Published in final edited form as:

Biometrics. 2020 June ; 76(2): 460–471. doi:10.1111/biom.13155.

Cox Regression with Survival-Time-Dependent Missing Covariate Values

Yanyao Yi^{1,2,3,*}, Ting Ye^{2,3,**}, Menggang Yu^{3,***}, Jun Shao^{1,2,****}

¹KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, China

²Department of Statistics, University of Wisconsin, Madison, Wisconsin, U.S.A

³Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin, U.S.A

Summary:

Analysis with time-to-event data in clinical and epidemiological studies often encounters missing covariate values and the missing at random assumption is commonly adopted (e.g., Qi et al., 2005), which assumes that missingness depends on the observed data, including the observed outcome which is the minimum of survival and censoring time. However, it is conceivable that in certain settings, missingness of covariate values is related to the survival time but not to the censoring time (Rathouz, 2007). This is especially so when covariate missingness is related with an unmeasured variable affected by patient's illness and prognosis factors at baseline. If this is the case, then the covariate missingness is not at random as the survival time is censored, and it creates a challenge in data analysis. In this article, we propose an approach to deal with such survival-time-dependent covariate missingness based on the well known Cox proportional hazard model. Our method is based on inverse propensity weighting with the propensity estimated by nonparametric kernel regression. Our estimators are consistent and asymptotically normal, and their finite-sample performance is examined through simulation. An application to a real-data example is included for illustration.

Keywords

Censoring; Missing not at random; Nonparametric kernel estimator; Propensity

1. Introduction

Cox regression is one of the most popular methods dealing with censored failure time in survival analysis. For a continuous failure time T and covariate vector V measured at baseline, in this paper we consider the following Cox proportional hazard model,

* yanyao.yi@wisc.edu. ** tye27@wisc.edu. *** meyu@biostat.wisc.edu. **** shao@stat.wisc.edu.

Supporting Information

In Example 1 of Section 1, the data was obtained from The National Cancer Database: <https://www.facs.org/quality-programs/cancer/ncdb>, and The American Joint Committee on Cancer: <https://cancerstaging.org/references-tools/quickreferences/Documents/LungMedium.pdf>. The proof of the fact that (5) implies (4) and (6), the proofs of Theorems 1-2, the figures referenced in Section 3, and R codes for numerical work are available with this paper at the Biometrics website on Wiley Online Library.

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

where $\lambda(t|\mathbf{V})$ is the hazard at time t given $\mathbf{V}$, $\lambda_0(t)$ 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. In many survival studies, there exists censoring. In this paper, we focus on right censoring, i.e., there is a censoring time C and what we observe is $(T \wedge C, \delta)$, where $T \wedge C = \min(T, C)$ and $\delta = I_{\{T \leq C\}}$ is the indicator of event $T \leq C$. A common assumption on censoring is

$$T \perp C | \mathbf{V}, \quad (2)$$

i.e., T and C are conditionally independent given $\mathbf{V}$. Based on a random sample from the distribution of $(T \wedge C, \delta, \mathbf{V})$, $\boldsymbol{\theta}$ can be estimated by maximizing the partial likelihood derived in Cox (1975) under (1) and (2), and the asymptotic properties of this estimator can be found in Andersen and Gill (1982).

In clinical and epidemiological studies some components of the covariate vector $\mathbf{V}$ may have missing data and the partial likelihood cannot be directly applied. Let $\mathbf{V} = (\mathbf{X}, \mathbf{Z})$ with $\mathbf{X}$ being the sub-vector that may have missing values and $\mathbf{Z}$ being the sub-vector that is always observed, and let R be the indicator equaling 1 if $\mathbf{X}$ is completely observed and 0 if at least one component of $\mathbf{X}$ is missing. As pointed out in Paik and Tsai (1997), Lipsitz and Ibrahim (1998), and Rathouz (2007), the complete-case analysis with the partial likelihood based only on subjects with $R = 1$ is valid if $R \perp (T, C) | \mathbf{V}$, i.e., missingness depends only on $\mathbf{V}$, not on outcome (T, C) , although missingness may be not at random and methods more efficient than complete-case analysis can be derived, e.g., Lin and Ying (1993) and Cook et al. (2011).

However, missingness of covariate values is often believed to be outcome related, either directly or indirectly. In survival studies, some researchers assume that missingness is $T \wedge C$ related, i.e.,

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

(e.g., Lipsitz and Ibrahim, 1998; Chen and Little, 1999; Herring and Ibrahim, 2001; Chen, 2002), which is a type of missing at random assumption (Rubin, 1976) because $T \wedge C$, $\mathbf{Z}$ and δ are all observed. Even if R is related with (T, C) , however, it is hard to imagine why R is related to $T \wedge C$, a very special function of the outcome (T, C) . We speculate that (3) is assumed for an easy analysis, since one can simply use methods valid under missing at random, e.g., inverse propensity weighting with estimated $\text{pr}(R = 1 | T \wedge C, \mathbf{Z}, \delta)$ based on always observed $(T \wedge C, \mathbf{Z}, \delta)$ (Wang and Chen, 2001; Qi et al., 2005; Xu et al., 2009).

Rathouz (2007) argues that the following survival-time-dependent missingness mechanism is more reasonable than assumption (3) in many biomedical studies,

$$R \perp (\mathbf{X}, C) | (T, \mathbf{Z}). \quad (4)$$

One scenario in which (4) holds is when there exists an unmeasured variable U prior to the measurement of X that satisfies

$$R \perp (T, C, X) | (U, Z) \quad \text{and} \quad U \perp (C, X) | (T, Z). \quad (5)$$

The first condition in (5) means that missingness of X is driven by U together with observed Z , whereas the second condition in (5) says that U is not related with (C, X) when (T, Z) is given. The directed acyclic graph in Figure 1 provides an illustration. It is also shown in the Supporting Information that (5) implies (4) and

$$\text{pr}(R = 1 | T, Z) = E\{\text{pr}(R = 1 | U, Z) | T, C, V\}. \quad (6)$$

Although $\text{pr}(R = 1 | T, Z)$ may be hard to interpret since T is a future observation while R is observed at baseline, it is a constructed missingness propensity for analysis in which T is used as a surrogate in (6) for the unmeasured variable U . An example of U could be a subjective assessment, as the following example indicates.

