Supporting Information for Inverse probability weighting methods for Cox regression with right-truncated data

Bella Vakulenko-Lagun*

e-mail: blagun@stat.haifa.ac.il

Micha Mandel†

e-mail: micha.mandel@mail.huji.ac.il

Rebecca A. Betensky‡

e-mail: rebecca.betensky@nyu.edu

October 11, 2019

*Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA; Present address: Department of Statistics, University of Haifa, Mount Carmel, Haifa 31905, Israel

†Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel

‡Department of Biostatistics, College of Global Public Health, 715 Broadway, New York, NY 10003, USA

Web Appendix A Asymptotic properties

Regularity conditions. To prove consistency and derive the asymptotic distribution we need the following regularity conditions:

A. (T_i, R_i, Z_i), $i = 1, ..., n$ are iid.

B. T is continuous.

C. Covariates $Z_i \in \mathcal{Z}$ are bounded.

D. (positivity of true and estimated weights in the support of T): $S_R(t) > 0$ and $\widehat{S}_R(t) > 0$ for all $t \in (0, \tau)$. Since $S_R(t)$ is a decreasing function, we can equivalently assume only that $S_R(\tau) > 0$ and $\widehat{S}_R(\tau) > 0$.

E. Let β_0 and Λ_0 are the true values of β and the baseline cumulative hazard function, and $s^{(k)}$, $k = 0, 1$, defined as in the Web Appendix A.
The matrix $\Gamma = E\left[-\int_0^\tau \left\{Z_1^* - \frac{s^{(1)}(\beta_0,t)}{s^{(0)}(\beta_0,t)}\right\}^{\otimes 2} I(T_1^* \geq t) S_R(T_1^*)^{-1} \exp(\beta_0 Z_1^*) \mathrm{d}\Lambda_0(t)\right]$ is positive definite.

The first three conditions impose standard requirements on the population variables. Condition D is the positivity assumption required for consistency. In order to eliminate the problem of zero estimated weights $\widehat{S}_R(t)$, we suggest to use $\widehat{S}_R(t) = \exp(-\widehat{\Lambda}_R(t))$ instead of the Kaplan-Meier estimator for $S_R(t)$, where $\widehat{\Lambda}_R(t)$ is a Nelson-Aalen estimator of the cumulative hazard function of R, defined below. Condition E is required for β_0 to be a unique solution of a limit of the estimating equation (3), needed in the proof of consistency of $\widehat{\beta}$, and for the proof of asymptotic normality and estimation of variance of $\widehat{\beta}$.

Definitions. Let

$$S_n^{(l)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} I(T_i^* \geq t) \{S_R(T_i^*)\}^{-1} Z_i^{*\otimes l} \exp(\beta Z_i^*),$$

$$\widetilde{S}_n^{(l)}(\beta, t) = \frac{1}{n} \sum_{i=1}^{n} I(T_i^* \geq t) \{\widehat{S}_R(T_i^*)\}^{-1} Z_i^{*\otimes l} \exp(\beta Z_i^*),$$

and define the population based quantity $s^{(l)}(\beta, t) = E\left\{I(T \geq t) Z^{\otimes l} \exp(\beta Z)\right\}$, $l = 0, 1, 2$.

Consistency. The proof is similar to the proof of consistency in Mandel et al. (2018). First, we construct a pseudo-log-likelihood function from the estimating equation (3):

$$PL(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[(\beta-\beta_0)Z_i^* - \log\left\{\frac{\widetilde{S}_n^{(0)}(\beta, T_i^*)}{\widetilde{S}_n^{(0)}(\beta_0, T_i^*)}\right\}\right].$$

Taking the second derivative of $PL(\beta)$ and noticing that the Hessian of $PL(\beta)$ is a negative definite matrix, we conclude that $PL(\beta)$ is a concave function with a unique maximum at $\widehat{\beta}$ (with probability tending to one).

Second, we need to show that $PL(\beta)$ converges pointwise to a nonrandom function of β, which is concave with a unique maximum at β_0. We will show this by proving that $PL(\beta)$ is asymptotically equivalent to

$$\frac{1}{n}\sum_{i=1}^{n}\left[(\beta-\beta_0)Z_i^* - \log\left\{\frac{s^{(0)}(\beta, T_i^*)}{s^{(0)}(\beta_0, T_i^*)}\right\}\right],$$

which is a sum of iid terms that, by the weak law of large numbers, converges (for each given β) to

$$\mathcal{L}(\beta) = E\left[(\beta-\beta_0)Z_1^* - \log\left\{\frac{s^{(0)}(\beta, T_1^*)}{s^{(0)}(\beta_0, T_1^*)}\right\}\right].$$

