Cox Regression with Survival-Time-Dependent Missing Covariate Values

Jun SHAO

Department of Statistics
University of Wisconsin-Madison

Joint work with Yanyao YI, Ting YE, and Menggang YU

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Overview

1. Introduction
2. Method
3. Simulation
4. Example

Cox PH model:

For any continuous failure time T and time-independent covariate vector $\boldsymbol{V}$, Cox proportional hazard model is specified by

$$\lambda(t|\boldsymbol{V}) = \lambda_0(t)\exp(\boldsymbol{\theta}^\top \boldsymbol{V}), \tag{1}$$

where $\lambda(t|\boldsymbol{V})$ is the hazard at time t given covariate $\boldsymbol{V}$, $\lambda_0(t)$ is an unspecified baseline hazard function common for all subjects, $\boldsymbol{\theta}$ is a vector of unknown parameters, and $\boldsymbol{\theta}^\top$ is its transpose.

Censoring Assumption:

In survival studies, there is usually a censoring time C. In our study, what we observe is $T \wedge C = min(T, C)$ and $\delta = I_{\{T \leq C\}}$, the indicator of the event $T \leq C$. A common censoring assumption is

$$T \perp C \mid \boldsymbol{V}, \tag{2}$$

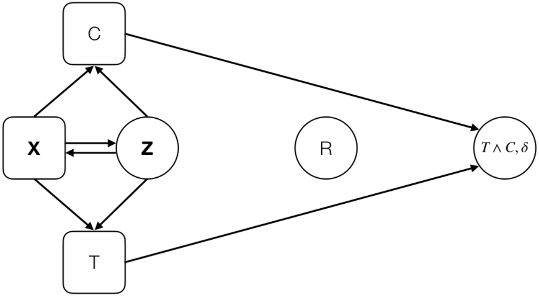
i.e., T and C are conditionally independent given $\boldsymbol{V}$.

Motivation

In clinical and epidemiological studies, it is frequent that some components of the covariate vector $\boldsymbol{V}$ may have missing values.

- $\boldsymbol{V} = (\boldsymbol{X}, \boldsymbol{Z})$: covariate vector measured at baseline,
- $\boldsymbol{X}$: component of $\boldsymbol{V}$ that may have missing values,
- $\boldsymbol{Z}$: component of $\boldsymbol{V}$ that is always observed,
- R: missing indicator that is 1 if and only if $\boldsymbol{X}$ is completely observed.

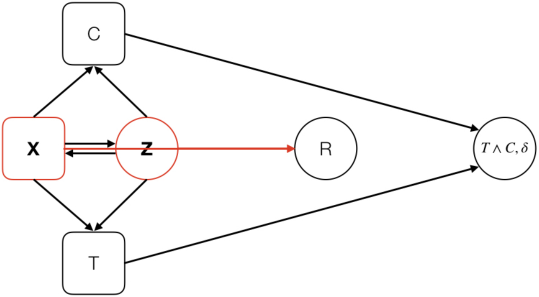
Motivation



$$R \perp (\boldsymbol{X}, \boldsymbol{Z}, T, C)$$

- missing completely at random,
- complete-case analysis can be applied,
- Lin and Ying (1993)

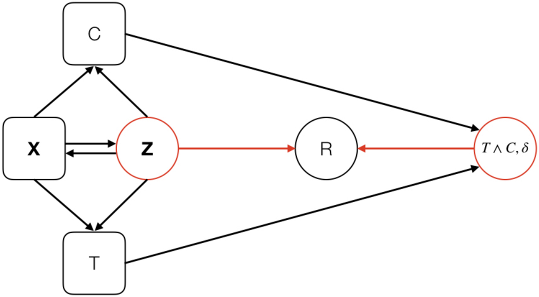
Motivation



$$R \perp (T, C) \mid (\boldsymbol{X}, \boldsymbol{Z})$$

- missingness depends only on covariates,
- complete-case analysis can be applied,
- Cook et al. (2011)

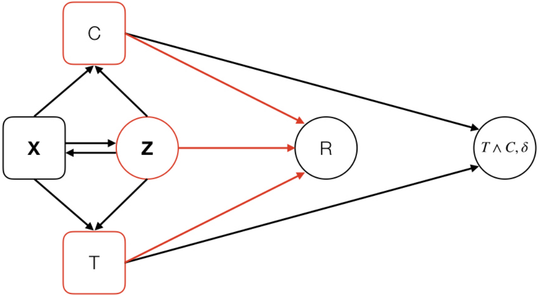
Motivation



$$R \perp (\boldsymbol{X}, T, C) \mid (\boldsymbol{Z}, T \wedge C, \delta) \quad (3)$$

- a type of missing at random assumption,
- inverse propensity weighting, imputation method, likelihood approach can be applied,
- Herring and Ibrahim (2001), Qi et al. (2005), Qiu (2017).

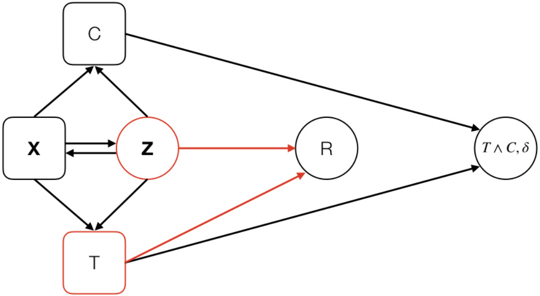
Motivation



$$R \perp \boldsymbol{X} \mid (\boldsymbol{Z}, T, C) \quad (4)$$

- (3) is a special case of (4),
- Inference is difficult to make without further assumption on the missingness or censoring mechanism,
- There is no existing literature.

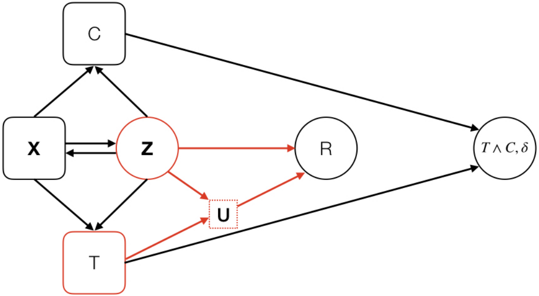
Motivation



$$R \perp (\boldsymbol{X}, C) \mid (\boldsymbol{Z}, T) \tag{5}$$

- **survival-time-dependent missingness**,
- a special case of (4),
- first introduced by Rathouz (2007).

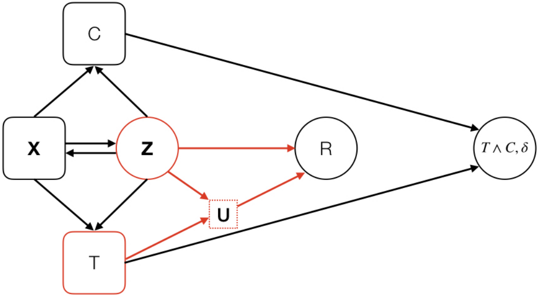
Motivation



$$R \perp (\boldsymbol{X}, T, C) \mid (\boldsymbol{Z}, U) \quad \text{and} \quad U \perp (C, \boldsymbol{X}) \mid (T, \boldsymbol{Z}) \qquad (6)$$

- (6) is an example of (5),
- $f(R, U, T, C, \boldsymbol{V}) = f(R \mid U, \boldsymbol{Z}) f(U \mid T, \boldsymbol{Z}) f(T \mid \boldsymbol{V}) f(C, \boldsymbol{V})$,
- $\mathrm{pr}(R = 1 \mid T, \boldsymbol{Z}) = E\{\mathrm{pr}(R = 1 \mid U, \boldsymbol{Z}) \mid T, C, \boldsymbol{V}\}$, i.e., T is a surrogate for U in the missingness propensity.

Motivation



- Stage III Non-Small Cell lung Cancer (NSCLC) data set from National Cancer Database (NCDB)
- X: more accurate stage III record, stage IIIA or IIIB,
- $\boldsymbol{Z}$: objective measurements, e.g., age, gender, treatment,
- U: physician's subjective assessment,
- R: physician's judgment on measuring X or not

Motivation

Under

(2) Independent Censoring: $T \perp C \mid \boldsymbol{V}$,

(5) Survival-time-dependent missingness: $R \perp (\boldsymbol{X}, C) \mid (\boldsymbol{Z}, T)$,

(1) Cox model: $\lambda(t|\boldsymbol{V}) = \lambda_0(t) \exp(\boldsymbol{\theta}^\top \boldsymbol{V})$.

