Published in final edited form as:

Lifetime Data Anal. 2014 July ; 20(3): 369–386. doi:10.1007/s10985-013-9268-2.

A Class of Semiparametric Transformation Models for Survival Data with a Cured Proportion

Sangbum Choi,

Department of Biostatistics, The University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 1411, Houston, TX 77030, U.S.A

Xuelin Huang, and

Department of Biostatistics, The University of Texas, MD Anderson Cancer Center, Houston, TX 77030, U.S.A

Yi-Hau Chen

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan

Abstract

We propose a new class of semiparametric regression models based on a multiplicative frailty assumption with a discrete frailty, which may account for cured subgroup in population. The cure model framework is then recast as a problem with a transformation model. The proposed models can explain a broad range of nonproportional hazards structures along with a cured proportion. An efficient and simple algorithm based on the martingale process is developed to locate the nonparametric maximum likelihood estimator. Unlike existing expectation-maximization based methods, our approach directly maximizes a nonparametric likelihood function, and the calculation of consistent variance estimates is immediate. The proposed method is useful for resolving identifiability features embedded in semiparametric cure models. Simulation studies are presented to demonstrate the finite sample properties of the proposed method. A case study of stage III soft-tissue sarcoma is given as an illustration.

Keywords

Discrete frailty; Nonparametric likelihood; Transformation model; Crossing survivals; Counting process; Survival analysis

1 Introduction

In many clinical and epidemiological settings, there is unobserved heterogeneity across individuals. For example, owing to varied lifestyles, experiences, and genetic traits, individuals have different risk factors associated with disease and death, and they differ in their susceptibility to illness and response to treatment. As a result, some individuals are more prone to developing a disease and to experiencing more disease recurrences, whereas others tend to have a lower risk of disease. Moreover, some individuals appear to be immune

or cured after therapy and thus do not again experience the disease of interest within an observable study window. Such individuals, referred to as “cured” or “long-term survivors,” are observed in various types of medical studies. Inferences about survival data when there is a cured subgroup can be conceptualized in a cure model framework.

Most survival cure rate models have been developed based on the assumption that the study population is a two-component mixture of cured and uncured patients (Peng and Dear, 2000; Sy and Taylor, 2000). Lu and Ying (2004) and Mao and Wang (2010) further introduced a class of semiparametric transformation models (Dabrowska and Doksum, 1988) to account for the uncured subgroup. Another class of cure rate models is the so-called promotion-time cure rate model (Chen et al., 1999; Broët et al., 2001; Tsodikov et al., 2003; Zeng et al., 2006). The two-mixture model usually allows for separate covariate formulations for the cure proportion and the survival function for uncured patients, whereas the promotion-time model has one covariate function that describes both the cure and survival components. Both models are based on an unobservable frailty assumption and thus share similar characteristics.

Semiparametric inference of cure rate models is inherently difficult owing to the identifiability feature that arises from the fact that cured patients cannot be identified with right-censored observations. Since most patients are censored near the end of the study, the survival estimates for uncured patients are indistinguishable from those for cured patients. This requires additional restrictions, such that the survival estimates for the uncured group are set to zero at the last observed failure time. Under this condition, Lu and Ying (2004) inferred the semiparametric two-mixture model through estimating equations. However, this is difficult to implement in the cure model setting and yields unsatisfactory convergence rates. An intuitive and popular alternative is to use expectation-maximization (EM) algorithm or its variations, treating the cure indicator as a missing value (Peng and Dear, 2000; Sy and Taylor, 2000; Mao and Wang, 2010). Even though EM-type algorithms provide stable estimates of both parametric and nonparametric components, they are generally slow, and statistical inferences cannot be made immediately.

In this article, we propose a class of transformation models with a cured proportion that includes two-mixture and promotion-time models as special cases. This class is induced from a frailty hazard model with a discrete frailty, which allows different cured proportions among subgroups and different formats for short-term survivors. Specifically, the regression of survival time T on a covariate path $\mathbf{Z}(t) = \{\mathbf{Z}(u), u \leq t\}$ has the following model:

$$\begin{aligned} \Lambda\{t|W, \bar{\mathbf{Z}}(t)\} &\equiv \lim_{\delta \rightarrow 0} Pr(t \leq T < t + \delta | T \geq t, W, \bar{\mathbf{Z}})/\delta, \\ &= W \cdot R \left(\int_0^t I(T \geq u) \exp\{\boldsymbol{\beta}^T \mathbf{Z}(u)\} d\Lambda(u) \right), \end{aligned} \quad (1.1)$$

where $\Lambda\{t|W, \bar{\mathbf{Z}}(t)\}$ is the conditional cumulative hazards function of T given the event, covariate history before t , and a random variable $W \geq 0$. The specified nonnegative function R is strictly increasing and continuously differentiable and satisfies $R(0) = 0$, $R(\infty) = \infty$, and $R'(0) > 0$. Given W , letting $R(x) = x$ leads to the Cox model, and $R(x) = \log(1 + x)$ corresponds to a proportional odds model. Also, $\boldsymbol{\beta}$ is an unknown parameter vector and $\Lambda(\cdot)$ is an increasing but arbitrarily unspecified function.

In model (1.1), a discrete random variable W is introduced such that $W = 0$ indicates the subjects who are cured and thus have zero risks; whereas $W > 0$ indicates the subjects who would experience the event of interest. To complete the model, we impose a distributional assumption for W , such as a Bernoulli or Poisson distribution that may depend on another set of covariates and unknown parameters. The frailty variable W can naturally represent latent clusters of patients with similar levels of risk. Chen et al. (1999) and Tsodikov et al. (2003) provided a biological justification for the discrete frailty. In their context, the frailty W represents the number of metastasis-competent carcinogenic cells in the subject that remain active after the initial treatment. Caroni et al. (2010) viewed a discrete frailty as the unknown number of flaws in an item being tested, and used parametric survival models with discrete frailty to analyze data sets.

We recognize that the population presentation of model (1.1) forms a new class of transformation models, and we base our inference procedure on the direct maximization of the observed likelihood function. Classical transformation models (Dabrowska and Doksum, 1988) have been extensively investigated in the literature. Recently, Zeng and Lin (2006) and Chen (2009) presented a novel and efficient algorithm for maximum likelihood estimation. Those methodologies, however, assume complete knowledge of the underlying link function and thus cannot be directly applicable to cure rate models, where the link function depends on an unknown link parameter. Motivated by Chen (2009), we propose a nonparametric maximum likelihood estimation (NPMLE) for semiparametric transformation models in the presence of a cured proportion. Our proposal contains a closed-form expression for the observed information matrix, thus standard errors are readily available for both the non-parametric and parametric components. Computation is substantially faster and more stable than either the EM-type or methods based on estimating equations.

The rest of the article is organized as follows. In Section 2, we describe the proposed model in detail. In Section 3, we introduce the proposed estimation and inference procedures. In Section 4, we provide some simulation studies to examine the performance of the proposed methods. In Section 5, we illustrate our methods using a soft-tissue sarcoma data sample. In Section 6, we give concluding remarks.