To prove this asymptotic equivalence, rewrite $PL(\beta)$ as follows

$$\begin{aligned} PL(\beta) = {} & \frac{1}{n}\sum_{i=1}^{n}\left[(\beta-\beta_0)Z_i^* - \log\left\{\frac{s^{(0)}(\beta, T_i^*)}{s^{(0)}(\beta_0, T_i^*)}\right\}\right] \\ & - \frac{1}{n}\sum_{i=1}^{n}\left[\log\left\{\frac{S_n^{(0)}(\beta, T_i^*)}{S_n^{(0)}(\beta_0, T_i^*)}\right\} - \log\left\{\frac{s^{(0)}(\beta, T_i^*)}{s^{(0)}(\beta_0, T_i^*)}\right\}\right] \quad (A.1) \\ & - \frac{1}{n}\sum_{i=1}^{n}\left[\log\left\{\frac{\widetilde{S}_n^{(0)}(\beta, T_i^*)}{\widetilde{S}_n^{(0)}(\beta_0, T_i^*)}\right\} - \log\left\{\frac{S_n^{(0)}(\beta, T_i^*)}{S_n^{(0)}(\beta_0, T_i^*)}\right\}\right]. \quad (A.2) \end{aligned}$$

Then, to prove that (A.1) is $o_p(1)$, rewrite (A.1) as

$$\frac{1}{n}\sum_{i=1}^{n}\left[\log\left\{\frac{S_n^{(0)}(\beta, T_i^*)}{s^{(0)}(\beta, T_i^*)}\right\} - \log\left\{\frac{S_n^{(0)}(\beta_0, T_i^*)}{s^{(0)}(\beta_0, T_i^*)}\right\}\right] \quad (A.3)$$

We need to show that each of the sums in (A.3) is $o_p(1)$. Using regularity conditions C and D we can show that for each β,

$$\sup_{t\in[0,\tau]} \log\left\{\frac{S_n^{(0)}(\beta,t)}{s^{(0)}(\beta,t)}\right\} = o_p(1). \tag{A.4}$$

Indeed (A.4) follows from the uniform convergence of a weighted sum $S_n^{(0)}(\beta,t)$ to its population counterpart $s^{(0)}(\beta,t)$ for each β:

$$\sup_{t\in[0,\tau]} \left\{S_n^{(0)}(\beta,t) - s^{(0)}(\beta,t)\right\} \xrightarrow{P} 0,$$

that holds by the uniform law of large numbers (e.g. Jennrich, 1969, Theorem 2) under conditions C and D. The latter condition of positive weights in the whole support of T is crucial for getting

$$E\left\{\frac{I(T_1^* \geq t)\exp(\beta Z_1^*)}{S_R(T_1^*)}\right\} = s^{(0)}(\beta,t),$$

which is the key condition in the proof of consistency. Here we also use that without loss of generality we can divide $S_R(T)$ by a constant c so that $E(S_R(T)/c) = 1$.

To prove that (A.2) is negligible, as previously, we shall work separately with the part of (A.2) involving only β and the part of (A.2) involving only β_0. Thus, we need to show that for each β

$$\frac{1}{n}\sum_{i=1}^{n}\left[\log\left\{\frac{\widetilde{S}_n^{(0)}(\beta,T_i^*)}{S_n^{(0)}(\beta,T_i^*)}\right\}\right] = o_p(1).$$

This follows from

$$\sup_{t\in[0,\tau]} \log\left\{\frac{\widetilde{S}_n^{(0)}(\beta,t)}{S_n^{(0)}(\beta,t)}\right\} = o_p(1). \tag{A.5}$$

Proving (A.5) is equivalent to showing that

$$\sup_{t\in[0,\tau]} \frac{\widetilde{S}_n^{(0)}(\beta,t) - S_n^{(0)}(\beta,t)}{S_n^{(0)}(\beta,t)} = o_p(1). \tag{A.6}$$

The ratio in (A.6) equals

$$\frac{\sum_{i=1}^{n} I(T_i^* \geq t)\exp(\beta Z_i^*)\left\{S_R(T_i^*)\right\}^{-1}\left\{\frac{S_R(t)}{\widehat{S}_R(t)} - 1\right\}}{\sum_{i=1}^{n} I(T_i^* \geq t)\exp(\beta Z_i^*)\left\{S_R(T_i^*)\right\}^{-1}} \leq \sup_{t\in[0,\tau]}\left|\frac{S_R(t)}{\widehat{S}_R(t)} - 1\right| = o_p(1),$$

which follows from condition D and the uniform convergence of the Kaplan-Meier estimator for truncated data (Woodroofe, 1985):

$$\sup_{t\in[0,\tau]}\left|\widehat{S}_R(t) - S_R(t)\right| = o_P(1).$$

Now, $\mathcal{L}(\beta)$ is concave with a unique maximum at β_0 since $\frac{\partial\mathcal{L}(\beta_0)}{\partial\beta} = 0$ and $\frac{\partial^2\mathcal{L}(\beta_0)}{\partial\beta^2} = -\Gamma$, and Γ is positive definite by condition E.

