A Liberating Framework from Truncation and Censoring, with Application to Learning Treatment Effects

Yuyao Wang

Department of Public Health and Health Sciences, Northeastern University
yuya.wang@northeastern.edu

Joint work with:

Andrew Ying, Google

Ronghui (Lily) Xu, UC San Diego

Left truncation – selection due to delayed entry

- Example: aging studies.
  - Age is the time scale of interest.
  - Subjects enrolled at various ages instead of at the time origin (time at birth).

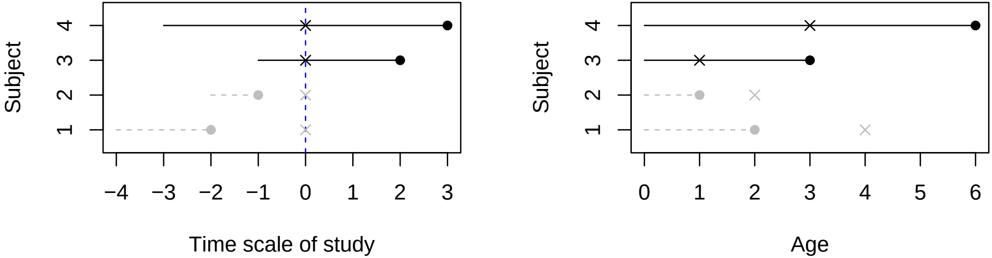


Figure: A toy example for aging study; '×' - enrollment times; dots - times to events.

Left truncation – mathematical formulation

- Time-to-event: T^*
- Left truncation time: Q^* – usually the study enrollment time

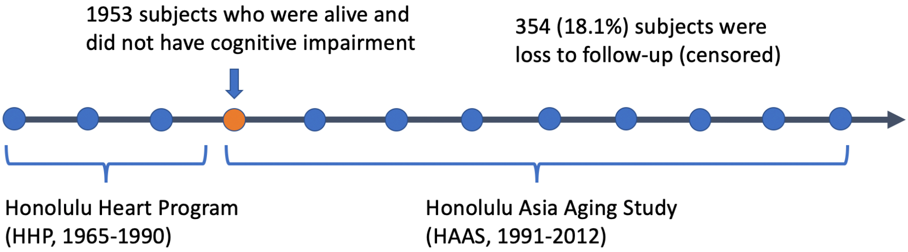
- T^* is **left truncated** by Q^* if only subjects with $T^* > Q^*$ are included in the data.

- Subjects with early event times tend not to be captured → **selection bias**

Examples:

- Aging studies – age is the time scale of interest
- Pregnancy studies
- Some cancer survivorship studies, e.g., SJLIFE.

HAAS data



1953 subjects who were alive and did not have cognitive impairment

354 (18.1%) subjects were loss to follow-up (censored)

Honolulu Heart Program
(HHP, 1965-1990)

Honolulu Asia Aging Study
(HAAS, 1991-2012)

- T^*: age to moderate cognitive impairment or death; Q^*: age at entry of HAAS.
- Causal questions:
  e.g., the effect of midlife alcohol exposure on late-life cognitive impairment.
- **Triple biases**:
  - **Selection bias from left truncation** – early event times are underrepresented.
  - Confounding in observational data.
  - Informative right censoring.

Literature on left truncation

Conventional methods on left truncation

- Under random left truncation/quasi-independence assumption
- Extended to conditional (quasi-)independence assumption under regression settings
  - when the dependence-inducing covariates are included as regressors

---

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Literature on left truncation

Conventional methods on left truncation

- Under random left truncation/quasi-independence assumption
- Extended to conditional (quasi-)independence assumption under regression settings
    - when the dependence-inducing covariates are included as regressors

For marginal estimands under covariate dependent left truncation

- Inverse probability of truncation weighting (Vakulenko-Lagun et al., 2022)
- Efficient influence function (EIF)-based doubly robust (DR) approaches (Wang et al., 2024)
    - Constructed model DR and rate DR estimators in the presence of left truncation – selection bias that does NOT fall under the established framework of coarsened data

---

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Literature on left truncation