Motivation

Under

(2) Independent Censoring: $T \perp C \mid \boldsymbol{V}$,

(5) Survival-time-dependent missingness: $R \perp (\boldsymbol{X}, C) \mid (\boldsymbol{Z}, T)$,

Rathouz (2007) has showed the survival function of T is identifiable, but there is no proposed method for estimating $\boldsymbol{\theta}$ in

(1) Cox model: $\lambda(t|\boldsymbol{V}) = \lambda_0(t) \exp(\boldsymbol{\theta}^\top \boldsymbol{V})$.

The main challenge is that the missingness mechanism (5) is missing not at random, since T may be censored.

Method

Let $\{(T_i, C_i, \boldsymbol{V}_i, R_i), i = 1, ..., n\}$ be independent and identically distributed samples from $(T, C, \boldsymbol{V}, R)$, where $\boldsymbol{V}_i = (\boldsymbol{X}_i, \boldsymbol{Z}_i)$, $\boldsymbol{Z}_i$ is always observed and $\boldsymbol{X}_i$ is completely observed if and only if $R_i = 1$. What we observe are $T_i \wedge C_i$ and $\delta_i = I_{\{T_i \leq C_i\}}$.

Define

$$\pi_1(T, \boldsymbol{Z}) = \mathrm{pr}(R = 1 \mid T, C, \boldsymbol{V}), \tag{7}$$

$$\psi(T, \boldsymbol{V}) = \mathrm{pr}(\delta = 1 \mid T, \boldsymbol{V}). \tag{8}$$

Doubly Weighted Estimator

$$M^*(t) = I_{\{T \le t\}} - \int_0^t I_{\{T \ge u\}} \lambda_0(u) \exp(\boldsymbol{\theta}^\top \boldsymbol{V}) du \qquad (9)$$

Doubly Weighted Estimation Equation:

$$\sum_{i=1}^n \int_0^\infty \frac{R_i \delta_i}{\pi_1(T_i, \boldsymbol{Z}_i)\psi(T_i, \boldsymbol{V}_i)} \left\{ \boldsymbol{V}_i - \frac{\boldsymbol{S}_1^{(1)}(\boldsymbol{\theta}, t)}{S_1^{(0)}(\boldsymbol{\theta}, t)} \right\} dN_i(t) = 0 \qquad (10)$$

where $N_i(t) = \delta_i I_{\{T_i \le t\}}$ is the counting process, and

$$\boldsymbol{S}_1^{(1)}(\boldsymbol{\theta}, t) = \frac{1}{n} \sum_{i=1}^n \frac{R_i \delta_i I_{\{T_i \ge t\}}}{\pi_1(T_i, \boldsymbol{Z}_i)\psi(T_i, \boldsymbol{V}_i)} \boldsymbol{V}_i \exp(\boldsymbol{\theta}^\top \boldsymbol{V}_i),$$

$$S_1^{(0)}(\boldsymbol{\theta}, t) = \frac{1}{n} \sum_{i=1}^n \frac{R_i \delta_i I_{\{T_i \ge t\}}}{\pi_1(T_i, \boldsymbol{Z}_i)\psi(T_i, \boldsymbol{V}_i)} \exp(\boldsymbol{\theta}^\top \boldsymbol{V}_i).$$

Doubly Weighted Estimator (DWE):

Estimator $\widehat{\boldsymbol{\theta}}_1$ of $\boldsymbol{\theta}$ can be obtained by solving (10).

Once $\boldsymbol{\theta}$ is estimated by $\widehat{\boldsymbol{\theta}}_1$, a semi-parametric estimator of $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ can be obtained by

$$\widehat{\Lambda}_{01}(t) = \int_0^t \frac{1}{nS_1^{(0)}(\widehat{\boldsymbol{\theta}}_1, u)} \sum_{i=1}^{n} \frac{R_i \delta_i dN_i(u)}{\pi_1(T_i, \boldsymbol{Z}_i)\psi(T_i, \boldsymbol{V}_i)}. \tag{11}$$

Estimation of $\pi_1(T, Z)$:

- Missingness assumption (5) implies that

$$\mathrm{pr}(R = 1 | T, Z) = \mathrm{pr}(R = 1 | T, Z, \delta = 1),$$

- A parametric form of $\mathrm{pr}(R = 1 | T, Z)$ is hard to specify, especially when the scenario (6) holds, i.e.,

$$\mathrm{pr}(R = 1 \mid T, \boldsymbol{Z}) = E\{\mathrm{pr}(R = 1 \mid U, \boldsymbol{Z}) \mid T, C, \boldsymbol{V}\}$$

- Product kernel regression (Racine and Li, 2004) is applied to handle both continuous and discrete covariates.

Estimation of $\pi_1(T, \boldsymbol{Z})$:

$$\widehat{\pi}_1(T, \boldsymbol{Z}) = \frac{\sum_i R_i \delta_i K_h \{(T_i, \boldsymbol{Z}_i), (T, \boldsymbol{Z})\}}{\sum_i \delta_i K_h \{(T_i, \boldsymbol{Z}_i), (T, \boldsymbol{Z})\}} \tag{12}$$

where

$$K_h \{(T_i, \boldsymbol{Z}_i), (T, \boldsymbol{Z})\} = L_{h_c} \{(T_i, \boldsymbol{Z}_i^c), (T, \boldsymbol{Z}^c)\} D_{h_d}(\boldsymbol{Z}_i^d, \boldsymbol{Z}^d),$$

$\boldsymbol{Z}_i = (\boldsymbol{Z}_i^c, \boldsymbol{Z}_i^d)$, $h = (h_c, h_d)$, L_{h_c} is a continuous kernel for T and $\boldsymbol{Z}^c$, continuous components of $\boldsymbol{Z}$, with h_c as the smoothing parameter, and D_{h_d} is a discrete kernel for $\boldsymbol{Z}^d$, discrete components of $\boldsymbol{Z}$, with $h_d \in [0, 1]$ as the smoothing parameter (Racine and Li, 2004).

Estimation of $\psi(T, \boldsymbol{V})$:

- Under Censoring Assumption (2):

$$\psi(T, \boldsymbol{V}) = \mathrm{pr}(\delta = 1 | T, \boldsymbol{V}) = S_{C|\boldsymbol{V}}(T)$$

where $S_{C|\boldsymbol{V}}(t) = \mathrm{pr}(C > t | \boldsymbol{V})$.

Estimation of $\psi(T, \boldsymbol{V})$:

- Under Censoring Assumption (2):

$$\psi(T, \boldsymbol{V}) = \mathrm{pr}(\delta = 1 | T, \boldsymbol{V}) = S_{C|\boldsymbol{V}}(T)$$

  where $S_{C|\boldsymbol{V}}(t) = \mathrm{pr}(C > t | \boldsymbol{V})$.

- By Proposition 1.11 in Shao (2003), assumption (5) implies that

$$C \perp R \mid (T, \boldsymbol{V}),$$

Estimation of $\psi(T, \boldsymbol{V})$:

- Under Censoring Assumption (2):

$$\psi(T, \boldsymbol{V}) = \mathrm{pr}(\delta = 1 | T, \boldsymbol{V}) = S_{C|\boldsymbol{V}}(T)$$

  where $S_{C|\boldsymbol{V}}(t) = \mathrm{pr}(C > t | \boldsymbol{V})$.

- By Proposition 1.11 in Shao (2003), assumption (5) implies that

$$C \perp R \mid (T, \boldsymbol{V}),$$

  together with assumption (2), we have

$$C \perp (R, T) \mid \boldsymbol{V}.$$

Estimation of $\psi(T, \boldsymbol{V})$:

$S_{C|\boldsymbol{V}}(t)$ can be estimated by using the subset of data with R=1 and treating $C \wedge T$ as the observed time and $1-\delta$ as the "event status" for censoring time C. Then,

$$\widehat{\psi}(T, \boldsymbol{V}) = \widehat{S}_{C|\boldsymbol{V}}(T). \tag{13}$$

Cox proportional hazard model is adopted here.

DWE

Theorem 1

Assume the following regularity conditions.

Condition 1: $\boldsymbol{V}$ is time-independent and bounded.