EXAMPLE 1: In our real data example analyzed in Section 4 based on the national cancer database, all patients were diagnosed with stage III Non-Small Cell Lung Cancer at baseline, but only about 40% patients had more accurate tumor stage X recorded as either stage IIIA or stage IIIB. From the lung cancer staging system provided by the American Joint Committee on Cancer, the main difference between stage IIIA and stage IIIB is the nodal involvement, i.e., stage IIIB has much more extensive metastasis in regional lymph nodes regardless of tumor size. Measuring X is difficult and may require invasive techniques. Recent advanced non-invasive methods such as positron emission tomography or compute tomography do not provide definitive stage confirmation, since the region around the lung to be examined for lymph node involvement includes the superior mediastinum, the lower mediastinum, the aortopulmonary window, and para-aorta (Teran and Brock, 2014). Thus, a major reason for missing X in this example is physician's assessment on whether an invasive test for a difficult measurement of X is worthwhile. This somewhat subjective assessment could be the variable U in (5) and Figure 1, which is affected by patient's prognosis factors in Z and illness related to the eventual survival time T , but is unlikely to be directly related with censoring time C . Thus, assumption (4) is more reasonable than assumption (3) for this real data example.

Since T may be censored, the missingness mechanism (4) is not missing at random. If (4) holds instead of (3), then estimators in Wang and Chen (2001), Qi et al. (2005) and Xu et al. (2009) will yield a biased result. Also, if (4) holds and T cannot be excluded from the missingness propensity, then the complete-case analysis is inconsistent.

Although the survival function is identifiable under (4) and censoring assumption (2) (Rathouz, 2007), there is no proposed method for estimating θ in Rathouz (2007) and afterwards. The main challenge is that $\text{pr}(R = 1 | T, Z)$ cannot be directly estimated when T is censored.

In this article, we construct two inverse probability weighting estimators of θ in the Cox proportional hazard model (1) under assumptions (2) and (4). The first one is based on a

weighted score function using only subjects with observed survival time and completely observed covariates. The first estimator may not be efficient since only data with $(\delta, R) = (1, 1)$ are involved in the score function, but this estimator is used as an initial estimator in the construction of a more efficient second estimator based on a weighted score function using all data with $R = 1$, censored or not. The major step in obtaining the second estimator is to overcome the difficulty in constructing suitable weights when T is censored with the help from the first estimator. In the construction of weights, we adopt nonparametric estimation of propensity, since a parametric form of propensity is hard to specify especially when the propensity is obtained through (6) by averaging over an unobserved variable U satisfying (5). The product kernel in Racine and Li (2004) is applied to handle both continuous and categorical covariates.

Both proposed estimators of θ are shown to be consistent and asymptotically normal under (1), (2), (4), and some regularity conditions, and their performance is examined in Section 3 by simulation. In Section 4, we return to the real data example of Non-Small Cell Lung Cancer to illustrate our procedures. All technical details are given in the Supporting Information.

2. Method

Under (2) and (4), we consider estimating θ in (1) based on a random sample $(T_i \wedge C_i, \delta_i, V_i, R_i)$, $i = 1, \dots, n$, from the population of $(T \wedge C, \delta, V, R)$. For each i , Z_i is observed but X_i is completely observed if and only if $R_i = 1$. To illustrate the idea, in this section we do not use partially observed X when X is multivariate. An extension of using partially observed X is given in Section 5.

2.1 Doubly Weighted Estimator

We first derive an estimator of θ that is consistent and asymptotically normal under some conditions. It serves as an initial estimator in constructing a more efficient estimator in Section 2.2. Under (4), $\text{pr}(R = 1|T, C, V)$ is a function of (T, Z) only. Define

$$\pi_1(T, Z) = \text{pr}(R = 1|T, C, V) \quad \text{and} \quad \psi(T, V) = \text{pr}(\delta = 1|T, V). \quad (7)$$

Assume first that π_1 and ψ in (7) are known functions. Based on the fact that, without censoring,

$$M^*(t) = I_{\{T \leq t\}} - \int_0^t I_{\{T \geq u\}} \lambda_0(u) \exp(\theta^\top V) du \quad (8)$$

is a mean zero martingale with respect to filtration $\mathcal{F}_t^* = \{I_{\{T \geq u\}}, V: 0 \leq u \leq t\}$ (Fleming and Harrington, 1991), our first proposed estimator $\hat{\theta}_1$ of θ is obtained by using data from subjects with observed X ($R = 1$) and non-censored T ($\delta = 1$), and solving the following weighted estimation equation:

$$\sum_{i=1}^n \int_0^\infty \frac{R_i \delta_i}{\pi_1(T_i, \mathbf{Z}_i) \psi(T_i, \mathbf{V}_i)} \left\{ \mathbf{V}_i - \frac{\mathbf{S}_1^{(1)}(\boldsymbol{\theta}, t, \pi_1, \psi)}{\mathbf{S}_1^{(0)}(\boldsymbol{\theta}, t, \pi_1, \psi)} \right\} dN_i(t) = 0, \quad (9)$$

where $N_i(t) = \delta_i I\{T_i \leq t\}$ is the counting process,

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

$$\mathbf{S}_1^{(0)}(\boldsymbol{\theta}, t, \pi_1, \psi) = \frac{1}{n} \sum_{i=1}^n \frac{R_i \delta_i}{\pi_1(T_i, \mathbf{Z}_i) \psi(T_i, \mathbf{V}_i)} I\{T_i \geq t\} \exp(\boldsymbol{\theta}^\top \mathbf{V}_i).$$

Two inverse probability weights are applied in (9), i.e., $\pi_1(T, \mathbf{Z})$ for missing $\mathbf{X}$ value, which is commonly used in missing data problems (Robins, 1997), and $\psi(T, \mathbf{V})$ for censoring (Robins and Finkelstein, 2000).

By differentiating the left hand side of (9) with respect to $\boldsymbol{\theta}$ and applying the Cauchy-Schwartz inequality, we can show that the left hand side of (9) is the derivative of a convex function of $\boldsymbol{\theta}$ and, hence, (9) can be easily solved. Once $\boldsymbol{\theta}$ is estimated by $\hat{\boldsymbol{\theta}}_1$, the following Breslow type semi-parametric estimator of the cumulative baseline hazard $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ can be obtained:

$$\hat{\Lambda}_{01}(t) = \int_0^t \frac{1}{n \mathbf{S}_1^{(0)}(\hat{\boldsymbol{\theta}}_1, u, \pi_1, \psi)} \sum_{i=1}^n \frac{R_i \delta_i dN_i(u)}{\pi_1(T_i, \mathbf{Z}_i) \psi(T_i, \mathbf{V}_i)} \quad (10)$$

(Fleming and Harrington, 1991).

