Doubly Robust Estimation under Covariate-induced Dependent Left Truncation

Yuyao Wang

Department of Mathematics, University of California San Diego
yuw079@ucsd.edu

Joint work with:

Andrew Ying, Google

Ronghui (Lily) Xu, University of California San Diego

Left truncation and selection bias

- Quantity of interest: time to event (T)
- In prevelant cohort studes:
  - Often only subjects with time to events greater than the enrollment times (Q) are included in the data
  - Subjects with early event times tend not to be captured

Example: CNS lymphoma data

- Data from a study on central nervous system (CNS) lymphoma (Wang et al., 2015) *[Publicly available in the supplement of Vakulenko-Lagun et al. (2022)]*

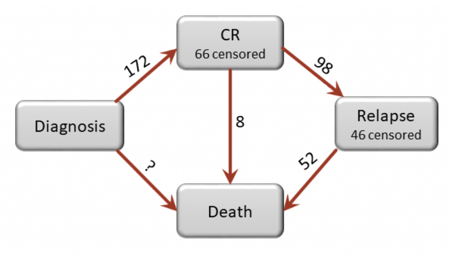
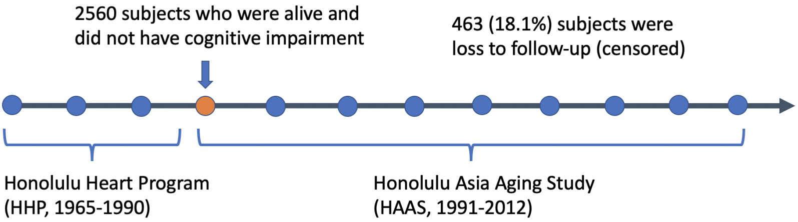


Figure from Vakulenko-Lagun et al., (2022)
CR: complete response.

→ Restrict to the 98 patients that were relapsed, for which the time is recorded.

- Quantity of interest: overall survival
- T - time to death
- For the original data set with 172 patients
  - Q - time to CR
- For the restricted data set with 98 patients
  - Q - time to relapse

Example: HAAS data



- Quantity of interest:
  Cognitive impairment-free survival on the age scale.
- T - age to moderate cognitive impairment or death
- Q - age at entry of HAAS

$\rightarrow$ **Selection bias**

Literature

Under the random left truncation assumption

- Likelihood-based approaches (Woodroofe, 1985; Wang et al., 1986; Wang, 1989, 1991; Qin et al. 2011)
- Random truncation assumption can be weakened to quasi-independence assumption (Tsai, 1990)

! The quasi-independence assumption may be violated.

- CNS lymohoma data:
  - It is plausible that time to death and time to relapse are dependent (Vakulenko-Lagun et al., 2022).
- HAAS data:
  - Violation of quasi-independence is detected by conditional Kendall's tau test (Tsai, 1990);
  - tau $= 0.0426$ with p-value 0.0014.

Literature

When the left truncation time and the event time are dependent:

- Copula models (Chaieb et al., 2006; Emura et al., 2011; Emura & Wang, 2012)
- Structural transformation models (Efron & Petrosian, 1994; Chiou et al., 2019)

! Depend on strong model assumptions

Literature

When the left truncation time and the event time are dependent:

- Copula models (Chaieb et al., 2006; Emura et al., 2011; Emura & Wang, 2012)
- Structural transformation models (Efron & Petrosian, 1994; Chiou et al., 2019)

! Depend on strong model assumptions

- Incorporate left truncation time as a covariate in the event time model (Gail et al., 2009; Mackenzie, 2012; Cheng & Wang 2015).
  - e.g., Entry-age adjusted age-scale model (Gail et al., 2009)

! Biologically unjustified; depend on model assumptions.

Literature

When the left truncation time and the event time are dependent:

- Copula models (Chaieb et al., 2006; Emura et al., 2011; Emura & Wang, 2012)
- Structural transformation models (Efron & Petrosian, 1994; Chiou et al., 2019)

! Depend on strong model assumptions

- Incorporate left truncation time as a covariate in the event time model (Gail et al., 2009; Mackenzie, 2012; Cheng & Wang 2015).
  - e.g., Entry-age adjusted age-scale model (Gail et al., 2009)

! Biologically unjustified; depend on model assumptions.

! Do not use covariate information.

Literature

When the dependence is captured by measured covariates:

In regression settings:

- Cox model with risk set adjustment

For marginal survival probabilities:

- Inverse probability weighting (IPW) estimators (Vakulenko-Lagun et al., 2022).

! Sensitive to misspecification of the truncation model; inefficient.

Literature

When the dependence is captured by measured covariates:

In regression settings:

- Cox model with risk set adjustment

For marginal survival probabilities:

- Inverse probability weighting (IPW) estimators (Vakulenko-Lagun et al., 2022).