2 A class of semiparametric cure rate models

Suppose that T has the decomposition $T = \varepsilon T^* + (1 - \varepsilon)\infty$, where $\varepsilon \in \{0, 1\}$ is a cure indicator whether a subject is cured ($\varepsilon = 0$) or not ($\varepsilon = 1$) and $T^* \in [0, \infty)$ denotes the survival time for an uncured patient. Let $\mathbf{Z}(\cdot)$ denote a p -vector of covariates related to the model of T^* and $\mathbf{X}$ denote a q -vector of covariates related to the cure probability. $\mathbf{X}$ includes 1 as its first entry and may share some time-independent components with $\mathbf{Z}$. The censoring time C is assumed to be independent of T given $(\mathbf{Z}, \mathbf{X})$. Define $\tilde{T} = \min(T, C)$ and $\delta = I(\tilde{T} = C)$. Let $N(t) = I(\tilde{T} \leq t, \delta = 1)$ denote the counting process for events and $Y(t) = I(\tilde{T} > t)$, the at-risk process. In practice, we observe $\{N_i(t), Y_i(t), \mathbf{Z}_i(t), \mathbf{X}_i; i = 1, \dots, n, 0 \leq t \leq \tau\}$, which are i.i.d. copies of $\{N(t), Y(t), \mathbf{Z}(t), \mathbf{X}\}$, where τ is the maximum follow-up time. We assume that $\mathbf{X}_i$ is bounded and that $\mathbf{Z}_i(\cdot)$ has a uniformly bounded variation on $[0, \tau]$.

Based on model (1.1), we can formulate population models for T by positing various parametric assumptions on the frailty variable. For example, W may have a binary distribution with $p = \Pr(\varepsilon = 1)$ or a Poisson distribution with mean θ among many other choices, which may depend on $\mathbf{X}$ and related unknown parameters γ . Define $G_\gamma(\cdot) \equiv G_\gamma(\cdot|\mathbf{X}) = E_\gamma[\exp\{-W \cdot R(\cdot)\}|\mathbf{X}]$, where the expectation is taken over W . It follows from (1.1) that regression model is specified as

$$\Pr(T \geq t|\mathbf{X}, \mathbf{Z}) = G_\gamma \left\{ \int_0^t Y(u) \exp\{\beta^T \mathbf{Z}(u)\} d\Lambda(u) \right\}, \quad (2.1)$$

which forms a new class of semiparametric transformation models. The survivorship function in (2.1) has a positive value in the limit, which facilitates modeling the proportion of cured patients.

Formulation (2.1) can be used to comprehensively model survival data with a cure fraction. When W_i is a binary variable taking on the values of 0 and 1 with probability $\Pr(W_i = 0|\mathbf{X}_i) = p_i(\mathbf{X}_i)$, we obtain the two-component mixture model (e.g., Peng and Dear, 2000; Sy and Taylor, 2000),

$$\Pr(T_i \geq t|\mathbf{X}_i, \mathbf{Z}_i) = 1 - p(\mathbf{X}_i) + p(\mathbf{X}_i) S(t|\mathbf{Z}_i), \quad (2.2)$$

where

$$S(t|\mathbf{Z}_i) = \exp \left[-R \left\{ \int_0^t \exp\{\beta^T \mathbf{Z}_i(u)\} d\Lambda(u) \right\} \right]$$

is the survival function for uncured subjects, referred to as a latency distribution for T_i^* . A conventional choice for $p(\mathbf{X}_i)$ is a logistic model, such as $\text{logit}\{p(\mathbf{X}_i)\} = \gamma^T \mathbf{X}_i$, which gives the cure rate, $\Pr(T_i = \infty|\mathbf{X}_i) = \{1 + \exp(\gamma^T \mathbf{X}_i)\}^{-1}$. When W_i has a Poisson distribution with mean $e^{\gamma^T \mathbf{X}_i}$, model (2.1) leads to the promotion-time model (e.g., Broët et al., 2001):

$$\Pr(T_i \geq t|\mathbf{X}_i, \mathbf{Z}_i) = \exp \left\{ -F(t|\mathbf{Z}_i) e^{\gamma^T \mathbf{X}_i} \right\}, \quad (2.3)$$

where $F(\cdot|\mathbf{Z}_i) = 1 - S(\cdot|\mathbf{Z}_i)$. Note that $\Pr(T = \infty|\mathbf{X}_i) = \exp\{-e^{\gamma^T \mathbf{X}_i}\}$, which corresponds to the complementary log-log format for the binary response. Besides covering two major types of cure models, we observe that if W_i has a geometric distribution with mean $e^{\gamma^T \mathbf{X}_i}$, model (1.1) implies the analogue of the proportional odds model, given by

$$\Pr(T_i \geq t|\mathbf{X}_i, \mathbf{Z}_i) = 1 / \{1 + F(t|\mathbf{Z}_i) e^{\gamma^T \mathbf{X}_i}\}. \quad (2.4)$$

An additional model can be made by alternating the frailty distribution. By our proposed models (2.2)-(2.4), treatment effects can be assessed through modeling the survivorship function $S(\cdot|\mathbf{Z}_i)$ and the cure probability $\Pr(T = \infty|\mathbf{X}_i)$, which separately evaluates the effects of $\mathbf{Z}_i$ and $\mathbf{X}_i$ within a single framework. This allows an assessment of time-varying treatment effects through $\mathbf{Z}_i$ and $\mathbf{X}_i$ in a parsimonious way.

Tsodikov et al. (2003) and Zeng et al. (2006) studied the bounded cumulative hazards model of the form $\Pr(T_i \leq t | \mathbf{X}_i) = \exp\{-F(t)e^{\boldsymbol{\gamma}^T \mathbf{X}_i}\}$, which is a subclass of model (2.3) under $\boldsymbol{\beta} \equiv 0$. This model has attracted much attention in the literature, because it is an analogue of the proportional hazards model for long-term survival data. However, in the absence of $\mathbf{Z}_i$ (i.e., $\boldsymbol{\beta} \equiv 0$), this class is no longer identifiable, and consequently the estimation procedure becomes complicated (Zeng et al., 2006). Therefore, throughout this paper we assume $\boldsymbol{\beta} \neq 0$ to ensure model identifiability.

3 Nonparametric maximum likelihood estimation

3.1 Direct maximization of nonparametric likelihood

Using the counting process notation, the logarithm of the observed likelihood function for model (2.1) can be written as the above likelihood function can be rewritten as

$$l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) = \sum_{i=1}^n \left\{ \int_0^\tau \log dH_i(t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) dN_i(t) - \int_0^\tau Y_i(t) dH_i(t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) \right\}, \quad (3.1)$$

where

$$dH_i(t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) = e^{\boldsymbol{\beta}^T \mathbf{Z}_i(t)} H'_{\boldsymbol{\gamma}}(\xi_i(t-; \boldsymbol{\beta}, \Lambda)) d\Lambda(t).$$

Here, we write $\xi_i(t; \boldsymbol{\beta}, \Lambda) = \int_0^t Y_i(u) \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i(u)\} d\Lambda(u)$, and let $H_{\boldsymbol{\gamma}} = -\log G_{\boldsymbol{\gamma}}$ denote the baseline cumulative hazards function with $H'_{\boldsymbol{\gamma}}$ and $H''_{\boldsymbol{\gamma}}$ being its first and second derivatives.