Since π_1 and ψ in (7) are usually unknown, in the rest of this subsection we consider their estimation. Since $\pi_1(T, \mathbf{Z})$ depends on $(T, \mathbf{Z})$, not $(C, \mathbf{X})$, the subset of non-censored data are used in estimating π_1 , which makes use of data from all subjects with $\delta = 1$ even if $\mathbf{X}$ is missing. A parametric form of $\text{pr}(R = 1 \mid T, \mathbf{Z})$ is hard to specify, especially when it is derived from $\text{pr}(R = 1 \mid T, \mathbf{Z}) = E\{\text{pr}(R = 1 \mid U, \mathbf{Z}) \mid T, \mathbf{Z}\}$ with an unobserved U and nonparametric $p(U \mid T, \mathbf{Z})$. To obtain robust estimators, here we consider nonparametric product kernel regression (Racine and Li, 2004) estimator of $\pi_1(T, \mathbf{Z})$, which is given by

$$\hat{\pi}_1(T, \mathbf{Z}) = \frac{\sum_i R_i \delta_i K_h((T_i, \mathbf{Z}_i), (T, \mathbf{Z}))}{\sum_i \delta_i K_h((T_i, \mathbf{Z}_i), (T, \mathbf{Z}))}, \quad (11)$$

where K_h is the product of a kernel for T and continuous components of $\mathbf{Z}$ and a kernel for discrete components of $\mathbf{Z}$ (Racine and Li, 2004), $h = (h_c, h_d)$, and h_c and h_d are bandwidths for continuous and discrete kernels, respectively, selected by cross validation according to Racine and Li (2004).

For the censoring propensity $\psi(T, \mathbf{V})$, it follows from assumption (2) that

$$\psi(T, \mathbf{V}) = \text{pr}(\delta = 1 | T, \mathbf{V}) = \text{pr}(C > T | \mathbf{V}) = S_C | \mathbf{V}(T),$$

where $S_C | \mathbf{V}(t)$ is the conditional “survival” function of the censoring time C given $\mathbf{V}$. Note that

$$\begin{aligned} \text{pr}(C \leq c, R = r, T \leq t | \mathbf{V}) &= E\{\text{pr}(C \leq c, R = r, T \leq t | T, \mathbf{V}) | \mathbf{V}\} \\ &= E\{\text{pr}(C \leq c | T, \mathbf{V}) \text{pr}(R = r, T \leq t | T, \mathbf{V}) | \mathbf{V}\} \\ &= E\{\text{pr}(C \leq c | \mathbf{V}) \text{pr}(R = r, T \leq t | T, \mathbf{V}) | \mathbf{V}\} \\ &= \text{pr}(C \leq c | \mathbf{V}) \text{pr}(R = r, T \leq t | \mathbf{V}), \end{aligned}$$

where the second equation is based on the fact that assumption (4) and Proposition 1.11 in Shao (2003) imply that $R \perp C | (T, \mathbf{X}, \mathbf{Z})$, and the third equation follows from assumption (2). Consequently, $C \perp (R, T) | \mathbf{V}$. Thus, the conditional survival function $S_C | \mathbf{V}(t)$ can be estimated by using the subset of data with $R = 1$. Treating $C \wedge T$ as the observed time and $1 - \delta$ as the “event status” for censoring time C , we estimate $S_C | \mathbf{V}(t)$ using a Cox proportional hazard model similar to Model (1): $\lambda_C(t | \mathbf{V}) = \lambda_{0C}(t) \exp(\beta^T \mathbf{V})$. Then β is estimated by $\hat{\beta}$ which is the solution to

$$\sum_{i=1}^n R_i \int_0^\infty \left[\mathbf{V}_i - \frac{S_\delta^{(1)}(\beta, t)}{S_\delta^{(0)}(\beta, t)} \right] dN_i^\delta(t) = 0,$$

with $N_i^\delta(t) = (1 - \delta_i)I\{C_i \leq t\}$, $S_\delta^{(0)}(\beta, t) = n^{-1} \sum_{i=1}^n R_i Y_i(t) \exp(\beta^T \mathbf{V}_i)$ and $S_\delta^{(1)}(\beta, t) = n^{-1} \sum_{i=1}^n R_i Y_i(t) \mathbf{V}_i \exp(\beta^T \mathbf{V}_i)$. The estimated cumulative baseline hazard is

$$\hat{\Lambda}_{0C}(t) = \int_0^t \frac{1}{n S_\delta^{(0)}(\hat{\beta}, u)} \sum_{i=1}^n R_i dN_i^\delta(u).$$

Consequently, the estimator of $\psi(T, \mathbf{V})$ has the form

$$\hat{\psi}(T, \mathbf{V}) = \hat{S}_C | \mathbf{V}(T) = \exp\left\{-\exp(\hat{\beta}^T \mathbf{V}) \hat{\Lambda}_{0C}(T)\right\}.$$

Estimators $\hat{\theta}_1$ of θ and $\hat{\lambda}_{01}$ of λ_0 can be obtained by using (9) and (10) with π_1 and ψ replaced by their estimators $\hat{\pi}_1$ and $\hat{\psi}$, respectively. We call $\hat{\theta}_1$ a doubly weighted estimator (DWE) as two weight functions $\hat{\pi}_1$ and $\hat{\psi}$ are used.

The following result establishes the consistency and asymptotic normality of $\hat{\theta}_1$. The proof is given in the Supporting Information.

THEOREM 1: *Assume the following regularity conditions.*

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

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

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 for $\int L(u) du = 1$, $\int u^m L(u) du = 0$ for $m = 1, \dots, (r-1)$, and $\int u^r L(u) du \neq 0$.

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

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

Then, as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\mathbf{\theta}}_1 - \mathbf{\theta}) \xrightarrow{d} N(0, \Sigma_1),$$

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

$$\begin{aligned} \Sigma_1 = & \Omega^{*-1} \text{var} \left[\frac{R\delta}{\pi_1(T, \mathbf{Z})\psi(T, \mathbf{V})} \mathbf{M}_{\mathbf{V}}^* + \left\{ 1 - \frac{R}{\pi_1(T, \mathbf{Z})} \right\} E(\mathbf{M}_{\mathbf{V}}^* | T, \mathbf{Z}) \right. \\ & \left. + \Gamma(R\mathbf{M}_{\mathbf{V}}^\delta) \right] \Omega^{*-1}, \end{aligned} \quad (12)$$

