\gamma}; F_{T^*})$ and the objective function $Q_n(\boldsymbol{\beta}; F_T)$ by the corresponding variate G_i and obtain

$$\mathbf{S}^*_n(\boldsymbol{\gamma}; F_{T^*}) = n^{-1} \sum_{i=1}^n \int_0^L \frac{\mathbf{Z}_i\{1 - \pi(\boldsymbol{\gamma}^\top \mathbf{Z}_i)\}}{1 - \pi(\boldsymbol{\gamma}^\top \mathbf{Z}_i)F_{T^*}(t|\mathbf{Z}_i)} \mathrm{d}M_i(t; \boldsymbol{\gamma}, F_{T^*})G_i$$

and

$$Q^*_n(\boldsymbol{\beta}; F_T) = n^{-1} \sum_{i=1}^n \left[\upsilon_i(F_T)\rho_\tau\{Y_i - \exp(\mathbf{Z}_i^\top \boldsymbol{\beta})\} + \{1 - \upsilon_i(F_T)\}\rho_\tau\{Y^\infty - \exp(\mathbf{Z}_i^\top \boldsymbol{\beta})\}\right] G_i,$$

respectively. By using this proposed iterative algorithm, finding the root of $\mathbf{S}^*_n(\boldsymbol{\gamma}; F_{T^*})$ and the minimum of $Q^*_n(\boldsymbol{\beta}; F_T)$, we obtain one resampling pair of estimates. If we repeat this procedure a large number of times we can obtain resampling variances for the proposed estimators.

3. ASYMPTOTIC PROPERTIES

Denote $F_C(t|\mathbf{z}) = P(C \le t|\mathbf{z})$, $\bar{F}_C(t|\mathbf{z}) = 1 - F_C(t|\mathbf{z})$, $f_{0T^*}(t|\mathbf{z}) = \mathrm{d}F_{0T^*}(t|\mathbf{z})/\mathrm{d}t$, and $f_C(t|\mathbf{z}) = \mathrm{d}F_C(t|\mathbf{z})/\mathrm{d}t$. Let $\mathcal{Z}$ be the domain of covariates $\mathbf{Z}$ and impose the following conditions throughout this derivation.

C1 Let $\mathcal{B} \subset \mathbb{R}^{p+1}$ and $\mathcal{K} \subset \mathbb{R}^{p+1}$ be compact sets that contain $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ as interior points, respectively. With probability one $\mathbf{Z}$ is bounded and $E(\mathbf{Z}\mathbf{Z}^\top) > 0$.

C2 With a compact support in $\mathbb{R}$, $K(\cdot)$ is an ℓth order kernel function satisfying $\int_{\mathbb{R}} K(u)\mathrm{d}u = 1$, $\int_{\mathbb{R}} K^2(u)\mathrm{d}u < \infty$, $\int_{\mathbb{R}} u^j K(u)\mathrm{d}u = 0$ for $1 \le j < \ell$, $\int_{\mathbb{R}} u^\ell K(u)\mathrm{d}u \ne 0$, and $\int_{\mathbb{R}} |u|^\ell K(u)\mathrm{d}u < \infty$. Moreover it is Lipschitz continuous of order ℓ with $\ell \ge 2$.

C3 The first ℓ partial derivatives of $f_{\mathbf{Z}}(\mathbf{z})$, the density function of $\mathbf{Z}$, are uniformly bounded for $\mathbf{z} \in \mathcal{Z}$, and $f_{0T^*}(t|\mathbf{z})$ and $f_C(t|\mathbf{z})$ are bounded away from infinity uniformly for $t \in (0, L]$ and $\mathbf{z} \in \mathcal{Z}$, and the first ℓ partial derivatives of $f_{0T^*}(t|\mathbf{z})$ and $f_C(t|\mathbf{z})$ with respect to $\mathbf{z}$ are uniformly bounded for t and $\mathbf{z}$. Moreover $\pi(\boldsymbol{\gamma}^\top \mathbf{z})$ is uniformly bounded away from 0 and 1 for $\mathbf{z} \in \mathcal{Z}$ and $\boldsymbol{\gamma} \in \mathcal{K}$, and τ is smaller than the lower bound of $\pi(\boldsymbol{\gamma}_0^\top \mathbf{z})$.