! Sensitive to misspecification of the truncation model; inefficient.

Motivate us to seek estimators that

- Have more protect against model misspecification;
- More efficient;
- Allow us to incorporate nonparametric methods (which are known to have slower than root-n convergence) to obtain root-n consistent estimators.

Our contributions

- Derive the efficient influence curve (EIC) for the expectation of an arbitrarily transformed survival time.
- Construct EIC-based estimators that are shown to have favorable properties:
  - Model double robustness
  - Rate double robustness
  - Semiparametric efficiency
- Provide technical conditions for the asymptotic properties that appear to not have been carefully examined in the literature for time-to-event data.
- Our work represents the **first attempt** to construct doubly robust estimators in the presence of left truncation.
  - Does NOT fall under the established framework of coarsened data where doubly robust approaches are developed.

Notation and estimand

- Q - left truncation time; T - event time; Z - covariates
- Full data - if there were no left truncation
- We observe $O = (Q, T, Z)$ only if $Q < T$

- F, G, H: the full data CDF's of $T|Z$, $Q|Z$ and Z, respectively.
- superscript *: quantities related to the full data distribution e.g., $\mathbb{P}^*$, $\mathbb{E}^*$, p^*, P^*
- without *: quantities related to the observed data distribution e.g., $\mathbb{P}$, $\mathbb{E}$, p, P

Notation and estimand

- Q - left truncation time; T - event time; Z - covariates
- Full data - if there were no left truncation
- We observe $O = (Q, T, Z)$ only if $Q < T$

- F, G, H: the full data CDF's of $T|Z$, $Q|Z$ and Z, respectively.
- superscript *: quantities related to the full data distribution
  e.g., $\mathbb{P}^*$, $\mathbb{E}^*$, p^*, P^*
- without *: quantities related to the observed data distribution
  e.g., $\mathbb{P}$, $\mathbb{E}$, p, P

- Estimand: $\theta := \mathbb{E}^*\{\nu(T)\}$, where ν is a given function.
  - e.g., when $\nu(t) = \mathbb{1}(t > t_0)$, $\theta = \mathbb{P}^*(T > t_0)$ (survival probability).
  - e.g., when $\nu(t) = \min(t, t_0)$, $\theta = \mathbb{E}^*\{\min(T, t_0)\}$ (RMST).

Assumptions

- **Conditional quasi-independence**:
  Q and T are "independent" given Z on the observed region $\{t > q\}$.
- **Positivity**: $\mathbb{P}^*(Q < T|Z) > 0$ a.s.
- **Overlap assumption for F and G**: Stronger than positivity.

EIC and double robustness

- EIC:
$$\varphi(O; \theta, F, G, H) = \beta \cdot U(\theta; F, G),$$
where
$$U(\theta; F, G) = \frac{\nu(T) - \theta}{G(T|Z)} - \int_0^\infty \mathbb{E}^*\{\nu(T) - \theta \mid T < v, Z\} \cdot \frac{F(v|Z)}{1 - F(v|Z)} \cdot \frac{d\bar{M}_Q(v; G)}{G(v|Z)},$$
and recall that $\beta = \mathbb{P}^*(Q < T)$.
- The semiparametric efficiency bound : $\mathbb{E}(\varphi^2)$.

EIC and double robustness

- EIC:
$$\varphi(O;\theta,F,G,H)=\beta\cdot U(\theta;F,G),$$
where
$$U(\theta;F,G)$$
$$=\frac{\nu(T)-\theta}{G(T|Z)}-\int_0^\infty \mathbb{E}^*\{\nu(T)-\theta\mid T<v,Z\}\cdot\frac{F(v|Z)}{1-F(v|Z)}\cdot\frac{d\bar{M}_Q(v;G)}{G(v|Z)},$$
and recall that $\beta=\mathbb{P}^*(Q<T)$.
- The semiparametric efficiency bound : $\mathbb{E}(\varphi^2)$.

Double robustness:

$\mathbb{E}\{U(\theta_0;F,G)\}=0$ if either $F=F_0$ or $G=G_0$

Estimation

Let $\{O_i\}_{i=1}^{n}$ be an observed random sample of size n; $\quad O_i = (Q_i, T_i, Z_i)$.

- First estimate F and G
- Then solve the following equation for θ:

$$\sum_{i=1}^{n} U_i(\theta; \hat{F}, \hat{G}) = 0$$

Estimation

Let $\{O_i\}_{i=1}^n$ be an observed random sample of size n; $\quad O_i = (Q_i, T_i, Z_i)$.

