Published in final edited form as:

Biometrics. 2015 June ; 71(2): 302–312. doi:10.1111/biom.12286.

Causal Estimation using Semiparametric Transformation Models under Prevalent Sampling

Yu-Jen Cheng^{1,*} and Mei-Cheng Wang²

¹Institute of Statistics, National Tsing Hua University, Hsin-Chu 300, Taiwan

²Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, U.S.A.

Summary

This paper develops methods and inference for causal estimation in semiparametric transformation models for prevalent survival data. Through estimation of the transformation models and covariate distribution, we propose analytical procedures to estimate the causal survival function. As the data are observational, the unobserved potential outcome (survival time) may be associated with the treatment assignment, and therefore there may exist a systematic imbalance between the data observed from each treatment arm. Further, due to prevalent sampling, subjects are observed only if they have not experienced the failure event when data collection began, causing the prevalent sampling bias. We propose a unified approach which simultaneously corrects the bias from the prevalent sampling and balances the systematic differences from the observational data. We illustrate in the simulation study that standard analysis without proper adjustment would result in biased causal inference. Large sample properties of the proposed estimation procedures are established by techniques of empirical processes and examined by simulation studies. The proposed methods are applied to the Surveillance, Epidemiology, and End Results (SEER) and Medicare linked data for women diagnosed with breast cancer.

Keywords

Causal estimation; Dependent truncation; Prevalent sampling; Survival analysis

1. Introduction

In a randomized study with a time-to-event outcome, the Kaplan-Meier estimator (Kaplan and Meier, 1958) is typically used to estimate the survival functions for treated and untreated groups in standard setting, where the difference between the two survival functions is referred to as the causal treatment effect. As a contrast, in observational studies, systematic differences could exist between treated and untreated groups at baseline, and therefore direct comparisons of observed outcomes from the two groups may not be

*ycheng@stat.nthu.edu.tw.

7. Supplementary Materials

Web Appendices A, B, C, D, and E referenced in Sections 2, 4, 5, appendix, and the R script to obtain the proposed estimators are available at the Biometrics website on Wiley Online Library.

appropriate (Holland, 1986; Rubin, 2004). To develop causal inferential methods, we consider the counterfactual or potential outcome models (Neyman, 1923; Rubin, 1974, 1978), in which a causal survival function is defined as the survival function of the potential outcome and the causal treatment effect is referred to as the difference between two causal survival functions. In such models, the difference between two potential outcome distributions is of interest, even though in reality only a single outcome is observable.

In the analysis of survival data, standard data are typically collected or observed with the *incident sampling* where the incidence of disease, or equivalently, time origin for measuring survival time, is randomly sampled from a pre-determined calendar time interval. For standard survival data, with the assumption of no unmeasured confounders, various approaches have been proposed to estimate the causal treatment effect. For example, Chen and Tsiatis (2001) specified the proportional hazards model as a type of potential outcome model and obtained the causal survival function by averaging out the covariate-specific survival functions over all subjects. Their approach can be considered as a model-based standardization procedure for estimating the population-average quantities (Lane and Nelder, 1982; Greenland, 2004). In contrast to an incident sampling scheme, which is known to be inefficient if failure times are long, a *prevalent sample* includes only those subjects who have experienced the incidence of disease but not the failure event at the time of recruitment. In many situations the prevalent sampling is more efficient and economical when compared with the incident sampling, and therefore could be preferable in biomedical studies (Lynden-Bell, 1971; Woodroffe, 1985; Wang et al., 1986; Brookmeyer and Gail, 1987).

It is well known that prevalent sampling is associated with left truncation, where population subjects with longer failure times are more likely observed. An example of such is the Surveillance, Epidemiology, and End Results (SEER)-Medicare data, the linkage of cancer registry data from the National Cancer Institute and Medicare claims from the Centers for Medicare and Medicaid Services (Warren et al., 2002). The survival times of subjects in the linkage data are left truncated and right censored, where left truncation arises because Medicare claims are available only for living patients in 1986 and, like most of the surveillance follow-up data, right censoring is caused by the administrative end-of-study.

In this paper, we consider a general class of semiparametric transformation models and develop a model-based standardization procedure to analyze prevalent survival data under the assumption of no unmeasured confounders. Section 2 introduces notation and the model framework. Section 3 proposes a pseudo-partial likelihood approach for estimating causal survival functions under dependent truncation, which relaxes the conventional independence assumption between the failure and truncation times. A specific challenge in the proposed methodology is to handle the biased distribution of observed covariates, where the bias arises because of the correlation between covariates and biased failure time. In Section 3, with nonrestrictive assumptions, a structural approach is proposed to correct the bias in observed covariates. In Section 4, a simulation study is presented to examine the appropriateness and efficiencies of the proposed methods. In Section 5, the proposed methods are applied to SEER-Medicare breast cancer data for causal analysis. A brief discussion is presented in Section 6 to conclude the paper.

2. Notation, Modeling Framework and Inference Procedures

For each subject, denote by T^* , W^* , A^* and $\mathbf{X}^*$ the failure time, truncation time, treatment status and $p \times 1$ vector of covariates $\mathbf{X}^* = (X_1^*, \dots, X_p^*)$, respectively. Let T_0^* or T_1^* be the potential outcome that would be observed if the unit received treatment “0” or “1”. The failure time T^* would be either T_0^* or T_1^* , depending on the treatment received. For example, in SEER-Medicare data, the potential outcomes T_0^* and T_1^* are the times from diagnosis with breast cancer to death for the same cancer patient with ($A^* = 1$) or without ($A^* = 0$) radiation therapy. A study subject’s failure time, T^* , is the time from diagnosis of breast cancer to death which equals either T_0^* or T_1^* , depending on treatment assignment A^* .

For clarity of presentation, variables with asterisk represent measurements from the target population, and variables without asterisk represent measurements subject to truncation sampling. With slight abuse of notation for continuous variables, the density function of $(T, W, A, \mathbf{X})$ is identified as $\text{pr}(T, W, A, \mathbf{X}) = \text{pr}(T^*, W^*, A^*, \mathbf{X}^* | T^* = W^*)$. For subjects in the diseased prevalent population, denote by C the residual censoring time (time from recruitment to censoring). Let $Y = \min(T, C + W)$ be the failure or censoring time and $I(T < C + W)$ the censoring indicator. The typical observable data from an observational study subject to prevalent sampling can be summarized as n independent random vectors

$\tilde{O}_i = (Y_i, \Delta_i, W_i, A_i, \mathbf{X}_i)$, $i=1, \dots, n$, where the observed Y_i is obtained from a two-step procedure: $T_i^* = (1 - A_i^*) T_{i0}^* + A_i^* T_{i1}^*$ and $Y_i = \min(T_i, C_i + W_i)$ if $T_i^* \geq W_i^*$. Thus, the observed Y_i contains two non-ignorable sources of missingness: 1) The two potential variables (T_{i0}^*, T_{i1}^*) are never simultaneously observed, since observation of one necessarily excludes observation of the other, so the concept of potential quantities implies missing information; and 2) Y_i is obtained only when $T_i^* \geq W_i^*$.