C4 For $\boldsymbol{\gamma}$ in a neighbourhood of $\boldsymbol{\gamma}_0$ the matrix

$$\begin{aligned}\boldsymbol{\Gamma}(\boldsymbol{\gamma}; F_{0T^*}) &= -\frac{\partial E\{\mathbf{S}_n(\boldsymbol{\gamma}; F_{0T^*})\}}{\partial \boldsymbol{\gamma}} \\ &= E\left[\int_0^L \frac{\mathbf{Z}^{\otimes 2}\pi(\boldsymbol{\gamma}^\top\mathbf{Z})\{1-\pi(\boldsymbol{\gamma}^\top\mathbf{Z})\}\{1-F_{0T^*}(t|\mathbf{Z})\}}{\{1-\pi(\boldsymbol{\gamma}^\top\mathbf{Z})F_{0T^*}(t|\mathbf{Z})\}^2}\mathrm{d}M(t; \boldsymbol{\gamma}, F_{0T^*})\right] \\ &\quad + E\left[\int_0^L \left(\frac{\mathbf{Z}\{1-\pi(\boldsymbol{\gamma}^\top\mathbf{Z})\}}{1-\pi(\boldsymbol{\gamma}^\top\mathbf{Z})F_{0T^*}(t|\mathbf{Z})}\right)^{\otimes 2} I(X \ge t)\mathrm{d}\Lambda_{T,\boldsymbol{\gamma}}(t|\mathbf{Z})\right] \qquad (6)\end{aligned}$$

is positive definite, where $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^\top$ for any column vector $\mathbf{a}$.

C5 For $\boldsymbol{\beta}$ in a neighbourhood of $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}$ in a neighbourhood of $\boldsymbol{\gamma}_0$

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

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 $\ell > 3p/2$.

The bounded aspects of the parameter spaces and covariates in condition C1 are common assumptions in survival analysis. Condition C3 enables Taylor's expansion to determine the order of convergence of the estimators. For ease of exposition we assume $\mathbf{Z}$ to be continuous, whereas discrete covariates can be handled by replacing integration with summation over the probability mass functions. Furthermore if Z_1, the first component of $\mathbf{Z}$, is a discrete covariate, for example, the corresponding kernel function is simply $K(u_1) = I(u_1 = 0)$. This is the routine treatment for the kernel smoothing method when discrete covariates are involved; this approach is also known as stratification by discrete covariates. Due to censoring as well as the existence of a cure fraction in the underlying population, we require that the quantile level τ be smaller than the lower bound of $\pi(\boldsymbol{\gamma}_0^\top \mathbf{z})$ to guarantee model identifiability. Conditions C4 and C5 are the positive definite assumptions for the "Hessian" matrices of the regression parameters. These two conditions obviously hold at the true parameter values. The theoretical verification of these two conditions requires further assumptions as discussed in Asgharian (2014). Condition C6 states the convergence rate of the bandwidth which is needed to obtain the consistency of the proposed estimators, whereas condition C7 is a strengthened version which is required to establish the

weak convergence properties. Due to the dependence of ℓ on p, a higher order kernel function is required to control the bias for a larger p (Wang, Zhou, & Li, 2013). Overall these conditions are mild and are commonly required in survival analysis and kernel smoothing theory.

Suppose that there exists another group of parameters $\boldsymbol{\beta}^\dagger$, $\boldsymbol{\gamma}^\dagger$, and $F^\dagger_{T^*}$ such that (i) $\pi(\mathbf{z}^\top\boldsymbol{\gamma}^\dagger)F^\dagger_{T^*}(t|\mathbf{z}) = \pi(\mathbf{z}^\top\boldsymbol{\gamma})F_{T^*}(t|\mathbf{z})$ and (ii) $\exp(\mathbf{z}^\top\boldsymbol{\beta}^\dagger) = \exp(\mathbf{z}^\top\boldsymbol{\beta})$. Obviously (ii) implies $\boldsymbol{\beta}^\dagger = \boldsymbol{\beta}$ under condition C1. And when $t \to \infty$ in (i) we have $\pi(\mathbf{z}^\top\boldsymbol{\gamma}^\dagger) = \pi(\mathbf{z}^\top\boldsymbol{\gamma})$. Then $\boldsymbol{\gamma}^\dagger = \boldsymbol{\gamma}$ under conditions C1 and C3 and $F^\dagger_{T^*} = F_{T^*}$ follows directly. Thus the identifiability of model parameters is guaranteed. The asymptotic results are summarized in the following two theorems.

Theorem 1. *Under conditions C1–C4 and C6 $\widehat{\boldsymbol{\gamma}}$ converges in probability to $\boldsymbol{\gamma}_0$ as $n \to \infty$. Under conditions C1–C4 and C7 we have*

$$n^{1/2}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\xi}(\mathcal{O}_i; \boldsymbol{\gamma}_0, F_{0T^*}) + o_P(1), \tag{7}$$

where $\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}(\boldsymbol{\gamma}_0, F_{0T^*})$, $\mathcal{O}_i = (Y_i, \Delta_i, \mathbf{Z}_i)$, and $\boldsymbol{\xi}$ is a measurable function from $\mathcal{O}$ to $\mathbb{R}^{p+1}$ with a zero mean and finite variance. Thus $n^{1/2}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ is asymptotically normal with a zero mean and variance–covariance matrix $\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{-1}$, where $\boldsymbol{\Omega}_0 = E\{\boldsymbol{\xi}(\mathcal{O}; \boldsymbol{\gamma}_0, F_{0T^*})^{\otimes 2}\}$.