- First estimate F and G
- Then solve the following equation for θ:

$$\sum_{i=1}^{n} U_i(\theta; \hat{F}, \hat{G}) = 0$$

- Closed-form solution:

$$\hat{\theta}_{dr} = \left(\sum_{i=1}^{n}\left[\frac{1}{\hat{G}(T_i|Z_i)} - \int_0^\infty \frac{\hat{F}(v|Z_i)}{\hat{G}(v|Z_i)\{1-\hat{F}(v|Z_i)\}} d\bar{M}_{Q,i}(v;\hat{G})\right]\right)^{-1}$$
$$\times \left(\sum_{i=1}^{n}\left[\frac{\nu(T_i)}{\hat{G}(T_i|Z_i)} - \int_0^\infty \frac{\int_0^v \nu(t) d\hat{F}(t|Z_i)}{\hat{G}(v|Z_i)\{1-\hat{F}(v|Z_i)\}} d\bar{M}_{Q,i}(v;\hat{G})\right]\right)$$

Model double robustness under asymptotic linearity

Suppose

- $\hat{F}$ and $\hat{G}$ uniformly converge to F^* and G^*, respectively;
- $\hat{F}$ and $\hat{G}$ are asymptotically linear.

If either $F^* = F_0$ or $G^* = G_0$, we have:

- $\sqrt{n}(\hat{\theta}_{dr} - \theta_0) \xrightarrow{d} N(0, \sigma^2)$.

Furthermore, when both $F^* = F_0$ and $G^* = G_0$,

- $\hat{\theta}_{dr}$ acheives the semiparametric efficiency bound;
- σ^2 can be consistently estimated by $\hat{\sigma}^2$, where

$$\hat{\sigma}^2 = \hat{\beta}^2 \cdot \frac{1}{n} \sum_{i=1}^{n} U_i^2(\hat{\theta}_{dr}, \hat{F}, \hat{G}), \quad \hat{\beta} = \left\{ n^{-1} \sum_{i=1}^{n} 1/\hat{G}(T_i|Z_i) \right\}^{-1}.$$

Rate double robustness with cross-fitting

K**-fold cross-fitting**:

$$\sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} U_i\{\theta, \hat{F}^{(-k)}, \hat{G}^{(-k)}\} = 0 \quad \rightarrow \quad \hat{\theta}_{cf}$$

Out-of-sample cross integral product:

$$\mathcal{D}_\dagger(\hat{F}, \hat{G}; F_0, G_0) := \mathbb{E}\left(\mathbb{E}_\dagger\left[\left|\int_{\tau_1}^{\tau_2}\left\{a(v, Z_\dagger; \hat{F}) - a(v, Z_\dagger; F_0)\right\} \cdot Y_\dagger(v)\, d\left\{\frac{1}{\hat{G}(v|Z_\dagger)} - \frac{1}{G_0(v|Z_\dagger)}\right\}\right|\right]\right),$$

where $a(v, Z; F) = \int_0^v \{\nu(t) - \theta\} dF(t|Z)/\{1 - F(v|Z)\}$
$Y_\dagger(v) = \mathbb{1}(Q_\dagger \leq v < T_\dagger)$.

Rate double robustness

Suppose

- $\hat{F}$ and $\hat{G}$ are uniformly consistent;
- $\mathcal{D}_\dagger(\hat{F}, \hat{G}; F_0, G_0) = o(n^{-1/2})$.

We have

- $\sqrt{n}(\hat{\theta}_{cf} - \theta_0) \xrightarrow{d} N(0, \sigma^2)$, where $\sigma^2 = \mathbb{E}(\varphi^2)$;
- $\hat{\theta}_{cf}$ achieves the semiparametric efficiency bound;
- σ^2 can be consistently estimated by $\hat{\sigma}^2_{cf}$, where

$$\hat{\sigma}^2_{cf} = \hat{\beta}^2_{cf} \cdot \frac{1}{n} \sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} U_i^2\{\hat{\theta}_{cf}, \hat{F}^{(-k)}, \hat{G}^{(-k)}\},$$

$$\hat{\beta}_{cf} = \left\{ \frac{1}{n} \sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} \frac{1}{\hat{G}^{(-k)}(T_i | Z_i)} \right\}^{-1}.$$

Nonparametric methods can be used to estimate F and G!

Application: CNS lymphoma data

- Data from a study on central nervous system (CNS) lymphoma (Wang et al., 2015)
  [Publicly available in the supplement of Vakulenko-Lagun et al. (2022)]

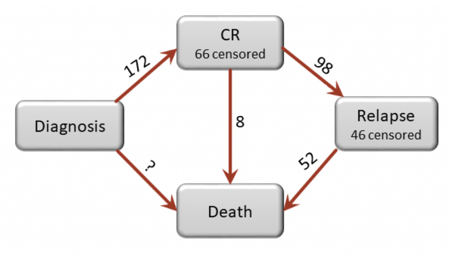


Figure from Vakulenko-Lagun et al., (2022)
CR: complete response.