Condition 2: $\boldsymbol{\Omega}^* = \int_0^\infty [\boldsymbol{s}_1^{(2)}(\boldsymbol{\theta}, t) - \boldsymbol{s}_1^{(1)}(\boldsymbol{\theta}, t)\{\boldsymbol{s}_1^{(1)}(\boldsymbol{\theta}, t)\}^\top / s_1^{(0)}(\boldsymbol{\theta}, t)]\lambda_0(t)dt$ is a positive definite matrix, where
$s_1^{(0)}(\boldsymbol{\theta}, t) = E[I_{\{T \geq t\}} \exp(\boldsymbol{\theta}^\top \boldsymbol{V})]$,
$\boldsymbol{s}_1^{(1)}(\boldsymbol{\theta}, t) = E[I_{\{T \geq t\}} \boldsymbol{V} \exp(\boldsymbol{\theta}^\top \boldsymbol{V})]$,
$\boldsymbol{s}_1^{(2)}(\boldsymbol{\theta}, t) = E[I_{\{T \geq t\}} \boldsymbol{V}\boldsymbol{V}^\top \exp(\boldsymbol{\theta}^\top \boldsymbol{V})]$.

Condition 3: Let L denote the kernel function for continuous components in K_h and define $L_{h_c}(\cdot) = L(\cdot/h_c)$. L is bounded, symmetric, Lipschitz continuous with $\int L(u)du = 1$, $\int u^m L(u)du = 0$ for $m = 1, \ldots, (r-1)$, and $\int u^r L(u)du \neq 0$.

Condition 4: $nh_c^{2r} \to 0$ and $nh_c^{2d} \to \infty$ as $n \to \infty$, where d denotes the dimension of continuous components of $(T, \boldsymbol{Z})$.

Condition 5: $\pi_1(T, \boldsymbol{Z})$ has r continuous and bounded partial derivatives with respect to the continuous components of $(T, \boldsymbol{Z})$ and $\pi_1(T, \boldsymbol{Z}) \geq \epsilon_0$ for any T and $\boldsymbol{V}$, where $\epsilon_0 > 0$ is a constant.

Theorem 1 (Cont.)

Then, as $n \to \infty$,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_1),$$

where $\xrightarrow{d}$ denotes convergence in distribution,

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Omega}^{*-1}\mathrm{var}\left[\frac{R\delta}{\pi_1(T,\boldsymbol{Z})\psi(T,\boldsymbol{V})}\boldsymbol{M}_{\boldsymbol{V}}^* + \left\{1 - \frac{R}{\pi_1(T,\boldsymbol{Z})}\right\}E(\boldsymbol{M}_{\boldsymbol{V}}^* \mid T, \boldsymbol{Z}) + \boldsymbol{\Gamma}(R\boldsymbol{M}_{\boldsymbol{V}}^{\delta})\right]\boldsymbol{\Omega}^{*-1},$$

$\boldsymbol{\Omega}^*$ is defined in Condition 2, $\boldsymbol{M}_{\boldsymbol{V}}^* = \int_0^\infty [\boldsymbol{V} - \boldsymbol{s}_1^{(1)}(\boldsymbol{\theta}, t)\{s_1^{(0)}(\boldsymbol{\theta}, t)\}^{-1}]dM^*(t)$ and $\boldsymbol{M}_{\boldsymbol{V}}^{\delta} = \int_0^\infty [\boldsymbol{V} - \boldsymbol{s}_\delta^{(1)}(\boldsymbol{\beta}, t)\{s_\delta^{(0)}(\boldsymbol{\beta}, t)\}^{-1}]dM^{\delta}(t)$ are both mean-zero martingale transformations with respect to martingale $M^*(t)$ and $M^{\delta}(t) = N^{\delta}(t) - \int_0^t Y(u)\lambda_{0C}(u)\exp(\boldsymbol{\beta}^\top \boldsymbol{V})du$ with $s_\delta^{(0)}(\boldsymbol{\beta}, t) = E\{RY(t)\exp(\boldsymbol{\beta}^\top \boldsymbol{V})\}$ and $\boldsymbol{s}_\delta^{(1)}(\boldsymbol{\beta}, t) = E\{RY(t)\boldsymbol{V}\exp(\boldsymbol{\beta}^\top \boldsymbol{V})\}$, respectively, and $\boldsymbol{\Gamma} = E\{\Lambda_{0C}(T)\exp(\boldsymbol{\beta}^\top \boldsymbol{V})\boldsymbol{M}_{\boldsymbol{V}}^* \boldsymbol{V}^\top\}\{\mathrm{var}(R\boldsymbol{M}_{\boldsymbol{V}}^{\delta})\}^{-1}$.

Method

In the construction of $\widehat{\boldsymbol{\theta}}_1$, we utilize some incomplete data in the estimator of π_1, but only data with $R\delta = 1$ is directly used by the doubly weighted score function (10), which is derived from the martingale $M^*(t)$ in (9) that only involves **non-censored** subjects.

- Can we make use of incomplete data more directly in the estimation of $\boldsymbol{\theta}$?
- Especially, can we derive a weighted score function from

$$M(t) = N(t) - \int_0^t I_{\{T \wedge C \geq t\}} \lambda_0(u) \exp(\boldsymbol{\theta}^\top \boldsymbol{V}) du, \qquad (14)$$

the martingale for Cox model with no missing data?

Method

In the construction of $\widehat{\boldsymbol{\theta}}_1$, we utilize some incomplete data in the estimator of π_1, but only data with $R\delta = 1$ is directly used by the doubly weighted score function (10), which is derived from the martingale $M^*(t)$ in (9) that only involves **non-censored** subjects.

- Can we make use of incomplete data more directly in the estimation of $\boldsymbol{\theta}$?
- Especially, can we derive a weighted score function from

$$M(t) = N(t) - \int_0^t I_{\{T \wedge C \geq t\}} \lambda_0(u) \exp(\boldsymbol{\theta}^\top \boldsymbol{V}) du, \qquad (14)$$

the martingale for Cox model with no missing data?

Compositely Weighted Estimation Equation:

$$\sum_{i=1}^{n} \int_0^{\tau} \frac{R_i}{\pi_1(T_i, \boldsymbol{Z}_i)} \left\{ \boldsymbol{V}_i - \frac{\boldsymbol{S}_2^{(1)}(\boldsymbol{\theta}, t)}{S_2^{(0)}(\boldsymbol{\theta}, t)} \right\} dN_i(t) = 0 \qquad (15)$$

where

$$\boldsymbol{S}_2^{(1)}(\boldsymbol{\theta}, t) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{R_i \delta_i I_{\{T_i \geq t\}}}{\pi_1(T_i, \boldsymbol{Z}_i)} + \frac{R_i(1-\delta_i) I_{\{C_i \geq t\}}}{\pi_0(C_i, \boldsymbol{V}_i)} \right\} \boldsymbol{V}_i \exp(\boldsymbol{\theta}^\top \boldsymbol{V}_i),$$

$$S_2^{(0)}(\boldsymbol{\theta}, t) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{R_i \delta_i I_{\{T_i \geq t\}}}{\pi_1(T_i, \boldsymbol{Z}_i)} + \frac{R_i(1-\delta_i) I_{\{C_i \geq t\}}}{\pi_0(C_i, \boldsymbol{V}_i)} \right\} \exp(\boldsymbol{\theta}^\top \boldsymbol{V}_i)$$

and

$$\pi_0(C_i, \boldsymbol{V}_i) = \mathrm{pr}(R = 1 | C, \boldsymbol{V}, \delta = 0).$$

Compositely Weighted Estimator (CWE):

Estimator $\widehat{\boldsymbol{\theta}}_2$ of $\boldsymbol{\theta}$ can be obtained by solving (10).

Once $\boldsymbol{\theta}$ is estimated by $\widehat{\boldsymbol{\theta}}_2$, a semi-parametric estimator of $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ can be obtained by

$$\widehat{\Lambda}_{02}(t) = \int_0^\infty \frac{1}{nS_2^{(0)}(\widehat{\boldsymbol{\theta}}_2, u)} \sum_{i=1}^n \frac{R_i dN_i(u)}{\pi_1(T_i, \boldsymbol{Z}_i)}. \tag{16}$$

Estimation of $\pi_0(C, \boldsymbol{V})$:

- Note that

$$\pi_0(C, \boldsymbol{V}) = \mathrm{pr}(R = 1 | C, \boldsymbol{V}, \delta = 0) = \frac{\mathrm{pr}(R = 1, \delta = 0 | C, \boldsymbol{V})}{\mathrm{pr}(\delta = 0 | C, \boldsymbol{V})}. \tag{17}$$

- Under assumption (2),

$$\mathrm{pr}(\delta = 0 | C, \boldsymbol{V}) = S_{T|\boldsymbol{V}}(C), \tag{18}$$

where $S_{T|\boldsymbol{V}}(\cdot)$ denotes the survival function of T given covariates $\boldsymbol{V}$.