Conventional methods on left truncation

- Under random left truncation/quasi-independence assumption
- Extended to conditional (quasi-)independence assumption under regression settings
  - when the dependence-inducing covariates are included as regressors

For marginal estimands under covariate dependent left truncation

- Inverse probability of truncation weighting (Vakulenko-Lagun et al., 2022)
- Efficient influence function (EIF)-based doubly robust (DR) approaches (Wang et al., 2024)
  - Constructed model DR and rate DR estimators in the presence of left truncation – selection bias that does NOT fall under the established framework of coarsened data

Complexities of right censoring in the presence of left truncation

- Informative right censoring for estimating a marginal parameter
- Truncation-induced dependence between the right censoring time and the event time

---

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Our contributions

- Introduce a liberating framework for handling covariate dependent **L**eft **T**runcation and **R**ight **C**ensoring (LTRC).
    - Allows existing estimating functions and loss functions from LTRC-free settings to be easily applied to the observed LTRC data
    - Maintain their desirable properties in LTRC-free data, e.g., orthogonality, double robustness (DR).

- As an illustration:
    - Average treatment effect (ATE): model DR and rate DR estimators.
    - Conditional Average Treatment Effect (CATE): orthogonal and doubly robust learners.

- The framework can also be applied to other problems with LTRC data.

Y. Wang, A. Ying, and R. Xu, (2024). A liberating framework from truncation and censoring, with application to learning treatment effects. *arXiv:2411.18879*.

Notation

- Q: left truncation time; T: event time;
- C: censoring time; $D = C - Q$: residual censoring time;
- A: binary treatment assignment; Z: covariates.

- Variables with '$*$' – truncation-free data; without '$*$' – truncated data.
- $T^*(a)$ – potential event time under treatment a.

- Observe $O = (Q, X, \Delta, A, Z)$ only if $Q^* < T^*$.
- $X = \min(T, C)$, $\Delta = I(T < C)$.

- Estimand:
  - ν: a given transformation; V^* a subset of covariates,
  - **CATE**: $\tau(v) = \mathbb{E}[\nu\{T^*(1)\} - \nu\{T^*(0)\} \mid V^* = v]$.
  - **ATE**: a sepcial case with $V^* = \varnothing$, $\theta = \tau(\varnothing)$.

Assumptions

1. SUTVA, no unmeasured confounding, consistency.
2. Conditional independent truncation: $Q^*(a) \perp\!\!\!\perp T^*(a) \mid Z^*$.
3. Conditional noninformative residual censoring: $D^*(a) \perp\!\!\!\perp (T^*(a), Q^*(a)) \mid Z^*$.
4. Strict positivity.

   $\pi(z) = \mathbb{P}(A^* = 1 | Z^* = z)$;
   F, G: conditional CDF of $T^* | A^*, Z^*$, and $Q^* | A^*, Z^*$, respectively;
   S_D: the conditional survival function of $D | Q, A, Z$.

   (i) There exist $\delta_1 > 0$ such that $\delta_1 \leq \pi(Z^*) \leq 1 - \delta_1$ a.s.;
   (ii) There exist $0 < \tau_1 < \tau_2 < \infty$ and $\delta_2 > 0$ such that $T^* \geq \tau_1$ a.s. and $Q^* < \tau_2$ a.s., $1 - F(\tau_2 | A^*, Z^*) \geq \delta_2$ a.s. and $G(\tau_1 | A^*, Z^*) \geq \delta_2$ a.s.;

   (iii) There exists $\delta_3 > 0$ such that $1 - F(t_{\max} | A, Z) \geq \delta_3$ a.s. and $S_D(t_{\max} - Q | Q, A, Z) \geq \delta_3$ a.s..

Doubly robust operator for left truncation

- Motivated by the EIF derived in Wang et al. (2024).
- For any function in the LTRC-free data $\zeta(T^*, A^*, Z^*)$,
- $\mathcal{V}_Q(\zeta; F, G) =$

$$\underbrace{\frac{\zeta(T, A, Z)}{G(T|A, Z)}}_{\text{IPW}} - \underbrace{\int_0^\infty m_\zeta(v, A, Z; F) \cdot \frac{F(v|A, Z)}{1 - F(v|A, Z)} \cdot \frac{d\bar{M}_Q(v; G)}{G(v|A, Z)}}_{\text{Augmentation}}$$

- F: the conditional CDF of $T^* \mid A^*, Z^*$.
- G: the conditional CDF of $Q^* \mid A^*, Z^*$.