→ Restrict to the 98 patients that were relapsed, for which the time is recorded.

- Quantity of interest: overall survival.
- It is plausible that time to death and time to relapse are dependent, and treatment is strongly associated with both.
- Binary treatment variables:
  - chemotherapy
  - radiation therapy

Application: CNS lymphoma data

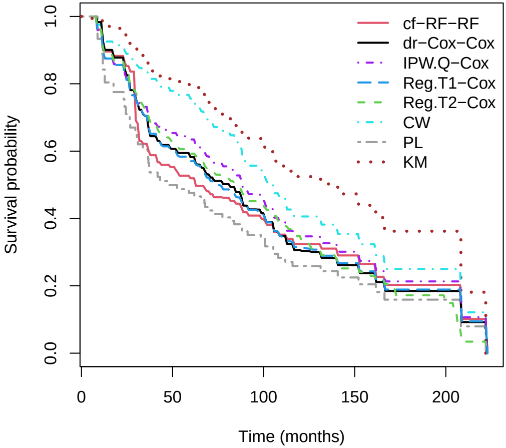
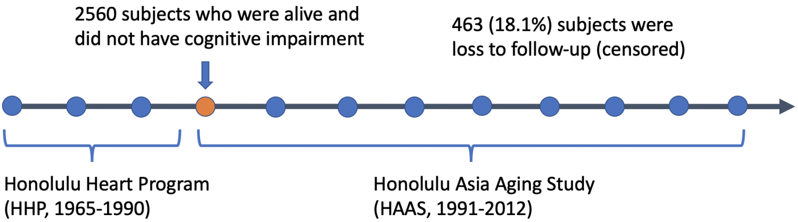


Figure: Estimates of the overall survival for the CNS lymphoma data.

Application: HAAS data



- Quantity of interest:
  Cognitive impairment-free survival on the age scale.
- Dependence detected: tau = 0.0426 with p-value 0.0014.
- Covariates:
  - education (years)
  - APOE positive (yes/no)
  - mid-life alcohol consumption (light/heavy)
  - mid-life cigarette consumption (yes/no)

Application: HAAS data

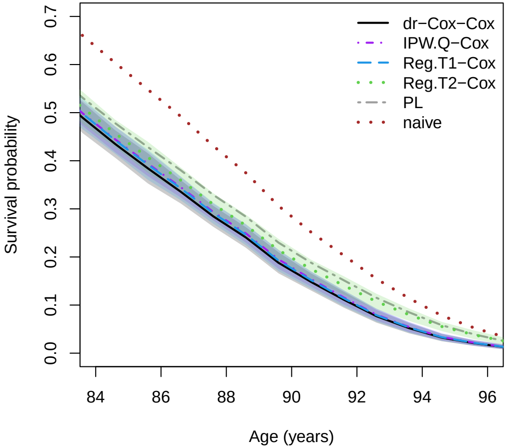


Figure: Estimated cognitive impairment-free survival and their 95% bootstrap confidence intervals (shaded, except for PL and naive) for the HAAS data.

Summary and discussion

- We derived the EIC for the mean of an arbitrarily transformed survival time and construct doubly robust estimators.
- Future direction:
  - Extension to also handle informative censoring.
  - Leverage domain knowledge to avoid the conditional quasi-independence assumption using the "proximal identification framework" (Tchetgen Tchetgen et al., 2020)

- ArXiv preprint: arXiv:2208.06836
- Code: https://github.com/wangyuyao98/left_trunc_DR

Questions & Comments?

Appendix

Assumptions

f, g, h: the densities of $T|Z$, $Q|Z$ and Z, respectively.

Assumption 1 (Conditional quasi-independence)

The observed data density for (Q, T, Z) *satisfies*

$$p_{Q,T,Z}(q,t,z) = \begin{cases} f(t|z)g(q|z)h(z)/\beta, & \textit{if } t > q, \\ 0, & \textit{otherwise}, \end{cases}$$

where $\beta = \mathbb{P}^*(Q < T) = \int \mathbb{1}(q < t) f(t|z) g(q|z) h(z)\ dt\ dq\ dz.$

Assumption 2 (Positivity)

$\mathbb{P}^*(Q < T|Z) > 0$ *a.s.*

Assumption 3 (Overlap)

There exists $0 < \tau_1 < \tau_2 < \infty$ *such that* $T \geq \tau_1$ *a.s.,* $Q \leq \tau_2$ *a.s.; also* $G(\tau_1|Z) \geq \delta_1$ *a.s. and* $F(\tau_2|Z) \leq 1 - \delta_2$ *a.s. for some constants* $\delta_1 > 0$ *and* $\delta_2 > 0$.