The consistency and asymptotic normality of Theorem 1 follows directly from Wu & Yin (2013) and thus the proofs are omitted. Under conditions C1–C4 and C6, and following Theorem 2.3 of Liang, de Una-Alvarez, & Iglesias-Perez (2012), we obtain the Bahadur representation of $\widehat{F}_{T^*}(t|\mathbf{z})$,

$$\widehat{F}_{T^*}(t|\mathbf{z}) - F_{0T^*}(t|\mathbf{z}) = \frac{1}{nh_n^p f_{\mathbf{Z}}(\mathbf{z})}\sum_{i=1}^{n} K_p\left(\frac{\mathbf{z} - \mathbf{Z}_i}{h_n}\right)\psi(\mathcal{O}_i, t, \mathbf{z}; \boldsymbol{\gamma}_0, F_{0T^*}) + O_P(\alpha_n), \tag{8}$$

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

Theorem 2. *Under conditions C1–C6 $\widehat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$ as $n \to \infty$. Under conditions C1–C5 and C7 we have*

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\Phi}_0^{-1}\boldsymbol{\phi}(\mathcal{O}_i; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, F_{0T^*}) + o_P(1), \tag{9}$$

where $\boldsymbol{\Phi}_0 = \boldsymbol{\Phi}(\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ and $\boldsymbol{\phi}$ is a measurable function from $\mathcal{O}$ to $\mathbb{R}^{p+1}$ with a zero mean and finite variance. Thus $n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically normal with a zero mean and variance–covariance matrix $\boldsymbol{\Phi}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Phi}_0^{-1}$, where $\boldsymbol{\Sigma}_0 = E\{\boldsymbol{\phi}(\mathcal{O}; \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, F_{0T^*})^{\otimes 2}\}$.

Obviously $\widehat{F}_T(t|\mathbf{z}) = \pi(\widehat{\boldsymbol{\gamma}}^\top\mathbf{z})\widehat{F}_{T^*}(t|\mathbf{z})$ is a uniformly consistent estimator for $F_{0T}(t|\mathbf{z})$ following Theorem 1 and the Bahadur representation in Equation (8). It follows that Theorem 2 holds immediately (Wang & Wang, 2009). Detailed formulations of the measurable functions $\boldsymbol{\xi}$, ψ, and $\boldsymbol{\phi}$ are provided in the Supplementary Material for this article.

4. SIMULATION STUDIES

We conducted extensive simulation studies to examine the effectiveness of the proposed cure rate quantile regression method. We first generated the susceptibility indicator η from the logistic regression model in Equation (2) with $\mathbf{Z} = (1, Z)^\top$ and the true parameter value $\boldsymbol{\gamma}_0^\top = (\gamma_0, \gamma_1) =$

$(1, -1)$, where $Z \sim \text{Unif}(0, 1)$. We generated the survival times T^* of susceptible subjects from the quantile regression model

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

where $\boldsymbol{\beta}_0^\top = (\beta_0, \beta_1) = (2, 1)$, $\tau^* = \tau/\pi(\boldsymbol{\gamma}_0^\top \mathbf{Z})$, and the error ϵ_{τ^*} follows a normal distribution with the τ^*th quantile of zero. If $\eta = 0$ we set T to be a very large number, say 10^9; otherwise $T = T^*$. We simulated the censoring time $C = \widetilde{C} \wedge L$, where $\widetilde{C}$ was generated from Unif$(l_0, L + 2)$ if $Z < 0.5$, and from Unif$(l_0 + 1, L + 2)$ otherwise. The constant l_0 and the study duration time L were chosen to yield a censoring rate of 45%. The cure rate was approximately 38%. We adopted the biquadratic kernel function and selected the bandwidth h_n via the proposed 10-fold cross-validation procedure. We replicated 1,000 simulations for each configuration.

Table 1 summarizes the simulation results for the proposed method. For the normal model error ϵ_τ we consider quantile levels $\tau = 0.1$ and 0.3, coupled with two different sample sizes $n = 200$ and 400. The biases of the estimates for the cure rate parameters are essentially negligible across the considered quantile levels. As expected more precise estimates are obtained as n increases to 400. Estimation of the quantile regression parameters is reasonably good. When the model error ϵ_τ follows the extreme value distribution or the heavy-tailed student t distribution with two degrees of freedom similar conclusions can be drawn.

We also conducted simulations to evaluate the bootstrap standard errors; these results are summarized in Table 2. Clearly the estimated standard errors (ESE) that we obtained using the bootstrap method agree with the sample standard errors (SE), and the coverage probabilities (CP) of the bootstrap confidence intervals are close to the nominal level of 95%.