Estimation of $\pi_0(C, \boldsymbol{V})$:

- Assumptions (5) implies that

$$R \perp C \mid (T, \boldsymbol{V}),$$

which further implies that

$$\mathrm{pr}(R = 1, \delta = 0 | C, \boldsymbol{V}) = -\int_C^\infty \pi_1(t, \boldsymbol{Z}) dS_{T|\boldsymbol{V}}(t). \qquad (19)$$

proof:

$$\begin{aligned}
\mathrm{pr}(R = 1, \delta = 0 | C, \boldsymbol{V}) =& E \{\mathrm{pr}(R = 1, T > C | C, \boldsymbol{V}, T) | C, \boldsymbol{V}\} \\
=& E \left\{ I_{\{T > C\}} \mathrm{pr}(R = 1 | C, \boldsymbol{V}, T) | C, \boldsymbol{V} \right\} \\
=& E \left\{ I_{\{T > C\}} \pi_1(T, \boldsymbol{Z}) | C, \boldsymbol{V} \right\} \\
=& -\int_C^\infty \pi_1(t, \boldsymbol{Z}) dS_{T|\boldsymbol{V}}(t)
\end{aligned}$$

Estimation of $\pi_0(C, \boldsymbol{V})$:

One way to estimate $S_{T|\boldsymbol{V}}(t)$ is using the DWE estimators $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\Lambda}_{01}(t)$. Then, it follows from (17)-(19) that we can estimate $\pi_0(C, \boldsymbol{V})$ by

$$\widehat{\pi}_0(C, \boldsymbol{V}) = \frac{-\int_C^\infty \widehat{\pi}_1(t, \boldsymbol{Z}) d\widehat{S}_{T|\boldsymbol{V}}(t)}{\widehat{S}_{T|\boldsymbol{V}}(C)}, \tag{20}$$

where

$$\widehat{S}_{T|\boldsymbol{V}}(t) = \exp\left\{-\exp(\widehat{\boldsymbol{\theta}}_1^\top \boldsymbol{V})\widehat{\Lambda}_{01}(t)\right\}.$$

Theorem 2

Assume that *Conditions 1-5* stated in Theorem 1 hold with $\boldsymbol{\Omega}^*$ in *Condition 2* replaced by
$\boldsymbol{\Omega} = \int_0^\infty [\boldsymbol{s}^{(2)}(\boldsymbol{\theta}, t) - \boldsymbol{s}^{(1)}(\boldsymbol{\theta}, t)\{\boldsymbol{s}^{(1)}(\boldsymbol{\theta}, t)\}^\top \{s^{(0)}(\boldsymbol{\theta}, t)\}^{-1}]\lambda_0(t)dt$, where
$s^{(0)}(\boldsymbol{\theta}, t) = E\{Y(t)\exp(\boldsymbol{\theta}^\top \boldsymbol{V})\}$, $\boldsymbol{s}^{(1)}(\boldsymbol{\theta}, t) = E[\{Y(t)\boldsymbol{V}\exp(\boldsymbol{\theta}^\top \boldsymbol{V})\}$, and
$\boldsymbol{s}^{(2)}(\boldsymbol{\theta}, t) = E\{Y(t)\boldsymbol{V}\boldsymbol{V}^\top \exp(\boldsymbol{\theta}^\top \boldsymbol{V})\}$.
Then, as $n \to \infty$,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_2),$$

where

$$\boldsymbol{\Sigma}_2 = \boldsymbol{\Omega}^{-1} + \boldsymbol{\Omega}^{-1} E\left[\left(\{\pi_1(T, \boldsymbol{Z})\}^{-1} - 1\right) \mathrm{var}(\boldsymbol{M}_{\boldsymbol{V}} \mid T, \boldsymbol{Z})\right] \boldsymbol{\Omega}^{-1},$$

and $\boldsymbol{M}_{\boldsymbol{V}} = \int_0^\infty [\boldsymbol{V} - \boldsymbol{s}^{(1)}(\boldsymbol{\theta}, t)\{s^{(0)}(\boldsymbol{\theta}, t)\}^{-1}]dM(t)$ is the martingale transformation with respect to mean-zero martingale $M(t)$.

Simulation

- **FULL**: Estimators computed without missing covariate values
- **CC**: Complete-case analysis
- **SWE**: Simple weighted estimator in Qi *et al.* (2005) valid under the MAR assumption (3)
- **DWE**: Doubly weighted estimator
- **CWE**: Compositely weighted estimator

Setting and Result 1

Table 1: Bias, SD, SE, and CP for Setting 1 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim \mathrm{N}(0,1)$, $Z \sim \text{binary}(0.5)$, $X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T - 0.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp(-T^{1/2})$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.002	0.075	0.079	0.945	0.000	0.131	0.135	0.946
	CC		0.151	0.115	0.122	0.663	0.149	0.195	0.204	0.864
	SWE	$T \wedge C$	0.092	0.113	0.118	0.845	0.087	0.189	0.196	0.925
	DWE	T	-0.015	0.131	0.134	0.950	-0.015	0.224	0.224	0.946
	CWE	T	0.018	0.114	0.120	0.947	0.009	0.185	0.192	0.947
	DWE	T, Z	-0.013	0.131	0.134	0.953	-0.005	0.219	0.222	0.953
	CWE	T, Z	0.020	0.115	0.120	0.942	0.016	0.177	0.188	0.951
1000	Full		0.001	0.055	0.055	0.939	-0.001	0.094	0.094	0.947
	CC		0.146	0.082	0.083	0.479	0.145	0.138	0.140	0.774
	SWE	$T \wedge C$	0.087	0.080	0.080	0.760	0.084	0.134	0.134	0.883
	DWE	T	-0.013	0.092	0.091	0.945	-0.011	0.158	0.155	0.943
	CWE	T	0.011	0.082	0.082	0.944	0.005	0.130	0.131	0.951

Missing Rate: 44.8%; Censoring Rate: 47.6%

Setting and Result 1

Table 1: Bias, SD, SE, and CP for Setting 1 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z),\ \ X \sim \mathrm{N}(0,1),\ \ Z \sim \text{binary}(0.5),\ \ X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z),\ \ \boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T - 0.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp(-T^{1/2})$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.002	0.075	0.079	0.945	0.000	0.131	0.135	0.946
	CC		0.151	0.115	0.122	0.663	0.149	0.195	0.204	0.864
	SWE	$T \wedge C$	0.092	0.113	0.118	0.845	0.087	0.189	0.196	0.925
	DWE	T	-0.015	0.131	0.134	0.950	-0.015	0.224	0.224	0.946
	CWE	T	0.018	0.114	0.120	0.947	0.009	0.185	0.192	0.947
	DWE	T, Z	-0.013	0.131	0.134	0.953	-0.005	0.219	0.222	0.953
	CWE	T, Z	0.020	0.115	0.120	0.942	0.016	0.177	0.188	0.951
1000	Full		0.001	0.055	0.055	0.939	-0.001	0.094	0.094	0.947
	CC		0.146	0.082	0.083	0.479	0.145	0.138	0.140	0.774
	SWE	$T \wedge C$	0.087	0.080	0.080	0.760	0.084	0.134	0.134	0.883
	DWE	T	-0.013	0.092	0.091	0.945	-0.011	0.158	0.155	0.943
	CWE	T	0.011	0.082	0.082	0.944	0.005	0.130	0.131	0.951

Missing Rate: 44.8%; Censoring Rate: 47.6%

Setting and Result 1

Table 1: Bias, SD, SE, and CP for Setting 1 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z),\ X \sim \mathrm{N}(0,1),\ Z \sim \text{binary}(0.5),\ X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z),\ \boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T - 0.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp(-T^{1/2})$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.002	0.075	0.079	0.945	0.000	0.131	0.135	0.946
	CC		0.151	0.115	0.122	0.663	0.149	0.195	0.204	0.864
	SWE	$T \wedge C$	0.092	0.113	0.118	0.845	0.087	0.189	0.196	0.925
	DWE	T	-0.015	0.131	0.134	0.950	-0.015	0.224	0.224	0.946
	CWE	T	0.018	0.114	0.120	0.947	0.009	0.185	0.192	0.947
	DWE	T, Z	-0.013	0.131	0.134	0.953	-0.005	0.219	0.222	0.953
	CWE	T, Z	0.020	0.115	0.120	0.942	0.016	0.177	0.188	0.951
1000	Full		0.001	0.055	0.055	0.939	-0.001	0.094	0.094	0.947
	CC		0.146	0.082	0.083	0.479	0.145	0.138	0.140	0.774
	SWE	$T \wedge C$	0.087	0.080	0.080	0.760	0.084	0.134	0.134	0.883
	DWE	T	-0.013	0.092	0.091	0.945	-0.011	0.158	0.155	0.943
	CWE	T	0.011	0.082	0.082	0.944	0.005	0.130	0.131	0.951

