Published in final edited form as:

Can J Stat. 2014 March ; 42(1): 18–35. doi:10.1002/cjs.11198.

A semiparametric linear transformation model to estimate causal effects for survival data

Huazhen LIN^{1,*}, Yi LI¹, Liang JIANG², Gang LI³

¹Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, P. R. China

²School of Mathematics, Sichuan University, Chengdu, Sichuan 610064, P. R. China

³Department of Biostatistics, School of Public Health, University of California, Los Angeles, California, USA

Abstract

Semiparametric linear transformation models serve as useful alternatives to the Cox proportional hazard model. In this study, we use the semiparametric linear transformation model to analyze survival data with selective compliance. We estimate regression parameters and the transformation function based on pseudo-likelihood and a series of estimating equations. We show that the estimators for the regression parameters and transformation function are consistent and asymptotically normal, and both converge to their true values at the rate of $n^{-1/2}$, the convergence rate expected for a parametric model. The practical utility of the methods is confirmed via simulations as well as an application of a clinical trial to evaluate the effectiveness of sentinel node biopsy in guiding the treatment of invasive melanoma.

Résumé:

Les modèles semi-paramétriques de transformation linéaire constituent des solutions de rechange utiles au modèle à risques proportionnels de Cox. Dans cet article, les auteurs utilisent un modèle semi-paramétrique de transformation linéaire pour analyser des données de survie affichant une observance sélective. Les auteurs estiment les paramètres de régression et la fonction de transformation de ce modèle au moyen de la pseudo-vraisemblance et d'une série d'équations d'estimation. Ils montrent que ces estimateurs convergent et sont asymptotiquement normaux. Leur vitesse de convergence de $n^{-1/2}$ correspond au taux attendu dans un modèle paramétrique. L'utilité pratique des méthodes est confirmée à l'aide de simulations et d'une application à un essai clinique visant à évaluer l'efficacité d'une biopsie des ganglions sentinelles dans le choix d'un traitement pour un mélanome invasif.

Keywords

Cox proportional hazard model; selective compliance; semiparametric linear transformation model; survival data

* Author to whom correspondence may be addressed. linhz@swufe.edu.cn.

1. INTRODUCTION

In controlled randomized trials, subjects are generally assigned a treatment regime. However, these subjects rarely exhibit perfect compliance with the treatment. Noncompliance typically renders intention-to-treat (ITT) estimators, which mix the effects of treatment in compliers with a lack of effect in noncompliers, incapable of measuring actual biological efficacy. On the other hand, the dangers of excluding noncompliant patients are well recognized (Altman, 1991). Such exclusion may result in biases because the subsets of compliers and noncompliers have different prognoses at baseline. Thus, the question about the causal effect of the treatment that is actually received becomes difficult to answer.

To handle survival outcomes with selective compliance, Frangakis & Rubin (1999), Loeys & Goetghebeur (2002), and Elashoff, Li, & Zhou (2012) proposed general nonparametric procedures to estimate the causal effect of treatment for data without covariates. For data with covariates, Robins & Tsiatis (1991), Altstein, Li, & Elashoff (2011), and Altstein & Li (2013) proposed accelerated failure time frameworks to quantify the causal effect of treatment. The most widely used regression method for analyzing survival data is the proportional hazards model (Cox, 1972), and work in this direction has been conducted by Robin & Finkelstein (2000), Hernan, Brumback, & Robins (2001), Loeys & Goetghebeur (2003), Loeys, Goetghebeur, & Vandebosch (2005), and Cuzick et al. (2007).

In this paper, we consider a semiparametric linear transformation model. The transformation of survival time is unknown but the distribution of the error term is specified. This class of regression models, which includes proportional hazards (Cox, 1972) and proportional odds models (Bennett, 1983), serves as a useful alternative to the Cox model in survival analysis. Pettitt (1982), Bennett (1983), Clayton & Cuzick (1985), Dabrowska & Doksum (1988), Cheng, Wei, & Ying (1995), Murphy, Rossini, & der Vaart (1997), Scharfstein, Tsiatis, & Gilbert (1998), Cai, Wei, & Wilcox (2000), Cai, Cheng, & Wei (2002), Chen, Jin, & Ying (2002), and Zeng & Lin (2006), among others, proposed estimators for this class of linear transformation models for the analysis of survival data without selective compliance.

Extending the partial likelihood method used in the Cox proportional hazards model to a general transformation model is difficult if is not impossible. In this paper, we propose a two-stage estimation procedure to estimate the parameters and the transformation function in the semiparametric linear transformation model for survival data with selective compliance. The proposed estimator is shown to be consistent and asymptotically normal. The estimator is also easy to compute because a likelihood function need not be maximized over an infinite dimensional parameter space, and our approach in estimating the transformation functions does not involve nonparametric smoothing. Thus, our estimator does not suffer from smoothing related problems, such as, the selection of a smoothing parameter.

The article is organized as follows. A detailed description of the aforementioned models is given in Section 2. A two-stage estimation procedure is described in Section 3. The asymptotic properties are derived in Section 4. An estimator of the variance is provided in Section 5. Section 6 contains the simulation results, and the application to a study, the

Multicenter Selective Lymphadenectomy Trial-I (MSLT-I), is detailed in Section 7. Section 8 provides the concluding remarks.

2. NOTATIONS AND MODEL

Consider a trial where n independent subjects are randomly assigned to an experimental treatment ($R_i = 1$) or a control group ($R_i = 0$). Let T_i and C_i denote survival and censoring time, respectively; let $W_i = \min(T_i, C_i)$ and δ_i indicate the corresponding observation time and censoring status, respectively. We assume uninformative censoring. However, not all randomized subjects receive their assigned treatment. The trial population is generally considered as being composed of three classes of individuals: (a) *compliers* which refers to subjects who accept whatever allocation they are offered; (b) *always takers*, which refers to subjects who demand treatment regardless of the group to which they are randomized; and (c) *never takers*, which refers to subjects who refuse the treatment, if and when offered. Let U_i be the indicator of the potential class of individuals for subject i . By convention, the values $U_i = 1, 2, 3$ refer to compliers, always takers and never takers, respectively. Statistical difficulties arise because determining the class to which the individuals belong is not always possible. The ITT principle contrasts $Pr(T_i > t | R_i = 1)$ with $Pr(T_i > t | R_i = 0)$ but does not estimate the casual effect of a treatment because a number of patients in the control and treatment groups do not comply for reasons related to outcome.

Let E_i be the indicator of the treatment that is actually received by subject i . $E_i = 1$ denotes that subject i actually received treatment; otherwise, $E_i = 0$. Hence, $E_i = R_i I(U_i = 2) + (1 - R_i) I(U_i = 3)$. U_i is not completely observed, but E_i is observable. Our goal is to identify the population averaged causal effect of treatment with adjustment for covariates. Specifically, given that $U_i = 2$ implies $E_i = 1$ and $U_i = 3$ implies $E_i = 0$, we contrast the survival probability in the treated group $Pr(T_i > t | E_i = 1, U_i = 1, \mathbf{X}_i = \mathbf{x})$ with its subpopulation-specific counterpart in the control group $Pr(T_i > t | E_i = 0, U_i = 1, \mathbf{X}_i = \mathbf{x})$. We assume that a monotone function H exists, such that the survival time T can be modeled as follows:

$$H(T_i) = c_u + \alpha_u E_i + \beta' \mathbf{X}_i + \varepsilon_i, \quad \text{for } u \in \{1, 2, 3\},$$

(1)

where H is the unknown monotonic transformation function; α_1 is the unknown parameter of interest used to measure the casual effect; $\mathbf{X}_i$ is the covariate vector measured prior to randomization; and ε_i is a random variable with mean zero, variance σ^2 , and a known distribution F , which is independent of $(E_i, \mathbf{X}_i)$. For model identification, the conditions $\alpha_2 = \alpha_3 = 0$ are required because $U_i = 2$ implies $E_i = 1$ and $U_i = 3$ implies $E_i = 0$. The condition $\sigma^2 = 1$ can be assumed without loss of generality because the model (1) holds if H is replaced by cH for any positive constant c . On the other hand, the condition $H(a_0) = b_0$ is required because the model (1) holds if H is replaced by $H - c$ for any constant c , where a_0 and b_0 are any given finite numbers. In practice, we can take a_0 as the median of the observed failure times among n observations to ensure that a_0 is within the support set of H ,

so that H can be completely identified. The choices of a_0 and b_0 have a negligible effect to the resulting estimators. A simulation study on sensitivity has been conducted by Lin, Fan, & Zhou (2013) for a more general transformation model for longitudinal data.