5. MELANOMA DATA EXAMPLE

We illustrate our proposed cure rate quantile regression method by applying it to data from a phase III melanoma clinical trial conducted by the Eastern Cooperative Oncology Group (Kirkwood et al., 2000). In this study there were 212 patients in the high-dose interferon arm and 205 patients in the control arm. The response variable was the relapse-free survival time (in years) and about 42.4% of the patients' survival times were censored. In the high-dose interferon arm 113 patients experienced the event, and in the control arm 127 patients relapsed. The covariates of interest included in this analysis were treatment (51% high-dose interferon = 1; 49% control = 0), age (ranging from 19.13 to 78.05 years with mean 48.05 years), sex (37% female = 1; 63% male = 0) and nodal status taking a value of 0 if there was no positive node or 1 if a patient had one or more positive nodes (27% nodal status = 0; 73% nodal status = 1). The covariate age was standardized to have mean 0 and variance 1. Our analysis focused on evaluating the treatment effect of the high-dose interferon on patients' relapse-free survival. The median follow-up time for this study was approximately 4 years, which was considered sufficient follow-up for this disease. Figure 1 shows the Kaplan–Meier survival curves for patients with melanoma in the high-dose interferon arm and the control arm, respectively. A stable plateau can be observed after approximately 5.5 years of relapse-free survival, which indicates the possible existence of a cure fraction.

We applied the proposed models outlined in (1) and (2) to fit the melanoma data by taking $\mathbf{Z} = (1, \text{Treatment}, \text{Age}, \text{Sex}, \text{Nodal})^\top$. Table 3 displays the estimates of covariate effects under the quantile and cure rate regression, and the corresponding P-values on the basis of 200 bootstrap samples. It can be seen that patients in the high-dose interferon arm survived significantly longer than those in the control arm at the $\tau = 0.1$ quantile level of the relapse-free survival time, whereas the adjusted treatment effects were not significant for $\tau = 0.3$ and 0.5. Patients with no positive node survived significantly longer than those with one or more positive nodes at all three quantiles

TABLE 1: Simulation results under the proposed cure rate quantile regression model with sample sizes $n = 200$ and 400 and three different distributions of model errors

				Intercept: $\gamma_0 = 1$ and $\beta_0 = 2$				Slope: $\gamma_1 = -1$ and $\beta_1 = 1$			
ϵ_τ	τ	n	Model parameter	Est.	Bias	SE	MSE	Est.	Bias	SE	MSE
Normal	0.1	200	$\boldsymbol{\gamma}$	1.0248	0.0248	0.3449	0.1194	−1.0398	−0.0398	0.5863	0.3450
			$\boldsymbol{\beta}$	1.9958	−0.0042	0.3246	0.1053	1.0210	0.0210	0.6994	0.4891
		400	$\boldsymbol{\gamma}$	1.0125	0.0125	0.2369	0.0562	−1.0118	−0.0118	0.3995	0.1595
			$\boldsymbol{\beta}$	1.9864	−0.0136	0.2264	0.0514	1.0141	0.0141	0.4769	0.2274
	0.3	200	$\boldsymbol{\gamma}$	1.0148	0.0148	0.3438	0.1183	−1.0211	−0.0211	0.5801	0.3366
			$\boldsymbol{\beta}$	2.0000	0.0000	0.3132	0.0980	0.9818	−0.0182	0.7226	0.5220
		400	$\boldsymbol{\gamma}$	1.0089	0.0089	0.2365	0.0559	−1.0071	−0.0071	0.3964	0.1571
			$\boldsymbol{\beta}$	1.9910	−0.0090	0.2218	0.0492	0.9945	−0.0055	0.5064	0.2563
Extreme	0.1	200	$\boldsymbol{\gamma}$	1.0374	0.0374	0.3320	0.1115	−1.0507	−0.0507	0.5749	0.3327
			$\boldsymbol{\beta}$	1.9907	−0.0093	0.5492	0.3014	0.9944	−0.0056	1.1960	1.4291
		400	$\boldsymbol{\gamma}$	1.0113	0.0113	0.2361	0.0558	−1.0092	−0.0092	0.3985	0.1588
			$\boldsymbol{\beta}$	1.9959	−0.0041	0.3845	0.1477	1.0087	0.0087	0.8003	0.6399
	0.3	200	$\boldsymbol{\gamma}$	1.0124	0.0124	0.3436	0.1181	−1.0162	−0.0162	0.5782	0.3342
			$\boldsymbol{\beta}$	1.9851	−0.0149	0.3802	0.1446	0.9696	−0.0304	0.8292	0.6878
		400	$\boldsymbol{\gamma}$	1.0065	0.0065	0.2357	0.0556	−1.0046	−0.0046	0.3952	0.1560
			$\boldsymbol{\beta}$	1.9806	−0.0194	0.2691	0.0727	0.9783	−0.0217	0.5819	0.3388
$t_{(2)}$	0.1	200	$\boldsymbol{\gamma}$	1.0192	0.0192	0.3650	0.1334	−1.0482	−0.0482	0.6047	0.3676
			$\boldsymbol{\beta}$	1.9742	−0.0258	0.6518	0.4250	0.9994	−0.0006	1.3166	1.7317
		400	$\boldsymbol{\gamma}$	1.0240	0.0240	0.2561	0.0661	−1.0557	−0.0557	0.4229	0.1818
			$\boldsymbol{\beta}$	1.9907	−0.0093	0.4316	0.1862	1.0076	0.0076	0.8522	0.7256
	0.3	200	$\boldsymbol{\gamma}$	1.0198	0.0198	0.3576	0.1282	−1.0364	−0.0364	0.5994	0.3602
			$\boldsymbol{\beta}$	1.9892	−0.0108	0.3636	0.1322	0.9889	−0.0111	0.8256	0.6810
		400	$\boldsymbol{\gamma}$	1.0134	0.0134	0.2558	0.0655	−1.0314	−0.0314	0.4190	0.1764
			$\boldsymbol{\beta}$	1.9838	−0.0162	0.2563	0.0659	1.0036	0.0036	0.5785	0.3343