Finally, by Theorem II.1 in Andersen and Gill (1982), *pointwise convergence in probability of $PL(\beta)$ implies uniform convergence of $PL(\beta)$ on a compact subspace for β.* Then, by Corollary II.2 in Andersen and Gill (1982), the uniform convergence of $PL(\beta)$ and the unique maximum of the limit implies that $\|\widehat{\beta} - \beta_0\| \xrightarrow{P} 0$.

Asymptotic distribution. The asymptotic distribution of $\widehat{\beta}$ can be derived as follows. First, using Taylor series expansion of $\widetilde{U}(\widehat{\beta})$ around β_0 we get

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \left\{\Gamma(\beta^*)\right\}^{-1}\frac{1}{\sqrt{n}}\widetilde{U}(\beta_0) + o_p(1)$$

where β^* is on the line segment between $\widehat{\beta}$ and β_0, and

$$\Gamma(\beta) = -\frac{1}{n}\frac{\partial\widetilde{U}(\beta)}{\partial\beta}.$$

Then, we can represent $\frac{1}{\sqrt{n}}\widetilde{U}(\beta_0)$ as a sum

$$\begin{aligned}\frac{1}{\sqrt{n}}\widetilde{U}(\beta_0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left\{Z_i - \frac{S_n^{(1)}(\beta_0,t)}{S_n^{(0)}(\beta_0,t)}\right\}\mathrm{d}N_i(t) \\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left\{\frac{S_n^{(1)}(\beta_0,t)}{S_n^{(0)}(\beta_0,t)} - \frac{\widetilde{S}_n^{(1)}(\beta_0,t)}{\widetilde{S}_n^{(0)}(\beta_0,t)}\right\}\mathrm{d}N_i(t) = I_1 + I_2,\end{aligned} \tag{A.7}$$

where $N_i(t) = I\{T_i^* \le t\} = I\{T_i \le t, T_i \le R_i\}$ is a counting process for the right-truncated lifetime T_i^*.

For the first part of (A.7), using the same technique like in Lin and Wei (1989) or Qin and Shen (2010) we can show that I_1 is asymptotically equivalent to a sum of n iid random vectors with zero mean and finite variance plus terms that converge in probability to 0:

$$\begin{aligned} I_1 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_i - \frac{S_n^{(1)}(\beta_0, t)}{S_n^{(0)}(\beta_0, t)} \right\} \mathrm{d}N_i(t) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \left\{ Z_i - \frac{s^{(1)}(\beta_0, t)}{s^{(0)}(\beta_0, t)} \right\} \mathrm{d}M_i(t) + o_p(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i(\beta_0) + o_p(1), \end{aligned}$$

where $M_i(t)$ is a mean zero stochastic process defined by $M_i(t) = N_i(t) - \int_0^t S_R(s) I(T_i^* \ge s)\{S_R(T_i^*)\}^{-1} \lambda(s \mid Z_i^*) \mathrm{d}s$, and $V_i(\beta_0) = \int_0^{\tau} \left\{ Z_i - \frac{s^{(1)}(\beta_0,t)}{s^{(0)}(\beta_0,t)} \right\} \mathrm{d}M_i(t)$ $(i = 1, ..., n)$ are iid with zero mean and finite variance.

For the second term of (A.7), it holds that

$$\begin{aligned} I_2 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{\tau} \left\{ \frac{S_n^{(1)}(\beta_0, t)}{S_n^{(0)}(\beta_0, t)} - \frac{\widetilde{S}_n^{(1)}(\beta_0, t)}{\widetilde{S}_n^{(0)}(\beta_0, t)} \right\} \mathrm{d}N_i(t) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\sum_{k=1}^{n} I(T_k^* \ge T_i^*) Z_k^* \exp(\beta_0 Z_k^*) \left\{ \frac{\widehat{S}_R(T_k^*) - S_R(T_k^*)}{S_R^2(T_k^*)} - \frac{\widehat{S}_R(T_i^*) - S_R(T_i^*)}{S_R(T_k^*) S_R(T_i^*)} \right\}}{\sum_{k=1}^{n} I(T_k^* \ge T_i^*) \exp(\beta_0 Z_k^*) \frac{1}{S_R(T_k^*)}} + o_p(1). \end{aligned}$$