For survival data without selective compliance, the model (1) has been extensively considered. The proportional hazards and the proportional odds models are special cases of (1), with ε_i following the extreme-value distribution and the standard logistic distribution, respectively. Thus, the Cox proportional hazard model proposed by Hernan, Brumback, & Robins (2001) and Cuzick et al. (2007) to analyze the causal effect of treatment for data with covariates and the model considered by Loeys & Goetghebeur (2003) to estimate the biological efficacy without covariates, are special cases of the model (1).

As shown in the literature, an implicit assumption in model (1) is “the exclusion restriction” used by Angrist, Imbens, & Rubin (1996) or the “absence of indirect effect” termed by Pearl (2002), which states that potential outcomes are unrelated to treatment assignment, that is, $Pr(R_i = 1 \mid T_i, U_i) = Pr(R_i = 1 \mid U_i)$.

Considering that different characteristics, such as disease status and prognostic factors, may affect the compliance status of subjects, a regression model for compliance status is required. We model the subgroup probability by using the logistic regression model

$$Pr(U_i = j \mid \mathbf{Z}_i) = \frac{\exp(\theta_j + \kappa'_j \mathbf{Z}_i)}{1 + \sum_{u=1}^2 \exp(\theta_u + \kappa'_u \mathbf{Z}_i)}, \quad \text{for } j = 1, 2,$$

(2)

where $\mathbf{Z}_i$ is the covariate, possibly different from $\mathbf{X}_i$. Thus, the subgroup probability depends on the subject's covariate vector but not on the randomization indicator R_i . Our simulation studies show that the choice of subgroup probability in model (2) is less important to the properties of the resulting estimators.

Although U_i is not completely observed, the survival distribution of compliers in the control and treatment groups can be identified by using the exclusion-restriction principle. First we note that only four groups can be observed from the data: always takers and compliers to treatment group ($R_i = 1, U_i = 2$); never takers to treatment group ($R_i = 1, U_i = 3$); always takers to control group ($R_i = 0, U_i = 2$); and compliers and never takers to control group ($R_i = 0, U_i = 3$). Considering that U_i and R_i are independent given $\mathbf{Z}_i$, $p_2(\mathbf{z}) = P(U_i = 2 \mid \mathbf{Z}_i = \mathbf{z})$ and $p_3(\mathbf{z}) = P(U_i = 3 \mid \mathbf{Z}_i = \mathbf{z})$ can be consistently estimated by maximizing the following likelihood function,

$$\prod_{i=1}^n p_2(\mathbf{Z}_i)^{I(U_i=2)} p_3(\mathbf{Z}_i)^{I(U_i=3)} \{1 - p_3(\mathbf{Z}_i)\}^{I(U_i=2)} \{1 - p_2(\mathbf{Z}_i)\}^{I(U_i \neq 2)},$$

so can $p_1(\mathbf{z}) = P(U_i = 1 \mid \mathbf{Z}_i = \mathbf{z})$. Furthermore, define $S_i(t \mid \mathbf{x}, \mathbf{z}) = P(T_i > t \mid R_i = r, \mathbf{X}_i = \mathbf{x}, \mathbf{Z}_i = \mathbf{z})$ and let $S_{ru}(t \mid \mathbf{x}, \mathbf{z}) = P(T_i > t \mid R_i = r, U_i = u, \mathbf{X}_i = \mathbf{x}, \mathbf{Z}_i = \mathbf{z})$. Under the exclusion restriction, $S_{12}(t \mid \mathbf{x}, \mathbf{z}) = S_{02}(t \mid \mathbf{x}, \mathbf{z})$ and $S_{13}(t \mid \mathbf{x}, \mathbf{z}) = S_{03}(t \mid \mathbf{x}, \mathbf{z})$, and then

$$\begin{aligned} S_0(t \mid \mathbf{x}, \mathbf{z}) &= p_1(\mathbf{z})S_{01}(t \mid \mathbf{x}, \mathbf{z}) + p_2(\mathbf{z})S_{02}(t \mid \mathbf{x}, \mathbf{z}) + \{1 - p_1(\mathbf{z}) - p_2(\mathbf{z})\}S_{13}(t \mid \mathbf{x}, \mathbf{z}), \\ S_1(t \mid \mathbf{x}, \mathbf{z}) &= p_1(\mathbf{z})S_{11}(t \mid \mathbf{x}, \mathbf{z}) + p_2(\mathbf{z})S_{02}(t \mid \mathbf{x}, \mathbf{z}) + \{1 - p_1(\mathbf{z}) - p_2(\mathbf{z})\}S_{13}(t \mid \mathbf{x}, \mathbf{z}). \end{aligned}$$

Because $p_1(\mathbf{z})$, $p_2(\mathbf{z})$, $S_0(t \mid \mathbf{x}, \mathbf{z})$, $S_1(t \mid \mathbf{x}, \mathbf{z})$, $S_{02}(t \mid \mathbf{x}, \mathbf{z})$, and $S_{13}(t \mid \mathbf{x}, \mathbf{z})$ can be estimated consistently, so can $S_{01}(t \mid \mathbf{x}, \mathbf{z})$ and $S_{11}(t \mid \mathbf{x}, \mathbf{z})$.

3. ESTIMATION

3.1. Estimation of the Parameters

Let $\Theta = (\alpha_1, \mathbf{c}', \boldsymbol{\beta}', \theta_1, \theta_2, \kappa'_1, \kappa'_2)' \equiv (\Theta_1, \dots, \Theta_d)'$, where $\mathbf{c} = (c_1, c_2, c_3)'$. Θ and H are the parameter and nonparametric function defined by the models (1) and (2). Under the model (1), noting that H is a monotonic function and that the subgroup probability is independent of the treatment assignment R_i , the likelihood for the observed data can be expressed as