Missing Rate: 44.8%; Censoring Rate: 47.6%

Setting and Result 1

Table 1: Bias, SD, SE, and CP for Setting 1 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim \mathrm{N}(0,1)$, $Z \sim \text{binary}(0.5)$, $X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T - 0.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp(-T^{1/2})$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.002	0.075	0.079	0.945	0.000	0.131	0.135	0.946
	CC		0.151	0.115	0.122	0.663	0.149	0.195	0.204	0.864
	SWE	$T \wedge C$	0.092	0.113	0.118	0.845	0.087	0.189	0.196	0.925
	DWE	T	-0.015	0.131	0.134	0.950	-0.015	0.224	0.224	0.946
	CWE	T	0.018	0.114	0.120	0.947	0.009	0.185	0.192	0.947
	DWE	T, Z	-0.013	0.131	0.134	0.953	-0.005	0.219	0.222	0.953
	CWE	T, Z	0.020	0.115	0.120	0.942	0.016	0.177	0.188	0.951
1000	Full		0.001	0.055	0.055	0.939	-0.001	0.094	0.094	0.947
	CC		0.146	0.082	0.083	0.479	0.145	0.138	0.140	0.774
	SWE	$T \wedge C$	0.087	0.080	0.080	0.760	0.084	0.134	0.134	0.883
	DWE	T	-0.013	0.092	0.091	0.945	-0.011	0.158	0.155	0.943
	CWE	T	0.011	0.082	0.082	0.944	0.005	0.130	0.131	0.951

Missing Rate: 44.8%; Censoring Rate: 47.6%

Setting and Result 1

Table 1: Bias, SD, SE, and CP for Setting 1 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim \mathrm{N}(0,1)$, $Z \sim$ binary(0.5), $X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T - 0.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp(-T^{1/2})$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.002	0.075	0.079	0.945	0.000	0.131	0.135	0.946
	CC		0.151	0.115	0.122	0.663	0.149	0.195	0.204	0.864
	SWE	$T \wedge C$	0.092	0.113	0.118	0.845	0.087	0.189	0.196	0.925
	DWE	T	-0.015	0.131	0.134	0.950	-0.015	0.224	0.224	0.946
	CWE	T	0.018	0.114	0.120	0.947	0.009	0.185	0.192	0.947
	DWE	T, Z	-0.013	0.131	0.134	0.953	-0.005	0.219	0.222	0.953
	CWE	T, Z	0.020	0.115	0.120	0.942	0.016	0.177	0.188	0.951
1000	Full		0.001	0.055	0.055	0.939	-0.001	0.094	0.094	0.947
	CC		0.146	0.082	0.083	0.479	0.145	0.138	0.140	0.774
	SWE	$T \wedge C$	0.087	0.080	0.080	0.760	0.084	0.134	0.134	0.883
	DWE	T	-0.013	0.092	0.091	0.945	-0.011	0.158	0.155	0.943
	CWE	T	0.011	0.082	0.082	0.944	0.005	0.130	0.131	0.951

Missing Rate: 44.8%; Censoring Rate: 47.6%

Setting and Result 1

Table 1: Bias, SD, SE, and CP for Setting 1 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim \mathrm{N}(0,1)$, $Z \sim \text{binary}(0.5)$, $X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T - 0.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp(-T^{1/2})$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.002	0.075	0.079	0.945	0.000	0.131	0.135	0.946
	CC		0.151	0.115	0.122	0.663	0.149	0.195	0.204	0.864
	SWE	$T \wedge C$	0.092	0.113	0.118	0.845	0.087	0.189	0.196	0.925
	DWE	T	-0.015	0.131	0.134	0.950	-0.015	0.224	0.224	0.946
	CWE	T	0.018	0.114	0.120	0.947	0.009	0.185	0.192	0.947
	DWE	T, Z	-0.013	0.131	0.134	0.953	-0.005	0.219	0.222	0.953
	CWE	T, Z	0.020	0.115	0.120	0.942	0.016	0.177	0.188	0.951
1000	Full		0.001	0.055	0.055	0.939	-0.001	0.094	0.094	0.947
	CC		0.146	0.082	0.083	0.479	0.145	0.138	0.140	0.774
	SWE	$T \wedge C$	0.087	0.080	0.080	0.760	0.084	0.134	0.134	0.883
	DWE	T	-0.013	0.092	0.091	0.945	-0.011	0.158	0.155	0.943
	CWE	T	0.011	0.082	0.082	0.944	0.005	0.130	0.131	0.951

Missing Rate: 44.8%; Censoring Rate: 47.6%

Setting and Result 2

Table 2: Bias, SD, SE, and CP for setting 2 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim \text{binary}(0.5)$, $Z \sim \text{binary}(0.5)$, $X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T + Z)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp\{-T^{1/2}\exp(-X/4)\}$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.005	0.122	0.124	0.946	0.004	0.120	0.124	0.945
	CC		0.082	0.150	0.151	0.896	0.155	0.153	0.157	0.804
	SWE	$T \wedge C, Z$	0.055	0.147	0.148	0.917	0.101	0.143	0.149	0.880
	DWE	T, Z	0.004	0.165	0.166	0.946	0.028	0.157	0.161	0.946
	CWE	T, Z	0.024	0.151	0.150	0.938	0.043	0.141	0.144	0.940
	DWE	T	0.024	0.170	0.167	0.928	0.094	0.176	0.176	0.892
	CWE	T	0.039	0.150	0.149	0.925	0.108	0.151	0.155	0.878
1000	Full		0.000	0.086	0.086	0.941	0.000	0.087	0.086	0.944
	CC		0.075	0.103	0.105	0.874	0.151	0.108	0.109	0.674
	SWE	$T \wedge C, Z$	0.045	0.099	0.102	0.927	0.091	0.101	0.102	0.833
	DWE	T, Z	-0.002	0.115	0.114	0.942	0.014	0.113	0.110	0.941
	CWE	T, Z	0.010	0.106	0.105	0.948	0.027	0.106	0.099	0.944

Missing Rate: 29.4%; Censoring Rate: 37.0%

Setting and Result 2

Table 2: Bias, SD, SE, and CP for setting 2 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim \text{binary}(0.5)$, $Z \sim \text{binary}(0.5)$, $X \perp Z$
True hazard of T:	$\lambda(t \mid \boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T + Z)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\mid \boldsymbol{V}}(T) = \exp\{-T^{1/2}\exp(-X/4)\}$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.005	0.122	0.124	0.946	0.004	0.120	0.124	0.945
	CC		0.082	0.150	0.151	0.896	0.155	0.153	0.157	0.804
	SWE	$T \wedge C, Z$	0.055	0.147	0.148	0.917	0.101	0.143	0.149	0.880
	DWE	T, Z	0.004	0.165	0.166	0.946	0.028	0.157	0.161	0.946
	CWE	T, Z	0.024	0.151	0.150	0.938	0.043	0.141	0.144	0.940
	DWE	T	0.024	0.170	0.167	0.928	0.094	0.176	0.176	0.892
	CWE	T	0.039	0.150	0.149	0.925	0.108	0.151	0.155	0.878
1000	Full		0.000	0.086	0.086	0.941	0.000	0.087	0.086	0.944
	CC		0.075	0.103	0.105	0.874	0.151	0.108	0.109	0.674
	SWE	$T \wedge C, Z$	0.045	0.099	0.102	0.927	0.091	0.101	0.102	0.833
	DWE	T, Z	-0.002	0.115	0.114	0.942	0.014	0.113	0.110	0.941
	CWE	T, Z	0.010	0.106	0.105	0.948	0.027	0.106	0.099	0.944