Ω^* is defined in Condition 2, $\mathbf{M}_{\mathbf{V}}^* = \int_0^\infty \left[\mathbf{V} - s_1^{(1)}(\mathbf{\theta}, t) \left\{ s_1^{(0)}(\mathbf{\theta}, t) \right\}^{-1} \right] dM^*(t)$ and $\mathbf{M}_{\mathbf{V}}^\delta = \int_0^\infty \left[\mathbf{V} - s_\delta^{(1)}(\mathbf{\beta}, t) \left\{ s_\delta^{(0)}(\mathbf{\beta}, t) \right\}^{-1} \right] dM^\delta(t)$ are both mean zero martingale transformations, $M^\delta(t) = N^\delta(t) - \int_0^t Y(u) \lambda_{0C}(u) \exp(\beta^\top \mathbf{V}) du$, $s_\delta^{(0)}(\mathbf{\beta}, t) = E \left\{ RY(t) \exp(\beta^\top \mathbf{V}) \right\}$, $s_\delta^{(1)}(\mathbf{\beta}, t) = E \left\{ RY(t) \exp(\beta^\top \mathbf{V}) \mathbf{V} \right\}$ and $\Gamma = E \left\{ \Lambda_{0C}(T) \exp(\beta^\top \mathbf{V}) \mathbf{M}_{\mathbf{V}}^* \mathbf{V}^\top \right\} \left\{ \text{var}(\mathbf{M}_{\mathbf{V}}^\delta) \right\}^{-1}$.

The quantity $\Gamma(R\mathbf{M}_{\mathbf{V}}^\delta)$ in (12) is due to the estimation of ψ defined in (7). Since the three quantities inside of the variance given in (12) can be correlated, the form of Σ_1 is complicated. Based on Theorem 1, Σ_1 can be estimated by the bootstrap, which is used in our simulation studies.

2.2 Compositely Weighted Estimator

In the construction of $\hat{\mathbf{\theta}}_1$, although we utilize some incomplete data in the estimator of π_1 given in (11), the weighted score function (9) is derived from the martingale $M^*(t)$ in (8) that

only involves non-censored subjects. Consequently, $\hat{\theta}_1$ may be inefficient and, in particular, $\hat{\theta}_1$ does not reduce to the usual maximum partial likelihood estimator of θ when there is no missing value.

The partial likelihood for the case of no missing data is based on the martingale

$$M(t) = N(t) - \int_0^t Y(u) \lambda_0(u) \exp(\theta^\top V) du \quad (13)$$

with respect to the filtration $\mathcal{F}_t = \{I_{\{T < u, \delta = 1\}}, I_{\{T \wedge C \geq u\}}, V: 0 \leq u \leq t\}$ (Fleming and Harrington, 1991), where $Y(t) = I_{\{T \wedge C \geq t\}}$ is the at-risk process. Unlike the martingale $M^*(t)$ in (8), $M(t)$ in (13) involves both non-censored and censored subjects. If we can derive a weighted score function for θ from $M(t)$ in the presence of missing covariate values, then the estimation may be more efficient than the one based on $M^*(t)$. Specifically, we consider the following weighted estimation equation,

$$\sum_{i=1}^n \int_0^\infty \frac{R_i}{\pi_1(T_i, Z_i)} \left\{ V_i - \frac{S^{(1)}(\theta, t, \pi_1, \pi_0)}{S^{(0)}(\theta, t, \pi_1, \pi_0)} \right\} dN_i(t) = 0 \quad (14)$$

where

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

$$S^{(0)}(\theta, t, \pi_1, \pi_0) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{R_i \delta_i I_{\{T_i \geq t\}}}{\pi_1(T_i, Z_i)} + \frac{R_i (1 - \delta_i) I_{\{C_i \geq t\}}}{\pi_0(C_i, V_i)} \right\} \exp(\theta^\top V_i)$$

$\pi_1(T, Z)$ is given in (7), and $\pi_0(C, V) = \text{pr}(R = 1 | C, V, \delta = 0)$ is the inverse weight for R when $\delta = 0$. Our proposed second estimator $\hat{\theta}_2$ of θ is a solution to (14) after we substitute π_1 and π_0 by their estimators $\hat{\pi}_1$ and $\hat{\pi}_0$, respectively. An estimator of $\Lambda_0(t)$ is then

$$\hat{\Lambda}_{02}(t) = \int_0^t \frac{1}{n S^{(0)}(\hat{\theta}_2, u, \hat{\pi}_1, \hat{\pi}_0)} \sum_{i=1}^n \frac{R_i dN_i(u)}{\hat{\pi}_1(T_i, Z_i)}.$$

The same estimator $\hat{\pi}_1$ defined in (11) can be used to estimate π_1 . It remains to derive an estimator of $\pi_0(C, V)$. Under assumption (4), missingness of X values depends on T which is not observed when $\delta = 0$. Thus, $\pi_0(C, V) = \text{pr}(R = 1 | C, V, \delta = 0)$ is a function of the entire V containing possibly missing values of X and its estimation is challenging. Note that

$$\pi_0(C, V) = \text{pr}(R = 1 | C, V, \delta = 0) = \frac{\text{pr}(R = 1, \delta = 0 | C, V)}{\text{pr}(\delta = 0 | C, V)}. \quad (15)$$

Under assumption (2),

$$\text{pr}(\delta = 0|C, \mathbf{V}) = \text{pr}(T > C|C, \mathbf{V}) = S_{T|\mathbf{V}}(C), \quad (16)$$

where $S_{T|\mathbf{V}}(\cdot)$ denotes the survival function of T given $\mathbf{V}$. Moreover, assumptions (2) and (4) imply $R \perp C|(T, \mathbf{V})$ and, therefore,

$$\text{pr}(R = 1, \delta = 0|C, \mathbf{V}) = - \int_C^\infty \pi_1(t, \mathbf{Z}) dS_{T|\mathbf{V}}(t). \quad (17)$$

Hence, the survival function $S_{T|\mathbf{V}}(\cdot)$ is the only unknown part we need to estimate in order to estimate $\pi_0(C, \mathbf{V})$. One way to do this is using the estimators $\hat{\theta}_1$ and $\hat{\Lambda}_{01}(t)$ in Section 2.1 to construct an initial estimator of the survival function,

$$\hat{S}_{T|\mathbf{V}}(t) = \exp\left\{-\exp\left(\hat{\theta}_1^\top \mathbf{V}\right) \hat{\Lambda}_{01}(t)\right\}. \quad (18)$$

Then, it follows from (15)-(18) that we can estimate $\pi_0(C, \mathbf{V})$ by

$$\hat{\pi}_0(C, \mathbf{V}) = - \frac{\int_C^\infty \hat{\pi}_1(t, \mathbf{Z}) d\hat{S}_{T|\mathbf{V}}(t)}{\hat{S}_{T|\mathbf{V}}(C)}. \quad (19)$$

In any application with a finite n , we set $\hat{S}_{T|\mathbf{V}}(t) = 0$ when $t \geq \max\{T_i \wedge C_i : i = 1, \dots, n\}$ so that the integral in (19) is finite. This does not affect the asymptotic result, since $\hat{S}_{T|\mathbf{V}}(t)$ converges to $S_{T|\mathbf{V}}(t)$ almost surely. If there is no missing value, $R_i = 1$ for all i , then it follows from (11) that $\hat{\pi}_1 \equiv 1$ and, hence, $\hat{\pi}_0 \equiv 1$ and $\hat{\theta}_2$ reduces to the usual maximum partial likelihood estimator.