- $m_\zeta(v, a, z; \theta, F) = \mathbb{E}\{\zeta(T^*, A^*, Z^*; \theta) \mid T^* \leq v, A^* = a, Z^* = z\}$.

Y. Wang, A. Ying, and R. Xu (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.

Doubly robust operator for censoring

- Adapt the **AIPCW** (Rotnitzky and Robins, 2005) to the **residual time scale**.
- Recall $D = C - Q$: the residual censoring time.
- For any function $\xi(Q, T, A, Z)$ in the truncated but censoring-free data,
- $\mathcal{V}_C(\xi; F, S_D) =$

$$\underbrace{\frac{\Delta\ \xi(Q, X, A, Z)}{S_D(X - Q|Q, A, Z)}}_{\text{IPCW}} + \underbrace{\int_0^\infty \bar{m}_\xi(u, Q, A, Z; F) \cdot \frac{dM_D(u; S_D)}{S_D(u|Q, A, Z)}}_{\text{Augmentation}}$$

- S_D: the conditional survival function of $D \mid Q, A, Z$.
- Recall F: the conditional CDF of $T^* \mid A^*, Z^*$.

$$\bar{m}_\xi(u, q, a, z; F) = \mathbb{E}\{\xi(Q, T, A, Z)|T - Q \geq u, Q = q, A = a, Z = z\}$$
$$= \frac{\int_{q+u}^\infty \xi(q, t, a, z) dF(t|a, z)}{1 - F(q + u|a, z)}.$$

General framework and double robustness

- **General framework**:

$$\begin{tikzcd}
\text{Full data} \arrow[d, "{\text{AIPTW}(F, \pi),\ \text{orthogonal loss}}"] \\
\text{LTRC-free data} \arrow[d, "{\mathcal{V}_Q(F, G)}"] \\
\text{Truncated \& cen-free data} \arrow[d, "{\mathcal{V}_C(F, S_D)}"] \\
\text{Observed data}
\end{tikzcd}$$

- **Double robustness**: The expectation is maintained (up to a constant factor) by $\mathcal{V}_C \circ \mathcal{V}_Q$ if either F or (G, S_D) is the truth.

ATE estimation

- Estimand: $\theta = \mathbb{E}[\nu\{T^*(1)\} - \nu\{T^*(0)\}]$. Propensity score: $\pi(z) = \mathbb{P}(A^* = 1|Z^* = z)$.

- AIPTW estimating function $U_0(\theta; F, \pi) \xrightarrow{\mathcal{V}_C \circ \mathcal{V}_Q} U(\theta; F, \pi, G, S_D)$

- **Double robustness**: $\mathbb{E}\{U(\theta; F, \pi, G, S_D)\} = 0$ if either F or (π, G, S_D) is the truth.

- Estimation:

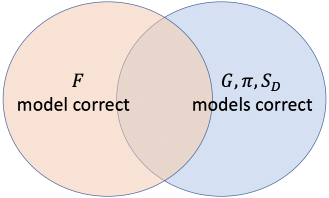
$$\sum_{i=1}^{n} U_i(\theta; \hat{F}, \hat{\pi}, \hat{G}, \hat{S}_D) = 0 \quad \Rightarrow \quad \text{closed-form solution } \hat{\theta}$$

Theorem 1 (Model double robustness)

Assume that $\hat{\pi}, \hat{F}, \hat{G}, \hat{S}_D$ *converge uniformly to* π^*, F^*, G^*, S_D^**, respectively, and that* π^*, F^*, G^*, S_D^* *also satisfy the strict positivity asssumption. Then*
(i) $\hat{\theta} \xrightarrow{p} \theta_0$ *if either* $F^* = F_0$ *or* $(\pi^*, G^*, S_D^*) = (\pi_0, G_0, S_{D0})$.
(ii) In addition, if $\hat{\pi}, \hat{F}, \hat{G}, \hat{S}_D$ *are asymptotically linear, then*