Missing Rate: 29.4%; Censoring Rate: 37.0%

Setting and Result 2

Table 2: Bias, SD, SE, and CP for setting 2 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim \text{binary}(0.5)$, $Z \sim \text{binary}(0.5)$, $X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T + Z)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp\{-T^{1/2}\exp(-X/4)\}$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.005	0.122	0.124	0.946	0.004	0.120	0.124	0.945
	CC		0.082	0.150	0.151	0.896	0.155	0.153	0.157	0.804
	SWE	$T \wedge C, Z$	0.055	0.147	0.148	0.917	0.101	0.143	0.149	0.880
	DWE	T, Z	0.004	0.165	0.166	0.946	0.028	0.157	0.161	0.946
	CWE	T, Z	0.024	0.151	0.150	0.938	0.043	0.141	0.144	0.940
	DWE	T	0.024	0.170	0.167	0.928	0.094	0.176	0.176	0.892
	CWE	T	0.039	0.150	0.149	0.925	0.108	0.151	0.155	0.878
1000	Full		0.000	0.086	0.086	0.941	0.000	0.087	0.086	0.944
	CC		0.075	0.103	0.105	0.874	0.151	0.108	0.109	0.674
	SWE	$T \wedge C, Z$	0.045	0.099	0.102	0.927	0.091	0.101	0.102	0.833
	DWE	T, Z	-0.002	0.115	0.114	0.942	0.014	0.113	0.110	0.941
	CWE	T, Z	0.010	0.106	0.105	0.948	0.027	0.106	0.099	0.944

Missing Rate: 29.4%; Censoring Rate: 37.0%

Setting and Result 2

Table 2: Bias, SD, SE, and CP for setting 2 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim \text{binary}(0.5)$, $Z \sim \text{binary}(0.5)$, $X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T + Z)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp\{-T^{1/2}\exp(-X/4)\}$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.005	0.122	0.124	0.946	0.004	0.120	0.124	0.945
	CC		0.082	0.150	0.151	0.896	0.155	0.153	0.157	0.804
	SWE	$T \wedge C, Z$	0.055	0.147	0.148	0.917	0.101	0.143	0.149	0.880
	DWE	T, Z	0.004	0.165	0.166	0.946	0.028	0.157	0.161	0.946
	CWE	T, Z	0.024	0.151	0.150	0.938	0.043	0.141	0.144	0.940
	DWE	T	0.024	0.170	0.167	0.928	0.094	0.176	0.176	0.892
	CWE	T	0.039	0.150	0.149	0.925	0.108	0.151	0.155	0.878
1000	Full		0.000	0.086	0.086	0.941	0.000	0.087	0.086	0.944
	CC		0.075	0.103	0.105	0.874	0.151	0.108	0.109	0.674
	SWE	$T \wedge C, Z$	0.045	0.099	0.102	0.927	0.091	0.101	0.102	0.833
	DWE	T, Z	-0.002	0.115	0.114	0.942	0.014	0.113	0.110	0.941
	CWE	T, Z	0.010	0.106	0.105	0.948	0.027	0.106	0.099	0.944

Missing Rate: 29.4%; Censoring Rate: 37.0%

Setting and Result 2

Table 2: Bias, SD, SE, and CP for setting 2 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim$ binary(0.5), $Z \sim$ binary(0.5), $X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T + Z)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp\{-T^{1/2}\exp(-X/4)\}$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.005	0.122	0.124	0.946	0.004	0.120	0.124	0.945
	CC		0.082	0.150	0.151	0.896	0.155	0.153	0.157	0.804
	SWE	$T \wedge C, Z$	0.055	0.147	0.148	0.917	0.101	0.143	0.149	0.880
	DWE	T, Z	0.004	0.165	0.166	0.946	0.028	0.157	0.161	0.946
	CWE	T, Z	0.024	0.151	0.150	0.938	0.043	0.141	0.144	0.940
	DWE	T	0.024	0.170	0.167	0.928	0.094	0.176	0.176	0.892
	CWE	T	0.039	0.150	0.149	0.925	0.108	0.151	0.155	0.878
1000	Full		0.000	0.086	0.086	0.941	0.000	0.087	0.086	0.944
	CC		0.075	0.103	0.105	0.874	0.151	0.108	0.109	0.674
	SWE	$T \wedge C, Z$	0.045	0.099	0.102	0.927	0.091	0.101	0.102	0.833
	DWE	T, Z	-0.002	0.115	0.114	0.942	0.014	0.113	0.110	0.941
	CWE	T, Z	0.010	0.106	0.105	0.948	0.027	0.106	0.099	0.944

Missing Rate: 29.4%; Censoring Rate: 37.0%

Setting and Result 2

Table 2: Bias, SD, SE, and CP for setting 2 based on 2000 simulation runs

Covariate Vector:	$\boldsymbol{V} = (X, Z)$, $X \sim$ binary(0.5), $Z \sim$ binary(0.5), $X \perp Z$
True hazard of T:	$\lambda(t\|\boldsymbol{V}) = \exp(\theta_x X + \theta_z Z)$, $\boldsymbol{\theta} = (\theta_x, \theta_z) = (1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{Z}) = 1 - \{1 + \exp(T + Z)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\|\boldsymbol{V}}(T) = \exp\{-T^{1/2}\exp(-X/4)\}$
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$

			Estimation of θ_x				Estimation of θ_z			
n	Method	Variables used in $\widehat{\pi}_1$	Bias	SD	SE	CP	Bias	SD	SE	CP
500	Full		0.005	0.122	0.124	0.946	0.004	0.120	0.124	0.945
	CC		0.082	0.150	0.151	0.896	0.155	0.153	0.157	0.804
	SWE	$T \wedge C, Z$	0.055	0.147	0.148	0.917	0.101	0.143	0.149	0.880
	DWE	T, Z	0.004	0.165	0.166	0.946	0.028	0.157	0.161	0.946
	CWE	T, Z	0.024	0.151	0.150	0.938	0.043	0.141	0.144	0.940
	DWE	T	0.024	0.170	0.167	0.928	0.094	0.176	0.176	0.892
	CWE	T	0.039	0.150	0.149	0.925	0.108	0.151	0.155	0.878
1000	Full		0.000	0.086	0.086	0.941	0.000	0.087	0.086	0.944
	CC		0.075	0.103	0.105	0.874	0.151	0.108	0.109	0.674
	SWE	$T \wedge C, Z$	0.045	0.099	0.102	0.927	0.091	0.101	0.102	0.833
	DWE	T, Z	-0.002	0.115	0.114	0.942	0.014	0.113	0.110	0.941
	CWE	T, Z	0.010	0.106	0.105	0.948	0.027	0.106	0.099	0.944

Missing Rate: 29.4%; Censoring Rate: 37.0%

Setting and Result 3

Table 3: Bias, SD, SE, and CP for Setting 3 based on 2000 simulation runs

Covariate vector:	$\boldsymbol{V} = (X, Z_1, Z_2)$, $X \mid Z_1, Z_2 \sim \text{logit}(0.5Z_1 - Z_2)$, $Z_1 \sim$ binary(0.5), $Z_2 \sim$ uniform(0, 0.5), $Z_1 \perp Z_2$
True hazard of T:	$\lambda(t \mid \boldsymbol{V}) = 0.5t \exp(\theta_x X + \theta_{z1} Z_1 + \theta_{z2} Z_2)$, $\boldsymbol{\theta} = (\theta_x, \theta_{z1}, \theta_{z2}) = (1, 1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{V}) = 1 - \{1 + \exp(2T - Z_2 - 1.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\mid\boldsymbol{V}}(T) = \exp\{-(\alpha T)^{1.2} \exp(0.2X + 0.1Z_1)\}$ α determines the censoring proportion
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$
Variables used in $\widehat{\pi}_1$:	$(T \wedge C, Z_1, Z_2)$ for SWE and (T, Z_1, Z_2) for DWE and CWE

$n = 1000$		Estimation of θ_x				Estimation of θ_{z1}				Estimation of θ_{z2}			
α	Method	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
1	Full	0.004	0.111	0.110	0.944	0.004	0.109	0.110	0.945	0.020	0.350	0.357	0.948
	CC	0.207	0.179	0.182	0.743	0.206	0.173	0.180	0.754	-0.002	0.587	0.594	0.948
	SWE	0.147	0.175	0.176	0.852	0.221	0.164	0.169	0.711	0.022	0.552	0.569	0.956
	DWE	-0.025	0.234	0.227	0.939	-0.022	0.229	0.221	0.939	-0.135	0.794	0.746	0.938
	CWE	0.018	0.170	0.176	0.946	0.025	0.170	0.170	0.950	-0.091	0.498	0.528	0.954
2	Full	0.004	0.161	0.163	0.950	0.010	0.162	0.164	0.944	0.007	0.520	0.526	0.947
	CC	0.261	0.297	0.312	0.836	0.265	0.292	0.311	0.826	-0.022	0.969	0.997	0.954
	SWE	0.229	0.299	0.306	0.862	0.290	0.285	0.300	0.806	0.013	0.964	0.972	0.947
	DWE	-0.015	0.622	0.548	0.928	-0.106	0.627	0.554	0.929	-0.427	1.829	1.635	0.952
	CWE	0.099	0.377	0.335	0.926	0.123	0.358	0.336	0.924	-0.105	1.033	0.999	0.958