Then, we use the following iid representation of the Kaplan-Meier estimator for left-truncated data (Wang, 1991):

$$\sqrt{n} \frac{\widehat{S}_R(t) - S_R(t)}{S_R(t)} = -\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \xi(T_j^*, R_j^*, t) + o_p(1),$$

where

$$\begin{aligned} \xi(T^*, R^*, t) &= \frac{I(R^* \le t)}{Q(t)} + \int_0^t \frac{I(R^* \le u)}{Q^2(u)} \mathrm{d}Q(u) - \int_0^t \frac{I(T^* \le u \le R^*)}{Q^2(u)} \mathrm{d}K(u), \\ Q(u) &= P(T^* \le u \le R^*), \\ K(u) &= P(R^* \le u). \end{aligned}$$

and obtain

$$I_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\sum_{k=1}^{n} I(T_k^* \geq T_i^*) Z_k^* \exp(\beta_0 Z_k^*) \frac{1}{S_R(T_k^*)} \frac{1}{n} \sum_{j=1}^{n} \left\{ \xi(T_j^*, R_j^*, T_i^*) - \xi(T_j^*, R_j^*, T_k^*) \right\}}{n S_n^{(0)}(\beta_0, T_i^*)} + o_p(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a(\beta_0, T_i^*, R_i^*) + o_p(1),$$

where

$$a(\beta_0, t, r) = \lim_{n\to\infty} \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{I(T_k^* \geq T_j^*) Z_k^* \exp(\beta_0 Z_k^*) \frac{S_R(T_j^*)}{S_R(T_k^*)} \left\{ \xi(t, r, T_j^*) - \xi(t, r, T_k^*) \right\}}{S^{(0)}(\beta_0, T_j^*)},$$

So, we can represent the normalized score (A.7) by the sum of iid random vectors with zero expectation:

$$\frac{1}{\sqrt{n}} \widetilde{U}(\beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\int_0^\tau \left\{ Z_i - \frac{s^{(1)}(\beta_0, t)}{s^{(0)}(\beta_0, t)} \right\} \mathrm{d}M_i(t) + a(\beta_0, T_i^*, R_i^*) \right] + o_p(1).$$

Then, by multivariate central limit theorem $\frac{1}{\sqrt{n}} \widetilde{U}(\beta_0) \rightsquigarrow N(0, \Sigma)$ with

$$\Sigma(\beta) = \frac{1}{n} \sum_{i=1}^{n} \{ V_i(\beta) + a(\beta, T_i^*, R_i^*) \}^{\otimes 2}$$

and $\Sigma = \lim_{n\to\infty} \Sigma(\beta_0)$. Therefore, $n^{-1/2}(\widehat{\beta} - \beta_0)$ is asymptotically normal with zero mean and covariance matrix $\Gamma^{-1}\Sigma\Gamma^{-1}$.

Web Appendix B Estimation of SE

Covariance matrix of $\widehat{\beta}$ can be consistently estimated by $\widehat{\Gamma}^{-1}\widehat{\Sigma}\widehat{\Gamma}^{-1}$, where

$$\widehat{\Gamma} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\widetilde{S}_n^{(2)}(\widehat{\beta}, T_i^*)}{\widetilde{S}_n^{(0)}(\widehat{\beta}, T_i^*)} - \frac{\widetilde{S}_n^{(1)}(\widehat{\beta}, T_i^*)}{\widetilde{S}_n^{(0)}(\widehat{\beta}, T_i^*)} \left\{ \frac{\widetilde{S}_n^{(1)}(\widehat{\beta}, T_i^*)}{\widetilde{S}_n^{(0)}(\widehat{\beta}, T_i^*)} \right\}^T \right],$$

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{k=1}^{n} \left\{ Z_i - \frac{\widetilde{S}_n^{(1)}(\widehat{\beta}, T_k^*)}{\widetilde{S}_n^{(0)}(\widehat{\beta}, T_k^*)} \right\} \mathrm{d}\widehat{M}_i(T_k^*) + \widehat{a}(\widehat{\beta}, T_i^*, R_i^*) \right]^{\otimes 2},$$

$$\widetilde{S}_n^{(l)}(\beta,t)=\frac{1}{n}\sum_{i=1}^{n}I(T_i^*\geq t)\widehat{W}(T_i^*)^{-1}Z_i^{*\otimes l}\exp(\beta Z_i^*),\quad l=0,1,2$$

$$\mathrm{d}\widehat{M}_i(T_k^*)=I(T_k^*=T_i^*)-\frac{\frac{1}{\widehat{S}_R(T_i^*)}\exp(\widehat{\beta}Z_i^*)I(T_i^*\geq T_k^*)}{n\widetilde{S}^{(0)}(\widehat{\beta},T_k^*)},$$

$$\widehat{a}(\widehat{\beta},T_i^*,R_i^*)=\frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{I(T_k^*\geq T_j^*)Z_k^*\exp(\widehat{\beta}Z_k^*)\frac{1}{\widehat{S}_R(T_k^*)}\left\{\widehat{\xi}(T_i^*,R_i^*,T_j^*)-\widehat{\xi}(T_i^*,R_i^*,T_k^*)\right\}}{\widetilde{S}_n^{(0)}(\widehat{\beta},T_j^*)},$$

$$\widehat{\xi}(T_i^*,R_i^*,t)=\frac{I(R_i^*\leq t)}{\widehat{Q}(t)}+\frac{1}{n}\sum_{q=1}^{n}\left\{\frac{I(R_i^*\leq T_q^*\leq t)}{\widehat{Q}^2(T_q^*)}-\frac{I(R_i^*\leq R_q^*\leq t)}{\widehat{Q}^2(R_q^*)}-\frac{I(T_i^*\leq R_q^*\leq R_i^*)I(R_q^*\leq t)}{\widehat{Q}^2(R_q^*)}\right\},$$