In order to maximize function (3.1) via the NPMLE approach, nonparametric component Λ should be seen as a nondecreasing step function, which has a jump of size Δ at distinct observed event times, say $J = \{t_1^*, \dots, t_{k_n}^*\}$, because Λ is identified only at these uncensored failure times. An efficient estimation for $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda)$ can be made by maximizing the observed log-likelihood over $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \{\Delta\Lambda(t_j^*); t_j^* \in J\})$. Function (3.1) is essentially a parametric log-likelihood function for independent censored data (Kalbfleisch and Prentice, 2002, Chapter 5.8; Chen, 2009) which treats the jump sizes of Λ for fixed n as a set of finite-dimensional parameters. From (3.1), it is implicit that $E\{dN_i(t) | F_{t-}\} = Y_i(t) dH_i(t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda)$, where F_t is the filtration up to t . Associated with each $N_i(t)$ is a F_{t-} predictable covariate process $\{\mathbf{Z}_i(t), \mathbf{X}_i\}$. The counting process $N_i(t)$ can be uniquely decomposed so that

$$M_i(t; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda) = N_i(t) - \int_0^t Y_i(u) dH_i(u; \boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda), \quad i=1, \dots, n,$$

is a local square integrable martingale process adapted to F_{t-} .

Within the NPMLE framework, we derive the likelihood score function by taking derivatives of function (3.1) with respect to each component of $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \boldsymbol{\gamma}, \Lambda)$. Denote the likelihood score equations associated with $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and Λ as $U_{\boldsymbol{\beta}}$, $U_{\boldsymbol{\gamma}}$ and U_{Λ} , and write $U(\boldsymbol{\theta}) = [U_{\boldsymbol{\beta}}(\boldsymbol{\theta}), U_{\boldsymbol{\gamma}}(\boldsymbol{\theta}), U_{\Lambda}(\boldsymbol{\theta})]$. It follows from the martingale process theory that

$$U(\boldsymbol{\theta}) = \sum_{i=1}^n \int_0^\tau \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log dH_i(t; \boldsymbol{\theta}) \right] dM_i(t; \boldsymbol{\theta}). \quad (3.2)$$

More precisely, equation (3.2) consists of the following components:

$$U_{\beta}(\boldsymbol{\theta}) = \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{Z}_i(t) + \psi_{\gamma}(\xi_i(t-; \beta, \Lambda)) \times \int_0^{t-} Y_i(u) \mathbf{Z}_i(u) e^{\beta^T \mathbf{Z}_i(u)} \Delta \Lambda(u) \right\} dM_i(t; \boldsymbol{\theta}) \quad (3.3)$$

$$U_{\gamma}(\boldsymbol{\theta}) = \sum_{i=1}^n \int_0^\tau \phi_{\gamma}(\epsilon_i(t; \beta, \Lambda)) dM_i(t; \boldsymbol{\theta}), \quad (3.4)$$

where $\psi_{\gamma}(x) = d \log H'_{\gamma}(x) / (dx)$ and $\phi_{\gamma}(x) = d \log H'_{\gamma}(x) / (d\gamma)$. In addition,

$$U_{\Lambda}(\boldsymbol{\theta}) = \sum_{i=1}^n \left\{ dN_i(t) - Y_i(t) e^{\beta^T \mathbf{Z}_i(t)} \pi_i(t; \boldsymbol{\theta}) H'_{\gamma}(\xi_i(t-; \beta, \Lambda)) \Delta \Lambda(t) \right\}, \quad (3.5)$$

where

$$\pi_i(t; \boldsymbol{\theta}) = 1 - \frac{\int_{t+}^{\tau} \psi_{\gamma}(\xi_i(u-; \beta, \Lambda)) dM_i(u; \boldsymbol{\theta})}{H'_{\gamma}(\xi_i(t-; \beta, \Lambda))}.$$

It can be checked that the solution to (3.5) is equivalent to the Breslow estimator of Chen (2009) for the jump size of Λ , which satisfies

$$\Delta \Lambda(t) = \frac{\sum_{i=1}^n dN_i(t)}{\sum_{i=1}^n Y_i(t) e^{\beta^T \mathbf{Z}_i(t)} \pi_i(t; \boldsymbol{\theta}) H'_{\gamma}(\xi_i(t-; \beta, \Lambda))}. \quad (3.6)$$

The NPMLEs $\hat{\boldsymbol{\theta}} = (\hat{\beta}, \hat{\gamma}, \hat{\Lambda})$ are then defined as the solutions to $U(\boldsymbol{\theta}) = 0$. We note that setting $\pi \equiv 1$, or equivalently, $\psi \equiv 0$ leads to the martingale-based estimating equations of Lu and Ying (2004) for β and Λ , which is consistent but not fully efficient.

3.2 Standard error estimation

In the direct maximization method, constructing a tractable and invertible Hessian matrix of the likelihood function is critical for estimating the NPMLEs and their standard errors. By the martingale theory, the observed information matrix, $I^*(\boldsymbol{\theta}) = -^2 l(\boldsymbol{\theta}) / \boldsymbol{\theta} \boldsymbol{\theta}^T = - U(\boldsymbol{\theta}) / \boldsymbol{\theta}$, is given by

$$I^*(\boldsymbol{\theta}) = \sum_{i=1}^n \int_0^\tau \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log dH_i(t; \boldsymbol{\theta}) \right]^{\otimes 2} dN_i(t) - \sum_{i=1}^n \int_0^\tau \frac{\partial^2 dH_i(t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} [dH_i(t; \boldsymbol{\theta})]^{-1} dM_i(t; \boldsymbol{\theta}), \quad (3.7)$$

where $a^{\otimes 2} = aa^T$ for a vector a . Note that the second term in the right side of (3.7) is of the order $o_p(n)$, and thus is asymptotically ignorable. Then the information matrix can be approximated by the predictable variation process

$$I(\theta) = \sum_{i=1}^n \int_0^\tau \left[\frac{\partial}{\partial \theta} \log dH_i(t; \theta) \right]^{\otimes 2} Y_i(t) dH_i(t; \theta). \quad (3.8)$$

Following the standard parametric information theory, the asymptotic variance-covariance matrix for $\hat{\theta}$ can be obtained from the inverse of either (3.7) or (3.8) that is evaluated at $\theta = \hat{\theta}$. In our estimating procedure, we utilized the process (3.8) since the computation of (3.7) requires the evaluation of the second derivatives of $dH_i(t; \theta)$ for θ that is considerably involved in this current semiparametric setting. The closed-form formula for the information matrix (3.8) is provided in the Appendix.