We call $\hat{\theta}_2$ a compositely weighted estimator (CWE) since the weight function π_0 involves both the weight function π_1 and the survival function of T . The following result for the consistency and asymptotic normality of $\hat{\theta}_2$ is proved in the Supporting Information.

THEOREM 2: Assume that Conditions 1–5 stated in Theorem 1 hold with Ω^* in Condition 2

replaced by $\Omega = \int_0^\infty \left[s^{(2)}(\boldsymbol{\theta}, t) - s^{(1)}(\boldsymbol{\theta}, t) \{s^{(1)}(\boldsymbol{\theta}, t)\}^\top \{s^{(0)}(\boldsymbol{\theta}, t)\}^{-1} \right] \lambda_0(t) dt$, where

$s^{(0)}(\boldsymbol{\theta}, t) = E\{Y(t) \exp(\boldsymbol{\theta}^\top \mathbf{V})\}$, $s^{(1)}(\boldsymbol{\theta}, t) = E\{Y(t) \mathbf{V} \exp(\boldsymbol{\theta}^\top \mathbf{V})\}$, and

$s^{(2)}(\boldsymbol{\theta}, t) = E\{Y(t) \mathbf{V} \mathbf{V}^\top \exp(\boldsymbol{\theta}^\top \mathbf{V})\}$. Besides, we assume that the support of survival time T is a subset of the support of censoring time C . Then, as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}) \xrightarrow{d} N(0, \Sigma_2),$$

where

$$\Sigma_2 = \Omega^{-1} + \Omega^{-1} E\{(\pi_1^{-1} - 1) \text{var}(\mathbf{M}_V | T, \mathbf{Z})\} \Omega^{-1}, \quad (20)$$

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

Note that $\Omega^{-1} = \{\text{var}(\mathbf{M}_V)\}^{-1}$ is the asymptotic covariance matrix of the estimator of $\boldsymbol{\theta}$ obtained under the usual Cox regression without missing data. Thus, the second term on the right hand side of (20) is the efficiency loss due to missing covariate values compared with the estimator derived from the Cox regression without missing data. This term does not depend on $\boldsymbol{\pi}_0$, i.e., as long as $\hat{\boldsymbol{\pi}}_0$ is consistent, the estimation of $\boldsymbol{\pi}_0$ does not affect the efficiency of $\hat{\boldsymbol{\theta}}_2$.

3. Simulation

Simulations are conducted in three different settings to examine the finite-sample performance of our proposed DWE and CWE of $\boldsymbol{\theta}$, and to compare them with the estimator computed without missing covariate values (Full) as a standard, the complete-case (CC) estimator, and the simple weighted estimator (SWE) in Qi et al. (2005). As we discussed in Section 1, CC estimator and SWE may be inconsistent under assumption (4).

3.1 Simulation Settings

The details about data generating for three simulation settings are given in Tables 1–3, respectively. The time-independent covariate vector is $\mathbf{V} = (X, Z)$ with independent univariate X and Z in settings 1–2, and $\mathbf{V} = (X, Z_1, Z_2)$ with correlated univariate X and bivariate (Z_1, Z_2) in setting 3. The survival time T follows the Cox proportional hazard model (1) with baseline hazard $\lambda_0(t)$, which is equal to 1 in settings 1–2 and $t/2$ in setting 3, and with true values of $\boldsymbol{\theta}$ specified in Tables 1–3. The true propensity for missing X values is logistic depending only on T in setting 1, on T and Z in setting 2, and on T and Z_2 in setting 3, as described in Tables 1–3. The censoring time C follows another Cox proportional hazard model, which depends on no covariate in setting 1, on X in setting 2, and on (X, Z_1) in setting 3. The rates of censoring and missing vary from 37% to 82% and 29% to 47%, respectively, and details are included in Tables 1–3.

The propensity for setting 1 is used to check the performances of DWE and CWE when Z is “unnecessarily” included in the estimation of $\boldsymbol{\pi}_1$, whereas the propensity for setting 2 is used to compare the performances of DWE and CWE with a misspecified propensity model including only T and a correct propensity including both T and Z .

3.2 Computation

To obtain the nonparametric kernel estimator of $\boldsymbol{\pi}_1$ defined by (11), the “*npreg*” function from the “*np*” package in R is used with its default continuous kernel, the second order Gaussian kernel, and the unordered categorical kernel (Racine and Li, 2004) achieved by setting option “*ukertype=liracine*”. By default, cross-validation is applied to select the

bandwidth parameter. As discussed in Section 2.1, the subset of data with $\delta = 1$ is used to estimate π_1 . Variables included in computing $\hat{\pi}_1$ are indicated in Tables 1–3.

The estimator of ψ is obtained by a Cox regression on V using the “*coxph*” function from the “*survival*” package in R. The subset of data with $R = 1$ is used to estimate ψ , according to the discussion in Section 2.1.

The score functions (9) and (14) are solved by the “*multiroot*” function from the “*rootSolve*” package in R with π_1 , ψ and π_0 replaced by their estimators, where π_0 is estimated according to equations (18) and (19) with DWE as an initial estimate.

Under assumption (3), the SWE can be calculated using (14) with both $\pi_1(T_i Z_i)$ and $\pi_0(C_i V_i)$ replaced by $\pi(T_i \wedge C_i, Z_i, \beta_i) = \text{pr}(R_i = 1 | T_i \wedge C_i, Z_i, \delta_i)$. Since $\pi(T_i \wedge C_i, Z_i, \delta_i)$ is a function of observed $T_i \wedge C_i Z_i$ and δ_i , it can be simply estimated using the kernel method (Qi et al., 2005), which is the Nadaraya-Watson estimator. For a fair comparison, the missing data propensity associated with SWE is also estimated by using “*npreg*” in R, instead of “*ksmooth*” and “*sm.regression*” mentioned in Qi et al. (2005). Variables included in estimating π are indicated in Tables 1–3.