Missing Rate: 46.9%; Censoring Rate: 60.2% ($\alpha = 1$), 81.5% ($\alpha = 2$)

Setting and Result 3

Table 3: Bias, SD, SE, and CP for Setting 3 based on 2000 simulation runs

Covariate vector:	$\boldsymbol{V} = (X, Z_1, Z_2)$, $X \mid Z_1, Z_2 \sim \mathrm{logit}(0.5Z_1 - Z_2)$, $Z_1 \sim$ binary(0.5), $Z_2 \sim$ uniform(0, 0.5), $Z_1 \perp Z_2$
True hazard of T:	$\lambda(t \mid \boldsymbol{V}) = 0.5t\exp(\theta_x X + \theta_{z1}Z_1 + \theta_{z2}Z_2)$, $\boldsymbol{\theta} = (\theta_x, \theta_{z1}, \theta_{z2}) = (1, 1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{V}) = 1 - \{1 + \exp(2T - Z_2 - 1.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\mid\boldsymbol{V}}(T) = \exp\{-(\alpha T)^{1.2}\exp(0.2X + 0.1Z_1)\}$ α determines the censoring proportion
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$
Variables used in $\widehat{\pi}_1$:	$(T \wedge C, Z_1, Z_2)$ for SWE and (T, Z_1, Z_2) for DWE and CWE

$n = 1000$		Estimation of θ_x				Estimation of θ_{z1}				Estimation of θ_{z2}			
α	Method	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
1	Full	0.004	0.111	0.110	0.944	0.004	0.109	0.110	0.945	0.020	0.350	0.357	0.948
	CC	0.207	0.179	0.182	0.743	0.206	0.173	0.180	0.754	-0.002	0.587	0.594	0.948
	SWE	0.147	0.175	0.176	0.852	0.221	0.164	0.169	0.711	0.022	0.552	0.569	0.956
	DWE	-0.025	0.234	0.227	0.939	-0.022	0.229	0.221	0.939	-0.135	0.794	0.746	0.938
	CWE	0.018	0.170	0.176	0.946	0.025	0.170	0.170	0.950	-0.091	0.498	0.528	0.954
2	Full	0.004	0.161	0.163	0.950	0.010	0.162	0.164	0.944	0.007	0.520	0.526	0.947
	CC	0.261	0.297	0.312	0.836	0.265	0.292	0.311	0.826	-0.022	0.969	0.997	0.954
	SWE	0.229	0.299	0.306	0.862	0.290	0.285	0.300	0.806	0.013	0.964	0.972	0.947
	DWE	-0.015	0.622	0.548	0.928	-0.106	0.627	0.554	0.929	-0.427	1.829	1.635	0.952
	CWE	0.099	0.377	0.335	0.926	0.123	0.358	0.336	0.924	-0.105	1.033	0.999	0.958

Missing Rate: 46.9%; Censoring Rate: 60.2% ($\alpha = 1$), 81.5% ($\alpha = 2$)

Setting and Result 3

Table 3: Bias, SD, SE, and CP for Setting 3 based on 2000 simulation runs

Covariate vector:	$\boldsymbol{V} = (X, Z_1, Z_2)$, $X \mid Z_1, Z_2 \sim \text{logit}(0.5Z_1 - Z_2)$, $Z_1 \sim$ binary(0.5), $Z_2 \sim$ uniform(0, 0.5), $Z_1 \perp Z_2$
True hazard of T:	$\lambda(t \mid \boldsymbol{V}) = 0.5t \exp(\theta_x X + \theta_{z1} Z_1 + \theta_{z2} Z_2)$, $\boldsymbol{\theta} = (\theta_x, \theta_{z1}, \theta_{z2}) = (1, 1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{V}) = 1 - \{1 + \exp(2T - Z_2 - 1.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C \mid \boldsymbol{V}}(T) = \exp\{-(\alpha T)^{1.2} \exp(0.2X + 0.1Z_1)\}$ α determines the censoring proportion
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$
Variables used in $\widehat{\pi}_1$:	$(T \wedge C, Z_1, Z_2)$ for SWE and (T, Z_1, Z_2) for DWE and CWE

$n = 1000$		Estimation of θ_x				Estimation of θ_{z1}				Estimation of θ_{z2}			
α	Method	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
1	Full	0.004	0.111	0.110	0.944	0.004	0.109	0.110	0.945	0.020	0.350	0.357	0.948
	CC	0.207	0.179	0.182	0.743	0.206	0.173	0.180	0.754	-0.002	0.587	0.594	0.948
	SWE	0.147	0.175	0.176	0.852	0.221	0.164	0.169	0.711	0.022	0.552	0.569	0.956
	DWE	-0.025	0.234	0.227	0.939	-0.022	0.229	0.221	0.939	-0.135	0.794	0.746	0.938
	CWE	0.018	0.170	0.176	0.946	0.025	0.170	0.170	0.950	-0.091	0.498	0.528	0.954
2	Full	0.004	0.161	0.163	0.950	0.010	0.162	0.164	0.944	0.007	0.520	0.526	0.947
	CC	0.261	0.297	0.312	0.836	0.265	0.292	0.311	0.826	-0.022	0.969	0.997	0.954
	SWE	0.229	0.299	0.306	0.862	0.290	0.285	0.300	0.806	0.013	0.964	0.972	0.947
	DWE	-0.015	0.622	0.548	0.928	-0.106	0.627	0.554	0.929	-0.427	1.829	1.635	0.952
	CWE	0.099	0.377	0.335	0.926	0.123	0.358	0.336	0.924	-0.105	1.033	0.999	0.958

Missing Rate: 46.9%; Censoring Rate: 60.2% ($\alpha = 1$), 81.5% ($\alpha = 2$)

Setting and Result 3

Table 3: Bias, SD, SE, and CP for Setting 3 based on 2000 simulation runs

Covariate vector:	$\boldsymbol{V} = (X, Z_1, Z_2)$, $X \mid Z_1, Z_2 \sim \mathrm{logit}(0.5Z_1 - Z_2)$, $Z_1 \sim$ binary(0.5), $Z_2 \sim$ uniform(0, 0.5), $Z_1 \perp Z_2$
True hazard of T:	$\lambda(t \mid \boldsymbol{V}) = 0.5t\exp(\theta_x X + \theta_{z1} Z_1 + \theta_{z2} Z_2)$, $\boldsymbol{\theta} = (\theta_x, \theta_{z1}, \theta_{z2}) = (1, 1, 1)$
True propensity:	$\pi_1(T, \boldsymbol{V}) = 1 - \{1 + \exp(2T - Z_2 - 1.5)\}^{-1}$
True censoring:	$\psi(T, \boldsymbol{V}) = S_{C\mid\boldsymbol{V}}(T) = \exp\{-(\alpha T)^{1.2}\exp(0.2X + 0.1Z_1)\}$ α determines the censoring proportion
Variables used in $\widehat{\psi}$:	entire $\boldsymbol{V}$
Variables used in $\widehat{\pi}_1$:	$(T \wedge C, Z_1, Z_2)$ for SWE and (T, Z_1, Z_2) for DWE and CWE

$n = 1000$		Estimation of θ_x				Estimation of θ_{z1}				Estimation of θ_{z2}			
α	Method	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
1	Full	0.004	0.111	0.110	0.944	0.004	0.109	0.110	0.945	0.020	0.350	0.357	0.948
	CC	0.207	0.179	0.182	0.743	0.206	0.173	0.180	0.754	-0.002	0.587	0.594	0.948
	SWE	0.147	0.175	0.176	0.852	0.221	0.164	0.169	0.711	0.022	0.552	0.569	0.956
	DWE	-0.025	0.234	0.227	0.939	-0.022	0.229	0.221	0.939	-0.135	0.794	0.746	0.938
	CWE	0.018	0.170	0.176	0.946	0.025	0.170	0.170	0.950	-0.091	0.498	0.528	0.954
2	Full	0.004	0.161	0.163	0.950	0.010	0.162	0.164	0.944	0.007	0.520	0.526	0.947
	CC	0.261	0.297	0.312	0.836	0.265	0.292	0.311	0.826	-0.022	0.969	0.997	0.954
	SWE	0.229	0.299	0.306	0.862	0.290	0.285	0.300	0.806	0.013	0.964	0.972	0.947
	DWE	-0.015	0.622	0.548	0.928	-0.106	0.627	0.554	0.929	-0.427	1.829	1.635	0.952
	CWE	0.099	0.377	0.335	0.926	0.123	0.358	0.336	0.924	-0.105	1.033	0.999	0.958

Missing Rate: 46.9%; Censoring Rate: 60.2% ($\alpha = 1$), 81.5% ($\alpha = 2$)

Simulation Summary

1. Both CC and SWE are biased, whereas DWE and CWE have negligible biases, except when π_1 is misspecified.

2. Over-fitting for estimation of π_1 and ψ is not a problem, but misspecification (e.g., under-fitting) will affect the performance of DWE and CWE.