Moreover, in this paper, we do not assume the conventional independent truncation assumption that T^* is independent of W^* , conditioning on $(A^*, \mathbf{X}^*)$. Instead, by treating a pre-specified transformation of W^* as one of the covariates, the truncation and failure times are allowed to be correlated in the transformation model. For simplicity, we consider only the case that the identity link as the pre-specified transformation. Let $\mathbf{Z}^* = (W^*, \mathbf{X}^*)$ and $\mathbf{Z} = (W, \mathbf{X})$. Consider the following assumptions:

(A1) *Stable Unit Treatment Value Assumption (SUTVA)*: Response of the i -th subject is not be affected by responses of other subjects (Noninterference). Treatment A^* could be assigned by different ways, but they all lead to the same outcome (Consistency).

(A2) *Strongly ignorable treatment assignment*: The treatment assignment A^* is independent of potential outcomes, (T_0^*, T_1^*) , given covariates $\mathbf{Z}^*$, i.e., $\text{pr}(A^* | \mathbf{Z}^*, T_0^*, T_1^*) = \text{pr}(A^* | \mathbf{Z}^*)$, where $0 < \text{pr}(A^* = a | \mathbf{Z}^*) < 1$.

(A3) *Conditional independent censoring*: Conditioning on $T^* = W^*$ and $(A, \mathbf{Z})$, the residual censoring time from recruitment to censoring is independent of the residual failure time from recruitment to failure event for subjects in the disease prevalent population.

For identifiability, we assume: $\tau_{W^*}^1 < \tau_{T^*}^1$ and $\tau_{W^*}^2 < \tau_{T^*}^2$, where the intervals $[\tau_{T^*}^1, \tau_{T^*}^2]$ and $[\tau_{W^*}^1, \tau_{W^*}^2]$ are supports of failure time and truncation time, respectively. A general class of semiparametric transformation models is considered in this paper, that is,

$$H(T^*) = -\theta'(A^*, \mathbf{Z}^*) + \epsilon \quad (1)$$

where $H(\cdot)$ is an unspecified monotone function satisfying $H(0) = -\infty$, ϵ is a random variable with a known distribution, $\theta = (\gamma, \beta)$ in which γ denotes the treatment effect with respect to A^* , and β is a $(p+1) \times 1$ vector of regression coefficients for covariates $\mathbf{Z}^*$. Special cases of the transformation models include the proportional hazards model and the proportional odds model, where ϵ corresponds to the extreme-value distribution and the standard logistic distribution, respectively (Cheng et al., 1995; Chen et al., 2002; Zeng and Lin, 2006). Define $R(u) = \exp\{H(u)\}$, $\Phi(s) = \Lambda_\epsilon(\log s)$, where $\Lambda(\cdot)$ and $\Lambda_\epsilon(\cdot)$ are the cumulative hazard functions of T^* and ϵ , respectively. Formula (1) can be equivalently expressed as $\Lambda(u|A^*=a, \mathbf{Z}^*=z) = \Phi\left\{R(u) \exp(\gamma a + \beta' z)\right\}$.

For simplicity of notation, let $G_{A^*, \mathbf{Z}^*}(a, z) = \text{pr}(A^* = a, \mathbf{Z}^* = z)$ and $G_{\mathbf{Z}^*}(z) = \text{pr}(\mathbf{Z}^* = z)$, $a = 0, 1$. If assumption (A3) holds, the full likelihood, L_F , for observed data $\{Y_i = y_i, \delta_i = \delta_i, A_i = a_i, \mathbf{Z}_i = \mathbf{z}_i; i = 1, \dots, n\}$ can be decomposed as the product of the marginal likelihood (L_M) of $(a_1, \mathbf{Z}_1, \dots, a_n, \mathbf{z}_n)$ and the conditional likelihood (L_C) of $(y_1, \delta_1, \dots, y_n, \delta_n)$ given $(a_1, \mathbf{z}_1, \dots, a_n, \mathbf{z}_n)$, where

$$L_M = \prod_{i=1}^n \frac{S(w_i | a_i, \mathbf{z}_i) dG_{A^*, \mathbf{Z}^*}(a_i, \mathbf{z}_i)}{\int_a \int_{\mathbf{z}} S(w | a, \mathbf{z}) dG_{A^*, \mathbf{Z}^*}(a, \mathbf{z})} \quad (2)$$

and

$$L_C \propto \prod_{i=1}^n \left\{ \lambda(y_i | a_i, \mathbf{z}_i)^{\delta_i} \frac{S(y_i | a_i, \mathbf{z}_i)}{S(w_i | a_i, \mathbf{z}_i)} \right\}, \quad (3)$$

in which $\lambda(u | A^*, \mathbf{Z}^*)$ and $S(u | A^*, \mathbf{Z}^*)$ are the hazard and survival function of T^* at u conditional on treatment assignment A^* and covariates $\mathbf{Z}^*$, respectively. Let superscript dot denote derivative and define $\phi(s) = d\Phi(s)/ds = \dot{\Phi}(s)$. Then

$$\lambda(u | A^*=a, \mathbf{Z}^*=z) = \phi\left\{R(u) \exp(\gamma a + \beta' z)\right\} \exp(\gamma a + \beta' z) \dot{R}(u) \text{ and}$$

$$S(u | A^*=a, \mathbf{Z}^*=z) = \exp\left[-\Phi\left\{R(u) \exp(\gamma a + \beta' z)\right\}\right].$$

Note that the conditional independence assumption between the truncation time W^* and failure time T^* , conditioning on $(A^*, \mathbf{Z}^*)$, is relaxed in the likelihood decomposition through the inclusion of the observed truncation time W as one component of the covariates when formulating the conditional likelihood L_C . Essentially, the truncation time not only plays a role in the sampling criterion for recruiting a member into the study, but also serves as a component of the covariates in L_C , and therefore possibly a component of the covariates in the transformation model.