$$\begin{aligned} L(\Theta; H) &\propto \prod_{i=1}^n \{dH(W_i)\}^{\delta_i} \\ &\times ([f\{H(W_i) - c_1 - \alpha_1 - \boldsymbol{\beta}'\mathbf{X}_i\}p_{1i}(\Theta) + f\{H(W_i) - c_2 - \boldsymbol{\beta}'\mathbf{X}_i\}p_{2i}(\Theta)]^{\delta_i} \\ &\times [S\{H(W_i) - c_1 - \alpha_1 - \boldsymbol{\beta}'\mathbf{X}_i\}p_{1i}(\Theta) + S\{H(W_i) - c_2 - \boldsymbol{\beta}'\mathbf{X}_i\}p_{2i}(\Theta)]^{(1-\delta_i)R_iI(U_i \leq 2)} \\ &\times ([f\{H(W_i) - c_3 - \boldsymbol{\beta}'\mathbf{X}_i\}]^{\delta_i} [S\{H(W_i) - c_3 - \boldsymbol{\beta}'\mathbf{X}_i\}]^{1-\delta_i} p_{3i}(\Theta))^{R_iI(U_i = 3)} \\ &\times ([f\{H(W_i) - c_2 - \boldsymbol{\beta}'\mathbf{X}_i\}]^{\delta_i} [S\{H(W_i) - c_2 - \boldsymbol{\beta}'\mathbf{X}_i\}]^{1-\delta_i} p_{2i}(\Theta))^{(1-R_i)I(U_i = 2)} \\ &\times ([f\{H(W_i) - c_1 - \boldsymbol{\beta}'\mathbf{X}_i\}p_{1i}(\Theta) + f\{H(W_i) - c_3 - \boldsymbol{\beta}'\mathbf{X}_i\}p_{3i}(\Theta)]^{\delta_i} \\ &\times [S\{H(W_i) - c_1 - \boldsymbol{\beta}'\mathbf{X}_i\}p_{1i}(\Theta) + S\{H(W_i) - c_3 - \boldsymbol{\beta}'\mathbf{X}_i\}p_{3i}(\Theta)]^{1-\delta_i})^{(1-R_i)I(U_i \neq 2)}, \end{aligned} \quad (3)$$

where f is the density function of F , $S(x) = 1 - F(x)$,

$$p_{1i}(\Theta) = \exp(\theta_1 + \kappa'_1 \mathbf{Z}_i) / (1 + \sum_{u=1}^2 \exp(\theta_u + \kappa'_u \mathbf{Z}_i)),$$

$p_{2i}(\Theta) = \exp(\theta_2 + \kappa'_2 \mathbf{Z}_i) / (1 + \sum_{u=1}^2 \exp(\theta_u + \kappa'_u \mathbf{Z}_i))$, and $p_{3i}(\Theta) = 1 - p_{1i}(\Theta) - p_{2i}(\Theta)$. The likelihood function involves both the finite dimensional parameter Θ and the infinite

dimensional parameter H . Maximization of the likelihood function over an infinite dimensional parameter space can be complicated, especially when the form of the objective function is as complicated as the one described in this paper. In fact, even for the simple case of survival data without selective compliance, the computation of the nonparametric maximum likelihood estimator for the transformation is difficult (Zeng & Lin, 2006). In addition, the nonparametric maximum likelihood estimator of H is not assured to be monotonic. In this article, we use a two-stage approach; our estimator for the transformation function is monotone and the computation is easy because the maximization of the likelihood function over an infinite dimensional parameter space is avoided. First, we use a series of estimating equations described in Section 3.2 to estimate the transformation function. Then, the parameter Θ is estimated by maximizing a pseudo likelihood, which is the likelihood function $L(\Theta; H)$ with H being replaced by the estimated values. We repeat the procedure until convergence.

3.2. Estimation of the Transformation Functions

In this section, we estimate the transformation function H given Θ . Throughout this article, let $0 < t_1 < \dots < t_K$ denote the observed K failure times among the n observations. Denote

$$\begin{aligned} \Lambda_i(x; \Theta) = & -R_i I(U_i \leq 2) \log \left\{ \frac{S(x - c_1 - \alpha_1 - \beta' \mathbf{X}_i) p_{1i}(\Theta) + S(x - c_2 - \beta' \mathbf{X}_i) p_{2i}(\Theta)}{p_{1i}(\Theta) + p_{2i}(\Theta)} \right\} \\ & - R_i I(U_i = 3) \log \{ S(x - c_3 - \beta' \mathbf{X}_i) \} - (1 - R_i) I(U_i = 2) \log \{ S(x - c_2 - \beta' \mathbf{X}_i) \} \\ & - (1 - R_i) I(U_i \neq 2) \log \left\{ \frac{S(x - c_1 - \beta' \mathbf{X}_i) p_{1i}(\Theta) + S(x - c_3 - \beta' \mathbf{X}_i) p_{3i}(\Theta)}{1 - p_{2i}(\Theta)} \right\}. \end{aligned}$$

Using the usual counting process notation, let $Y_i(t) = I(W_i \leq t)$, $N_i(t) = \delta_i I(W_i \leq t)$, $M_i(t) = N_i(t) - \int_0^t Y_i(s) d\Lambda_i(H_0(s); \Theta_0)$. Motivated by the fact that $M_i(t)$ is a martingale process, we consider the following estimating equation to estimate $H(t)$ for any $t \geq 0$,

$$\sum_{i=1}^n \{ dN_i(t) - Y_i(t) d\Lambda_i(H(t); \Theta) \} = 0,$$

(4)

where $H(0) = -\infty$. We denote by $\hat{H}(t; \Theta)$ the solution of (4) with respect to $H(t)$ given Θ . It is clear that $\hat{H}(\cdot; \Theta)$ is nondecreasing step function on $[0, \infty)$ with $H(0) = -\infty$ and with jumps only at the observed failure times $t_1, \dots, t_K$. Then, solving the infinite system of estimating equations defined by (4) is equivalent to solving the infinite system of equations. Denote by $\lambda_{\hat{H}}(x; \Theta) = \Lambda_{\hat{H}}(x; \Theta) / x$. Following the discussions in Chen, Jin, & Ying (2002) and using $H(0) = -\infty$ and (4), we propose to estimate $H(t_1)$ by solving

$$\sum_{i=1}^n \{dN_i(t_1) - Y_i(t_1)\lambda_i(H(t_1); \Theta)\} = 0,$$

and obtain $\hat{H}(t_k)$ for $k = 2, \dots, K$, one-by-one by the equations

$$\hat{H}(t_k) = \hat{H}(t_{k-1}) + \frac{\sum_{i=1}^n dN_i(t_k)}{\sum_{i=1}^n Y_i(t_k)\lambda_i(\hat{H}(t_{k-1}); \Theta)}.$$

The computation of $\hat{H}(t)$ is considerably simple. In addition, unlike a traditional nonparametric approach used to estimate the transformation function (Horowitz, 1996; Zhou, Lin, & Johnson, 2009), our approach does not involve nonparametric smoothing.

4. INFERENCE IN LARGE SAMPLES

In this section, we present the large sample properties of the proposed estimators. Let $\hat{\Theta}$ and $\hat{H}(t)$ denote the estimators of Θ and $H(t)$, respectively; and let Θ_0 and H_0 be the true values of Θ and H , respectively. Some notation and regular conditions are needed. Let $\tau = \inf\{t: \Pr(W_i > t) = 0\}$ and $\psi(t) = (-1/t) \log(f(t)/S(t))$. We assume $f(t)/S(t) > 0$, $\psi(t)$ to be continuous, and $\lim_{t \rightarrow -\infty} f(t)/S(t) = \lim_{t \rightarrow -\infty} \psi(t) = 0$. Furthermore, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are bounded, and H_0 has continuous and positive derivatives. Define $\lambda_{f(x)} = \lambda_f(x; \Theta_0)$,

$$\dot{\lambda}_f(x) = \frac{d\lambda_f(x)}{dx}, \quad \lambda^*(H_0(t)) = \exp \left[\int_0^t \frac{E\{Y_i(x)\dot{\lambda}_f(H_0(x))\}dH_0(x)}{E\{Y_i(x)\lambda_f(H_0(x))\}} \right],$$

$$\begin{aligned} \Sigma &= E \left[\frac{\partial^2 \log L_i(\Theta_0; H_0)}{\partial \Theta \partial \Theta'} - \frac{\partial^2 \log L_i(\Theta_0; H_0)}{\partial \Theta \partial H(W_i)} \frac{1}{\lambda^*\{H_0(W_i)\}} \right. \\ &\quad \times \left. \int_0^\tau Y_i(s) \frac{\lambda^*\{H_0(s)\}}{E[Y_j(s)\lambda_j(H_0(s))]} E \left[Y_j(s) \frac{\partial \lambda_j(H_0(s); \Theta_0)}{\partial \Theta'} \right] dH_0(s) \right], \end{aligned}$$

$$\begin{aligned} \Delta &= E \left[\frac{\partial \log L_i(\Theta_0; H_0)}{\partial \Theta} \right. \\ &\quad \left. + \int_0^\tau E \left[\frac{Y_j(s)}{\lambda^*(H_0(W_j))} \frac{\partial^2 \log L_j(\Theta_0; H_0)}{\partial \Theta \partial H(W_j)} \right] \frac{\lambda^*(H_0(s))}{E[Y_j(s)\lambda_j(H_0(s))]} dM_{t(s)} \right] \otimes 2, \end{aligned}$$

where $L_f(\Theta; H)$ is the contribution of subject i to the likelihood $L(\Theta; H)$ and $a^{\otimes 2} = aa'$ for any vector a . Assume that Σ and Δ are finite and nondegenerate. Some additional technical conditions are presented in the Appendix.