"Est." is the average value of the parameter estimates, "Bias" is the average difference between Est. and the true value, "SE" is the sample standard error of the estimates, and "MSE" is the mean squared error of the parameter estimates.

of the response variable fitted. Female patients tended to have significantly longer survival than male patients when $\tau = 0.5$, but similar effects were not detected when $\tau = 0.1$ or 0.3. The results suggested that patients' age had no apparent effect on their relapse-free survival for all values of τ that we considered. With respect to the cure rate the covariates sex and nodal status were not significantly associated with susceptibility status, whereas patient age was significantly associated

TABLE 2: Simulation results for the bootstrap estimated standard errors and the estimated coverage probabilities when the sample size $n = 200$

		Model	Intercept: $\gamma_0 = 1$ and $\beta_0 = 2$			Slope: $\gamma_1 = -1$ and $\beta_1 = 1$		
ϵ_τ	τ	parameter	ESE	ESE/SE	CP (%)	ESE	ESE/SE	CP (%)
Normal	0.1	$\boldsymbol{\gamma}$	0.3481	1.0003	96.4	0.5881	0.9959	95.6
		$\boldsymbol{\beta}$	0.3244	1.0217	95.0	0.7087	1.0004	94.8
	0.3	$\boldsymbol{\gamma}$	0.3465	0.9943	96.0	0.5796	1.0019	95.6
		$\boldsymbol{\beta}$	0.3193	1.0234	95.4	0.7558	1.0321	94.8
Extreme	0.1	$\boldsymbol{\gamma}$	0.3466	1.0043	96.2	0.5860	1.0012	95.6
		$\boldsymbol{\beta}$	0.5723	1.0823	95.2	1.2098	1.0380	95.4
	0.3	$\boldsymbol{\gamma}$	0.3440	0.9908	96.0	0.5752	0.9995	95.4
		$\boldsymbol{\beta}$	0.3829	1.0181	94.8	0.8587	1.0200	94.6
$t_{(2)}$	0.1	$\boldsymbol{\gamma}$	0.4014	1.0878	97.0	0.6693	1.0658	96.4
		$\boldsymbol{\beta}$	0.7004	1.1960	96.6	1.4177	1.1359	96.6
	0.3	$\boldsymbol{\gamma}$	0.3963	1.0722	96.8	0.6584	1.0338	96.0
		$\boldsymbol{\beta}$	0.3727	1.0307	96.0	0.8787	1.0583	95.4

"ESE" is the average of the bootstrap estimated standard errors, "SE" is the sample standard error of the estimates, "ESE/SE" is the ratio of ESE to SE, and "CP" is the estimated coverage probability of the 95% confidence intervals obtained using the bootstrap.

with the cure rate, indicating that younger patients were more likely to be long-term relapse-free survivors of melanoma. There also was a trend for patients in the high-dose interferon arm to experience longer relapse-free survival.