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

Furthermore, when both $F^* = F_0$ *and* $(\pi^*, G^*, S_D^*) = (\pi_0, G_0, S_{D0})$*, we have* $\sigma^2 = \beta^2 \mathbb{E}\{U(\theta_0; \pi_0, F_0, G_0, S_{D0})^2\}$*, which can be consistently estimated by* $\hat{\sigma}^2 = n \sum_{i=1}^n U_i^2(\hat{\theta}; \hat{\pi}, \hat{F}, \hat{G}, \hat{S}_D) / \{\sum_{i=1}^n \mathcal{V}_i(1; \hat{F}, \hat{G}, \hat{S}_D)\}^2$.



Rate double robustness

Cross-fitted estimator:

$$\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k} U_i\left(\theta; \hat{\pi}^{(-k)}, \hat{F}^{(-k)}, \hat{G}^{(-k)}, \hat{S}_D^{(-k)}\right) = 0 \quad \Rightarrow \quad \hat{\theta}_{cf}$$

Assumptions:

- Uniform consistency:
  $\|\hat{\pi}-\pi_0\|_2 = o_p(1)$, $\|\hat{F}-F_0\|_{\sup,2} = o_p(1)$, $\|\hat{G}-G_0\|_{\sup,2} = o_p(1)$, $\|\hat{S}_D - S_{D0}\|_{\sup,2} = o_p(1)$.
- Product rate condition:

$$\|\hat{F}-F_0\|_{\sup,2}\cdot\left\{\|\hat{\pi}-\pi_0\|_2 + \|\hat{G}-G_0\|_{\sup,2} + \|\hat{S}_D - S_{D0}\|_{\sup,2}\right\} \\ + \|K_1(\hat{g},g_0)\|_1 + \|K_2(\hat{g},g_0)\|_1 + \|K_3(\hat{g},g_0)\|_1 = o_p(n^{-1/2}).$$

$g = (F, G, S_D)$, $\hat{g} = (\hat{F}, \hat{G}, \hat{S}_D)$.

$K_1(g,g_0)$, $K_2(g,g_0)$, and $K_3(g,g_0)$ are integral products between the estimation errors of F and (G, S_D), containing terms such as

$$\int_Q^X \left\{\frac{\int_0^v \nu(t)dF(t|A,Z)}{1-F(v|A,Z)} - \frac{\int_0^v \nu(t)dF_0(t|A,Z)}{1-F_0(v|A,Z)}\right\}\cdot d\left\{\frac{1}{G(v|A,Z)} - \frac{1}{G_0(v|A,Z)}\right\}.$$

Rate double robustness

Theorem 2 (Rate double robustness)

(i) Under uniform consistency of the nuisance estimators, $\hat{\theta}_{cf} \xrightarrow{p} \theta_0$.

(ii) In addition, if the product rate condition holds,

$$n^{1/2}(\hat{\theta}_{cf} - \theta_0) \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = \beta^2 \mathbb{E}\{U(\theta_0; \pi_0, F_0, G_0, S_{D0})^2\}$, *which can be consistently estimated by*

$$\hat{\sigma}^2_{cf} = n \sum_{i,k} U_i^2 \left(\hat{\theta}_{cf}; \hat{\pi}^{(-k)}, \hat{F}^{(-k)}, \hat{G}^{(-k)}, \hat{S}_D^{(-k)}\right) \Bigg/ \left\{ \sum_{i,k} \mathcal{V}_i(1; \hat{F}^{(-k)}, \hat{G}^{(-k)}, \hat{S}_D^{(-k)})/n \right\}^2 .$$

CATE estimation

- Estimand: $\tau(v) = \mathbb{E}[\nu\{T^*(1)\} - \nu\{T^*(0)\} \mid V^* = v]$, where $V^* \subseteq Z^*$
- Loss function $\ell(T^*, A^*, Z^*; \tau, F, \pi) \xrightarrow{\mathcal{V}_C \circ \mathcal{V}_Q} \tilde{\ell}(O; \tau, F, \pi, G, S_D)$
  - R-loss → ltrcR-loss, for the special case with $V^* = Z^*$.
  - DR-loss → ltrcDR-loss
- **Neyman orthogonality** and **double robustness** are maintained.