Theorem 1.

As $n_1 = \sum_{i=1}^n R_i \rightarrow \infty$ and $n_2 = \sum_{i=1}^n (1 - R_i) \rightarrow \infty$, we have

$$\|\hat{\Theta} - \Theta_0\| \rightarrow 0, \quad \sup_{t \in [a, \tau]} |\hat{H}(t) - H_0(t)| \rightarrow 0$$

in probability for any fixed $a \in (0, \tau]$.

Theorem 2.

As $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$, we have

$$\sqrt{n}(\hat{\Theta} - \Theta_0) \rightarrow N(0, \Sigma^{-1} \Delta (\Sigma^{-1})').$$

(5)

Theorem 3.

As $n_1 \rightarrow \infty$ and $n_2 \rightarrow \infty$, for any $t \in (0, \tau)$, we have

$$\sqrt{n}(\hat{H}(t) - H_0(t)) \rightarrow N\left(0, \frac{\Gamma(t)}{\lambda^*{}^2(H_0(t))}\right),$$

where

$$\begin{aligned} \Gamma(t) = & E \left[\mu'(t) \Sigma^{-1} \frac{\partial \log L_i(\Theta_0; H_0)}{\partial \Theta} + \mu'(t) \Sigma^{-1} \int_0^\tau E \left[\frac{Y_j(s)}{\lambda^*(H_0(w_j))} \frac{\partial^2 \log L_j(\Theta_0; H_0)}{\partial \Theta \partial H(w_j)} \right] \right. \\ & \left. \times \frac{\lambda^*(H_0(s))}{E[Y_j(s) \lambda_j(H_0(s))]} dM_{t^-(s)} + \int_0^t \frac{\lambda^*(H_0(s))}{E[Y_j(s) \lambda_j(H_0(s))]} dM_{t^-(s)} \right]^2, \end{aligned}$$

$$\text{and } \mu(t) = \int_0^t \frac{\lambda^*(H_0(s))}{E[Y_j(s) \lambda_j(H_0(s))]} E \left[Y_j(s) \frac{\partial \lambda_j(H_0(s); \Theta_0)}{\partial \Theta} \right] dH_0(s).$$

From Theorem 3, we see that $\hat{H}(t)$ converges to $H(t)$ at the rate of $n^{-1/2}$; this result shows that the nonparametric function $H(\cdot)$ has a parametric convergence rate. The conclusion, that the transformation function can be estimated with $n^{-1/2}$ rate of convergence, has also been reached by Horowitz (1996), Chen, Jin, & Ying (2002), Ye & Duan (1997), Zhou, Lin, & Johnson (2009), and Lin & Zhou (2009), among others.

5. ESTIMATION OF ASYMPTOTIC VARIANCE OF $\hat{\Theta}$

As shown in Theorem 2, the asymptotic variance of $\hat{\Theta}$ has a standard sandwich form $\Sigma^{-1} (\Sigma^{-1})'$. However, the matrices Σ and Σ^{-1} exhibit complicated analytic forms that require complicated computation. Therefore, a feasible computation approach to approximate the asymptotic variance of $\hat{\Theta}$ is needed. In this section, we use the resampling scheme proposed by Jin, Ying, and Wei (2001) to approximate the asymptotic distribution of $\hat{\Theta}$. Compared with the simple bootstrap resampling method, Jin, Ying, and Wei's resampling scheme, that uses continuous weights, is more stable. The resampling algorithm proceeds as follows. First, we generate n standard exponential random variables ξ_i , $i = 1, \dots, n$. Fixing the data at their observed values, we solve the following ξ_i -weighted estimation equations and denote the solutions as Θ^* and $H^*(t)$ for any $t > 0$:

$$\sum_{i=1}^n \xi_i \{dN_i(t) - Y_i(t)d\Lambda_i(H(t); \Theta)\} = 0,$$

(6)

$$\sum_{i=1}^n \xi_i \log(L_i(\Theta; H)) = 0,$$

(7)

where $H(0) = -\infty$. Here, the exponential random variable is unnecessary, any positive random variable with mean 1 and variance 1 can be used. The estimates Θ^* and $H^*(\cdot)$ can be obtained by using the iterative algorithm proposed in Section 3. Following the method of Jin, Ying, & Wei (2001) and by using Equations (A.1) and (A.4) in the Appendix, we establish the validity of the proposed resampling method.

Proposition:

Under the conditions given in the Appendix, the conditional distribution of $n^{1/2}(\Theta^ - \hat{\Theta})$, given the observed data, converges almost surely to the asymptotic distribution of $n^{1/2}(\hat{\Theta} - \Theta_0)$.*

Based on this proposition, by repeatedly generating $\xi_1, \dots, \xi_n$ numerous times, we may obtain a large number of realizations of Θ^* . The variance estimate of $\hat{\Theta}$ can then be approximated by the empirical variance of Θ^* .

6. SIMULATION

In this section, we describe the simulation studies conducted to assess the finite-sample performance of the proposed method by comparing it with other existing methods. The existing approaches to analyze survival data with noncompliance and adjustment for covariates include (1) accelerated failure time frameworks (AFT, Robins & Tsiatis, 1991; Altstein, Li, & Elashoff, 2011; Altstein & Li, 2013) and (2) the Cox proportional hazards model (Robin & Finkelstein, 2000; Hernan, Brumback, & Robins, 2001; Loeys & Goetghebeur, 2003; Loeys, Goetghebeur, & Vandebosch, 2005; Cuzick et al., 2007). Considering that the existing literature on the AFT model assumes that two subgroups exist in the population, whereas we assume three subgroups, the direct comparison of our method and the existing AFT model is infeasible. Given that the AFT model can be regarded as the model (1) with the specified transformation function, we evaluate the AFT model by investigating the performance of the model (1) with misspecified transformation function (indicated by “MT”). We choose the full likelihood method developed by Cuzick et al. (2007) as a representative of the Cox proportional hazards model (indicated by “Cox”).