The conditional likelihood L_C can be further decomposed as a product of a partial likelihood and the remaining residual likelihood, where the residual likelihood is provided in the Web Appendix A. The partial likelihood is formulated as

$$L_P(\theta; R(\cdot)) = \prod_{(i)} \frac{\phi\{R(y_{(i)}) \exp(\gamma a_{(i)} + \beta' z_{(i)})\} \exp(\gamma a_{(i)} + \beta' z_{(i)})}{\sum_{j=1}^n Y_j(y_{(i)}) \phi\{R(y_{(i)}) \exp(\gamma a_j + \beta' z_j)\} \exp(\gamma a_j + \beta' z_j)}, \quad (4)$$

where $y_{(1)}, \dots, y_{(k)}$ are distinct ordered uncensored failure times and $Y_j(u) = I(W_j < u - Y_j)$ is an at-risk indicator. Using martingale decomposition, one can construct the following estimating equation:

$$\sum_{i=1}^n \left\{ dN_i(u) - Y_i(u) d\Phi\{R(u) \exp(\gamma a_i + \beta' z_i)\} \right\} = 0, \quad (5)$$

where $N_i(u) = I(Y_i \leq u, i = 1)$. While the estimating equation in (5) is used to estimate $R(u)$ with fixed θ , we maximize the partial likelihood in (4) to derive the maximum likelihood estimate of θ with fixed $R(u)$. Moreover, when θ is fixed, the estimating equation in (5) can be expressed as the Breslow-type estimator for $R(u)$:

$$\hat{R}(u; \theta) = \int_0^u \frac{\sum_{i=1}^n dN_i(s)}{\sum_{j=1}^n Y_j(s) \phi\{\hat{R}(s; \theta) \exp(\gamma a_j + \beta' z_j)\} \exp(\gamma a_j + \beta' z_j)},$$

which provides a more efficient forward recursive formula to compute $\hat{R}(u; \theta)$ with $\hat{R}(0; \theta) = 0$. Thus, the estimators for θ and $R(u)$ can be obtained through an iterative algorithm until a predetermined convergence criterion is met. When $\Phi(u) = u$, the proportional hazards model, the foregoing procedure is precisely equivalent to the Cox partial likelihood procedure (Wang et al., 1993) for left-truncated and right-censored data. In the absence of left-truncation, the estimating procedures above reduce to those in the articles of Zucker (2005) and Martinussen and Scheike (2006). Define the martingale process $M(u) = N(u) - \int_0^u Y(s) d\Phi\{R_0(s) \exp(\gamma_0 a + \beta'_0 z)\}$, where $(\theta_0, R_0(\cdot))$ are the true parameters. Let the normalized partial score vector of $L_P\{\theta, \hat{R}(\cdot; \theta)\}$ be

$$U\{\theta, \hat{R}(\cdot; \theta)\} = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \psi_i(s; \theta) - \frac{S^{(1)}(s; \theta)}{S^{(0)}(s; \theta)} \right\} dN_i(s),$$

where $\tau = \inf\{u : \text{pr}(Y > u) = 0\}$,

$\psi(s; \theta) = \partial \log [\phi\{\hat{R}(s; \theta) \exp(\gamma a + \beta' z)\} \exp(\gamma a + \beta' z)] / \partial \theta$ and

$S^{(q)}(s; \theta) = n^{-1} \sum_{i=1}^n Y_i(s) \phi\{\hat{R}(s; \theta) \exp(\gamma a_i + \beta' z_i)\} \exp(\gamma a_i + \beta' z_i) \psi_i(s; \theta)^{\otimes q}$, $q = 0, 1, 2$ with $\psi(s; \theta)^{\otimes 0} = 1$, $\psi(s; \theta)^{\otimes 1} = \psi(s; \theta)$ and $\psi(s; \theta)^{\otimes 2} = \psi(s; \theta) \psi(s; \theta)'$. Also define

$V^{(q)}(s; \theta) = n^{-1} \sum_{i=1}^n \left[Y_i(s) \dot{\phi} \left\{ R_0(s) \exp(\gamma_0 a_i + \beta'_0 z_i) \right\} \exp \left\{ 2(\gamma_0 a_i + \beta'_0 z_i) \right\} \right] \psi_i(s; \theta)^{\otimes q}$, Under regularity conditions, $S^{(q)}(s; \theta_0)$ and $V^{(q)}(s; \theta_0)$ uniformly converge in probability to $s^{(q)}(s)$ and $v^{(q)}(s)$, respectively. In the Web Appendix B, we show that $\hat{\theta}$ and $\hat{R}(u; \hat{\theta})$ are consistent and asymptotically normal and their asymptotic properties are summarized in Proposition 1 and 2, respectively.

In addition, the parameters, θ and $R_0(u)$, have causal interpretations if assumptions (A1) and (A2) also hold, since

$$\begin{aligned} S_a(u | \mathbf{Z}^*) &= S_a(u | A^*=a, \mathbf{Z}^*) \\ &= S(u | A^*=a, \mathbf{Z}^*), \end{aligned}$$

where $S_a(u | \mathbf{Z}^*) = \text{pr}(T_a^* \geq u | \mathbf{Z}^*)$, $a=0$ or 1 . In the equation above, the first equality comes from Assumption (A2) and the second equality holds because of Assumption (A1). In the Web Appendix B, we summarize the asymptotic normality property of

$\exp \left[-\Phi \left\{ \hat{R}(u; \hat{\theta}) \exp(\hat{\gamma}a + \hat{\beta}'z) \right\} \right]$, the estimator of $S(u | A^*=a, \mathbf{Z}^*=z)$, in Proposition 3.