Assumptions

Assumption 4 (Uniform Convergence)

There exist F^* and G^* such that

$$\left\|\hat{F}(\cdot|Z) - F^*(\cdot|Z)\right\|_{\sup,2} = o(1), \quad \left\|\hat{G}(\cdot|Z) - G^*(\cdot|Z)\right\|_{\sup,2} = o(1).$$

Assumptions

Assumption 5 (Asymptotic Linearity)

For fixed $(t,z) \in [\tau_1, \tau_2] \times \mathcal{Z}$, $\hat{F}(t|z)$ *and* $\hat{G}(t|z)$ *are regular and asymptotically linear estimators for* $F(t|z)$ *and* $G(t|z)$ *with influence functions* $\xi_1(t,z,O)$ *and* $\xi_2(t,z,O)$, *respectively. In addition, denote*

$$R_1(t,z) = \hat{F}(t|z) - F^*(t|z) - \frac{1}{n}\sum_{i=1}^{n} \xi_1(t,z,O_i),$$

$$R_2(t,z) = \hat{G}(t|z) - G^*(t|z) - \frac{1}{n}\sum_{i=1}^{n} \xi_2(t,z,O_i).$$

Suppose $\|R_1(\cdot,Z)\|_{\text{sup},2} = o(n^{-1/2})$, $\|R_2(\cdot,Z)\|_{\text{sup},2} = o(n^{-1/2})$, *and either* $\|R_1(\cdot,Z)\|_{TV,2} = o(1)$ *or* $\|R_2(\cdot,Z)\|_{TV,2} = o(1)$.

Inverse probability weighting (IPW) identification

- Under Assumptions 1 and 2,

$$\theta = \mathbb{E}\left\{\frac{\nu(T)}{G(T|Z)}\right\} \Big/ \mathbb{E}\left\{\frac{1}{G(T|Z)}\right\}.$$

- Let α be the *reverse time hazard function* of Q given Z in the full data:

$$\begin{aligned}\alpha(q|z) &:= \lim_{h\to 0+} \frac{\mathbb{P}^*\left(q-h < Q \le q | Q \le q, Z = z\right)}{h} \\ &= \lim_{h\to 0+} \frac{\mathbb{P}^*\left(q-h < Q \le q | Z = z\right)}{h\ \mathbb{P}^*\left(Q \le q | Z = z\right)} = \frac{\partial G(q|z)/\partial q}{G(q|z)}.\end{aligned}$$

$\implies$ $G(q|z) = \exp\{-\int_q^\infty \alpha(t|z)dt\}$.

- α can be identified:

$$\alpha(q|z) = \frac{p_{Q|Z}(q|z)}{\mathbb{P}\left(Q \le q < T | Z = z\right)}.$$

$\implies$ G can be identified from the observed data distribution.

Reverse time counting process and backwards martingale

For $t \geq 0$, let

$$\bar{N}_Q(t) = \mathbb{1}(t \leq Q < T),$$
$$\bar{\mathcal{F}}_t = \sigma\left\{Z, \mathbb{1}(Q < T), \mathbb{1}(s \leq Q < T), \mathbb{1}(s \leq T) : s \geq t\right\}.$$

Define

$$\bar{A}_Q(t; G) = \int_t^\infty \mathbb{1}(Q \leq s < T)\alpha(s|Z)ds = \int_t^\infty \mathbb{1}(Q \leq s < T)\frac{dG(s|Z)}{G(s|Z)}.$$

Then

$$\bar{M}_Q(t; G) := \bar{N}_Q(t) - \bar{A}_Q(t; G)$$

is a backwards martingale with respect to $\{\bar{\mathcal{F}}_t\}_{t\geq 0}$ in the observed data.

Steps for constructing the proposed estimators

Inverse probability of truncation weighting identification

↓

Derive an IC

↓

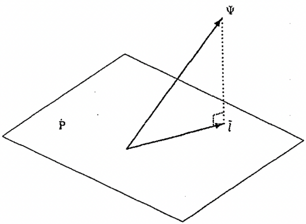
Find the EIC

↓

Construct EIC-based estimator

$$\theta = \mathbb{E}\left\{\frac{\nu(T)}{G(T|Z)}\right\} \Big/ \mathbb{E}\left\{\frac{1}{G(T|Z)}\right\}$$

$G \leftarrow$ conditional reverse time hazard
$\uparrow$
observed data distribution.



(Figure from Bickel et al., 1993)

$\dot{P}$ - tangent space; Ψ - IC; $\tilde{\ell}$ - EIC.