We present four simulations in this paper. In all of the simulations, we choose $n = 400$ and generate T_i by using the model (1) with $\alpha_1 = 2$, $\beta = (1, 1)'$ and $\mathbf{X}_i = (X_{i1}, X_{i2})'$, where X_{i1} and X_{i2} are independently generated by the standard normal distribution. We take the transformation functions $H(t) = \log(t)$ and $H(t) = \log\{\log(t)\}$ and generate ε_i in the model (1) by using the hazard function of the form $\lambda(t) = \exp(t)/(1 + r \exp(t))$ with $r = 0$ and 1 . Notably, the proportional hazards and proportional odds models correspond to $r = 0$ and $r = 1$, respectively. We divide the n observations equally into the treatment and control groups. Subsequently, we generate R_i , then, $E_i = R_i I(U_i \leq 2) + (1 - R_i) I(U_i > 2)$, where U_i is generated by the model (2) with $\theta_1 = \theta_2 = 0$, $\kappa_1 = (0.5, -0.5)'$ and $\kappa_2 = (0.5, -0.5)'$. Considering that the model assumption on U_i used by Cuzick et al. (2007) differs from that used in this paper for the case of class membership correlated with covariates, to compare the proposed method with that of Cuzick et al. (2007), we also generate U_i by using the model (2) with $\theta_1 = \theta_2 = 0$, $\kappa_1 = \kappa_2 = (0, 0)'$, that is, the class membership is independent of covariates. The independent censoring time C_i is generated from the uniform distribution on $(0.5, a)$. The choice of a yields censoring proportions of approximately 30–40% for each case. For simplicity in presentation, we list the simulation settings in Table 1. We simulated 500 datasets for each setting.

For Cases 1 and 2, we analyze each data of 500 datasets by using the proposed method and the MT estimator, where the transformation function is misspecified by the identity function. However, the MT estimator has great difficulty in estimating the regression coefficients properly, thus causing all runs to fail to converge when the initial value is selected as the true values and the estimator based on the linear regression model with noncompliance to be disregarded, as initial values. Convergence failure implies that the MT estimator is far away from the true values and is severely biased. Tables 2 and 3 present the biases and empirical standard deviations (SDs) of the parameter estimators for Cases 1 and 2, respectively, obtained from the 500 simulations by using the proposed method. Tables 4 and 5 present the biases and SD for Cases 3 and 4, respectively, which were obtained by using the proposed method and the Cox estimator. Figure 1a–d display the average of the estimated

transformation functions and their empirical pointwise 95% confidential intervals for Cases 1 to 4, respectively, which were derived by using the proposed method.

From Tables 2–5 and Figure 1, we observed the following:

- Figure 1 shows that the proposed method produces reasonable estimates of the transformation functions.
- A useful rule of thumb for evaluating the bias is that it should not have a substantial negative impact on inferences unless the standardized bias (bias as a percent of the SD) exceeds 40% (Olsen & Schafer, 2001). Tables 2 and 3 show that the proposed estimators are unbiased in all cases. By contrast, the MT estimators may have significantly large bias, suggesting that the misspecification of the transformation function yields a biased estimator.
- Interestingly, Table 4 shows that the proposed estimator has less bias and variance than the full likelihood method developed by Cuzick et al. (2007) although a Cox model is appropriate. This result is surprising. By further checking the Cuzick et al.'s estimator, we find that the full likelihood function developed by Cuzick et al. (2007) involves $\pi_{Hi} = \min(N_i^{CT} \rho / N_i^{TT}, 1)$ and $\pi_{Ri} = \min(N_i^{TC} \rho^{-1} / N_i^{CC}, 1)$. The π_{Hi} is estimate of the proportion of individuals complying with randomization to treatment and at risk at time t_i — who would have insisted on receiving new treatment even if they had been randomized to control. π_{Ri} has a similar interpretation as the estimated proportion of refusers among the CC group at t_i , where the definitions of CC, CT, TC, and TT can be found in Cuzick et al. (2007). π_{Hi} and π_{Ri} are not convincing estimators when N_i^{TT} and N_i^{CC} become small, which actually occurs as t_i becomes large.
- Table 5 shows that the Cox estimator is biased and unstable when a Cox model is inappropriate, whereas the proposed estimators appear consistently better than the Cox method.

We also test the accuracy of estimation of the standard error given in Section 5. The SDs of the 500 estimators for parameters, based on 500 simulations and also presented in Tables 2–5, can be regarded as the true standard errors. The average and the standard deviation of 500 estimated bootstrap standard errors, denoted by SE_{ave} and SE_{sd} summarizes the overall performance of the standard error estimator. Tables 2–5 present the results, which showing that the standard error estimators are generally close to the empirical standard errors and are reasonably adequate as a resampling method for estimating the standard errors. We also display the empirical coverage probabilities (CP) of 95% confidence intervals in Tables 2–5.

We investigate the issue of whether the misspecification of the model on the subgroup probability yields a biased estimator by testing the performance of the proposed estimators with the misspecification of the subgroup probability model (indicated by “MSP”) for Case 1. Table 2 presents the biases and SDs of the MSP estimators, where κ_1 and κ_2 are misspecified as zero, that is, the class membership is independent of the covariates. The

results in Table 2 suggest that the misspecification of the subgroup probability model has a little effect on the parameters of interest.

7. EXAMPLE

The proposed approach is applied to MSLT-I, an ongoing clinical trial, to evaluate the effectiveness of sentinel node biopsy (SNB) in guiding the treatment of invasive melanoma (Morton et al., 2005). At the time the trial began in 1994, the standard method of care consisted of wide excision of the primary melanoma, followed by observation, with lymphadenectomy (surgical excision of a lymph node or nodes) only undertaken on clinical evidence of recurrence. However, patients with nodal metastases have a poor prognosis and may stand to benefit from immediate lymphadenectomy. The MSLT-I study randomized patients who received an experimental course of therapy (treatment arm) or the standard care (control arm). All patients in the treatment arm received SNB (Morton et al., 1999), and those with biopsies positive for metastasis underwent an immediate (elective) lymphadenectomy. Node-negative patients received the same observational course of treatment as the controls.

The ITT analysis of MSLT-I determines whether the strategy of SNB-plus-immediate-lymphadenectomy improves survival in the diseased aggregate population. The analysis does not estimate the biological efficacy of the surgery on the subgroup of patients with nodal metastases. The third interim analysis of the trial (Morton et al., 2006) attempted to address this estimate by comparing node-positive patients undergoing treatment with control patients who subsequently developed recurrence. However, this procedure does not produce an unbiased estimate of biological efficacy because of the lack of one-to-one correspondence between positive SNB results at the time of randomization and the eventual development of a clinically evident recurrence.

We analyzed the data by using the proposed semiparametric transformation model to estimate the biological efficacy of the surgery on the subgroup of patients with nodal metastases. Following the suggestion of Altstein, Li, & Elashoff (2011), we considered distant disease free survival (DDFS) as the end-point. For these data, several explanatory variables, including the Breslow score (X_1), an indicator of the presence of ulceration (X_2), as well as an indicator that the melanoma is located on the trunk of the body (X_3), was recorded. The Breslow score measures the thickness of the primary melanoma. We modeled the data using the proposed model with $\mathbf{X} = \mathbf{Z} = (X_1, X_2, X_3)'$, and the hazard function of ε_i was assumed to take the form of $\lambda(t) = \exp(t)/(1 + r \exp(t))$. We chose $r = 0$ from $r = 0, 0.5, 1, 1.5, 2, 2.5, 3$ by comparing the distribution of $\hat{\varepsilon}_i = \hat{H}(T_i) - \hat{c}_{\tilde{U}_i} - \hat{\alpha}_{\tilde{U}_i} E_i - \hat{\beta}' \mathbf{X}_i$ and the real distribution of ε_i based on the Q-Q plot, where $\tilde{U}_i$ is U_i if observed, or otherwise, obtained by the most probable class based on the conditional probability given the observed data. The resulting estimates of the regression coefficients and the transformation function with $a_0 = 32$ and $b_0 = 3.5$ are displayed in Table 6 and Figure 2a. The standard errors were calculated according to the method discussed in Section 5. Table 6 shows that the treatment has no significant effect on DDFS. However, a long DDFS is strongly associated with a low Breslow score, the absence of ulceration and the absence of the melanoma on the trunk of

the body. In addition, Table 6 also suggests that the node status depends on the Breslow score. To compare, we also display the estimation results from the accelerated failure time mixture model used by Altstein & Li (2013), and a similar conclusion is obtained.