3. The SD of CC, SWE and CWE are comparable, howeve, the SD of DWE is largest as only data with $R\delta = 1$ is directly used by the doubly weighted estimation equation (10).

Simulation Summary

4. The relative efficiency of DWE with respect to CWE (variance ratio) decreases as the rate of censoring increases.

Table 4: Relative Efficiency of DWE with respect to CWE

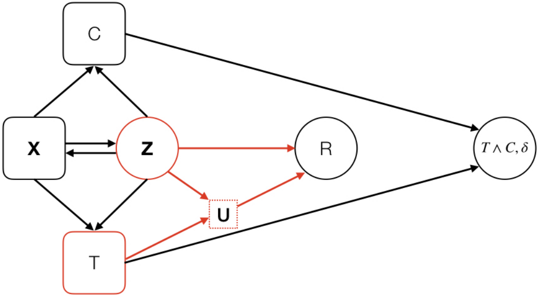
	Censoring Rate	Estimation of θ_x	Estimation of θ_z	
Setting 2	37.0%	0.857	0.887	
Setting 1	47.6%	0.796	0.675	
		Estimation of θ_x	Estimation of θ_{z1}	Estimation of θ_{z2}
Setting 3 ($\alpha = 1$)	60.2%	0.526	0.744	0.627
Setting 3 ($\alpha = 2$)	81.5%	0.367	0.572	0.565

5. The bootstrap SE is close to SD, even in the cases where point estimators are biased.

Example

We analyze the Non-Small Cell Lung Cancer (NSCLC) data set from National Cancer Database (NCDB). We focus on 1642 stage III patients who had private insurance and were diagnosed during 2004-2006 with 70-80 years of age at the time of diagnosis. The research interest is the overall survival of patients with stage III NSCLC under adjustment of **age, gender, tumor stage, and two treatments: the Stereotactic Body Radiation Therapy (SBRT) and surgery**. In our data, all patients were diagnosed with stage III at baseline, but only about 40% patients had **more accurate tumor stage** recorded as either stage IIIA or stage IIIB. Thus, we treat the tumor stage as covariate X having missing values.

Example



- X = more accurate stage III record,
- Z_1 = treatment, Z_2 = age, Z_3 = gender,
- R: physician's judgment on measuring X or not,
- U: physician's subjective assessments.

Example

Table 5: Estimates of coefficients of covariates for stage III NSCLC data

	X = tumor stage 0 = stage IIIA 1 = stage IIIB		Z_1 = treatment 0 = surgery 1 = SBRT		Z_2 = age		Z_3 = gender 0 = male 1 = female	
Method	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
CC	0.318	0.112	0.662	0.143	0.025	0.016	-0.435	0.099
SWE	0.297	0.112	0.679	0.139	0.042	0.012	-0.436	0.077
DWE	0.354	0.126	0.298	0.163	0.008	0.025	-0.602	0.119
CWE	0.278	0.105	0.573	0.150	0.049	0.017	-0.270	0.091

Missing Rate: 59.4%; Censoring Rate: 38.5%

Example

Table 6: Mutual comparison between CC, SWE and CWE

	X = tumor stage			Z_1 = treatment		
Comparing Methods	Diff	SE	Corr	Diff	SE	Corr
CC - SWE	0.021	0.021	0.983	-0.016	0.018	0.993
CC - CWE	0.040	0.068	0.842	0.090	0.058	0.928
SWE - CWE	0.019	0.070	0.830	0.106	0.058	0.928
	Z_2 = age			Z_3 = gender		
	Diff	SE	Corr	Diff	SE	Corr
CC - SWE	-0.017	0.004	0.968	0.001	0.023	0.974
CC - CWE	-0.024	0.022	0.581	-0.165	0.062	0.841
SWE - CWE	-0.007	0.022	0.578	-0.166	0.065	0.827

Note:

Diff is the difference of estimated coefficient between two methods,

SE is obtained by 1000 bootstrap replicates of the difference,

Corr is the sample correlation coefficient based on 1000 bootstrap replicates of estimated coefficients from two different methods.

Example

Table 6: Mutual comparison between CC, SWE and CWE

	X = tumor stage			Z_1 = treatment		
Comparing Methods	Diff	SE	Corr	Diff	SE	Corr
CC - SWE	0.021	0.021	0.983	-0.016	0.018	0.993
CC - CWE	0.040	0.068	0.842	0.090	0.058	0.928
SWE - CWE	0.019	0.070	0.830	0.106	0.058	0.928
	Z_2 = age			Z_3 = gender		
	Diff	SE	Corr	Diff	SE	Corr
CC - SWE	-0.017	0.004	0.968	0.001	0.023	0.974
CC - CWE	-0.024	0.022	0.581	-0.165	0.062	0.841
SWE - CWE	-0.007	0.022	0.578	-0.166	0.065	0.827

Note:

Diff is the difference of estimated coefficient between two methods,
SE is obtained by 1000 bootstrap replicates of the difference,
Corr is the sample correlation coefficient based on 1000 bootstrap replicates of estimated coefficients from two different methods.

Example

Table 6: Mutual comparison between CC, SWE and CWE

	X = tumor stage			Z_1 = treatment		
Comparing Methods	Diff	SE	Corr	Diff	SE	Corr
CC - SWE	0.021	0.021	0.983	-0.016	0.018	0.993
CC - CWE	0.040	0.068	0.842	0.090	0.058	0.928
SWE - CWE	0.019	0.070	0.830	0.106	0.058	0.928
	Z_2 = age			Z_3 = gender		
	Diff	SE	Corr	Diff	SE	Corr
CC - SWE	-0.017	0.004	0.968	0.001	0.023	0.974
CC - CWE	-0.024	0.022	0.581	-0.165	0.062	0.841
SWE - CWE	-0.007	0.022	0.578	-0.166	0.065	0.827

Note:

Diff is the difference of estimated coefficient between two methods,

SE is obtained by 1000 bootstrap replicates of the difference,

Corr is the sample correlation coefficient based on 1000 bootstrap replicates of estimated coefficients from two different methods.

Example

1. DWE may be not reliable for this example. We treat it as an initial estimator for CWE.
2. Estimates by SWE are very close to those by CC, except for age.
3. Among the differences of estimated coefficients between CWE and SWE (or CC), the difference for gender Z_3 is most significant.
4. Overall, the tumor stage, treatment, age, and gender all have significant association with the survival time.

Thank you so much!

Reference

Cook, V. J., Hu, X. J., and Swartz, T. B. (2011). Cox regression with covariates missing not at random. *Statistics in Biosciences* **3,** 208–222.

Herring, A. H. and Ibrahim, J. G. (2001). Likelihood-based methods for missing covariates in the cox proportional hazards model. *Journal of the American Statistical Association* **96,** 292–302.

Lin, D. Y. and Ying, Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association* **88,** 1341–1349.

Qi, L., Wang, C. Y., and Prentice, R. L. (2005). Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association* **100,** 1250–1263.

Qiu, Z. (2017). Statistical inference under imputation for proportional hazard model with missing covariates. *Communications in Statistics - Theory and Methods* **46,** 11575–11590.

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* **119,** 99–130.

Rathouz, P. J. (2007). Identifiability assumptions for missing covariate data in failure time regression models. *Biostatistics* **8,** 345–356.

Shao, J. (2003). *Mathematical Statistics*. Springer.