Deriving the EIC

- Derive an influence curve (IC)

$$\frac{\partial}{\partial \epsilon}\theta(P_\epsilon)\bigg|_{\epsilon=0} = \mathbb{E}\left\{\varphi(O)\mathcal{S}(O)\right\}, \quad \mathcal{S}(O) = \frac{\partial}{\partial \epsilon}\log p_\epsilon(O)\bigg|_{\epsilon=0}$$

- Project the IC onto the tangent space $\rightarrow$ EIC

$$\varphi(O;\theta,F,G,H) = \beta \cdot U(\theta;F,G)$$

where

$$U(\theta;F,G) = \frac{\nu(T)-\theta}{G(T|Z)} - \int_0^\infty \mathbb{E}^*\{\nu(T)-\theta \mid T < v, Z\} \cdot \frac{F(v|Z)}{1-F(v|Z)} \cdot \frac{d\bar{M}_Q(v;G)}{G(v|Z)}$$

Recall that $\beta = \mathbb{P}^*(Q < T)$

- The semiparametric efficiency bound : $\mathbb{E}(\varphi^2)$.

Two special cases

- By setting $\hat{F} \equiv 0 \;\; \rightarrow$ IPW estimator

$$\hat{\theta}_{\text{IPW.Q}} = \left\{ \sum_{i=1}^{n} \frac{\nu(T_i)}{\hat{G}(T_i|Z_i)} \right\} \Big/ \left\{ \sum_{i=1}^{n} \frac{1}{\hat{G}(T_i|Z_i)} \right\},$$

- By setting $\hat{G} \equiv 1 \;\; \rightarrow$ Regression-based estimator

$$\hat{\theta}_{\text{Reg.T1}} = \left\{ \sum_{i=1}^{n} \frac{1}{1 - \hat{F}(Q_i|Z_i)} \right\}^{-1} \left[\sum_{i=1}^{n} \frac{\nu(T_i)\{1 - \hat{F}(Q_i|Z_i)\} + \int_0^{Q_i} \nu(t) d\hat{F}(t|Z_i)}{1 - \hat{F}(Q_i|Z_i)} \right].$$

  - $[\nu(T)\{1 - F(Q|Z)\} + \int_0^{Q} \nu(t) dF(t|Z)]$ identifies $\mathbb{E}^* \{\nu(T)|Q, Z\}$.

- Another regression based estimator is

$$\hat{\theta}_{\text{Reg.T2}} = \left[\sum_{i=1}^{n} \frac{\hat{\mathbb{E}}^* \{\nu(T_i)|Z_i\}}{1 - \hat{F}(Q_i|Z_i)} \right] \Big/ \left\{ \sum_{i=1}^{n} \frac{1}{1 - \hat{F}(Q_i|Z_i)} \right\},$$

where $\hat{\mathbb{E}}^* \{\nu(T_i)|Z_i\} = \int_0^{\infty} \nu(t) d\hat{F}(t|Z_i)$.

Some norm notation

For a random function $X(t, z)$ with $t \in [\tau_1, \tau_2]$ and $z \in \mathcal{Z}$, define

$$\|X(\cdot, Z)\|^2_{\text{sup},2} = \mathbb{E}\left\{ \sup_{t \in [\tau_1, \tau_2]} |X(t, Z)|^2 \right\},$$
$$\|X(\cdot, Z)\|^2_{\text{TV},2} = \mathbb{E}\left[\text{TV}\{X(\cdot, Z)\}^2 \right],$$

- $\text{TV}\{X(\cdot, z)\} = \sup_{\mathcal{P}} \sum_{j=1}^{J} |X(x_j, z) - X(x_{j-1}, z)|$ is the total variation of $X(\cdot, z)$ on the interval $[\tau_1, \tau_2]$
- $\mathcal{P}$ is the set of all possible partitions $\tau_1 = x_0 < x_1 < ... < x_J = \tau_2$ of $[\tau_1, \tau_2]$

Assumptions for $\hat{F}$ and $\hat{G}$ for $\hat{\theta}_{dr}$

- **Uniform convergence**: There exist F^* and G^* such that

$$\left\|\hat{F}(\cdot|Z) - F^*(\cdot|Z)\right\|_{\text{sup},2} = o(1), \quad \left\|\hat{G}(\cdot|Z) - G^*(\cdot|Z)\right\|_{\text{sup},2} = o(1).$$

- **Asymptotic linearity**:

$$\hat{F}(t|z) - F^*(t|z) = \frac{1}{n}\sum_{i=1}^{n} \xi_1(t, z, O_i) + R_1(t, z),$$

$$\hat{G}(t|z) - G^*(t|z) = \frac{1}{n}\sum_{i=1}^{n} \xi_2(t, z, O_i) + R_2(t, z).$$