3.3 Simulation Results

Simulation results are reported in Tables 1–3 for settings 1–3, respectively, with 2000 simulation runs and sample sizes $n = 500$ or 1000 as specified in Tables 1–3. The results include the simulation bias and standard deviation (SD) of estimators, the standard error (SE) of estimators based on the bootstrap with 1000 replications, and the coverage probability (CP) of 95% confidence intervals based on the bootstrap percentile. In the Supporting Information, we also include some figures to show the averages of estimated π_1 and ψ over simulation runs against the true curves.

The simulation results can be summarized as follows.

1. Both CC and SWE are substantially biased in settings 1–2, which leads to substantially low CP, e.g., CP is much smaller than 95% in many cases. In setting 3, the biases of CC and SWE are also serious except for the case of estimating θ_{22} . Overall, SWE is less biased than CC. The proposed DWE and CWE have negligible biases in all cases and good CP, except for the case where π_1 is misspecified (the middle block of Table 2 with $n = 500$).
2. The SD of SWE and CWE are comparable, but having a small SD for a biased estimator SWE is not an advantage. In fact, the SD of CC is also comparable with those of SWE and CWE. The SD of DWE is largest as the estimation equation (9) for DWE does not directly use data from censored units. The relative efficiency of DWE with respect to CWE (variance ratio) is 0.85–0.89 in setting 2, 0.67–0.80 in setting 1, 0.53–0.74 in setting 3 with $\alpha = 1$, and 0.37–0.57 in setting 3 with $\alpha = 2$, which decreases as the rate of censoring increases.
3. When Z is “unnecessarily” included in the estimation of π_1 (the last block of Table 1 when $n = 500$), the SD of CWE (or DWE) is comparable with that in the

case where Z is correctly not included. When an important variable is excluded in estimating π_1 , however, the performances of DWE and CWE are affected (the middle block of Table 2 with $n = 500$). Similarly, including a few unnecessary covariates in the estimation of ψ does not lead to any problem, as we use the entire V in all cases.

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

4. Real Data Example

We analyze the Non-Small Cell Lung Cancer (NSCLC) data set introduced in Example 1 of Section 1. 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. Surgery is the standard first line treatment for operable NSCLC. For stage III NSCLC patients whose lung tumor can not be easily or cleanly removed due to size, location, and nodal involvement around the lung, however, SBRT is a new promising radiation therapy that can be a good alternative (Kumar et al., 2017). Older patients may also opt for non-surgical interventions due to complications commonly associated with 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. As we argue in Example 1 of the introduction section, it is reasonable to assume that missingness of stage record X is related to survival time T and possibly other covariates, $Z_1 =$ the treatment choice of SBRT or surgery, $Z_2 =$ age (treated as continuous), and $Z_3 =$ gender, where $Z = (Z_1, Z_2, Z_3)$ is all observed; on the other hand, we think missing X is unlikely to be related to C given T and Z .

We apply the CC, SWE, DWE, and CWE as previously described in Section 3. The propensity is estimated by the product kernel regression based on $(T \wedge C, Z)$ for SWE, and (T, Z) for DWE and CWE. We use the entire Z in estimating the propensity and entire V in estimating ψ , as the simulation results in Section 3 indicate that overfitting is not a problem.

The results of estimating coefficients of covariates in Cox regression and some of their differences are given in Table 4. The SEs are provided using 1000 bootstrap replications. The marginal quantiles of observed survival and follow-up times, and censoring and missing rates are also included in Table 4. We can draw the following conclusions based on these results.

1. Estimates by SWE are very close to those by CC except for the coefficient of covariate age Z_2 . In fact, the sample correlation coefficients between CC and SWE based on 1000 bootstrap replicates ranges from 97% to 99% for four covariate components. For X or Z_2 , the difference between CWE and SWE is close to 0, whereas the difference is beyond $2 \times \text{SE}$ for Z_3 and slightly smaller than $2 \times \text{SE}$ for Z_1 .

2. Generally, CC, SWE and CWE have similar SE, while DWE has the largest SE. This is consistent with the simulation results in Section 3. DWE estimates are very different from others. Since DWE is less stable than CWE according to our simulation results, we believe that results from DWE are not reliable in this example.
3. Overall, the tumor stage, treatment, age, and gender all have significant association with the survival time. The difference between the proposed CWE and SWE or CC is in the magnitude of covariate effects.

5. Discussion

Under Cox proportional hazard model (1) with right censoring and survival-time-dependent missing baseline covariate values, we propose two consistent and asymptotically normal estimators of the Cox regression parameters, based on inverse probability weighting. The first is an initial estimator and the second estimator is more efficient and recommended.

In general, assumption (3) and assumption (4) do not have definite relationship, except that they are the same when there is no censoring. Both assumptions are special cases of the following assumption,

$$R \perp X(T, C, Z). \quad (21)$$

Because both T and C are in the conditioning in (21) and they are not observed simultaneously, inference is difficult to make without any further assumption on the missingness or censoring mechanism. Some more discussions about missingness or censoring mechanism can be found in Rathouz (2007).

In Section 2, $R = 1$ if and only if X is completely observed. This means that some incompletely observed data for a multivariate X are not used in our DWE and CWE. Here, we extend our method to the case of components of X have item missingness. To illustrate, consider the case where $X = (X_1, X_2)$ is bivariate. Define $R^{(1,1)}$ as the indicator of observing both X_1 and X_2 , $R^{(1,0)}$ as the indicator of observing X_1 but not X_2 , and $R^{(0,1)}$ as the indicator of observing X_2 but not X_1 . We still assume (4) with $R = (R^{(1,1)}, R^{(1,0)}, R^{(0,1)})$. Let R_i be R from subject i . Then an extended DWE is obtained by solving (9) with R_i replaced by $R_i^{(k,l)}$,

$\pi_i(T_i, Z_i)$ replaced by $\hat{\pi}_1^{(k,l)}(T_i, Z_i)$, and $\Sigma_{i=1}^n$ replaced by $\Sigma_{i=1}^n \Sigma_{(k,l)}$ with $(k, l) = (1, 1), (1, 0), (0, 1)$, where $\hat{\pi}_1^{(k,l)}(T, Z)$ is defined by (11) with R_i replaced by $R_i^{(k,l)}$. An extended CWE can be obtained similarly. For a general q -dimensional X with item missingness, there are $2^q - 1$ missingness patterns and our DWE and CWE can be extended with R_i defined to be the $(2^q - 1)$ -dimensional indicator vector for missingness patterns. Of course, this extension may be infeasible if q is large or there are not enough subjects in one particular missingness pattern, in which case more research is needed.