We assess the validity of the logistic regression for the susceptible indicators and the proposed local Nelson–Aalen type estimator using the predicted survival curves under the model specified in Equation (2). These are obtained by averaging the estimated survival curves over all the covariates except the treatment indicator. Figure 1 shows that the predicted survival curves are very similar to the corresponding Kaplan–Meier survival curves, indicating that our model fits the data very well. To quantify the overall fit of the logistic regression model, we further consider the cumulative residuals

$$\mathcal{T}_n(\mathbf{z}) = n^{-1/2} \sum_{i=1}^{n} \int_0^L \frac{1 - \pi(\widehat{\boldsymbol{\gamma}}^\top \mathbf{Z}_i)}{1 - \pi(\widehat{\boldsymbol{\gamma}}^\top \mathbf{Z}_i) \widehat{F}_{T^*}(t|\mathbf{Z}_i)} \mathrm{d}M_i(t; \widehat{\boldsymbol{\gamma}}, \widehat{F}_{T^*}) I(\mathbf{Z}_i \le \mathbf{z})$$

over the covariates, where $I(\mathbf{Z}_i \le \mathbf{z})$ is the indicator that each of the $p + 1$ components of $\mathbf{Z}_i$ is no larger than the corresponding component of $\mathbf{z}$. The null distribution of $\mathcal{T}_n(\mathbf{z})$ can be approximated by the zero-mean process

$$\mathcal{T}_n^*(\mathbf{z}) = n^{-1/2} \sum_{i=1}^{n} \int_0^L \frac{1 - \pi(\widehat{\boldsymbol{\gamma}}^\top \mathbf{Z}_i)}{1 - \pi(\widehat{\boldsymbol{\gamma}}^\top \mathbf{Z}_i) \widehat{F}_{T^*}(t|\mathbf{Z}_i)} \mathrm{d}M_i(t; \widehat{\boldsymbol{\gamma}}, \widehat{F}_{T^*}) I(\mathbf{Z}_i \le \mathbf{z}) G_i,$$

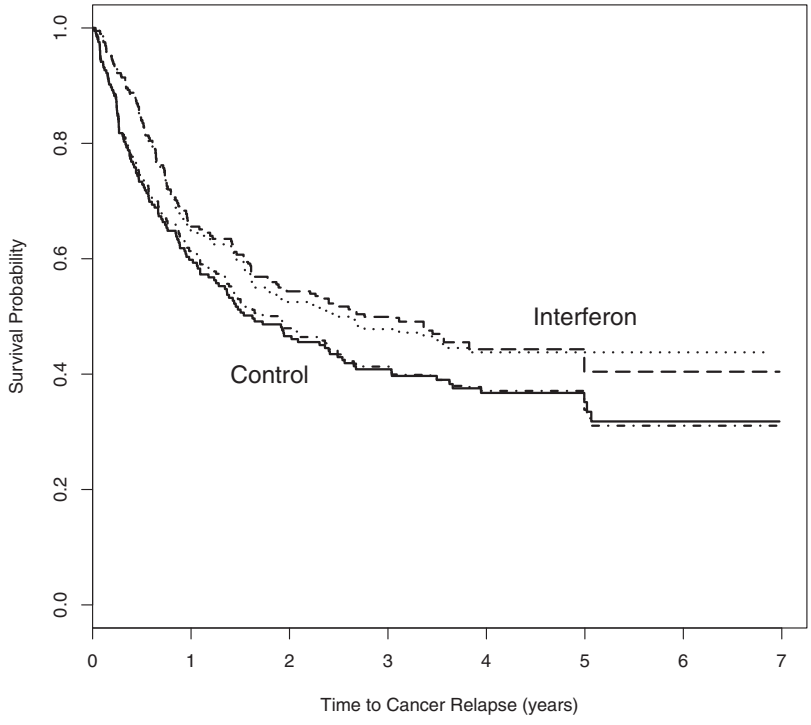

FIGURE 1: Kaplan–Meier survival curves (dotted and solid lines) and the corresponding predicted survival curves (dashed and dot-dashed lines) under the proposed model for patients with melanoma in the interferon and control arms, respectively

where $(G_1, \ldots, G_n)$ are generated independently from the standard normal distribution while fixing the data $\{(Y_i, \Delta_i, \mathbf{Z}_i), i = 1, \ldots, n\}$ at their observed values. The supremum statistic $\sup_{\mathbf{z}} |\mathcal{T}_n(\mathbf{z})|$ can be used to test the overall fit of our logistic regression model for the susceptibility status. We generated a large number of realizations from $\sup_{\mathbf{z}} |\mathcal{T}_n^*(\mathbf{z})|$, say 1,000, and obtained its empirical 95th percentile as 1.122. The observed value of $\sup_{\mathbf{z}} |\mathcal{T}_n(\mathbf{z})|$ is 0.2, which indicates that the logistic regression model fits the susceptible indicators very well. To evaluate the fit of the cure rate quantile regression model we define another version of the cumulative residuals over the covariates, namely

$$\mathcal{R}_n(\mathbf{z}, \tau) = n^{-1/2} \sum_{i=1}^{n} \{\tau - \upsilon_i(\widehat{F}_T) I(\log Y_i \leq \widehat{\boldsymbol{\beta}}^\top \mathbf{Z}_i)\} I(\mathbf{Z}_i \leq \mathbf{z}).$$