3.3 Implementation

Implementation of the proposed direct maximization can be accomplished via some large-scale quasi Newton algorithms, such as a subspace trust-region procedure based on the interior-reflective Newton method (Coleman and Li, 1994; 1996). Such algorithms are available in scientific optimization routines such as `fminunc` in MATLAB or the contributed R-package `trust` by Geyer (2009). Efficient and reliable performance of these algorithms requires explicit expressions of the gradient and Hessian of the objective function. Following Zeng and Lin (2006), we utilized the MATLAB optimization `fminunc` to locate the likelihood maximizer, providing the likelihood scores (3.2) and the information matrix (3.8). The variance estimates are obtained from the inversion of the observed information matrix. The search may not guarantee the global maximum, but this algorithm generally works well in our problem, as long as the initial values are reasonably chosen. We used $(\beta^{(0)}, \gamma^{(0)}) = (0, 0)$ and $\lambda^{(0)} = 1/n$ as initial values.

As we discussed, semiparametric cure models generally suffer from a lack of identifiability, because the latency distribution is not clearly separated from the cure probability in the limit. A common remedy is to assign an artificially large value on the last jump size $\Delta \lambda(t_{k_n}^*)$ (e.g., Sy and Taylor, 2000; Mao and Wang, 2010). However, our experience is that such a restriction increases the bias of estimates, especially for $\hat{\lambda}$. In contrast, our algorithm works well without such a constraint unless the sample size is too small. Compared to the EM-type algorithms, the proposed method is simple and shows high convergence rates in substantially less computing time. In the simulations studies in Section 4, we demonstrate the advantages of the proposed method over the EM methods.

3.4 Asymptotic results

We impose the following regularity conditions:

- a. The covariate $\mathbf{X}_i$ and the covariate process $\mathbf{Z}_i(\cdot)$ are uniformly bounded, have nondegenerate variance-covariance matrices, and are cadlag for $i = 1, \dots, n$.

- b. With probability one, there exists a positive constant ε such that $\Pr(C_i \leq \tau | \mathbf{Z}_i, \mathbf{X}_i) > \varepsilon > 0$.
- c. (β_0, γ_0) is an interior point in a compact and convex subset R^{p+q} . $A_0 \in A_0$, where A_0 is the collection of functions that are strictly increasing and continuously differentiable with $A(0) = 0$ and $A(\tau) < \infty$.
- d. The survivorship function $G_{\gamma}(t)$ is at least thrice differentiable with respect to t , and at least twice differentiable with respect to γ . Also, there exists a constant $c > 0$ such that $\sup_{t \in [0, \tau]} |t^{1+c} G'_{\gamma}(t)| < \infty$. The functions ψ_{γ} and ϕ_{γ} are both Lipschitz.

Sketches of proofs for the asymptotic results are provided here. Following the theories by Murphy (1994, 1995) and Zeng and Lin (2006), it can be shown that $(\hat{\beta}, \hat{\gamma})$ is strongly consistent to (β_0, γ_0) and $\hat{A}(\cdot)$ uniformly converges to $A_0(\cdot)$ in the interval $[0, \tau]$ with probability one. To claim asymptotic normality, it suffices to consider the linear functional

$$n^{1/2} \left[\int_0^{\tau} h_{\Lambda}(t) d\{\hat{\Lambda}(t) - \Lambda_0(t)\} + h_{\beta}^T (\hat{\beta} - \beta_0) + h_{\gamma}^T (\hat{\gamma} - \gamma_0) \right], \quad (3.9)$$

where h_{Λ} is an arbitrary nonnegative bounded function on $[0, \tau]$, and h_{β} and h_{γ} are bounded vectors in R^p and R^q , respectively. Since the asymptotic variance for (3.9) is equal to the asymptotic variance of $n^{1/2} \sum_{t^* \in J} h_{\Lambda}(t^*) \Delta \Lambda(t^*) + n^{1/2} h_{\beta}^T (\hat{\beta} - \beta_0) + n^{1/2} h_{\gamma}^T (\hat{\gamma} - \gamma_0)$, it follows from the martingale central limit theorem that the linear functional (3.9) is asymptotically normal with variance matrix $\zeta^T I_0^{-1} \zeta$, where I_0 is the variance matrix of $n^{1/2} U(\theta_0)$ and $\zeta = (\{h_{\Lambda}^*\}, h_{\beta}, h_{\gamma})^T$ with $\{h_{\Lambda}^*\}$ the vector of $h_{\Lambda}(t)$ evaluated at $t^* \in J$. A consistent variance estimator can be obtained by replacing I_0 with $n^{-1} I(\hat{\theta})$, that is, $n \zeta^T I(\hat{\theta})^{-1} \zeta$. Its consistency can be verified along the lines of Parner (1998). The weak convergence of $n^{1/2} \{\hat{A}(\cdot) - A_0(\cdot), \hat{\beta} - \beta_0, \hat{\gamma} - \gamma_0\}$ to a mean-0 Gaussian process can be justified by standard parametric likelihood theory. In addition, $(\hat{\beta}, \hat{\gamma})$ is an asymptotically linear estimator for (β_0, γ_0) and its influence function lies on the linear space spanned by the score functions. Thus, it follows that the limiting variance of $n^{1/2}(\hat{\beta} - \beta_0, \hat{\gamma} - \gamma_0)$ attains the semiparametric efficiency bound in the sense of Bickel et al. (1998).

4 Simulation studies

In simulation studies, we examined the finite sample performance of the estimator from the proposed method. We generated T from the following models

- i. Two-mixture cure model (binary frailty):

$$\Pr(T \geq t | Z_1, Z_2, X_1) = \frac{1}{1 + e^{\gamma_0 + \gamma_1 X_1}} + \frac{e^{\gamma_0 + \gamma_1 X_1}}{1 + e^{\gamma_0 + \gamma_1 X_1}} S(t | Z_1, Z_2);$$

- ii. Promotion-time cure model (Poisson frailty):

$$\Pr(T \geq t | Z_1, Z_2, X_1) = \exp[-e^{\gamma_0 + \gamma_1 X_1} \{1 - S(t | Z_1, Z_2)\}];$$

iii. Proportional-odds type cure model (geometric frailty):

$$Pr(T \geq t | Z_1, Z_2, X_1) = \frac{1}{1 + e^{\gamma_0 + \gamma_1 X_1} \{1 - S(t | Z_1, Z_2)\}},$$

where $S(t | Z_1, Z_2) = \exp \{-R(\Lambda(t)e^{\beta_1 Z_1 + \beta_2 Z_2})\}$. Two covariates Z_1 and Z_2 were generated respectively from the standard normal distribution truncated at ± 2 and the Bernoulli distribution with a success probability of 0.5; $X_1 = Z_2$ was included for the cure probability. We chose $\Lambda(t) = t/4$, $(\beta_1, \beta_2) = (1, -1)$, $(\gamma_0, \gamma_1) = (1, -1)$, and $\tau = 12$. By setting $(\gamma_0, \gamma_1) = (1, -1)$, the cure rates for the control group ($X_1 = 0$) and for the treatment group ($X_1 = 1$) are 26.9% and 50.0%, respectively, in both models (i) and (iii), whereas they correspond to 6.6% and 36.8% in model (ii). The censoring time follows an exponential distribution with the hazard rate $\lambda_0 e^{0.693 Z_2}$, where the value of λ_0 is varied to make the censoring proportions higher than the cure rates by 5%–10%. For each model we considered both $R(x) = x$ and $R(x) = \log(1+x)$, corresponding, respectively, to the proportional hazards and proportional odds models for uncured patients.