3. Causal Estimation

In this article, one of the aims is to estimate the causal survival functions,

$S_a(u) = \text{pr}(T_a^* \geq u)$, $a=0$ or 1 . In the literature of causal inference, population-average quantities such as the causal survival function are commonly estimated by model-based standardization procedures (Lane and Nelder, 1982; Greenland, 2004) for analyzing standard survival data with right censoring, that is, expressing $S_a(u)$ as $S_a(u) = \int_z S_a(u | z) dG_{\mathbf{Z}^*}(z)$. In case that $G_{\mathbf{Z}^*}(\cdot)$ is known, one can immediately estimate $S_a(u)$, though $G_{\mathbf{Z}^*}(\cdot)$ is typically unknown in practice and needs to be estimated. When analyzing prevalent survival data, covariates accompanied by survival times are both biased because of prevalent sampling, where the bias of covariates is caused by its association with biased failure times (Chan and Wang, 2012; Cheng and Wang, 2012). Thus, the use of empirical measure to estimate $G_{\mathbf{Z}^*}(\cdot)$ is statistically inappropriate, and therefore further modification is needed in the estimation procedures. We consider the marginal likelihood in (2) and treating it as a multinomial distribution with the selection probability $S(w | a, z)$:

$$dG_{A, \mathbf{Z}}(a, z) = \frac{S(w | a, z) dG_{A^*, \mathbf{Z}^*}(a, z)}{\alpha}$$

in which $G_{A, \mathbf{Z}}(a, z) = \text{pr}(A=a, \mathbf{Z}=z)$ and $\alpha = \int_z \int_a S(w | a, z) dG_{A^*, \mathbf{Z}^*}(a, z)$. Substituting S with $L_M(S, G_{\mathbf{Z}^*})$ in the marginal likelihood $G_{\mathbf{Z}^*}(\cdot)$, the estimator of $G_{\mathbf{Z}^*}(\cdot)$ can be derived by maximizing the marginal pseudo likelihood $L_M(\hat{S}, G_{\mathbf{Z}^*})$ without parametric distributional assumptions on truncation time and covariates:

$$\hat{G}_{A^*, Z^*}(a_0, z_0; \hat{S}) = \left(\sum_{i=1}^n \frac{1}{\hat{S}(w_i | a_i, z_i)} \right)^{-1} \sum_{i=1}^n \frac{I(a_i \leq a_0, z_i \leq z_0)}{\hat{S}(w_i | a_i, z_i)},$$

where $\hat{S}(w | a, z)$ is described in Section 2. Note that $\hat{\alpha}(\hat{S}) = n \left\{ \sum_{i=1}^n 1/\hat{S}(w_i | a_i, z_i) \right\}^{-1}$ is an estimator of α . Summing over all possible values of a_i , an estimator of $G_{Z^*}(z_0)$ can be obtained as

$$\hat{G}_{Z^*}(z_0; \hat{S}) = \left(\sum_{i=1}^n \frac{1}{\hat{S}(w_i | a_i, z_i)} \right)^{-1} \sum_{i=1}^n \frac{I(z_i \leq z_0)}{\hat{S}(w_i | a_i, z_i)}.$$

Note that this result explicitly provides an estimate of the distribution of the truncation time since the variable of truncation times is also one of the covariates Z^* . Here we indicate that the estimation of $G_{Z^*}(\cdot)$ is more general than the one proposed by Wang (1991), since in the latter case the failure and truncation times are assumed independent, whereas in our procedures they are not.

Theorem 1

Under suitable regularity conditions,

$$\hat{G}_{Z^*}(z_0; \hat{S}) - G_{Z^*}(z_0) = n^{-1} \sum_{i=1}^n \int_0^\tau \xi_{G1}(\tilde{O}_i, z_0, s) dM_i(s) + n^{-1} \sum_{i=1}^n \xi_{G2}(\tilde{O}_i, z_0) + o_p(n^{-1/2}),$$

where

$$\begin{aligned} \xi_{G1}(\tilde{O}, z_0, s) &= \alpha f \{G_{Z^*}(z_0) - I(z \leq z_0)\} \left\{ \xi_{CS}(\tilde{O}, w, a, z, s) / S^2(w | a, z) \right\} dG_{A,Z}(a, z) \\ \text{and } \xi_{G2}(\tilde{O}, z_0) &= \alpha \{I(Z \leq z_0) - G_{Z^*}(z_0)\} / S(W | A, Z). \end{aligned}$$

From Theorem 1, one can show the weak convergence of $n^{1/2} \{ \hat{G}_{Z^*}(z_0; \hat{S}) - G_{Z^*}(z_0) \}$ through the martingale theory (Rebolledo, 1980) and empirical process (Hahn, 1977)

representations. Moreover, the expression of $\hat{G}_{Z^*}(z_0)$ shows that this estimator has an Inverse-Probability-Weight (IPW) format where the probability weight is inversely proportional to the survival function, that is, $1/\hat{S}(w | a, z)$. In the absence of truncation (i.e., $W^* = 0$ with probability 1), $\hat{S}(w | a, z) = 1$ and the estimator reduces to an empirical estimator with equal weights.

In the case $G_{Z^*}(z)$ is estimated, the causal survival functions, $S_a(u)$, $a=0$ or 1 , can be estimated by the model-based standardization procedures via calculation of the convolution integral,

$$\hat{S}_a(u) = \int_z \hat{S}_a(u|z) d\hat{G}_{z^*}(z) = \left(\sum_{i=1}^n \frac{1}{\hat{S}(w_i|a_i, z_i)} \right)^{-1} \sum_{i=1}^n \frac{\hat{S}_a(u|z_i)}{\hat{S}(w_i|a_i, z_i)}, \quad a=0, 1. \quad (6)$$

In equation (6), the same distribution estimate $\hat{G}_{z^*}(z_0)$ on the left-hand side is used for *both* treatment groups, since the truncation distribution G does not involve treatment assignment.

Note that in the absence of truncation, $\hat{S}(w_i|a_i, z_i) = \hat{S}(0|a_i, z_i) = 1$ and the estimator $\hat{S}_a(u)$ reduces to the average of $\hat{S}_a(u|z_i)$ over the empirical measure on z_i . Essentially, as an interesting structure of $\hat{S}_a(u)$, the fraction $1/\hat{S}(w_i|a_i, z_i)$ in equation (6) serves to correct selection bias and systematic imbalance simultaneously in estimation of the causal survival function when left truncation is present in the observational study. Note that the estimator in (6) is structural and has a simple closed form.