The null distribution of $\mathcal{R}_n(\mathbf{z}, \tau)$ can be similarly approximated by the zero-mean process

$$\mathcal{R}_n^*(\mathbf{z}, \tau) = n^{-1/2} \sum_{i=1}^{n} \{\tau - \upsilon_i(\widehat{F}_T) I(\log Y_i \leq \widehat{\boldsymbol{\beta}}^\top \mathbf{Z}_i)\} I(\mathbf{Z}_i \leq \mathbf{z}) G_i.$$

For $\tau = 0.1$ the 95th percentile of 1,000 realizations of $\sup_{\mathbf{z}} |\mathcal{R}_n^*(\mathbf{z}, 0.1)|$ is 0.822, whereas the observed value of $\sup_{\mathbf{z}} |\mathcal{R}_n(\mathbf{z}, 0.1)|$ is 0.176. The corresponding two quantities at $\tau = 0.3$ are 1.233 and 0.43, and those at $\tau = 0.5$ are 1.316 and 0.313. As a result we conclude that the cure rate quantile regression model also fits these melanoma data very well.

TABLE 3: Analysis results for the melanoma data using the cure rate quantile regression model

Parameter	Covariate	Estimate	Std. Error	P-value
$\boldsymbol{\beta}\,(\tau = 0.1)$	Intercept	−0.9599	0.2638	0.0003
	Treatment	0.5534	0.2640	0.0361
	Age	−0.0176	0.1202	0.8839
	Sex	−0.2341	0.2827	0.4076
	Nodal	−0.8598	0.2421	0.0004
$\boldsymbol{\beta}\,(\tau = 0.3)$	Intercept	0.0293	0.2855	0.9182
	Treatment	0.2859	0.2410	0.2354
	Age	−0.1341	0.1111	0.2275
	Sex	0.0230	0.2712	0.9325
	Nodal	−0.7397	0.2919	0.0113
$\boldsymbol{\beta}\,(\tau = 0.5)$	Intercept	0.8358	0.2117	0.0001
	Treatment	0.3635	0.2549	0.1539
	Age	−0.1915	0.1327	0.1492
	Sex	0.6100	0.2741	0.0260
	Nodal	−0.7479	0.2681	0.0053
$\boldsymbol{\gamma}$	Intercept	1.3135	0.4835	0.0066
	Treatment	−0.5820	0.3224	0.0710
	Age	0.4611	0.1632	0.0047
	Sex	−0.1634	0.3311	0.6217
	Nodal	−0.0344	0.4678	0.9414

As a comparison we also fitted the melanoma data using the Cox PH mixture and non-mixture cure rate regression methods and the traditional Cox PH regression model. The results summarized in Table 4 show there is no significant association between treatment and the relapse-free survival times of either the entire population or the susceptible subjects, whereas our fitted cure rate quantile regression model indicates that when $\tau = 0.1$ patients in the two treatment arms experienced significantly different relapse-free survival. As one might expect, the Cox PH non-mixture cure rate regression model produces results that are very similar to those derived from a traditional Cox PH regression model that ignores the cure fraction. But the former model can provide an estimate of cure probability via the quantity $\exp\{-\exp(\widehat{\boldsymbol{\beta}}^{\top}\mathbf{z})\}$ for each patient with covariate $\mathbf{z}$, which is not furnished by the traditional Cox PH model. We also applied the quantile regression method of Wang & Wang (2009) to the melanoma data; it also does not include a cure fraction. Those results are summarized in Table 5. The estimates for $\boldsymbol{\beta}(\tau)$ found therein are very similar to those obtained using our proposed cure rate quantile regression model. The reason is that no matter which quantile is fitted, subjects with survival times of ∞ must lie to the right of the fitted quantile regression line. Nonetheless through the logistic regression, our proposed cure rate quantile regression can provide an in-depth analysis of covariate effects on the cure rate; these are often viewed as long-term effects of the covariates. This kind

TABLE 4: Analysis results for the melanoma data using the Cox PH mixture and non-mixture cure rate regression models and the traditional Cox PH regression model