$\widehat{S}_R(t)=\exp\left\{-\widehat{\Lambda}_R(t)\right\}$, $\widehat{\Lambda}_R(t)=\int_0^t\frac{\mathrm{d}\widehat{K}(u)}{\widehat{Q}(u)}$ is the Nelson-Aalen estimator of the cumulative hazard for left-truncated R, $\widehat{Q}(s)=\frac{1}{n}\sum_{k=1}^{n}I(T_k^*\leq s\leq R_k^*)$ and $\widehat{K}(s)=\frac{1}{n}\sum_{k=1}^{n}I(R_k^*\leq s)$.

Some remarks on asymptotic standard error estimator.

1. When there is no right truncation ($R=\infty$), all the weights equal 1 and a block of I_2 is zero. In this case the estimating equation (3) reduces to the score equation of the partial likelihood, and the SE reduces to the robust estimator derived by Lin and Wei (1989) for the robust inference under the Cox model when it is probably misspecified. It can be obtained from `coxph` function from `survival` package in `R` with `robust` option. In this case, the naive estimator (inverse of minus Hessian of the partial likelihood) $\widehat{\Gamma}^{-1}$ is also a valid estimator of $Var(\widehat{\beta})$, since the robust and naive variance estimators are asymptotically equivalent here.

2. When there is right truncation and the weights W are known, we can correctly estimate the variance by using `robust` option and an `offset` in `coxph` function

   ```
   > coxph(Surv(T,rep(1,n)) ~ Z1 + Z2 + offset(-log(W)), robust=TRUE)
   ```

   correctly assuming by this that $I_2=0$, and the standard `coxph` function with `robust` option will result in a valid sandwich estimator that accounts for biased sampling, but not for the estimation of weights.

3. When there is right truncation and the weights are unknown, the standard `coxph` function

with `robust` option incorrectly assumes that $I_2 = 0$ and, hence, does not account for the variability created by estimation of weights. Such ignoring I_2 for smaller samples ($n \leq 1000$) can lead to considerable underestimation of $Var(\widehat{\beta})$ which leads to tests with supranominal size and confidence intervals with poor coverage probability.

4. Slight overestimation of the asymptotic variance using our suggested estimator $\widehat{\Gamma}^{-1}\widehat{\Sigma}\widehat{\Gamma}^{-1}$ may happen due to near-zero weights, and this reflects the true instability of IPW estimators. This overestimation comes from I_1 part even for large samples in the settings characterized by high likelihood of near-zero weights.

Web Appendix C Hypothesis testing using IPW-S without positivity

Here we show that the IPW method that solves equation (3) in the main text still can be used for global and partial null hypotheses testing for β when the positivity assumption does not hold and the two covariates $Z = (Z_1, Z_2)$ are independent.

Proof of Proposition 2. The joint distribution of (T^*, Z^*) is

$$f_{T^*,Z^*}(t,z) = \frac{P(R > t) f_{T|Z}(t \mid z) dF_Z(z)}{\int_s \int_{u=0}^{\infty} P(R > u) f_{T|Z}(u \mid s) dF_Z(s) du} = \frac{P(R > t) f_{T|Z}(t \mid z) dF_Z(z)}{P(R > T)}.$$

The limit of the estimating equation (2) is given by

$$E_{T_i^*} E_{Z_i^*}(Z_i^* \mid T_i^*) - E_{T_i^*,Z_i^*}\left[\frac{E_{T_j^*,Z_j^*|T_i^*}\left\{Z_j^* \exp(\beta^* Z_j^*) S_R(T_j^*)^{-1} I(T_j^* \geq T_i^*)\right\}}{E_{T_j^*,Z_j^*|T_i^*}\left\{\exp(\beta^* Z_j^*) S_R(T_j^*)^{-1} I(T_j^* \geq T_i^*)\right\}}\right] = 0, \quad \text{(C.1)}$$

where β^* is a solution of it.