CATE estimation

- Estimand: $\tau(v) = \mathbb{E}[\nu\{T^*(1)\} - \nu\{T^*(0)\} \mid V^* = v]$, where $V^* \subseteq Z^*$
- Loss function $\ell(T^*, A^*, Z^*; \tau, F, \pi) \xrightarrow{\mathcal{V}_C \circ \mathcal{V}_Q} \tilde{\ell}(O; \tau, F, \pi, G, S_D)$
  - R-loss → ltrcR-loss, for the special case with $V^* = Z^*$.
  - DR-loss → ltrcDR-loss
- **Neyman orthogonality** and **double robustness** are maintained.
- Empirical risk minimization with sample splitting:

$$\hat{\tau} = \operatorname*{arg\,min}_{\tau \in \mathcal{T}} \left[\frac{1}{m} \sum_{i=1}^{m} \tilde{\ell}\{O_i; \tau, \hat{F}, \hat{\pi}, \hat{G}, \hat{S}_D\} \right],$$

where $\hat{\pi}, \hat{F}, \hat{G}$ and $\hat{S}_D$ are estimated from data $\{O_i : i = m+1, \dots, n\}$, and m is roughly $n/2$.

Oracle rate results

Assume the nuisance estimators are consistent.

- Using the ltrcR-loss, $\hat{\tau}_{\mathrm{R}}$ achieves the oracle error rate if

$$\|\hat{\pi}-\pi_0\|_4^2+\|\hat{F}-F_0\|_{\sup,4}\cdot\left\{\|\hat{\pi}-\pi_0\|_4+\|\hat{G}-G_0\|_{\sup,4}+\|\hat{S}_D-S_{D0}\|_{\sup,4}\right\}$$
$$+\|K_1(\hat{g},g_0)\|_2+\|K_2(\hat{g},g_0)\|_2+\|K_3(\hat{g},g_0)\|_2=o_p(n^{-1/2}).$$

- Using the ltrcDR-loss, $\hat{\tau}_{\mathrm{DR}}$ achieves the oracle error rate if

$$\|\hat{F}-F_0\|_{\sup,4}\cdot\left\{\|\hat{\pi}-\pi_0\|_4+\|\hat{G}-G_0\|_{\sup,4}+\|\hat{S}_D-S_{D0}\|_{\sup,4}\right\}$$
$$+\|K_1(\hat{g},g_0)\|_2+\|K_2(\hat{g},g_0)\|_2+\|K_3(\hat{g},g_0)\|_2=o_p(n^{-1/2}).$$

$g=(F,G,S_D)$, $\hat{g}=(\hat{F},\hat{G},\hat{S}_D)$.

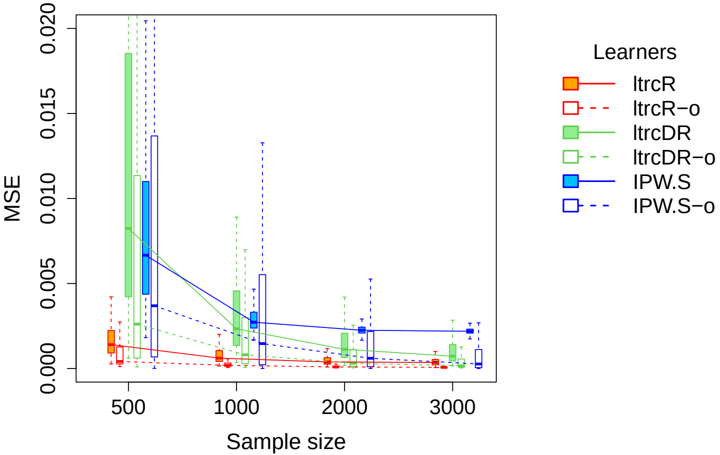
$K_1(g,g_0)$, $K_2(g,g_0)$, and $K_3(g,g_0)$ are integral products between the estimation errors of F and (G,S_D), containing terms such as

$$\int_Q^X\left\{\frac{\int_0^v \nu(t)dF(t|A,Z)}{1-F(v|A,Z)}-\frac{\int_0^v \nu(t)dF_0(t|A,Z)}{1-F_0(v|A,Z)}\right\}\cdot d\left\{\frac{1}{G(v|A,Z)}-\frac{1}{G_0(v|A,Z)}\right\}.$$