We obtained the NPMLE through the MATLAB optimization toolbox `fminunc`, using the initial values of 0 for β and γ and $1/n$ for Λ with n samples. Other initial values were also tested in the simulation study and the results were not sensitive to those choices. The simulation results are summarized in Table 1. The parameter estimators appear virtually unbiased, the standard error estimates (SEEs) reasonably reflect the true variations (SEs), and the confidence intervals have correct coverage probabilities. The proposed method has a convergence rate of 99.5% for $n = 100$ and 100% for $n = 200$. Under the present simulation settings, the censoring rates were approximately 50%–55%, and the effective sample size is nearly half of the overall sample size. Because of this, there are some discrepancies between the SE and SEE and the downward coverage probabilities for the estimator of $\Lambda(\cdot)$, but the results improve as the sample size increases. Our experience is that when censoring rates are too low, the computation becomes less stable. Many censored observations beyond some event time and sufficiently long follow-up times are needed to gain satisfactory precision.

In the second set of simulations, we compared our direct maximization method with the EM algorithm (Peng and Dear, 2000; Sy and Taylor, 2000) for the two-mixture model (2.2) with $R(x) = x$. The simulation setting is the same as that in the first experiment. Because the SEs of the nonparametric parts are not available with the EM method, we display only the estimates and SEs of the parametric components in Table 2. The relative efficiency, represented by “MSE ratio,” ranges between 0.90 and 0.99 with $n = 200$, indicating that our direct maximization method is preferable. The performance of our proposal is good with $n = 100$, especially for estimating the parameters in the cure probability. In Figure 1, we plot the mean estimates of the nonparametric component $\Lambda_0(t) = t/4$ for $n = 200$. Owing to the sparsity of observations near the end, upward bias is inevitable, and our method presents an improved fit. In addition, direct maximization is substantially effective in operation time: to complete 1,000 repetitions with $n = 200$, the proposed approach took only approximately 900 seconds, whereas the EM algorithm required more than 4 hours using an Intel 3 GHz Core 2 processor.

5 Data analysis

As an illustration, we consider a study conducted to assess the role of chemotherapy on stage III soft-tissue sarcoma (STS). Patients with large ($> 5\text{cm}$), highgrade, deep-extremity STS, whose disease is classified as stage III (American Joint Committee on Cancer, 2002), are at significant risk for distant tumor recurrence and subsequent sarcoma-associated death. The primary treatment for these patients is surgical resection of the tumor. The use of systemic chemotherapy as adjuvant treatment remains controversial, despite its having been evaluated in several randomized trials and a meta-analysis. To evaluate the impact of chemotherapy on sarcoma-related death while accounting for known prognostic variables, Cormier et al. (2004) performed a cohort analysis of 680 patients with primary stage III STS of an extremity who were treated between 1984 and 1999 at two comprehensive cancer centers. Medical records were examined retrospectively for verification of known prognostic factors, treatments, and survival outcomes. Among 680 patients, 364 (53.5%) were treated with local therapy only, and 316 (46.5%) were treated with local therapy plus chemotherapy.

A noticeable feature of the data is its high censoring rate (57.4%). Figure 2 displays the Kaplan-Meier estimates for patients who received chemotherapy as adjuvant treatment compared with those who did not receive chemotherapy. Both curves show stable plateaus away from zero in the tail. Furthermore, a crossing of the survival estimates is observed at approximately 2 years after the initiation of treatment. This implies that the observed effects of chemotherapy vary over time. The effect of chemotherapy does not appear to be significant, with a large p -value (0.73) in the log-rank test, which has lower power under this clear violation of the proportional hazards. Given this pattern, Cormier et al. (2004) applied Cox methods separately over two distinct time periods, before and after the crossing of the survival curves. It is clear that the analysis of the divided data set leads to a loss of efficiency, compared to the integrated approach.

Table 3 summarizes the estimation results of the proposed models with $R(x; \rho) = (1/\rho) \log(1 + \rho x)$, $\rho = 0, 1, 2$, which corresponds to the proportional hazards ($\rho = 0$) and the proportional odds ($\rho = 1$) models for the latency distribution. Thus, the coefficient β may be interpreted as a relative risk or odds ratio, depending on the choice of ρ . By contrast, the cure probability is $\Pr(\text{cured}) = 1/(1 + e^{\gamma^T \mathbf{X}})$ with models (2.2) and (2.4) and $\Pr(\text{cured}) = \exp\{-e^{\gamma^T \mathbf{X}}\}$ with model (2.3). These respectively correspond to a logistic regression and a complementary log-log model; thus, γ may have the interpretation of a (negative) odds ratio and rate ratio, respectively. According to the observed likelihoods, none of the models are superior to the others. Based on the two-mixture model with $\rho = 0$, the effects of chemotherapy appear to be more significant in the short term. In this model, the effects have opposing directions: $e^{\beta} = 0.742$ ($p = 0.065$) may represent the short-term relative risk for chemotherapy vs. no chemotherapy, whereas $e^{-\gamma} = 0.673$ ($p = 0.068$) may represent the odds ratio of the long-term effect of chemotherapy. These effects can be more correctly assessed by the likelihood-ratio test. The use of chemotherapy is seemingly beneficial early on but is not beneficial or maybe even harmful in the long term. This is possible because chemotherapy may damage normal cells in addition to killing cancer cells. If so, the opposite effects for different groups may partly explain the crossed survival curves. However, this interpretation should be made with caution as the study was not randomized. A naive Cox method gives a coefficient of

-0.021 (SE, 0.119) for chemotherapy. This shows that a comprehensive overall picture requires the proper separation of short-term and long-term effects in the working model; otherwise, the effects could be misleadingly mitigated. In Figure 3, the survival estimates for patients treated with/without chemotherapy both with radiation (Fig. 3a) and without radiation (Fig. 3b) are plotted from the Kaplan-Meier and proposed methods. The use of radiation prolongs patient survival times (median survival times are 5.78 for patients who received no radiation and 11.6 years for those who received radiation). For these two groups, the effect of chemotherapy was reversed at approximately 2 and 7 years after observation, implying that it is beneficial in the short term only. In both settings, our estimates follow the nonparametric estimates well.