Using the density f_{T^*,Z^*} and writing down the expectations in (C.1) we get

$$\int_{t=0}^{\tau \wedge r_*} \int_z z P(R > t) f(t \mid z) dF_Z(z) dt$$
$$- \int_{t=0}^{\tau \wedge r_*} \int_z \frac{\int_s \int_u s \exp(\beta^* s) I(u > t) f(u \mid s) dF_Z(s) du}{\int_s \int_u \exp(\beta^* s) I(u > t) f(u \mid s) dF_Z(s) du} P(R > t) f(t \mid z) dF_Z(z) dt = 0.$$

After some simplification the estimating equation becomes

$$\int_{t=0}^{\tau\wedge r_*}\int_z zP(R>t)f(t\mid z)dF_Z(z)dt \tag{C.2}$$
$$-\int_{t=0}^{\tau\wedge r_*}\int_z \frac{\int_s s\exp(\beta^* s)(S(t\mid s)-S(\tau\wedge r_*\mid s))dF_Z(s)}{\int_s \exp(\beta^* s)(S(t\mid s)-S(\tau\wedge r_*\mid s))dF_Z(s)}P(R>t)f(t\mid z)dF_Z(z)dt=0.$$

For partial testing of the hypothesis $H_0:\beta_1=0$, under H_0 the first equation corresponding to the first component of the vector of covariates is given by

$$\begin{aligned}&\int_{t=0}^{\tau\wedge r_*}\int_z z_1P(R>t)h_0(t)\frac{\exp(\beta_2 z_2)}{S(t\mid 0)^{\exp(\beta_2 z_2)}}dF_Z(z)dt\\&-\int_{t=0}^{\tau\wedge r_*}\int_z \frac{\int_s s_1\exp(\beta_1^* s_1+\beta_2^* s_2)\left\{S(t\mid 0)^{\exp(\beta_2 s_2)}-S(\tau\wedge r_*\mid 0)^{\exp(\beta_2 s_2)}\right\}dF_Z(s)}{\int_s \exp(\beta_1^* s_1+\beta_2^* s_2)\left\{S(t\mid 0)^{\exp(\beta_2 s_2)}-S(\tau\wedge r_*\mid 0)^{\exp(\beta_2 s_2)}\right\}dF_Z(s)}\\&\times P(R>t)h_0(t)\frac{\exp(\beta_2 z_2)}{S(t\mid 0)^{\exp(\beta_2 z_2)}}dF_Z(z)dt=0.\end{aligned} \tag{C.3}$$

Denote by $A(F_{Z_2})=\int_{t=0}^{\tau\wedge r_*}P(R>t)h_0(t)\int_{z_2}\frac{\exp(\beta_2 z_2)}{S(t|0)^{\exp(\beta_2 z_2)}}dF_{Z_2}(z_2)dt$. Then the equation can be written as

$$A(F_{Z_2})\left(E(Z_1)-\frac{\int_{s_1}s_1\exp(\beta_1^* s_1)dF_{Z_1}(s_1)}{\int_{s_1}\exp(\beta_1^* s_1)dF_{Z_1}(s_1)}\right)=0.$$

Then $A(F_{Z_2})>0$ implies that

$$E(Z_1)-\frac{\int_{s_1}s_1\exp(\beta_1^* s_1)dF_{Z_1}(s_1)}{\int_{s_1}\exp(\beta_1^* s_1)dF_{Z_1}(s_1)}=0,$$

which is solved by $\beta_1^*=0$, since $Cov(e^{\beta Z_1},Z_1)=0$ holds only for $\beta_1^*=0$. □

We conducted a small simulation for the case where two covariates $Z=(Z_1,Z_2)$ are not independent. In this experiment, our aim was to compare IPW-S to the likelihood method of Finkelstein et al. (1993). We recreated their setting for two binary correlated covariates, $Z_1\sim Bern(0.5)$ and Z_2 with $P(Z_2=0\mid Z_1=0)=2/3=P(Z_2=1\mid Z_1=1)$, $T\sim Exp(\beta_1Z_1+\beta_2Z_2)$ with $\beta_1=-1$ and $\beta_2=0$, and $R=3-X$ with $X\sim Exp(1)$. This resulted in $P(T>R)=0.34$ and $P(T>r_*\mid Z=0)=0.2$, where $r_*=3$ and $Z=(Z_1,Z_2)$. Although Finkelstein's simulation is based on 500 replications and samples of size $n=800$, we used 2000 replications to reduce variability in the estimated type I error. Following Finkelstein et al. (1993), we tested $H_0 : \beta_2 = 0$ and compared tests based on IPW-S to their Table 4 where they assumed positivity, i.e. $P(T > r_* \mid Z = 0) = 0$. For a 5% level test, the size of the test based on IPW-S is 133/2000=0.0665 (compared to 25/500=0.05 size of Finkelstein et al. (1993)), and for a 1% level of test, the size of the test based on IPW-S is 24/2000=0.012 (compared to 5/500=0.01 size of Finkelstein et al. (1993)). These results demonstrate that IPW-S can be used for testing even in the absence of positivity in some cases that do not satisfy Proposition 2.

Web Appendix D Simulations under positivity

Since in the simulation experiment that uses an unstable scenario of Shen et al. (2017), bootstap estimators of SE do not perform well and IPW-NS has only a bootstrap-based estimator of SE, we cannot compare two IPW approaches in that setting. The aim of this simulation is to compare them in a stable scenario with positivity and no heavy weights.