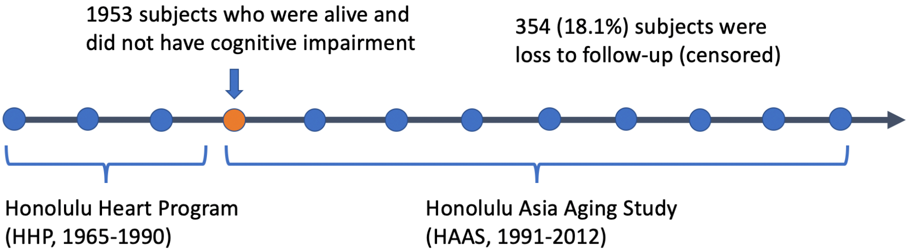
Simulation - CATE

- 500 simulated data sets. Estimand: $\tau(z) = \mathbb{E}\{\log T^*(1) - \log T^*(0) \mid Z^* = z\}$.
- Truncation rate: around 30%; treatment rate: around 50%; censoring rate: around 50%.

  $MSE = \frac{1}{n}\sum_{i=1}^{n}\{\hat{\tau}(V_i) - \tau_0(V_i)\}^2$.



HAAS data



1953 subjects who were alive and did not have cognitive impairment

354 (18.1%) subjects were loss to follow-up (censored)

Honolulu Heart Program (HHP, 1965-1990)

Honolulu Asia Aging Study (HAAS, 1991-2012)

- T^*: age to moderate cognitive impairment or death; Q^*: age at entry of HAAS.
- The effect of midlife alcohol exposure (heavy/non-heavy) on late-life cognitive impairment-free survival.
- Covariates:
  - Education ($\leq$ 12 years or otherwise)
  - ApoE genotype (positive/negative)
  - Systolic blood pressure (mmHg)
  - Heart rate (beats per minute)

HAAS: ATE

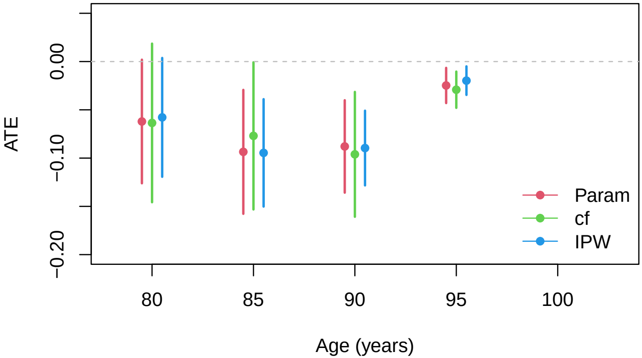


Figure: Estimates of ATE as difference of the potential DFS probabilities at various ages for heavy versus non-heavy drinkers in the HAAS data, with 95% confidence intervals.