6 Discussion

In this article, we present a class of semiparametric regression models for long-term survival data. A cure rate model framework is desirable when there exists a substantial cured subgroup of patients who do not experience the event of interest. The new model is motivated by a frailty model with a discrete frailty; it makes intuitive sense in that zero hazard is assigned to cured patients, whereas uncured patients have a certain amount of risk. By recasting the cure model problem as a transformation model framework, we have provided a simple but efficient algorithm for calculating maximum likelihood estimates. We base the inference procedure on directly maximizing the nonparametric likelihood in the spirit of a parametric case. Existing methodologies for cure models are often based on the EM algorithm or its variations, viewing the discrete frailty as a missing random variable. The survival cure model, however, is in general a marginal model in that the frailty variable is integrated out. Our proposed method can be applied as long as the population survival model has a tractable formulation that is asymptotically efficient and achieves high convergence rates even in a small sample. It has many computational advantages over existing EM-based methods in that it is less computationally demanding, requires only a short operation time, and immediately provides variance estimates.

Appendix

Information matrix

Let $\{t_j\}$ denote the sets of the observed failure times and $d\Lambda_j = \Lambda(t_j)$. Define $h_i(t; \theta) = H'_\gamma(\xi_i(t; \beta, \Lambda))$, $\psi_i(t; \theta) = \psi_\gamma(\xi_i(t; \beta, \Lambda))$, $\varphi_i(t; \theta) = \varphi_\gamma(\xi_i(t; \beta, \Lambda))$, and

$$\mathbf{Z}_i(t; \theta) = \mathbf{Z}_i(t) + \psi_i(t - ; \theta) \int_0^{t-} Y_i(u) \mathbf{Z}_i(u) e^{\beta^T \mathbf{Z}_i(u)} d\Lambda(u).$$

The observed information matrix I for the NPMLE can be approximated by (3.8). The closed-form expressions for submatrices for I , obtained from the idea of Chen (2009), are presented as follows;

$$\begin{aligned}
 I_{d\Lambda_j \, d\Lambda_j} &= \frac{-\partial U_{d\Lambda_j} / \partial (d\Lambda_j)}{\sum_{i=1}^n dN_i(t_j) + \left\{ \sum_{i=1}^n Y_i(t_j) e^{2\beta^T \mathbf{Z}_i(t_j)} \mu_i(t_j; \boldsymbol{\theta}) \right\}} (d\Lambda_j)^2, \\
 &= \sum_{i=1}^n dN_i(t_j) + \left\{ \sum_{i=1}^n Y_i(t_j) e^{2\beta^T \mathbf{Z}_i(t_j)} \mu_i(t_j; \boldsymbol{\theta}) \right\} (d\Lambda_j)^2,
 \end{aligned}$$

where

$$\mu_i(t_j; \boldsymbol{\theta}) = \int_{t_j+}^{\tau} Y_i(u) e^{\beta^T \mathbf{Z}_i(u)} h_i(u - ; \boldsymbol{\theta}) \psi_i^2(u - ; \boldsymbol{\theta}) d\Lambda(u).$$

For $j < k$

$$\begin{aligned}
 I_{d\Lambda_j \, d\Lambda_k} &= \frac{-\partial U_{d\Lambda_j} / \partial (d\Lambda_k)}{\left[\sum_{i=1}^n Y_i(t_j) e^{\beta^T \{\mathbf{Z}_i(t_k) + \mathbf{Z}_i(t_j)\}} \{h_i(t_j - ; \boldsymbol{\theta}) \psi_i(t_j - ; \boldsymbol{\theta}) + \mu_i(t_j; \boldsymbol{\theta})\} \right]} (d\Lambda_j \, d\Lambda_k). \\
 &= \left[\sum_{i=1}^n Y_i(t_j) e^{\beta^T \{\mathbf{Z}_i(t_k) + \mathbf{Z}_i(t_j)\}} \{h_i(t_j - ; \boldsymbol{\theta}) \psi_i(t_j - ; \boldsymbol{\theta}) + \mu_i(t_j; \boldsymbol{\theta})\} \right] (d\Lambda_j \, d\Lambda_k).
 \end{aligned}$$

By symmetry, $I_{d\Lambda_k \, d\Lambda_j}$ can also be obtained from the above expression.

$$\begin{aligned}
 I_{\beta d\Lambda_j} &= \frac{-\partial U_{\beta} / \partial (d\Lambda_j) = -\partial U_{d\Lambda_j} / \partial \beta = I_{d\Lambda_j \, \beta}}{\left[\sum_{i=1}^n Y_i(t_j) e^{\beta^T \mathbf{Z}_i(t_j)} \{\mathbf{Z}_i(t_j; \boldsymbol{\theta}) h_i(t_j - ; \boldsymbol{\theta}) + v_i(t_j; \boldsymbol{\theta})\} \right]} (d\Lambda_j), \\
 &= \left[\sum_{i=1}^n Y_i(t_j) e^{\beta^T \mathbf{Z}_i(t_j)} \{\mathbf{Z}_i(t_j; \boldsymbol{\theta}) h_i(t_j - ; \boldsymbol{\theta}) + v_i(t_j; \boldsymbol{\theta})\} \right] (d\Lambda_j),
 \end{aligned}$$

where

$$v_i(t_j; \boldsymbol{\theta}) = \int_{t_j+}^{\tau} Y_i(u) \mathbf{Z}_i(u; \boldsymbol{\theta}) e^{\beta^T \mathbf{Z}_i(u)} h_i(u - ; \boldsymbol{\theta}) \psi_i(u - ; \boldsymbol{\theta}) d\Lambda(u).$$

$$\begin{aligned}
 I_{\gamma d\Lambda_j} &= \frac{-\partial U_{\gamma} / \partial (d\Lambda_j) = -\partial U_{d\Lambda_j} / \partial \gamma = I_{d\Lambda_j \, \gamma}}{\left[\sum_{i=1}^n Y_i(t_j) e^{\beta^T \mathbf{Z}_i(t_j)} \{\phi_i(t_j; \boldsymbol{\theta}) h_i(t_j - ; \boldsymbol{\theta}) + w_i(t_j; \boldsymbol{\theta})\} \right]} (d\Lambda_j), \\
 &= \left[\sum_{i=1}^n Y_i(t_j) e^{\beta^T \mathbf{Z}_i(t_j)} \{\phi_i(t_j; \boldsymbol{\theta}) h_i(t_j - ; \boldsymbol{\theta}) + w_i(t_j; \boldsymbol{\theta})\} \right] (d\Lambda_j),
 \end{aligned}$$

where

$$w_i(t_j; \boldsymbol{\theta}) = \int_{t_j+}^{\tau} Y_i(u) \phi_i(u; \boldsymbol{\theta}) e^{\beta^T \mathbf{Z}_i(u)} h_i(u - ; \boldsymbol{\theta}) \psi_i(u - ; \boldsymbol{\theta}) d\Lambda(u).$$