Theorem 2

Under the regularity conditions, $n^{1/2} \{ \hat{S}_1(u) - S_1(u) \}$ converges weakly to a mean zero Gaussian process with variance function $\sigma_1^2(u)$, where

$$\sigma_1^2(u) = \int_0^\tau E \left[\xi_C^2(\tilde{O}, u, s) \phi \left\{ R_0(s) \exp(\gamma_0 A + \beta'_0 \mathbf{Z}) \right\} \exp(\gamma_0 A + \beta'_0 \mathbf{Z}) Y(s) \right] dR_0(s) + \alpha^2 \text{var} \left\{ \frac{S_1(u|\mathbf{Z})}{S(W|A, \mathbf{Z})} \right\}$$

can be consistently estimated by

$$\begin{aligned} \hat{\sigma}_1^2(u) = & n^{-1} \sum_{i=1}^n \int_0^\tau \hat{\xi}_C^2(\tilde{O}_i, u, s) \phi \left\{ \hat{R}(s; \hat{\theta}) \exp(\hat{\gamma} a_i + \hat{\beta}' z_i) \right\} \exp(\hat{\gamma} a_i \\ & + \hat{\beta}' z_i) Y_i(s) d\hat{R}(s, \hat{\theta}) \\ & + \hat{\alpha}^2 \hat{\text{var}} \left\{ \frac{S_1(u|\mathbf{Z})}{S(W|A, \mathbf{Z})} \right\} \end{aligned}$$

in which $\hat{\xi}_C(\tilde{O}, u, s)$ is obtained by plugging in the consistent estimators of the parameters,

θ and $R(s)$, into $\xi_C(\tilde{O}, u, s)$ and $\hat{\text{var}} \{ S_1(u|\mathbf{Z}) / S(W|A, \mathbf{Z}) \}$ is a moment estimator of $\{ S_1(u|\mathbf{Z}) / S(W|A, \mathbf{Z}) \}$.

4. Simulation Study

Simulation studies are conducted to examine performance of the proposed estimators for covariates distribution and causal survival functions based on prevalent survival data. To see the role of truncation and causality, the proposed estimators are compared to two naïve estimators: Naïve estimator 1 (N1) is derived by treating the data as the standard survival data subject to right censoring; Naïve estimator 2 (N2) accounts for the bias generated by independent truncation but ignores the systematic imbalance between covariates and dependence between truncation and failure times. More precisely,

$$\hat{G}_{\mathbf{z}^*, N_1}(\mathbf{z}_0) = n^{-1} \sum_{i=1}^n I(\mathbf{z}_i \leq \mathbf{z}_0), \hat{S}_{a, N_1}(u) = n^{-1} \sum_{i=1}^n \hat{S}_{a, N_1}(u | \mathbf{z}_i)$$

and $\hat{S}_{a, N_1}(u | \mathbf{z})$ is obtained using the same procedure in Section 2 but ignoring truncation effect, that is, setting $w_i = 0$ while

$$\hat{G}_{\mathbf{z}^*, N_2}(\mathbf{z}_0) = \left(\sum_{i=1}^n \frac{1}{\hat{S}_{N_2}(w_i)} \right)^{-1} \sum_{i=1}^n \frac{I(\mathbf{z}_i \leq \mathbf{z}_0)}{\hat{S}_{N_2}(w_i)}, \hat{S}_{a, N_2}(u) = \prod_{s \leq u} \left(1 - \frac{\sum_{i=1}^n dN_i(s) I(a_i = a)}{\sum_{i=1}^n Y_i(s) I(a_i = a)} \right)$$

and $\hat{S}_{a, N_2}(\cdot)$ is exactly the truncated type of Kaplan-Meier estimator (Lynden-Bell, 1971).

In our simulation setting, we consider $Z^* = W^*$ which is generated from an uniform distribution(0,3). Conditional on W^* , A^* was generated from Bernoulli distribution with treatment selection probability $\exp(\eta_0 + \eta_1 W^*) / \{1 + \exp(\eta_0 + \eta_1 W^*)\}$. Let $\eta_0 = 1$ and $\eta_1 = -1$. Conditional on W^* , the potential survival times, T_0^* and T_1^* , are simulated from the transformation models with survival functions $\exp[-\Phi\{R(u) \exp(\beta W^*)\}]$ and $\exp[-\Phi\{R(u) \exp(\gamma + \beta W^*)\}]$, respectively, with $(\gamma, \beta) = (1, 1)$ and $\Phi(u) = u$, $\log(1+u)$, and $2^{-1} \log(1+2u)$, where $\Phi(u) = u$ and $\Phi(u) = \log(1+u)$ respectively correspond to the proportional hazards (PH) model and the proportional odds (PO) model. In all cases, let $R(t) = t/10$. The survival times were simulated from the distribution of $T^* = A^* T_1^* + (1 - A^*) T_0^*$. For each such generated (T^*, W^*) , the data were kept only when $T^* \leq W^*$. Finally, after a further step of censoring, the observed data include iid copies of (Y, A, W) , where $Y = \min(T, W + C)$, $I = I(T < W + C)$, and C is the residual censoring time from a Uniform $(0, c)$ distribution. The constant c was chosen to satisfy the censoring rates around 20% and 60%, respectively. Using this algorithm, we simulated $M = 500$ data sets, with 1000 observed subjects in each scenario.

In Figures 1~3, the simulation results for the causal survival functions and truncation distribution are displayed from the top panel to bottom panel along with two different censoring rates from the left panel to right panel, respectively, where the solid lines indicate the true functions. In each simulation scenario, the proposed estimators are compared with two naïve estimators in Figure 1~3, and it is clear that the estimated survival curves based on the proposed estimator are close to the true curve in all simulation scenarios. As expected, the 95% confidence interval widens as the censoring rate increases. On the other hand, it also shows that there is a tendency that the two naïve estimators are biased. An explanation of the bias for naïve estimator 1 (N1) is that subjects with longer survival times are more likely to be sampled with the prevalent sampling scheme. For the naïve estimator 2 (N2), the bias is not only due to the systematic difference between treated and untreated groups, but also due to the dependence structure between the failure and truncation times. Similar explanation applies to the proposed estimator $\hat{G}_{\mathbf{z}^*}(\mathbf{z}_0)$ and the naïve estimators $\hat{G}_{\mathbf{z}^*, N_1}(\mathbf{z}_0)$ and $\hat{G}_{\mathbf{z}^*, N_2}(\mathbf{z}_0)$.

We also compare the proposed approach with the propensity score approach of Cheng and Wang (2012), with the results provided in Table 1. To summarize the comparison, note that the approach of Cheng and Wang (2012) was proposed under the proportional hazards model, where the validity of their analytical approach requires the independent truncation assumption as well as the correct specification of the propensity score model. As a result, the bias of the estimator of Cheng and Wang (2012) becomes substantial when model assumptions are violated, as shown in Table 1. In the situation where the independent truncation assumption is satisfied, such as the example illustrated in Web Appendix C, both approaches perform well as expected, but the variation of the currently proposed approach is smaller. In contrast with the more direct approach proposed in this paper, the two-stage estimating procedure of Cheng and Wang (2012) tends to generate larger variation in estimation. In particular, the estimation of propensity scores in the two-stage procedure requires an additional ‘offset term’ to adjust for the prevalent sampling, which is expected to lead to increased estimation error.