HAAS – CATE

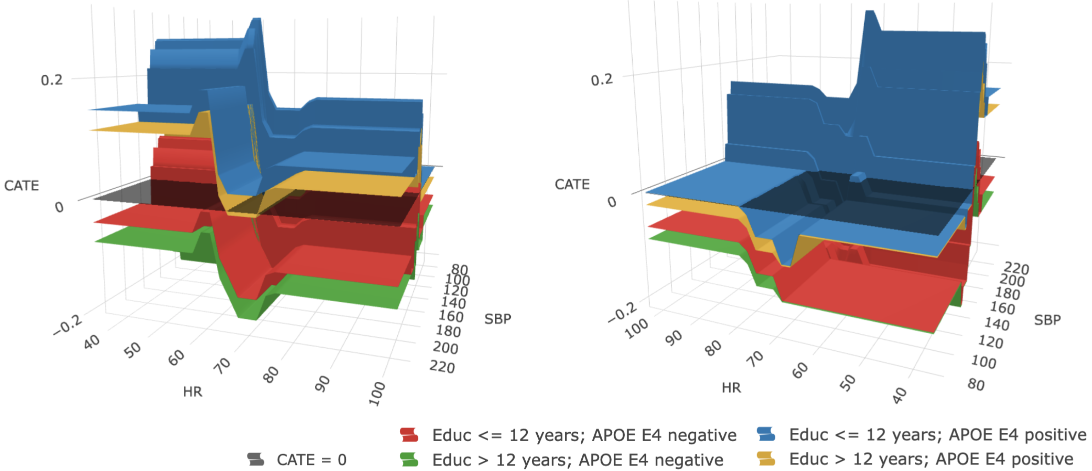


Figure: Estimated CATE surfaces of DFS at age 90 from the ltrcR-learner for different education and ApoE genotype subgroups (views from four different angles).

HAAS – CATE

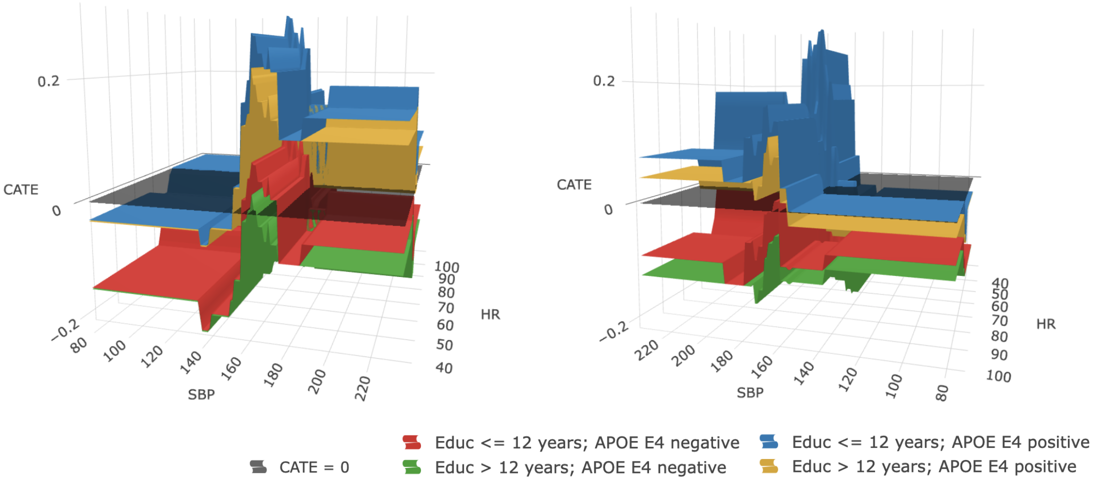


Figure: Estimated CATE surfaces of DFS at age 90 from the ltrcR-learner for different education and ApoE genotype subgroups (views from four different angles).

Summary:

- We developed a general orthogonal and doubly robust framework for handling covariate dependent LTRC, and applied it to estimate ATE and CATE.
- The framework can be applied to a broad range of estimation problems with LTRC data.
  - e.g., survival integral considered in Morenz et al. (2024),
  - e.g., weighted orthogonal learners developed in Morzywolek et al. (2023).

Acknowledgment:

- We thank Drs. Steve Edland, Jue (Marquis) Hou and Kendrick Li for helpful discussions.

References and contact info:

- **Wang, Y.**, Ying, A. and Xu, R. (2024). Doubly robust estimation under covariate-induced dependent left truncation. *Biometrika* 111: 789–808.
- **Wang, Y.**, Ying, A. and Xu, R. (2024). A liberating framework from truncation and censoring, with application to learning treatment effects. *arXiv:2411.18879*.