$$\begin{aligned}
 I_{\beta\beta} &= -\partial U_{\beta} / \partial \beta = \sum_{i=1}^n \int_0^{\tau} Y_i(t) \mathbf{Z}_i(t; \boldsymbol{\theta})^{\otimes 2} e^{\beta^T \mathbf{Z}_i(t)} h_i(t - ; \boldsymbol{\theta}) d\Lambda(t), \\
 I_{\gamma\gamma} &= -\partial U_{\gamma} / \partial \gamma = \sum_{i=1}^n \int_0^{\tau} Y_i(t) \phi_i(t; \boldsymbol{\theta})^{\otimes 2} e^{\beta^T \mathbf{Z}_i(t)} g_i(t - ; \boldsymbol{\theta}) d\Lambda(t),
 \end{aligned}$$

and

$$\begin{aligned}
 I_{\beta\gamma} &= -\partial U_{\beta}/\partial \gamma = -\partial U_{\gamma}/\partial \beta = -I_{\gamma\beta} \\
 &= \sum_{i=1}^n \int_0^{\tau} Y_i(t) \mathbf{Z}_i(t; \boldsymbol{\theta}) \phi_i(t; \boldsymbol{\theta}) e^{\beta^T \mathbf{Z}_i(t)} h_i(t-; \boldsymbol{\theta}) d\Lambda(t).
 \end{aligned}$$

References

1. American Joint Committee on Cancer. Cancer Staging Manual. 6. New York: Springer; 2002.
2. Bickel, PJ.; Klaassen, CA.; Ritov, Y.; Wellner, JA. Efficient and Adaptive Estimation for Semiparametric Models. 2. Baltimore: Johns Hopkins University Press; 1998.
3. Broët P, Rycke YD, Tubert-Bitter P, Lellouch J, Asselain B, Moreau T. A semiparametric approach for the two-sample comparison of survival times with longterm survivors. *Biometrics*. 2001; 57:844–852. [PubMed: 11550936]
4. Caroni C, Crowder M, Kimber A. Proportional hazards models with discrete frailty. *Lifetime Data Anal.* 2010; 16:374–384. [PubMed: 20111904]
5. Chen MN, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. *J Am Stat Assoc.* 1999; 94:909–919.
6. Chen Y-H. Weighted Breslow-type estimator and maximum likelihood estimation in semiparametric transformation models. *Biometrika*. 2009; 96:591–600.
7. Coleman TF, Li Y. On the convergence of reflective Newton methods for largescale nonlinear minimization subject to bounds. *Math Prog.* 1994; 67:189–224.
8. Coleman TF, Li Y. An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM J Optimiz.* 1996; 6:418–445.
9. Cormier JN, Huang X, Xing Y, Thall PF, Wang X, Benjamin RS, Pollock RE, Antonescu CR, Maki RG, Brennan MF, Pisters PWT. Cohort analysis of patients with localized high-risk extremity soft tissue sarcoma treated at two cancer centers: Chemotherapy-associated outcomes. *J Clin Oncol.* 2004; 22:4567–4574. [PubMed: 15542808]
10. Dabrowska DM, Doksum KA. Estimation and testing in the two-ample generalized odds-rate model. *J Am Stat Assoc.* 1988; 83:744–749.
11. Geyer GJ. Trust: Trust Region Optimization. R package 0.1-2. 2009
12. Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. 2. New York: Wiley; 2002.
13. Lu W, Ying Z. On semiparametric transformation cure models. *Biometrika*. 2004; 91:331–343.
14. Mao M, Wang J-L. Semiparametric efficient estimation for a class of generalized proportional odds cure models. *J Am Stat Assoc.* 2010; 105:302–311. [PubMed: 22865944]
15. Murphy SA. Consistency in a proportional hazards model incorporating a random effect. *Ann Stat.* 1994; 22:712–731.
16. Murphy SA. Asymptotic theory for the frailty model. *Ann Stat.* 1995; 23:182–198.
17. Parner E. Asymptotic theory for the correlated gamma-frailty models. *Ann Stat.* 1998; 26:183–214.
18. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics*. 2000; 56:237–243. [PubMed: 10783801]
19. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics*. 2000; 56:227–236. [PubMed: 10783800]
20. Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data: an alternative to two-component mixture models. *J Am Stat Assoc.* 2003; 98:1063–1079. [PubMed: 21151838]
21. Zeng D, Lin DY. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*. 2006; 93:627–640.
22. Zeng D, Yin G, Ibrahim JG. Semiparametric transformation models for survival data with a cure fraction. *J Am Stat Assoc.* 2006; 101:670–684.

Fig. 1.
Mean of the estimated $\Lambda_0(t) = t/4$ by the proposed and EM-type approach.

Fig. 2.

Kaplan-Meier estimates of patients with soft tissue sarcoma in two treatment groups (chemotherapy vs. no chemotherapy).

Fig. 3.

Estimated survival curves by the Kaplan-Meier estimates (solid) and the two-mixture cure model (2.2) with $\rho = 1$ (dashed) for chemotherapy treatment (a) with radiation received and (b) without radiation.

Table 1

Bias, sampling standard error (SE), mean of standard error estimator (SEE), and coverage probability of 95% confidence interval (CP) for the semiparametric transformation cure models, based on 1000 data sets with $n = 100$ and $n = 200$

Par	True	n = 100				n = 200			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
Model 1: two-mixture model with $R(x) = x$									
$\mathcal{A}(\pi/4)$	0.75	0.013	0.238	0.200	0.894	0.003	0.156	0.143	0.924
$\mathcal{A}(\pi/2)$	1.50	0.052	0.484	0.408	0.919	0.049	0.322	0.293	0.938
β_1	1.0	0.073	0.246	0.221	0.924	0.042	0.168	0.154	0.936
β_2	-1.0	-0.038	0.394	0.372	0.941	-0.041	0.282	0.263	0.934
γ_0	1.0	0.094	0.580	0.557	0.971	0.028	0.356	0.353	0.960
γ_1	-1.0	-0.007	0.654	0.647	0.963	-0.001	0.422	0.415	0.944
Model 2: two-mixture model with $R(x) = \log(1 + x)$									
$\mathcal{A}(\pi/4)$	0.75	0.014	0.325	0.296	0.918	0.007	0.221	0.225	0.928
$\mathcal{A}(\pi/2)$	1.50	0.096	0.753	0.677	0.925	0.051	0.531	0.519	0.940
β_1	1.0	0.081	0.390	0.332	0.927	0.058	0.257	0.233	0.939
β_2	-1.0	-0.013	0.557	0.536	0.962	-0.018	0.399	0.383	0.943
γ_0	1.0	0.113	0.662	0.810	0.965	0.101	0.558	0.636	0.961
γ_1	-1.0	-0.046	0.816	0.912	0.962	-0.040	0.574	0.624	0.964
Model 3: promotion-time model with $R(x) = x$									
$\mathcal{A}(\pi/4)$	0.75	0.054	0.244	0.220	0.932	-0.003	0.164	0.156	0.936
$\mathcal{A}(\pi/2)$	1.50	0.120	0.574	0.500	0.949	0.020	0.366	0.348	0.935
β_1	1.0	0.053	0.216	0.196	0.935	0.025	0.147	0.136	0.934
β_2	-1.0	-0.070	0.358	0.333	0.946	-0.029	0.135	0.146	0.941
γ_0	1.0	0.010	0.232	0.218	0.955	0.035	0.166	0.165	0.952
γ_1	-1.0	-0.025	0.284	0.268	0.943	-0.015	0.189	0.187	0.949
Model 4: promotion-time model with $R(x) = \log(1 + x)$									
$\mathcal{A}(\pi/4)$	0.75	0.036	0.305	0.280	0.921	0.057	0.233	0.228	0.939
$\mathcal{A}(\pi/2)$	1.50	0.085	0.751	0.668	0.929	0.052	0.560	0.557	0.964
β_1	1.0	0.045	0.258	0.245	0.939	0.041	0.185	0.175	0.944