Data were generated as follows. Truncation times R were simulated from *Gamma* distribution with shape=1.4 and scale=1. Lifetimes T followed Weibull proportional hazards regression $h(t; z) = \frac{\kappa}{\rho}\left(\frac{t}{\rho}\right)^{\kappa-1} \exp(\beta z)$ with shape parameter $\kappa = 2$ and scale parameter $\rho = 1$ and two covariates: $Z_1 \sim N(0, 1)$ and $P(Z_2 = 0 \mid Z_1 < 0) = 2/3 = P(Z_2 = 1 \mid Z_1 > 0)$. The true covariate effects $(\beta_1, \beta_2) = (1, 0)$. These setting resulted in relatively high proportion of truncation (40%), but almost no truncation of support of T, i.e. the average of $P(T > r_{(n)}) = 0.00015$ where r_n is an n^{th}-order statistic of $R_1^*, ..., R_n^*$ for $n = 1000$, meaning that the positivity assumption holds for this case.

Results are shown in Web Table 1. In this setting, the SE estimates and their appropriate CPs for β_1 and sizes of the test $H_0 : \beta_2 = 0$ are similar when using analytic SE or bootstrap distribution. Both IPW methods are unbiased, but IPW-S has lower variance. Looking at relative efficiencies, IPW-S performed on average 1.5 times more efficiently than IPW-NS. The inference based on both IPW methods is valid, but IPW-S is again the recommended method to use. We note also that the overall truncation probability $P(T > R) = 0.4$ is close to that probability in the setting of heavy truncation in the first simulation experiment, $P(T > R) = 0.45$, but this is probably not the most fundamental factor that affects the performance of IPW estimators.

Web Table 1: Comparison of two IPW methods using equations (3) and (4). bias($\widehat{\beta}$) is a bias of $\widehat{\beta}$, SD, SE$_a$, SE$_{bs}$, CP, CP$_{bs}$, $r_{(1000)}$ and $\bar{r}_{(1000)}$ are defined as in Table 1. Size and size$_{bs}$ are, respectively, size of 0.05% test for $H_0 : \beta_2 = 0$ based on normal approximation and SE$_a$, and size of 0.05% test for $H_0 : \beta_2 = 0$ based on bootstrap distribution with 500 replications. RE is the relative efficiency of two IPW methods defined as a ratio of the MSE of IPW-NS to the MSE of IPW-S. Results are based on 1000 replications.

Dependent covariates: $Z_1 \sim N(0,1)$, $P(Z_2 = 0 \mid Z_1 < 0) = 2/3 = P(Z_2 = 1 \mid Z_1 > 0)$
$\mathbf{R} \sim \boldsymbol{\Gamma}(\mathbf{1.4}, \mathbf{1})$, $\mathbf{P}(\mathbf{T} > \mathbf{R}) = \mathbf{0.4}$, average $\mathbf{P}(\mathbf{T} > \mathbf{r}_{(1000)}) = \mathbf{0.00015}$, $\bar{\mathbf{r}}_{(1000)} = \mathbf{9.07}$

		$\beta_1 = 1$							$\beta_2 = 0$						
	n	bias($\widehat{\beta}_1$)	SD	SE$_a$	SE$_{bs}$	CP	CP$_{bs}$	RE	bias($\widehat{\beta}_2$)	SD	SE$_a$	SE$_{bs}$	size	size$_{bs}$	RE
IPW-S	100	0.025	0.185	0.166	0.177	0.924	0.925		-0.017	0.279	0.282	0.283	0.049	0.063	
	300	0.008	0.102	0.097	0.098	0.942	0.943		-0.006	0.165	0.167	0.159	0.049	0.059	
	500	0.005	0.076	0.076	0.076	0.956	0.948		-0.007	0.126	0.129	0.123	0.045	0.062	
	1000	0.001	0.055	0.055	0.054	0.954	0.949		0.000	0.087	0.093	0.088	0.041	0.053	
IPW-NS	100	0.021	0.207		0.194		0.930	1.242	-0.009	0.331		0.319		0.058	1.400
	300	0.003	0.121		0.111		0.943	1.415	-0.001	0.207		0.185		0.066	1.556
	500	0.002	0.095		0.088		0.945	1.567	-0.006	0.164		0.146		0.067	1.681
	1000	-0.001	0.070		0.064		0.942	1.642	0.002	0.117		0.108		0.065	1.817

Web Table 2: Comparison of two types of confidence intervals for IPW-S approach for the setting of Shen et al. (2017) based on 1000 replications. CP is the coverage probability of 95% confidence interval using normal approximation and SE$_a$, and CP$_{bs}$ is the coverage probability of 95% confidence interval $(\widehat{\beta}^{bs}_{0.025}, \widehat{\beta}^{bs}_{0.925})$ based on bootstrap distribution with 500 replications.