In Table 1, the average of the standard error estimates from the bootstrap with the empirical standard deviation of estimated parameters are presented. We found that the two types of standard error estimate are very close to each other. Moreover, the coverage rates of the 95% confidence intervals based on the bootstrapped standard errors are very close to the nominal level (0.95).

5. Data Analysis

Breast radiation therapy has been used as a standard care for patients with early breast cancer for many years. Considerable effort had been made to assess the efficacy of breast radiation therapy after breast-conserving surgery (lumpectomy, with and without dissection of auxiliary lymph nodes) for older women with early breast cancer (Buchholz, 2009). To draw a causal conclusion regarding treatment efficacy, we integrated incidence and demographical information from the SEER database for breast cancer with treatment information from the Medicare database to identify women aged over 65 years with stage 0, 1 or 2 breast cancer. While the diagnosis data from SEER for breast cancer have been collected and recorded since 1970, the Medicare claims data became available in 1986. As a consequence, we can only use data from those patients who remained alive in 1986. Thus, by defining the failure time T^* as the time from breast cancer diagnosis to death in the population, the failure time is observed subject to left truncation and right censoring, where the truncation time is defined as the time from the date of diagnosis to 1986 and right censoring occurred due to administrative closure of data collection or death due to other causes. It turns out that allowing the dependency between failure and truncation time implies that the survival function can vary with date of diagnosis, even after adjusting for covariates. This situation is commonly seen when the treatment or the technology of screening is improved over years and, for example, residual lifetime since diagnosis is prolonged over calendar time.

To better examine the truncation effect and causality, we used the proposed estimators $\hat{S}_0(u)$ and $\hat{S}_1(u)$, compared to two naïve estimators, and the analyzed data only include information from patients diagnosed with breast cancer before 1986, and exclude those who

were diagnosed with breast cancer after 1986. This includes data from 2000 living women diagnosed with breast cancer before 1986, where a total of 969 patients were treated with radiation therapy. The overall censoring is about 84% from the observed data. Using the results from Rubin (1997), the potential confounding factors include age, race (2 levels), marital status (2 levels), stage (3 levels), laterality (2 levels) and metropolitan status (2 levels). Since the treatment and truncation times are two important variables of interest, additional interaction between these two variables is also included in the model.

The proposed estimating procedure in Section 3 is based on the specification of the link function $\Phi(\cdot)$. To select an appropriate link function in the transformation model with prevalent survival data, we can use a simple graphical method by comparing the fitted residual distribution with the corresponding error distribution. To do so, let

$\hat{\epsilon}(X_i) = \hat{H}(X_i) + \hat{\beta}' Z_i$ be the residual failure time and $\hat{\epsilon}(A_i) = \hat{H}(A_i) + \hat{\beta}' Z_i$ the residual truncation time, $i = 1, \dots, n$. Then the fitted residual distribution can be obtained by using a truncated Kaplan-Meier estimator on the data $\{\hat{\epsilon}(A_i), \hat{\epsilon}(X_i), \Delta_i, i=1, \dots, n\}$. If the underlying model assumption is correctly specified, we then expect that the fitted residual distribution is close to the error distribution (Cheng et al., 1997). In this data analysis, we compared the proportional hazards model with the proportional odds model. The result in the Web Appendix D suggests that the proportional odds model fits the data slightly better than the proportional hazards model although the difference is not significant.

Figure 4 displayed the estimation results of causal survival functions under proportional odds model. The pointwise 95% confidence interval of estimated causal survival functions are derived by using a bootstrap procedure. Without adjusting for the truncation effect in the model, we find that $\hat{S}_{0,N1}(u)$ and $\hat{S}_{1,N1}(u)$ have values higher than the proposed estimates in the with and without radiation therapy groups, showing the bias of overweighing longer survival times. Further, due to ignoring the systematic difference between treated and untreated groups and ignoring the dependence between failure and truncation times, the value of $\hat{S}_{0,N2}(u)$ is higher than the proposed estimate while the value of $\hat{S}_{1,N2}(u)$ is slightly lower than the proposed estimate. Also, although the proposed survival rates suggest a larger difference between groups, the overall difference between the treated and untreated groups does not seem to suggest a significant improvement from patients receiving the radiation therapy. Note that in comparison with the two stage approach proposed by Cheng and Wang (2012), under the proportional hazards model, we obtain a similar conclusion though the new approach proposed in this paper provides a more precise estimation result. Specifically, the 95% confidence interval of the proposed estimate of the difference (treatment versus control) in survival at 5 years is $(-0.01, 0.04)$ while Cheng and Wang (2012) provides an almost 2.5 times wider 95% confidence interval.

6. Discussion

In this paper, two aspects of the prevalent survival data are addressed. First, a general class of semiparametric transformation models is considered for estimation of causal survival functions. A pseudo-partial likelihood approach is proposed to estimate regression parameters based on prevalent survival data, and the model has the characteristic to allow

for dependence between failure and truncation times conditioning on other covariates. Secondly, we consider the estimation of causal survival functions from prevalent survival data and explain why the classical model-based standardization procedure is biased.

A specific challenge in the proposed methodology is to handle the biased distribution of observed covariates, where the bias arises because of the correlation between covariates and biased failure time (Chan and Wang, 2012; Cheng and Wang, 2012). This forms a contrast to incident sampling in which the empirical distribution of the observed covariates leads to a consistent estimator. With much relaxed model assumptions, an inverse weighting approach is proposed in Section 3 to correct the sampling bias, where the weights are associated with the distribution of failure times. Also, since the truncation time can be treated as a covariate under the modeling framework, it implies that the truncation time distribution can be estimated without strict independence assumption between the truncation and failure times. Moreover, once the distribution of failure times is fitted, the proposed model-based standardization procedure of causal survival functions can simultaneously correct the selection bias from the prevalent sampling and balance the systematic differences due to the observational study.