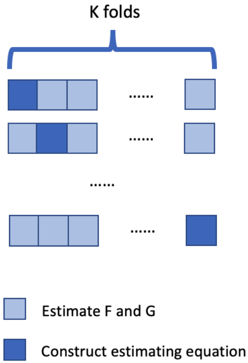
  where $\|R_1(\cdot, Z)\|_{\text{sup},2} = o(n^{-1/2})$, $\|R_2(\cdot, Z)\|_{\text{sup},2} = o(n^{-1/2})$, and either $\|R_1(\cdot, Z)\|_{\text{TV},2} = o(1)$ or $\|R_2(\cdot, Z)\|_{\text{TV},2} = o(1)$.

- e.g., it is satisfied when Cox model is used to estimate F and G.

K-fold cross-fitting

1: Split the data into K folds of (almost) equal size with the index sets $\mathcal{I}_1, ..., \mathcal{I}_K$.
2: **for** $k = 1$ to K **do**
3: Estimate F and G with the out-of-k-fold data $\implies$ $\hat{F}^{(-k)}$ and $\hat{G}^{(-k)}$
4: **end for**
5: Obtain $\hat{\theta}_{cf}$ by solving

$$\sum_{k=1}^{K} \sum_{i \in \mathcal{I}_k} U_i\{\theta, \hat{F}^{(-k)}, \hat{G}^{(-k)}\} = 0.$$



Norm notation and cross integral product

Let $\mathcal{O} = \{(Q_i, T_i, Z_i) : i = 1, ..., m\}$ denote the data used to obtain $\hat{F}$ and $\hat{G}$, and let $O_\dagger = (Q_\dagger, T_\dagger, Z_\dagger)$ be an copy of the data that is independent of, but from the same distribution as $\mathcal{O}$.

$$\|\hat{F} - F_0\|^2_{\dagger,\sup,2} := \mathbb{E}\left(\mathbb{E}_\dagger\left[\left\{\sup_{t\in[\tau_1,\tau_2]}\left|\hat{F}(t|Z_\dagger) - F_0(t|Z_\dagger)\right|\right\}^2\right]\right),$$

$$\|\hat{G} - G_0\|^2_{\dagger,\sup,2} := \mathbb{E}\left(\mathbb{E}_\dagger\left[\left\{\sup_{t\in[\tau_1,\tau_2]}\left|\hat{G}(t|Z_\dagger) - G_0(t|Z_\dagger)\right|\right\}^2\right]\right).$$

Out-of-sample cross integral product:

$$\mathcal{D}_\dagger(\hat{F}, \hat{G}; F_0, G_0) := \mathbb{E}\left(\mathbb{E}_\dagger\left[\left|\int_{\tau_1}^{\tau_2}\left\{a(v, Z_\dagger; \hat{F}) - a(v, Z_\dagger; F_0)\right\}\cdot Y_\dagger(v)\, d\left\{\frac{1}{\hat{G}(v|Z_\dagger)} - \frac{1}{G_0(v|Z_\dagger)}\right\}\right|\right]\right),$$

where $a(v, Z; F) = \int_0^v \{\nu(t) - \theta\} dF(t|Z)/\{1 - F(v|Z)\}$
$Y_\dagger(v) = \mathbb{1}(Q_\dagger \le v < T_\dagger)$.

Assumptions on $\hat{F}$ and $\hat{G}$ for $\hat{\theta}_{cf}$

- **Uniform Consistency**:

$$\|\hat{F} - F_0\|_{\dagger,\text{sup},2} = o(1), \quad \|\hat{G} - G_0\|_{\dagger,\text{sup},2} = o(1)$$

- **Product rate condition**: $\mathcal{D}_{\dagger}(\hat{F}, \hat{G}; F_0, G_0) = o(n^{-1/2})$.

Extensions to handle right censoring

C: censoring time; $X := \min(T, C)$; $\Delta := \mathbb{1}(T < C)$

$S_c(t) := \mathbb{P}(C > t)$, $S_D(t) := \mathbb{P}(D > t)$, where $D = C - Q$

Assume noninformative censoring.