Finally, we verified the validity of the assumed semiparametric transformation model (1) by examining how well the distribution of the estimated error $\hat{\varepsilon}_i$ fitted the given distribution.

Given that the treatment is insignificant and the estimators of c_1 , c_2 , and c_3 are close, we computed the estimated error by $\hat{\varepsilon}_i = \hat{H}(T_i) - c - \hat{\beta}'\mathbf{X}_i$, where $c = (c_1 + c_2 + c_3)/3$. Then, using the Kaplan–Meier estimator, the distribution of the estimated error was obtained based on the data $\{\hat{\varepsilon}_i, \delta_i\}$ for $i = 1, \dots, n$. Figure 2b displays the quantile–quantile plot of the distribution of the estimated error $\hat{\varepsilon}_i$ versus the given distribution. A perfect fit corresponds to the solid diagonal line. Points above or below this diagonal line indicate over- or underprediction by the model. Figure 2b suggests that the goodness-of-fit of the proposed model (1) is reasonable.

8. DISCUSSION

We use a semiparametric linear transformation model to analyze survival data with selective compliance. This model is more robust and flexible than other existing models. Although we have focused on independent data, our method can be extended to multivariate failure times by incorporating random effects or frailties. In our model, the regression coefficient β and the variance of the error are assumed to be the same for the groups $U_i = 1, 2$, and 3. Our method can be extended to allow different β and variance for different groups $U_i = 1, 2, 3$.

ACKNOWLEDGEMENTS

Lin's research was supported by the National Natural Science Funds for Distinguished Young Scholar (No. 11125104), the National Natural Science Foundation of China (No. 11071197) and Program for New Century Excellent Talents in University.

APPENDIX

Regularity conditions for ensuring the central limit theorem for counting process martingales such as those assumed in Fleming & Harrington (1991) are also assumed here without specific statement. In particular, we assume that τ is finite, $Pr(T_i > \tau) > 0$ and $Pr(C_i = \tau) > 0$. This is to avoid a lengthy technical discussion about the tail behaviour. The consistency and asymptotic normality stated in theorems 1 and 2 are proved using similar arguments to those of Chen, Jin, & Ying (2002), so we only highlight the steps that are different.

Proof of Theorems 1 and 2.

Step 1. Using similar arguments to Step A1 of Chen, Jin, & Ying (2002), it can be shown that $d\{\hat{H}(\cdot, \Theta_0), H_0(\cdot)\} \rightarrow 0$ almost surely, where $\hat{H}(\cdot, \Theta)$ is the function implicitly defined as the unique solution of (7) for fixed Θ and $d(\cdot, \cdot)$ is a distance defined as follows. For any two nondecreasing functions H_1 and H_2 on $[0, \tau]$ such that $H_1(0) = H_2(0) = -\infty$, define $d(H_1, H_2) = \sup\{|\exp\{H_1(t)\} - \exp\{H_2(t)\}| : t \in [0, \tau]\}$.

Step 2. Constructing the expression of $\hat{H}(t; \Theta_0)$. Let $a > 0$ and b be fixed finite numbers and define $B_1(t) = \int_a^t E[Y_i(s)\dot{\lambda}_i(H_0(s))]dH_0(s)$, $B_2(t) = E[Y_i(t)\lambda_i(H_0(t))]$, $\lambda^*(H_0(t)) = \exp(\int_0^t dB_1(x)/B_2(x))$, $\Lambda^*(x) = \int_b^x \lambda^*(s)ds$, for $t > 0$ and $x \in (-\infty, \infty)$. We choose finite $a > 0$ and b as the lower limits of the integration to ensure that the integrals are finite. It is easy to see that $d\lambda^*\{H_0(t)\} = [\lambda^*\{H_0(t)\}/B_2(t)]dB_1(t)$ and write

$$\begin{aligned} -\frac{1}{n} \sum_{i=1}^n M_i(t) &= -\frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(s) d \left\{ \frac{\lambda_i(H_0(s))}{\lambda^*(H_0(s))} (\Lambda^*(\hat{H}(s; \Theta_0)) - \Lambda^*(H_0(s))) \right\} + o_p(n^{-1/2}) \\ &= -\int_0^t (\Lambda^*(\hat{H}(s; \Theta_0)) - \Lambda^*(H_0(s))) \\ &\quad \times \frac{1}{n} \sum_{i=1}^n Y_i(s) \left\{ \frac{\dot{\lambda}_i(H_0(s))dH_0(s)}{\lambda^*(H_0(s))} - \frac{\lambda_i(H_0(s))[\lambda^*\{H_0(s)\}/B_2(s)]dB_1(s)}{(\lambda^*(H_0(s)))^2} \right\} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(s) \frac{\lambda_i(H_0(s))}{\lambda^*(H_0(s))} d(\Lambda^*(\hat{H}(s; \Theta_0)) - \Lambda^*(H_0(s))) + o_p(n^{-1/2}) \\ &= -\int_0^t \frac{B_2(s)}{\lambda^*(H_0(s))} d(\Lambda^*(\hat{H}(s; \Theta_0)) - \Lambda^*(H_0(s))) + o_p(n^{-1/2}). \end{aligned}$$

Therefore, for $t \in [0, \tau]$,

$$\Lambda^*(\hat{H}(t; \Theta_0)) - \Lambda^*(H_0(t)) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\lambda^*(H_0(s))}{B_2(s)} dM_i(s) + o_p(n^{-1/2}). \quad (\text{A.1})$$

Step 3. Denote $U(\Theta; H) = -\log L(\Theta; H)/n$. In the step, we compute $V(\Theta) = U(\Theta; \hat{H}(\cdot; \Theta))/n$ at $\Theta = \Theta_0$. By differentiating both sides of (7) with respect to Θ , we obtain the identity

$$\sum_{i=1}^n \int_0^t Y_i(s) d \left\{ \lambda_i(\hat{H}(s; \Theta_0)) \frac{\partial \hat{H}(s; \Theta_0)}{\partial \Theta} + \tilde{\lambda}_i(\hat{H}(s; \Theta_0); \Theta_0) \right\} = 0, \quad (\text{A.2})$$

where $\tilde{\lambda}_i(H(s); \Theta) = \frac{\partial \lambda_i(H(s); \Theta)}{\partial \Theta}$. Similarly to Step 2, we have that

$$\frac{\partial \hat{H}(t; \Theta_0)}{\partial \Theta} = -\frac{1}{\lambda^* \{H_0(t)\}} \int_0^t \frac{\lambda^* \{H_0(s)\}}{B_2(s)} E \left[Y_i(s) \frac{\partial \lambda_i(H_0(s); \Theta)}{\partial \Theta} \right] dH_0(s) + o_p(1).$$

(A.3)

It follows from the law of large numbers that

$$\begin{aligned} v(\Theta_0) &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\partial^2 \log L_i(\Theta_0; H_0)}{\partial \Theta \partial \Theta'} - \frac{\partial^2 \log L_i(\Theta_0; H_0)}{\partial \Theta \partial H(W_i)} \frac{1}{\lambda^* \{H_0(W_i)\}} \right. \\ &\quad \times \left. \int_0^\tau Y_i(s) \frac{\lambda^* \{H_0(s)\}}{B_2(s)} E \left[Y_j(s) \frac{\partial \lambda_j(H_0(s); \Theta_0)}{\partial \Theta} \right] dH_0(s) \right\} + o_p(1) \\ &= \Sigma + o_p(1), \end{aligned}$$

where $L_i(\Theta; H)$ is the contribution of subject i to the likelihood function $L(\Theta; H)$