Our approach is inverse probability weighting, and our contribution is to overcome the difficulty in constructing appropriate weights under assumption (4). An imputation approach under assumption (4) may be applied but further research is needed. A likelihood approach

may not work under assumption (4) without an additional assumption such as a parametric form on the missingness propensity, because assumption (4) is not a missing at random assumption due to censoring. Under a missing at random assumption such as (3), a likelihood approach may be applied but it requires a correct specification of covariate distribution (Chen and Little, 1999) or a stronger assumption on censoring (e.g., Lipsitz and Ibrahim, 1998; Herring and Ibrahim, 2001).

If $\mathbf{Z}$ has high dimensional continuous components, dimension reduction or variable selection can be applied prior to the kernel estimation (11). For example, if all components of $\mathbf{Z}$ are continuous, then applying an existing dimension reduction or variable selection method leads to $\pi_1(T, \mathbf{Z}) = \pi_1(T, \phi(\mathbf{Z}))$ with $\phi(\mathbf{Z})$ being a low dimensional linear function or subset of $\mathbf{Z}$. Then (11) can be applied with $\mathbf{Z}$ replaced by $\phi(\mathbf{Z})$.

When $\mathbf{V}$ in (1)–(2) is replaced by a time-varying covariate vector $\mathbf{V}(t) = (\mathbf{X}(t), \mathbf{Z}(t))$, limited publications about handling missing time-varying $\mathbf{X}(t)$ values can be found, under very strong assumptions on missingness. For example, Lin and Ying (1993) assumed missing $\mathbf{X}(t)$ is completely at random; Paik and Tsai (1997) assume that

$$R(t) \perp \mathbf{X}(t) | (T \wedge C, \mathbf{Z}(t), \delta), \quad (22)$$

where $R(t)$ is the indicator of whether $\mathbf{X}(t)$ is completely observed at time t . Assumption (22) is strong because it ignores the possible effect of $\mathbf{Z}(s)$, $s < t$, on $R(t)$.

We now discuss the possibility of extension of our work to time-dependent covariates. If $\mathbf{V}(t) = (\mathbf{X}, \mathbf{Z}(t))$, i.e., the always observed $\mathbf{Z}(t)$ is time-dependent but $\mathbf{X}$ having missing values is time-independent, then our DWE and CWE can be extended under two assumptions. The first one is to replace (4) by

$$R \perp (\mathbf{X}, \{\mathbf{Z}(t), t > 0\}, C) | (T, \mathbf{Z}(0)),$$

i.e., the missingness of baseline $\mathbf{X}$ depends only on T and the baseline $\mathbf{Z}(0)$. The second one is the main assumption (1) in Robins and Finkelstein (2000), i.e., conditional on the recorded history $\mathbf{V}(s)$, $s \leq t$, the hazard of censoring C at time t does not further depend on the possibly unobserved T . If the covariate vector with missing values is time-dependent, then the missingness mechanism can be very complicated, since missing $\mathbf{X}(t)$ at time t may depend on all history information up to time t on covariates, survival, and censoring. One possible modification of assumption (4) is

$$R(t) \perp (\{\mathbf{X}(s), s \geq 0\}, C) | (T, \{\mathbf{Z}(s), s \geq 0\}).$$

Along with the censoring assumption (1) in Robins and Finkelstein (2000), estimation equations (9) and (14) are still valid with $\pi_1(T_i, \mathbf{Z}_i)$, $\psi(T_i, \mathbf{V}_i)$, and $\pi_0(C_i, \mathbf{V}_i)$ replaced by $\pi_1(T_i, \{\mathbf{Z}_i(s), s \geq 0\})$, $\psi(T_i, \mathbf{V}_i(t))$, and $\pi_0(C_i, \mathbf{X}_i(t), \{\mathbf{Z}_i(s), s \geq 0\})$, respectively. Obviously the survival-time-dependent time-varying missingness propensity $\pi_1(T_i, \{\mathbf{Z}_i(s), s \geq 0\})$ is difficult to estimate although it is an interesting problem. A more careful research is needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We are grateful to the associate editor and four referees for comments and suggestions that led to significant improvements of the paper. The authors' research was partially supported by the National Natural Science Foundation of China grant 11831008, the U.S. National Science Foundation grants DMS-1612873 and DMS-1914411, the University of Wisconsin Carbone Cancer Center Support Grant [P30 CA014520] from the U.S. National Institute of Health (NIH), and the University of Wisconsin Head and Neck Specialized Program of Research Excellence grant [P50 DE026787] from NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

- Andersen PK and Gill RD. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist* 10, 1100–1120.
- Chen HY. (2002). Double-semiparametric method for missing covariates in cox regression models. *Journal of the American Statistical Association* 97, 565–576.
- Chen HY and Little RJA. (1999). Proportional hazards regression with missing covariates. *Journal of the American Statistical Association* 94, 896–908.
- Cook VJ, Hu XJ, and Swartz TB. (2011). Cox regression with covariates missing not at random. *Statistics in Biosciences* 3, 208–222.
- Cox DR. (1975). Partial likelihood. *Biometrika* 62, 269–276.
- Fleming TR and Harrington D. (1991). *Counting Processes & Survival Analysis*, volume 157 John Wiley & Sons, Inc.
- Herring AH and Ibrahim JG. (2001). Likelihood-based methods for missing covariates in the cox proportional hazards model. *Journal of the American Statistical Association* 96, 292–302.
- Kumar SS, Higgins KA, and McGarry RC. (2017). Emerging therapies for stage iii non-small cell lung cancer: Stereotactic body radiation therapy and immunotherapy. *Frontiers in Oncology* 7, 197. [PubMed: 28929083]
- Lin DY and Ying Z. (1993). Cox regression with incomplete covariate measurements. *Journal of the American Statistical Association* 88, 1341–1349.
- Lipsitz SR and Ibrahim JG. (1998). Estimating equations with incomplete categorical covariates in the cox model. *Biometrics* 54, 1002–1013. [PubMed: 9750248]
- Paik MC and Tsai W-Y. (1997). On using the cox proportional hazards model with missing covariates. *Biometrika* 84, 579–593.
- Qi L, Wang CY, and Prentice RL. (2005). Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association* 100, 1250–1263.
- Racine J and Li Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics* 119, 99–130.
- Rathouz PJ. (2007). Identifiability assumptions for missing covariate data in failure time regression models. *Biostatistics* 8, 345–356. [PubMed: 16840561]
- Robins JM. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine* 16, 21–37. [PubMed: 9004381]
- Robins JM and Finkelstein DM. (2000). Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. *Biometrics* 56, 779–788. [PubMed: 10985216]
- Rubin DB. (1976). Inference and missing data. *Biometrika* 63, 581–592.
- Shao J. (2003). *Mathematical Statistics* Springer.
- Teran MD and Brock MV. (2014). Staging lymph node metastases from lung cancer in the mediastinum. *Journal of Thoracic Disease* 6,.