It should be indicated that the validity of the proposed approaches require the three assumptions described in Section 2. Among these three assumptions, SUTVA and strongly ignorable treatment assignment are typical assumptions associated with the potential outcomes framework while the conditional independent censoring assumption is mainly used for survival analysis. The assumption SUTVA, including noninterference and consistency, serves to avoid the interference between units and multiple versions of treatment. Thus, under SUTVA, the potential outcomes is well defined for each unit under each possible treatment. In practice, SUTVA would be more plausible in well-designed experiments because an unit's treatment status can be manipulated hypothetically without interfering with other units (Rubin, 1990). An example of such conditions was considered by Cox (1958) and Rubin (1980) in agricultural experiments. In recent years, the assumption SUTVA has been extended in different directions; for example, Sobel (2006) and Hudgens and Halloran (2008) considered the the "spillover" effect in the presence of interference, while Cole and Frangakis (2009) and VanderWeele (2009) expanded the definition of potential outcomes to include multiple version of treatment. For strongly ignorable treatment assignment, this assumption tends to be true if there is no unmeasured confounding variables. To address the issue of hidden bias due to unmeasured confounding, typically, sensitivity parameters together with the propensity scores can be used (Rosenbaum and Rubin, 1983; Tan, 2006; VanderWeele and Arah, 2011). However, given that the propensity score is a function of covariates and that the distribution of observed covariates is biased (Cheng and Wang, 2012), standard approaches based on propensity score would not work properly and will need adjustment to deal with bias caused by prevalent sampling. Thus, sensitivity analysis becomes more complicated than the usual approach for analyzing standard survival data. Furthermore, we point out that Assumption (A3) is the typical independent censoring condition and, in our work, the standard independent truncation assumption is relaxed through the inclusion of the observed truncation time, W , as one component of the covariates. In real data applications, the standard independence truncation

assumption can be validated when the coefficient of truncation time W in the semiparameter transformation model equals zero.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported in part by the Taiwan National Science Council NSC101-2118-M-007-004 (for first author) and National Institute of Health grant R01AI078835 (to second author). The authors thank the Editor, the Associate Editor and two referees for their helpful and constructive comments which have led to an improved version of the paper.

Appendix

In this section, we provide a sketch of the proofs of Theorems 1-2. Detailed regularity conditions and the proofs are given in the Web Appendix E.

A Sketch of Proof of Theorem 1

To develop asymptotic properties of $\hat{G}_{Z^*}(z_0; \hat{S})$, we decompose $\hat{G}_{Z^*}(z_0; \hat{S}) - G_{Z^*}(z_0)$ as the sum of $\hat{G}_{Z^*}(z_0; \hat{S}) - \hat{G}_{Z^*}(z_0; S)$ and $\hat{G}_{Z^*}(z_0; S) - G_{Z^*}(z_0)$, in which $\hat{G}_{Z^*}(z_0; \hat{S}) - \hat{G}_{Z^*}(z_0; S)$ corresponds to the variation term due to error from estimation of $S(u | a, z)$ and, with $S(u | a, z)$ treated as a known function, $\hat{G}_{Z^*}(z_0; S) - G_{Z^*}(z_0)$ is the variation term due to error from estimation of G_{Z^*} . As seen in the proof of Theorem 1, the two terms are orthogonal to each other, and, therefore, the asymptotic variance is the sum of variances of the two variation terms.

A Sketch of Proof of Theorem 2

The asymptotic property of $\hat{S}_a(u)$ is summarized as follows. We shall express $\hat{S}_1(u) - S_1(u)$ as the sum of three processes, and show that the sum of the three processes are asymptotically equivalent to the sum of two processes which are asymptotically orthogonal to each other, for which one is expressed as a linear component of a martingale process and the other is the sum of independent and identically distributed processes. For the martingale process, a weak convergence property is achieved by the martingale central limit theorem (Rebolledo, 1980). The second process is tight according to Hahn (1977), that leads to its weak convergence. Using similar arguments, we can also derive the weak convergence property of $\hat{S}_0(u)$.

References

- Brookmeyer R, Gail MH. Biases in prevalent cohorts. *Biometrics*. 1987; 43:739–749. [PubMed: 3427161]
- Buchholz TA. Radiation therapy for early-stage breast cancer after breast-conserving surgery. *The New England Journal of Medicine*. 2009; 360:63–70. [PubMed: 19118305]

- Chan KCG, Wang M-C. Estimating incident population distribution from prevalent data. *Biometrics*. 2012; 68:521–531. [PubMed: 22313264]
- Chen K, Jin Z, Ying Z. Semiparametric analysis of transformation models with censored data. *Biometrika*. 2002; 89:659–668.
- Chen P-Y, Tsiatis A. Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics*. 2001; 57:1030–1038. [PubMed: 11764241]
- Cheng SC, Wei LJ, Ying Z. Analysis of transformation models with censored data. *Biometrika*. 1995; 82:867–878.
- Cheng SC, Wei LJ, Ying Z. Predicting survival probabilities with semiparametric transformation models. *Journal of the American Statistical Association*. 1997; 92:227–235.
- Cheng Y-J, Wang M-C. Estimating propensity scores and causal survival functions using prevalent survival data. *Biometrics*. 2012; 68:707–716. [PubMed: 22834993]
- Cole SR, Frangakis CE. The consistency assumption in causal inference: a definition or an assumption ? *Epidemiology*. 2009; 20:3–5. [PubMed: 19234395]
- Cox, DR. *Planning of Experiments*. John Wiley & Sons; New York: 1958.
- Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *American Journal of Epidemiology*. 2004; 160:301–305. [PubMed: 15286014]
- Hahn M. Conditions for sample-continuity and central limit theorem. *Annals of Probability*. 1977; 5:351–360.
- Holland P. Statistics and causal inference. *Journal of the American Statistical Association*. 1986; 81:945–970.
- Hudgens MG, Halloran ME. Toward causal inference with interference. *Journal of the American Statistical Association*. 2008; 103:832–842. [PubMed: 19081744]
- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958; 53:457–481.
- Lane PW, Nelder JA. Analysis of covariance and standardization as instances of prediction. *Biometrics*. 1982; 38:613–621. [PubMed: 7171691]
- Lynden-Bell D. A method of allowing for known observational selection in small samples applied to 3cr quasars. *Monographs of the Monthly Notices of the Royal Astronomical Society*. 1971; 155:95–118.
- Martinussen, T.; Scheike, TH. *Dynamic Regression Models for Survival Data*. Springer; 2006.
- Neyman J. On the application of probability theory to agricultural experiments. essay on principles. section 9. (In Polish) *Roczniki Nauk Rolniczych*, Tom X. 1923:1–51.
- Rebolledo R. Central limit theorems for local martingales. *Z Wahrsch Verw Gebiete*. 1980; 51:269–86.
- Rosenbaum PR, Rubin DB. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society Series B*. 1983; 45:212–218.
- Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*. 1974; 66:688–701.
- Rubin DB. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*. 1978; 6:34–58.
- Rubin DB. Comment on: “randomisation analysis of experimental data in the fisher randomisation test” by D. Basu. *Journal of the American Statistical Association*. 1980; 75:591–593.
- Rubin DB. Formal models of statistical inference for causal effects. *Journal of Statistical Planning and Inference*. 1990; 25:279–292.
- Rubin DB. Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine*. 1997; 127:757–763. [PubMed: 9382394]
- Rubin DB. Teaching statistical inference for causal effects in experiments and observational studies. *Journal of Educational and Behavioral Statistics*. 2004; 29(3):343–367.
- Sobel ME. What do randomized studies of housing mobility demonstrate ? *Journal of the American Statistical Association*. 2006; 101:1398–1407.