Step 4. In this step, we show the asymptotic normality of $U(\Theta_0; \hat{H}(\cdot; \Theta_0))$. Using the results of Steps 1 and 2 and some empirical process approximation techniques, we can write

$$\begin{aligned} \frac{1}{n} U(\Theta_0; \hat{H}(\cdot; \Theta_0)) &= \frac{1}{n} \sum_{i=1}^n \frac{\partial \log L_i(\Theta_0; H_0)}{\partial \Theta} + \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log L_i(\Theta_0; H_0)}{\partial \Theta \partial H(W_i)} (\hat{H}(W_i; \Theta_0) - H_0(W_i)) \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\partial \log L_i(\Theta_0; H_0)}{\partial \Theta} \right. \\ &\quad \left. + \int_0^\tau E \left[\frac{Y_j(s)}{\lambda^* \{H_0(W_j)\}} \frac{\partial^2 \log L_j(\Theta_0; H_0)}{\partial \Theta \partial H(W_j)} \right] \frac{\lambda^* \{H_0(s)\}}{B_2(s)} dM_i(s) \right\} + o_p(n^{-1/2}). \end{aligned}$$

It then follows that $n^{-1/2} U(\Theta_0; \hat{H}(\cdot; \Theta_0)) \rightarrow N(0, \Sigma)$.

The rest of the proof essentially proceeds along the lines of Chen, Jin, & Ying (2002) and is omitted here. ■

Proof of Theorem 3.

By Taylor series expansions, (A.1) and (A.3), we get,

$$\begin{aligned}
& \Lambda^*(\hat{H}(t)) - \Lambda^*(H_0(t)) \\
&= \Lambda^*(\hat{H}(t; \hat{\Theta})) - \Lambda^*(\hat{H}(t; \Theta_0)) + \Lambda^*(\hat{H}(t; \Theta_0)) - \Lambda^*(H_0(t)) \\
&= \lambda^*(H_0(t)) \frac{\partial \hat{H}(t; \Theta_0)}{\partial \Theta'} (\hat{\Theta} - \Theta_0) + \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\lambda^*(H_0(s))}{B_2(s)} dM_i(s) + o_p(n^{-1/2}) \\
&= - \int_0^t \frac{\lambda^*\{H_0(s)\}}{B_2(s)} E \left[Y_j(s) \frac{\partial \lambda_j(H_0(s); \Theta_0)}{\partial \Theta'} \right] dH_0(s) (\hat{\Theta} - \Theta_0) \\
&\quad + \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\lambda^*(H_0(s))}{B_2(s)} dM_i(s) + o_p(n^{-1/2}).
\end{aligned}
\tag{A.4}$$

Substituting $\hat{\Theta} - \Theta_0 = -\Sigma^{-1} n^{-1} U(\Theta_0; \hat{H}(\cdot; \Theta_0)) + o_p(n^{-1/2})$ into the equation above, we complete the proof of Theorem 3. ■

BIBLIOGRAPHY

- Angrist JD, Imbens GW, & Rubin DB (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444–445.
- Altman DG (1991). *Practical Statistics for Medical Research*, CRC Press, London.
- Altstein LL & Li G (2013). Latent subgroup analysis of a randomized clinical trial through a semiparametric accelerated failure time mixture model. *Biometrics*, 69, C52–C61.
- Altstein LL, Li G, & Elashoff RM (2011). A method to estimate treatment efficacy among latent subgroups of a randomized clinical trial. *Statistics in Medicine*, 30, 709–717. [PubMed: 21394747]
- Bennett S (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2, 273–277. [PubMed: 6648142]
- Cai T, Cheng S, & Wei LJ (2002). Semi-parametric random effects models for clustered failure time data. *Journal of the American Statistical Association*, 97, 514–522.
- Cai T, Wei LJ, & Wilcox M (2000). Semi-parametric regression analysis for clustered failure time data. *Biometrika*, 87, 867–878.
- Chen K, Jin Z, & Ying Z (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, 89, 659–668.
- Cheng SC, Wei LJ, & Ying Z (1995). Analysis of transformation models with censored data. *Biometrika*, 82, 835–845.
- Clayton DG & Cuzick J (1985). Multivariate generalizations of the proportional hazards model. *Journal of Royal Statistical Society Series B*, 34, 187–220.
- Cox DR (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistical Society Series B*, 34, 187–220.
- Cuzick J, Sasieni P, Myles J, & Tyrer J (2007). Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *Journal of Royal Statistical Society Series B*, 69, 565–588.
- Dabrowska DM & Doksum KA (1988). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics*, 15, 1–23.

- Elashoff RM, Li G, & Zhou Y (2012). Nonparametric inference for assessing treatment efficacy in randomized clinical trials with a time-to-event outcome and all-or-none compliance. *Biometrika*, 99, 393–404. [PubMed: 23843664]
- Fleming TR & Harrington DP (1991). *Counting processes and survival analysis*. John Wiley & Sons, Inc., New York
- Frangakis CE & Rubin DB (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86, 365–379.
- Hernan MA, Brumback B, & Robins JM (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of American Statistical Association*, 96, 440–448.
- Horowitz JL (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica*, 64, 103–137.
- Jin Z, Ying Z, & Wei LJ (2001). A simple resampling method by perturbing the minimand. *Biometrika*, 88, 381–390.
- Lin HZ, Zhou XH, & Zhou L (2013). Semiparametric regression analysis of longitudinal skewed data. *Scandinavian Journal of Statistics*, in revision.
- Lin HZ & Zhou XH (2009). A semi-parametric two-part mixed-effects heteroscedastic transformation model for correlated right-skewed semi-continuous data. *Biostatistics*, 10, 640–658. [PubMed: 19546166]
- Loeys T & Goetghebeur E (2002). Baseline information in structural failure time estimators for the effect of observed treatment compliance. *Statistics in Medicine*, 21, 1173–1180. [PubMed: 12111873]
- Loeys T & Goetghebeur E (2003). A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics*, 59, 100–105. [PubMed: 12762446]
- Loeys T, Goetghebeur E, & Vandebosch A (2005). Causal proportional hazards models and time-constant exposure in randomized clinical trials. *Lifetime Data Analysis*, 11, 435–449. [PubMed: 16328570]
- Morton DL, Thompson JF, Essner R, Elashoff R, Stern SL, Nieweg OE, Roses DF, Karakousis CP, Mozzillo N, Reintgen D, Wang HJ, Glass EC, & Cochran AJ. (1999). Validation of the accuracy of intraoperative lymphatic mapping and sentinel lymphadenectomy for early-stage melanoma. *Annals of Surgery*, 230(4), 453–463. [PubMed: 10522715]
- Morton DL, Cochran AJ, Thompson JF, Elashoff R, Essner R, Glass EC, Mozzillo N, Nieweg OE, Roses DF, Hoekstra HJ, Karakousis CP, Reintgen DS, Coventry BJ, Wang HJ, & the Multicenter Selective Lymphadenectomy Trial Group. (2005). Sentinel node biopsy for early-stage melanoma: Accuracy and morbidity in MSLT-I, an international multicenter trial. *Annals of Surgery*, 242, 302–313. [PubMed: 16135917]
- Morton D, Thompson J, Cochran A, Mozzillo N, Elashoff R, Essner R, Nieweg O, Roses D, Hoekstra H, Karakousis C, Reintgen D, Coventry B, Glass E, Wang H, & the Multicenter Selective Lymphadenectomy Trial Group. (2006). Sentinel-node biopsy or nodal observation in melanoma. *New England Journal of Medicine*, 355, 1307–1317. [PubMed: 17005948]
- Murphy SA, Rossini AJ, & van der Vaart AW (1997). MLE in the proportional odds model. *Journal of the American Statistical Association*, 92, 968–976.
- Olsen MK & Schafer JL (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association*, 96, 730–745.
- Pearl J (2002). Reasoning with cause and effect. *AI Magazine*, 23, 95–111.
- Pettitt AN (1982). Inference for the linear model using a likelihood based on ranks. *Journal of Royal Statistical Society Series B*, 44, 234–243.
- Robins JM & Finkelstein DM (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring (IPCW) logrank-tests. *Biometrics*, 56, 779–788. [PubMed: 10985216]

- Robins JM and Tsiatis AA (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics—Theory and Methods*, 20, 2609–2631.
- Scharfstein DO, Tsiatis AA, & Gilbert PB (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis*, 4, 355–391. [PubMed: 9880995]
- Ye JM & Duan NH (1997). Nonparametric $n^{-1/2}$ -consistent estimation for the general transformation models. *The Annals of Statistics*, 25, 2682–2717.
- Zeng D & Lin DY (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93, 627–640.
- Zhou XH, Lin HZ, & Johnson E (2009). Nonparametric heteroscedastic transformation regression models for skewed data with an application to health care costs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, 1029–1047.