Parameter	Covariate	Estimate	Std. Error	P-value
	Cox PH mixture cure rate regression model			
$\boldsymbol{\beta}$	Treatment	−0.0615	0.1747	0.7249
	Age	−0.0910	0.0901	0.3129
	Sex	−0.0297	0.1909	0.8762
	Nodal	0.5803	0.1689	0.0006
$\boldsymbol{\gamma}$	Intercept	0.7613	0.3365	0.0237
	Treatment	−0.3239	0.3095	0.2954
	Age	0.3729	0.1560	0.0168
	Sex	−0.3689	0.2880	0.2001
	Nodal	0.3551	0.3435	0.3013
	Cox PH non-mixture cure rate regression model			
$\boldsymbol{\beta}$	Intercept	−0.0985	0.1576	0.5319
	Treatment	−0.2342	0.1365	0.0862
	Age	0.1630	0.0711	0.0219
	Sex	−0.2524	0.1479	0.0880
	Nodal	0.5618	0.1659	0.0007
	Traditional Cox PH regression model			
$\boldsymbol{\beta}$	Treatment	−0.2287	0.1302	0.0790
	Age	0.1510	0.0663	0.0230
	Sex	−0.2332	0.1378	0.0910
	Nodal	0.5501	0.1602	0.0006

of information cannot be inferred by using a quantile regression model that ignores the cure fraction.

6. REMARKS

We have proposed a cure rate quantile regression method to analyze censored data when there is a possibility of cure. Without requiring global linearity, our cure rate model can directly examine the effects of covariates on the survival times, finite or infinite, of the entire underlying population at any specific quantile level. Due to its robust features quantile regression is better suited to extremely right-skewed survival data, especially when the data involve infinite survival times. In theory we require that $\tau < \pi(\boldsymbol{\gamma}_0^\top \mathbf{Z})$ to ensure model identifiability. For a given data set we can estimate the quantity $\pi(\boldsymbol{\gamma}_0^\top \mathbf{Z})$ for all subjects. If we first set τ to be close to $\max_{1\le i\le n} \pi(\widehat{\boldsymbol{\gamma}}^\top \mathbf{Z}_i)$, we can then select the final τ in an adaptive manner. If the regression quantiles at τ can be estimated we try to increase τ by some small step size, for example, 0.05 or 0.1; otherwise we decrease τ

TABLE 5: Analysis results for the melanoma data using the quantile regression model of Wang & Wang (2009)

Parameter	Covariate	Estimate	Std. Error	P-value
$\boldsymbol{\beta}\,(\tau = 0.1)$	Intercept	−0.9599	0.2534	0.0002
	Treatment	0.5534	0.2625	0.0350
	Age	−0.0176	0.1228	0.8860
	Sex	−0.2341	0.2751	0.3948
	Nodal	−0.8598	0.2366	0.0003
$\boldsymbol{\beta}\,(\tau = 0.3)$	Intercept	0.0288	0.2512	0.9087
	Treatment	0.2854	0.2202	0.1949
	Age	−0.1353	0.1122	0.2279
	Sex	0.0093	0.2511	0.9705
	Nodal	−0.7384	0.2648	0.0053
$\boldsymbol{\beta}\,(\tau = 0.5)$	Intercept	0.8533	0.2041	<0.0001
	Treatment	0.3547	0.2432	0.1447
	Age	−0.1915	0.1285	0.1362
	Sex	0.6013	0.2549	0.0183
	Nodal	−0.7566	0.2405	0.0017

by a small value. Through such a trial-and-error approach we can push τ to the largest value at which all the model parameters can be identified.

ACKNOWLEDGEMENTS

The authors thank the editor, the associate editor, and the referees 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.

BIBLIOGRAPHY

Asgharian, M. (2014). On the singularities of the information matrix and multipath change-point problems. *Theory of Probability and Its Applications*, 58, 546–561.

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

Chen, M. H., Ibrahim, J. G., & Sinha, D. (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94, 909–919.

Chen, Y. Q. & Wang, M. C. (2000). Analysis of accelerated hazards models. *Journal of the American Statistical Association*, 95, 608–618.

Choi, S., Huang, X., Cormier, J. N., & Doksum, K. A. (2014). A semiparametric inverse-Gaussian model and inference for survival data with a cured proportion. *The Canadian Journal of Statistics*, 42, 635–649.

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. & Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Kirkwood, J. M., Ibrahim, J. G., Sondak, V. K., Richards, J., Flaherty, L. E., Ernstoff, M. S., Smith, T. J., Rao, U., Steele, M., & Blum, R. H. (2000). High- and low-dose interferon alfa-2b in high-risk melanoma: First analysis of intergroup trial E1690/S9111/C9190. *Journal of Clinical Oncology*, 18, 2444–2458.

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

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

Liang, H.-Y., de Una-Alvarez, J., & 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., & 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. & Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika*, 91, 331–343.

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

Tsodikov, A. D. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, 54, 1508–1516.

Tsodikov, A. D., Ibrahim, J. G., & Yakovlev, A. Y. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association*, 98, 1063–1078.

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

Wang, H., Zhou, J., & Li, Y. (2013). Variable selection for censored quantile regression. *Statistica Sinica*, 23, 145–167.

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

Yakovlev, A. Y. & Tsodikov, A. D. (1996). *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific, Singapore.

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

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

Received 29 August 2015
Accepted 8 September 2016