Par	True	$n = 100$				$n = 200$			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
β_2	-1.0	-0.044	0.435	0.406	0.949	-0.037	0.305	0.292	0.947
γ_0	1.0	0.073	0.245	0.246	0.953	-0.003	0.172	0.178	0.963
γ_1	-1.0	-0.034	0.297	0.282	0.940	-0.010	0.196	0.196	0.946
<i>Model 5: proportional odds type model with $R(x) = x$</i>									
$\Lambda(\pi/4)$	0.75	0.014	0.280	0.237	0.928	-0.012	0.187	0.178	0.918
$\Lambda(\pi/2)$	1.50	0.101	0.647	0.527	0.923	0.015	0.433	0.395	0.927
β_1	1.0	0.070	0.269	0.244	0.944	0.038	0.172	0.170	0.944
β_2	-1.0	-0.072	0.487	0.424	0.936	-0.029	0.315	0.301	0.948
γ_0	1.0	0.092	0.359	0.343	0.956	0.068	0.254	0.253	0.948
γ_1	-1.0	-0.030	0.410	0.400	0.958	-0.021	0.272	0.278	0.960
<i>Model 6: proportional odds type model with $R(x) = \log(1 + x)$</i>									
$\Lambda(\pi/4)$	0.75	0.073	0.365	0.324	0.919	0.015	0.262	0.240	0.922
$\Lambda(\pi/2)$	1.50	0.103	0.825	0.788	0.945	0.093	0.608	0.571	0.938
β_1	1.0	0.096	0.352	0.326	0.931	0.051	0.211	0.211	0.946
β_2	-1.0	-0.082	0.576	0.533	0.932	-0.037	0.352	0.361	0.952
γ_0	1.0	0.079	0.331	0.347	0.953	0.054	0.254	0.261	0.964
γ_1	-1.0	-0.047	0.409	0.410	0.951	-0.016	0.292	0.284	0.942

Table 2

Results of simulations for the two-mixture model using the proposed method and the EM algorithm. Empirical bias (Bias), sampling standard error (SE), mean of standard error estimator (SEE), and mean squared error (MSE).

<i>n</i>	Par	Direct maximization					EM algorithm					MSE ratio (MSE _{DM} /MSE _{EM})
		Bias	SE	SEE	MSE _{DM}		Bias	SE	SEE	MSE _{em}		
100	β_1	0.032	0.236	0.222	0.057		0.047	0.240	0.224	0.060		0.948
	β_2	-0.053	0.470	0.461	0.224		-0.042	0.444	0.429	0.199		1.124
	γ_0	0.093	0.496	0.513	0.255		0.086	0.616	0.606	0.387		0.658
	γ_1	-0.021	0.591	0.611	0.350		-0.058	0.670	0.682	0.452		0.773
200	β_1	0.022	0.161	0.158	0.074		0.038	0.160	0.157	0.027		0.976
	β_2	-0.036	0.273	0.268	0.076		-0.040	0.286	0.269	0.083		0.909
	γ_0	0.061	0.372	0.368	0.142		0.055	0.374	0.381	0.143		0.994
	γ_1	-0.006	0.412	0.415	0.170		-0.063	0.410	0.423	0.172		0.987

Table 3

Estimation results for the soft-tissue sarcoma study: nonparametric maximum likelihood estimates and standard errors (in parentheses), using the transformation models with various values of $\rho = 0, 1$ and 2 .

Covariate		$R(x; \rho) = (1/\rho) \log(1 + \rho x)$		
		$\rho = 0$	$\rho = 1$	$\rho = 2$
<i>(i) Model 1: two-mixture cure model</i>				
β	Chemotherapy	-0.299 (0.162)	-0.554 (0.217)	-0.744 (0.275)
	Radiation	-0.283 (0.163)	-0.501 (0.218)	-0.551 (0.272)
	log (Tumor size)	0.373 (0.179)	0.321 (0.246)	0.634 (0.303)
γ	Intercept	-1.385 (0.640)	-1.005 (0.564)	-1.838 (0.676)
	Chemotherapy	0.396 (0.217)	0.316 (0.191)	0.482 (0.239)
	Radiation	-0.262 (0.222)	-0.381 (0.195)	-0.348 (0.232)
	log (Tumor size)	0.688 (0.253)	0.532 (0.229)	0.932 (0.277)
Log-likelihood ($\times 10^{-3}$)		-2.9490	-2.9434	-2.9368
<i>(ii) Model 2: promotion-time cure model</i>				
β	Chemotherapy	-0.579 (0.341)	-0.441 (0.285)	-0.763 (0.265)
	Radiation	-0.683 (0.411)	-0.679 (0.300)	-0.529 (0.265)
	log (Tumor size)	0.627 (0.322)	0.721 (0.315)	0.154 (0.171)
γ	Intercept	-1.138 (0.713)	-1.151 (0.543)	-1.338 (0.362)
	Chemotherapy	0.414 (0.283)	0.232 (0.173)	0.171 (0.120)
	Radiation	0.162 (0.344)	0.038 (0.189)	-0.241 (0.291)
	log (Tumor size)	0.364 (0.264)	0.418 (0.196)	0.462 (0.142)
Log-likelihood ($\times 10^{-3}$)		-2.9386	-2.9356	-2.9646
<i>(iii) Model 3: proportional odds type cure model</i>				
β	Chemotherapy	-0.563 (0.339)	-0.781 (0.329)	-0.734 (0.296)
	Radiation	-0.243 (0.336)	-0.371 (0.324)	-0.297 (0.296)
	log (Tumor size)	0.163 (0.336)	0.236 (0.365)	0.347 (0.338)
γ	Intercept	-1.708 (0.796)	-1.698 (0.691)	-1.184 (0.562)
	Chemotherapy	0.411 (0.288)	0.458 (0.246)	0.281 (0.197)
	Radiation	-0.253 (0.284)	-0.213 (0.242)	-0.356 (0.194)
	log (Tumor size)	0.856 (0.316)	0.833 (0.278)	0.676 (0.255)
Log-likelihood ($\times 10^{-3}$)		-2.9412	-2.9382	-2.9424