		$\beta_1 = 1$		$\beta_2 = 1$	
Light: $\mathbf{P}(\mathbf{T} > \mathbf{R}) = \mathbf{0.16}$, $\mathbf{R} \sim \mathbf{Exp(2.5)}$					
	n	CP	CP$_{bs}$	CP	CP$_{bs}$
IPW-S	100	0.960	0.928	0.946	0.942
	300	0.976	0.95	0.942	0.936
	500	0.980	0.955	0.950	0.938
Moderate: $\mathbf{P}(\mathbf{T} > \mathbf{R}) = \mathbf{0.32}$, $\mathbf{R} \sim \mathbf{Exp(7.5)}$					
	n	CP	CP$_{bs}$	CP	CP$_{bs}$
IPW-S	100	0.975	0.933	0.948	0.940
	300	0.977	0.923	0.943	0.936
	500	0.978	0.928	0.945	0.920
Heavy: $\mathbf{P}(\mathbf{T} > \mathbf{R}) = \mathbf{0.45}$, $\mathbf{R} \sim \mathbf{Exp(15)}$					
	n	CP	CP$_{bs}$	CP	CP$_{bs}$
IPW-S	100	0.970	0.898	0.949	0.923
	300	0.965	0.867	0.949	0.911
	500	0.960	0.858	0.952	0.908

Web Appendix E Analysis of original AIDS data

The original data set, collected by the Centers of Disease Control (CDC), includes 295 patients who were infected with HIV and developed AIDS by June 30, 1986, the time of data extraction.

The maximum observed truncation time in the sample is 8.5 years, meaning that $\widehat{P}(R > t) = 0$ for $t > 8.5$. Despite violation of positivity, we nonetheless try to fit the Cox model by both IPW methods in order to compare results to other methods used for analysis of this data set.

The results of the two methods of Shen et al. (2017) and the two IPW methods are shown in Web Table 3. All four methods give similar results: the hazard of developing AIDS for the youngest children is about 12 times that of senior adults, and the hazard of middle-aged patients is not significantly different from that of senior adults.

Web Table 3: Results of analysis of the original AIDS data set comprising 295 observations, assuming positivity. Comparison to Shen et al. (2016). SE_a is the analytical estimate of SE, SE_{bs} is a bootstrap-based estimate of SE, *pvalue* is calculated using normal approximation and SE_a, and $pvalue_{bs}$ uses bootstrap distribution based on 1000 bootstrap replications.

	covariate	$\widehat{\beta}$	SE_a	SE_{bs}	*pvalue*	$pvalue_{bs}$
Shen-EE	≤ 4y	3.606	0.789		< 0.001	
	$4 - 59$y	0.801	0.963		0.414	
Shen-cMLE	≤ 4y	2.656	0.674		< 0.001	
	$4 - 59$y	0.582	0.792		0.462	
IPW-S	≤ 4y	2.381	0.627	0.390	< 0.001	< 0.001
	$4 - 59$y	-0.605	1.199	0.549	0.613	0.270
IPW-NS	≤ 4y	2.307		0.444		< 0.001
	$4 - 59$y	-0.748		0.741		0.313

Web Appendix F Additional results from analysis of right-truncated AIDS data

Web Table 4: Number of failures in bootstrap replications in the sensitivity analysis of AIDS data based on 116 observations and with one age indicator Z =I(age$\leq$59). FAIL-SA and FAIL-NSA are the number of bootstrap samples, out of 1000 bootstrap replications, that had divergence issues when using IPW-SA or IPW-NSA, respectively.

a_0	FAIL-SA	FAIL-NSA
0-0.35	0	0
0.4	0	4
0.45	0	250
0.5	0	687
0.55	1	829
0.6	0	932
0.65	0	969
0.7	0	984
0.75	0	994
0.8	2	993
0.85	13	991
0.9	39	997

References

P.K. Andersen and R.D. Gill. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10(4):1100–1120, 1982.

D. M. Finkelstein, D. F. Moore, and D. A. Schoenfield. A proportional hazards model for truncated AIDS data. *Biometrics*, 49:731–740, 1993.

R.I. Jennrich. Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2):633–643, 1969.

D. Y. Lin and L. J. Wei. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, 84(408):1074–1078, 1989.

M. Mandel, J. de Uña Álvarez, D. K. Simon, and R. A. Betensky. Inverse probability weighted Cox regression for doubly truncated data. *Biometrics*, 74(2):481–487, 2018. doi: 10.1111/biom.12771.

J. Qin and Y. Shen. Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics*, 66:382–392, 2010.

P.-S. Shen, Y. Liu, D.-P. Maa, and Y. Ju. Analysis of transformation models with right-truncated data. *Statistics*, 51(2):404–418, 2017. doi: http://dx.doi.org/10.1080/02331888.2016.1268617.

M.-C. Wang. Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*, 86:130–143, 1991.

M. Woodroofe. Estimating a distribution function with truncated data. *The Annals of Statistics*, 13(1):163–177, 1985.