Two scenarios:

- Censoring can happen before truncation
  - $\mathbb{P}^*(C < Q) > 0$, subjects with $Q < X$ are included.

$$U_{c1}(\theta; F_x, G, S_c) = \frac{\Delta\{\nu(X) - \theta\}}{S_c(X) G(X|Z)} - \int_0^\infty \frac{\int_0^v \Delta\{\nu(x) - \theta\}/S_c(x) dF_x(x|Z)}{1 - F(v|Z)} \cdot \frac{d\tilde{M}_Q(v; G)}{G(v|Z)}.$$

- Censoring always after truncation
  - $\mathbb{P}^*(Q < C) = 1$, subjects with $Q < T$ are included.

$$U_{c2}(\theta; F, G, S_D) = \frac{\Delta}{S_D(X - Q)} \left[\frac{\nu(X) - \theta}{G(X|Z)} - \int_0^\infty \frac{\int_0^v \{\nu(t) - \theta\} dF(t|z)}{1 - F(v|Z)} \cdot \frac{d\tilde{M}_Q(v; G)}{G(v|Z)} \right].$$

Simulation results

500 simulated data sets each with sample size 1000.

Truncation rate: 29.5%; $\theta_0 = P^*(T > 3) = 0.576$.

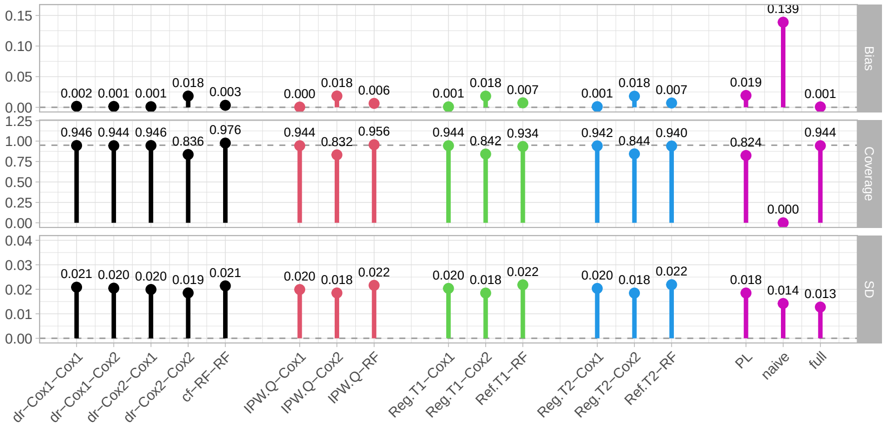


Figure: Absolute bias, coverage rate of 95% confidence intervals with bootstrap standard errors, and empirical standard deviation for different estimators.

Simulation

The models in red are misspecified.

SD: standard deviation, SE: standard error, CP: coverage probability.

Estimator	bias	SD	SE/boot SE	CP/boot CP
dr-Cox1-Cox1	-0.0016	0.021	0.020/0.020	0.948/0.946
dr-Cox1-Cox2	-0.0014	0.020	0.019/0.020	0.930/0.944
dr-Cox2-Cox1	-0.0010	0.020	0.019/0.020	0.938/0.946
dr-Cox2-Cox2	0.0184	0.019	0.018/0.019	0.838/0.836
cf-RF-RF	0.0032	0.021	0.023/0.025	0.966/0.976
IPW.Q-Cox1	-0.0004	0.020	0.018/0.020	0.924/0.944
IPW.Q-Cox2	0.0184	0.018	0.017/0.019	0.814/0.832
IPW.Q-RF	-0.0064	0.022	0.019/0.022	0.886/0.956
Reg.T1-Cox1	-0.0008	0.020	- /0.020	- /0.944
Reg.T1-Cox2	0.0183	0.018	- /0.019	- /0.842
Reg.T1-RF	-0.0073	0.022	- /0.022	- /0.934
Reg.T2-Cox1	-0.0010	0.020	- /0.020	- /0.942
Reg.T2-Cox2	0.0181	0.018	- /0.019	- /0.844
Reg.T2-RF	-0.0070	0.022	- /0.022	- /0.940
PL	0.0193	0.018	- /0.018	- /0.824
naive	0.1389	0.014	0.014/0.014	0.000/0.000
full data	-0.0007	0.013	0.013/0.013	0.956/0.944

CNS lymphoma data

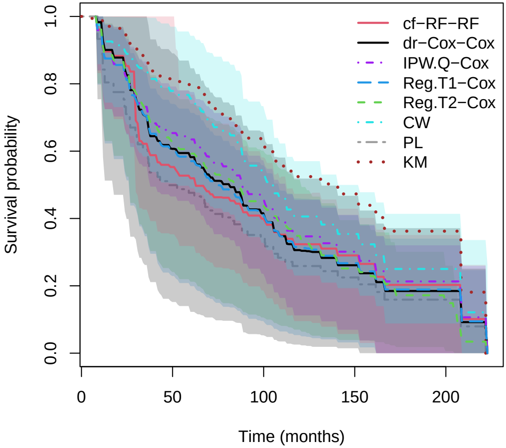


Figure: Estimates of the overall survival for the CNS lymphoma data with their 95% bootstrap confidence intervals.