Figure 1: Estimated transformation functions for (a) Case 1; (b) Case 2; (c) Case 3; and (d) Case 4.

Figure 2:

(a) The estimated transformation function (dotted) and its 95% confidence limits (dotted-lined) for the MSLT-I data. (b) The quantile–quantile plot of the distribution of the estimated error $\hat{\varepsilon}_i$ versus the given distribution.

Table 1:

The four simulation settings generated from varying $H(\cdot)$, r , κ_1 , and κ_2 .

	$H(t)$	r	κ_1	κ_2
Case 1	$\log(t)$	0	$(0.5, -0.5)'$	$(0.5, -0.5)'$
Case 2	$\log\{\log(t)\}$	1	$(0.5, -0.5)'$	$(0.5, -0.5)'$
Case 3	$\log(t)$	0	$(0, 0)'$	$(0, 0)'$
Case 4	$\log\{\log(t)\}$	1	$(0, 0)'$	$(0, 0)'$

The proportional hazards and proportional odds models correspond to $r=0$ and $r=1$, respectively.

The class membership is independent of covariates when $\kappa_1 = \kappa_2 = (0, 0)'$.

Table 2:

The bias, SD, and estimated SE of the estimators based on the 500 simulations for Case 1.

Parameter	Proposed					MSP	
	Bias	SD	SE _{ave}	SE _{sd}	CP	Bias	SD
α_1	0.0259	0.3885	0.3925	0.0543	0.940	-0.0132	0.4006
c_1	0.0066	0.2573	0.2560	0.0386	0.948	0.0168	0.2447
c_2	-0.0009	0.1396	0.1237	0.0152	0.912	-0.0055	0.1411
c_3	0.0072	0.1545	0.1657	0.0278	0.956	-0.0016	0.1637
β_1	0.0014	0.0836	0.0915	0.0234	0.946	-0.0021	0.0894
β_2	0.0059	0.0834	0.1006	0.0106	0.978	0.0111	0.0969
θ_1	-0.0020	0.2384	0.2151	0.02540	0.912	0.0018	0.2333
θ_2	0.0181	0.1461	0.1630	0.0146	0.962	0.0090	0.1558
κ_{11}	0.0058	0.2229	0.2236	0.0289	0.942	—	—
κ_{12}	-0.0372	0.2210	0.2348	0.0418	0.948	—	—
κ_{21}	0.0090	0.1673	0.1636	0.0179	0.932	—	—
κ_{22}	-0.0118	0.1562	0.1682	0.0217	0.958	—	—

Table 3:

The bias, SD, and estimated SE of the estimators for Case 2.

Parameter	Bias	SD	SE _{ave}	SE _{sd}	CP
α_1	-0.0073	0.6820	0.6842	0.0924	0.938
c_1	0.0461	0.3785	0.3971	0.0615	0.952
c_2	0.0194	0.2145	0.2095	0.0295	0.926
c_3	0.0128	0.2259	0.2229	0.0393	0.930
β_1	-0.0067	0.1185	0.1176	0.0128	0.944
β_2	0.0026	0.1244	0.1352	0.0198	0.956
θ_1	-0.0074	0.2674	0.2546	0.0448	0.926
θ_2	0.0079	0.1514	0.1392	0.0103	0.940
κ_{11}	0.0221	0.2341	0.2155	0.0249	0.914
κ_{12}	-0.0144	0.2413	0.2412	0.0255	0.944
κ_{21}	0.0091	0.1682	0.1648	0.0190	0.938
κ_{22}	-0.0040	0.1655	0.1705	0.0140	0.962

Table 4:

The bias, SD, and estimated SE of the estimators for Case 3.

	Proposed					Cox	
	Bias	SD	SE _{ave}	SE _{sd}	CP	Bias	SD
α_1	0.0194	0.4175	0.4325	0.0555	0.944	0.0602	1.3246
c_1	0.0029	0.2719	0.2570	0.0405	0.930	—	—
c_2	0.0064	0.1453	0.1441	0.0153	0.924	0.0179	1.2739
c_3	-0.0026	0.1557	0.1416	0.0136	0.932	0.0515	1.2899
β_1	0.0045	0.0841	0.0839	0.0075	0.950	0.0753	0.0775
β_2	0.0106	0.0885	0.0770	0.0074	0.910	0.0713	0.0759
θ_1	0.0020	0.2174	0.2210	0.0218	0.944	—	—
θ_2	0.0014	0.1359	0.1585	0.0394	0.954	—	—

Table 5:

The bias and SD of the estimators for Case 4.

	Proposed					Cox	
	Bias	SD	SE _{ave}	SE _{sd}	CP	Bias	SD
α_1	-0.0427	0.6866	0.6903	0.0863	0.932	0.2774	1.4585
c_1	0.0774	0.3683	0.3642	0.0416	0.942	—	—
c_2	0.0368	0.2264	0.2317	0.0214	0.948	0.1972	1.4426
c_3	0.0050	0.2152	0.2042	0.0181	0.946	0.1643	1.4694
β_1	-0.0180	0.1071	0.1204	0.0116	0.978	0.3603	0.0801
β_2	-0.0173	0.1215	0.1342	0.0255	0.952	0.3498	0.0849
θ_1	0.0010	0.2164	0.2295	0.0305	0.946	—	—
θ_2	-0.0121	0.1374	0.1526	0.0189	0.964	—	—

Table 6:

The estimation results for the MSLT-I data using the proposed method.

	Proposed			AFT			
	Estimate	SD	<i>P</i> -value	Node+		Node−	
				Estimate	<i>P</i> -value	Estimate	<i>P</i> -value
α_1	0.0177	0.1553	0.9093	0.358	0.28	—	—
β_1	−0.5476	0.1372	0.0000	−0.619	0.017	−0.654	0.0003
β_2	−0.4171	0.1398	0.0028	0.411	0.29	−0.530	0.012
β_3	−0.6058	0.0968	0.0000	−0.355	0.32	−0.615	0.0018
	Proposed			Proposed			
	Estimate	SD	<i>P</i> -value		Estimate	SD	<i>P</i> -value
c_1	7.4373	0.3149	0.0000	κ_{11}	−1.2211	0.3631	0.0008
c_2	7.0197	0.4885	0.0000	κ_{12}	−0.5256	0.3713	0.1569
c_3	7.4862	0.4629	0.0000	κ_{13}	0.1129	0.3232	0.7268
θ_1	3.5549	0.6813	0.0000	κ_{21}	−1.4890	0.6050	0.0138
θ_2	0.6753	0.9101	0.4581	κ_{22}	−0.2430	0.5814	0.6760
				κ_{23}	0.0582	0.4551	0.8982