- Wang CY and Chen HY. (2001). Augmented inverse probability weighted estimator for cox missing covariate regression. *Biometrics* 57, 414–419. [PubMed: 11414564]
- Xu Q, Paik MC, Luo X, and Tsai W-Y. (2009). Reweighting estimators for cox regression with missing covariates. *Journal of the American Statistical Association* 104, 1155–1167.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Directed acyclic graph for a scenario where (5) holds

Table 1Simulation bias, SD, SE, and CP based on 2000 runs for estimation of θ under setting 1

n	Method	Variables used in $\widehat{\pi}_1$	Estimation of θ_x				Estimation of θ_z			
			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
	SWE		0.094	0.112	0.118	0.849	0.114	0.187	0.194	0.904
	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	0.940	−0.001	0.093	0.094	0.952
	CC		0.146	0.083	0.082	0.467	0.146	0.138	0.140	0.777
	SWE	$T \wedge C, Z$	0.088	0.082	0.081	0.760	0.108	0.131	0.132	0.852
	DWE	T, Z	−0.014	0.092	0.091	0.946	−0.009	0.154	0.153	0.945
	CWE	T, Z	0.012	0.083	0.081	0.942	0.006	0.123	0.128	0.957
Covariate Vector:		$V = (X, Z), X \sim N(0, 1), Z \sim \text{binary}(0.5), X \perp Z$								
True hazard of T :		$\lambda(t V) = \exp(\theta_x X + \theta_z Z), \theta = (\theta_x, \theta_z) = (1, 1)$								
True propensity:		$\pi_1(T, Z) = 1 - \{1 + \exp(T - 0.5)\}^{-1}$								
True censoring:		$\psi(T, V) = S_{C V}(T) = \exp(-T^{1/2})$								
Variables used in $\widehat{\psi}$:		entire V								
Censoring and missing rate:		$\delta = 0$	$R = 0$	$R = 1, \delta = 1$		$R = 1, \delta = 0$		$R = 0, \delta = 1$		$R = 0, \delta = 0$
		0.476	0.448	0.249		0.303		0.275		0.173
Unconditional quantile of T :		25%	50%	75%						
		0.122	0.382	1.096						

Table 2Simulation bias, SD, SE, and CP based on 2000 runs for estimation of θ under setting 2

<i>n</i>	Method	Variables used in $\widehat{\pi}_1$	Estimation of θ_x				Estimation of θ_z			
			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
	SWE	$T \wedge C$	0.058	0.149	0.148	0.920	0.127	0.148	0.153	0.858
	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
Covariate Vector:		$V = (X, Z), X \sim \text{binary}(0.5), Z \sim \text{binary}(0.5), X \perp Z$								
True hazard of T :		$\lambda(t V) = \exp(\theta_x X + \theta_z Z), \theta = (\theta_x, \theta_z) = (1, 1)$								
True propensity:		$\pi_1(T, Z) = 1 - \{1 + \exp(T + Z)\}^{-1}$								
True censoring:		$\psi(T, V) = S_C V(T) = \exp(-T^{1/2}\exp(-X/4))$								
Variables used in $\widehat{\psi}$:		entire V								
Censoring and missing rate:		$\delta = 0$	$R = 0$	$R = 1, \delta = 1$		$R = 1, \delta = 0$		$R = 0, \delta = 1$		$R = 0, \delta = 0$
		0.370	0.294	0.434		0.272		0.195		0.098
Unconditional quantile of T :		25%	50%	75%						
		0.090	0.240	0.573						

Table 3

Simulation bias, SD, SE, and CP based on 2000 runs for estimation of θ under setting 3, $n = 1000$

α	Method	Estimation of θ_x				Estimation of θ_{z1}				Estimation of θ_{z2}			
		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
Covariate vector:		$V = (X, Z_1, Z_2), X Z_1, Z_2 \sim \text{logit}(0.5Z_1 - Z_2), Z_1 \sim \text{binary}(0.5), Z_2 \sim \text{uniform}(0, 0.5), Z_1 \perp Z_2$											
True hazard of T :		$\lambda(t V) = 0.5t \exp(\theta_x X + \theta_{z1} Z_1 + \theta_{z2} Z_2), \theta = (\theta_x, \theta_{z1}, \theta_{z2}) = (1, 1, 1)$											
True propensity:		$\pi_1(T, V) = 1 - \{1 + \exp(2T - Z_2 - 1.5)\}^{-1}$											
True censoring:		$\psi(T, V) = S_C V(T) = \exp\{-(\alpha T)^{1/2} \exp(0.2X + 0.1Z_1)\}$ α determines the censoring proportion Variables used in ψ : entire V											
Variables used in $\hat{\psi}$		entire V											
Variables used in $\hat{\pi}_1$		$(T \wedge C, Z_1, Z_2)$ for SWE and (T, Z_1, Z_2) for DWE and CWE											
Censoring and missing rate:		$\delta = 0$	$R = 0$	$R = 1, \delta = 1$		$R = 1, \delta = 0$		$R = 0, \delta = 1$		$R = 0, \delta = 0$			
$\alpha = 1$		0.602	0.469	0.157		0.374		0.241		0.228			
$\alpha = 2$		0.815	0.469	0.058		0.473		0.127		0.342			
Unconditional quantile of T :		25%		50%		75%							
		0.522		0.860		1.349							

Table 4

Estimates of coefficients of covariates in Cox regression with stage III NSCLC data

Method	$X = \text{stage record}$ $0 = \text{stage IIIA}$ $1 = \text{stage IIIB}$		$Z_1 = \text{treatment}$ $0 = \text{surgery}$ $1 = \text{SBRT}$		$Z_2 = \text{age}$		$Z_3 = \text{gender}$ $0 = \text{male}$ $1 = \text{female}$	
	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
CC-SWE	0.021	0.021	-0.016	0.018	-0.017	0.004	0.001	0.023
SWE-CWE	0.019	0.070	0.106	0.058	-0.007	0.022	-0.166	0.065
Censoring and missing rate								
$\delta = 0$	$R = 0$	$R = 1, \delta = 1$	$R = 1, \delta = 0$	$R = 0, \delta = 1$	$R = 0, \delta = 0$			
0.385	0.594	0.262	0.144	0.353	0.241			
			25%	50%	75%			
Observed marginal quantile of T when $\delta = 1$			18.1	38.6	66.1			
Observed marginal quantile of $T \wedge C$			27.3	62.4	91.5			