- Tan Z. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*. 2006; 101:1619–1637.
- VanderWeele TJ. Concerning the consistency assumption in causal inference ? *Epidemiology*. 2009; 20:880–883. [PubMed: 19829187]
- VanderWeele TJ, Arah OA. Unmeasured confounding for general outcomes, treatments, and confounders: Bias formulas for sensitivity analysis. *Epidemiology*. 2011; 22:42–52. [PubMed: 21052008]
- Wang M-C. Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*. 1991; 86:130–143.
- Wang M-C, Brookmeyer R, Jewell NP. Statistical models for prevalent cohort data. *Biometrics*. 1993; 49:1–11. [PubMed: 8513095]
- Wang M-C, Jewell NP, Tsai W-Y. Asymptotic properties of the product limit estimate under random truncation. *The Annals of Statistics*. 1986; 14:1597–1605.
- Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the seer-medicare data: content, research applications, and generalizability to the united states elderly population. *Med Care*. 2002; 40(8):3–18.
- Woodroffe M. Estimating a distribution function with truncated data. *Annals of statistics*. 1985; 13:163–177.
- Zeng D, Lin DY. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*. 2006; 93:627–640.
- Zucker DM. A pseudo-partial likelihood method for semiparametric survival regression with covariate errors. *Journal of the American Statistical Association*. 2005; 100:1264–1277.

Figure 1.

Simulation results for the proportional hazards (PH) model. The top and middle panels display the estimators of the causal survival functions; The bottom panel displays the estimator of truncation function; The left-side panel displays the results under 20% censoring rate, and the right side under 60% censoring rate. Solid line indicates the true curve. The dashed line indicates the replication mean of the proposed estimator, with dotted line indicating pointwise 95% confidence interval. The black dotdash line indicates the replication mean of naïve estimator 1 (N1); The grey dotdash line indicates the replication mean of naïve estimator 2 (N2).

Figure 2.

Simulation results for the proportional odds (PO) model. The top and middle panels display the estimators of the causal survival functions; The bottom panel displays the estimator of truncation function; The left-side panel displays the results under 20% censoring rate, and the right side under 60% censoring rate. Solid line indicates the true curve. The dashed line indicates the replication mean of the proposed estimator, with dotted line indicating pointwise 95% confidence interval. The black dotdash line indicates the replication mean of naïve estimator 1 (N1); The grey dotdash line indicates the replication mean of naïve estimator 2 (N2).

Figure 3.

Simulation results for semiparametric transformation models, $\Phi(u) = 2^{-1} \log(1+2u)$. The top and middle panels display the estimators of the causal survival functions; The bottom panel displays the estimator of truncation function; The left-side panel displays the results under 20% censoring rate, and the right side under 60% censoring rate. Solid line indicates the true curve. The dashed line indicates the replication mean of the proposed estimator, with dotted line indicating pointwise 95% confidence interval. The black dotdash line indicates the replication mean of naïve estimator 1 (N1); The grey dotdash line indicates the replication mean of naïve estimator 2 (N2).

Figure 4.

The panel displays estimation results of causal survival functions under the proportional odds model (left: $S_0(u)$; right: $S_1(u)$); Solid line indicates the mean of our estimator. The black dotdash line indicates the mean of naïve estimator 1 (N1). The grey dotdash line indicates the mean of naïve estimator 2 (N2). The dotted line indicates the pointwise 95% confidence interval.

Table 1

Summary of simulation studies when the survival functions and truncation distribution are evaluated at time 3 and 1.5, respectively. The superscript PS of the estimator denotes Cheng and Wang (2012) propensity scores approach. Bias and SD, empirical bias ($\times 1000$) and empirical standard deviation ($\times 1000$) of 500 parameter estimates; ASE, average of estimated standard errors from bootstrap (500 bootstrapping samples for each simulated data) ($\times 1000$); CR, coverage rate of the 95% confidence interval.

Model	Censoring rate	Estimator	Bias	SD(ASE)	CR
PH	20%	$\hat{S}_0(3)$	6	71(63)	0.93
		$\hat{S}_0^{PS}(3)$	200	43	
		$\hat{S}_1(3)$	3	29(26)	0.93
		$\hat{S}_1^{PS}(3)$	92	45	
		$\hat{G}_{Z^*}(1.5)$	8	111(96)	0.92
	60%	$\hat{S}_0(3)$	1	90(82)	0.96
		$\hat{S}_0^{PS}(3)$	189	48	
		$\hat{S}_1(3)$	0	34(33)	0.96
		$\hat{S}_1^{PS}(3)$	97	57	
		$\hat{G}_{Z^*}(1.5)$	5	145(128)	0.94
PO	20%	$\hat{S}_0(3)$	2	76(71)	0.92
		$\hat{S}_1(3)$	3	58(54)	0.91
		$\hat{G}_{Z^*}(1.5)$	5	66(64)	0.94
	60%	$\hat{S}_0(3)$	6	86(84)	0.94
		$\hat{S}_1(3)$	0	65(61)	0.92
		$\hat{G}_{Z^*}(1.5)$	10	79(79)	0.94
$\Phi(u) = 2^{-1} \log(1 + 2u)$	20%	$\hat{S}_0(3)$	16	90(93)	0.93
		$\hat{S}_1(3)$	8	77(78)	0.92
		$\hat{G}_{Z^*}(1.5)$	15	64(70)	0.92
	60%	$\hat{S}_0(3)$	19	96(102)	0.92
		$\hat{S}_1(3)$	11	82(83)	0.92
		$\hat{G}_{Z^*}(1.5)$	19	68(77)	0.91