- Github: https://github.com/wangyuyao98/truncAC   R-package: `truncAIPW`
- Email: yuya.wang@northeastern.edu

Appendix

Nuisance parameter estimation

$$\begin{tikzcd}
F & S_D \arrow[d, "\text{(Eliminate censoring, truncation still remains)}"'] \arrow[dd, bend left] \\
& G \arrow[d, "\text{(Eliminate truncation)}"'] \\
& \pi
\end{tikzcd}$$

- F: use existing software for LTRC data.
- S_D: use existing software for right censored data.
- G: use IPCW weights $\Delta/\hat{S}_D(X-Q|Q,A,Z)$ $+$ use existing software for left truncated data on the reversed time scale.
- π: use weights $\Delta/\{\hat{G}(X|A,Z)\hat{S}_D(X-Q|Q,A,Z)\}$ $+$ existing software for regressing A on Z.

Nuisance parameter estimation

$$\begin{tikzcd}
F & S_D \arrow[d, "\text{(Eliminate censoring, truncation still remains)}"'] \arrow[dd, bend left] \\
 & G \arrow[d, "\text{(Eliminate truncation)}"] \\
 & \pi
\end{tikzcd}$$

$$\begin{tikzcd}
F \arrow[dr, "\text{(Eliminate truncation)}"'] & S_D \arrow[dd, bend left] \\
 & G \arrow[d, "\text{(Eliminate truncation)}"] \\
 & \pi
\end{tikzcd}$$

$$\begin{tikzcd}
 & F \arrow[dl] \arrow[d, "\text{(Eliminate truncation)}"] & S_D \\
\pi & G &
\end{tikzcd}$$

$$\begin{tikzcd}
F \arrow[d, "\text{(Eliminate truncation)}"'] & S_D \arrow[d, "\text{(Eliminate censoring, truncation still remains)}"] \\
\pi & G
\end{tikzcd}$$

Figure: Different schemes for nuisance parameter estimation.

Instantiation for the ltrcR- and ltrcDR-learners

- **ltrcR-learner**:

$$
\begin{aligned}
&\|\hat{\pi} - \pi_0\|_4^2 + \|\hat{F} - F_0\|_{\text{sup},4} \cdot \|\hat{\pi} - \pi_0\|_4 \\
&\quad + \|\hat{F} - F_0\|_{\text{sup},4} \cdot \left\{ \|\hat{G} - G_0\|_{\text{sup},4} + \|\hat{S}_D - S_{D0}\|_{\text{sup},4} \right\} \\
&\quad + \|K_1(\hat{g}, g_0)\|_2 + \|K_2(\hat{g}, g_0)\|_2 + \|K_3(\hat{g}, g_0)\|_2
\end{aligned}
$$

- $\hat{g} = (\hat{F}, \hat{G}, \hat{S}_D)$; $g_0 = (F_0, G_0, S_{D0})$.
- $K_1(\hat{g}, g_0)$: integral product between $\hat{F} - F_0$ and $\hat{G} - G_0$;
- $K_2(\hat{g}, g_0)$: integral product between $\hat{F} - F_0$ and $\hat{S}_D - S_{D0}$;
- $K_2(\hat{g}, g_0)$: higher order integral products.

Instantiation for the ltrcR- and ltrcDR-learners

- **ltrcR-learner**:

$$\begin{aligned}&\|\hat{\pi}-\pi_0\|_4^2+\|\hat{F}-F_0\|_{\text{sup},4}\cdot\|\hat{\pi}-\pi_0\|_4\\&\quad+\|\hat{F}-F_0\|_{\text{sup},4}\cdot\left\{\|\hat{G}-G_0\|_{\text{sup},4}+\|\hat{S}_D-S_{D0}\|_{\text{sup},4}\right\}\\&\quad+\|K_1(\hat{g},g_0)\|_2+\|K_2(\hat{g},g_0)\|_2+\|K_3(\hat{g},g_0)\|_2\end{aligned}$$

- **ltrcDR-learner**:

$$\begin{aligned}&\|\hat{F}-F_0\|_{\text{sup},4}\cdot\|\hat{\pi}-\pi_0\|_4\\&\quad+\|\hat{F}-F_0\|_{\text{sup},4}\cdot\left\{\|\hat{G}-G_0\|_{\text{sup},4}+\|\hat{S}_D-S_{D0}\|_{\text{sup},4}\right\}\\&\quad+\|K_1(\hat{g},g_0)\|_2+\|K_2(\hat{g},g_0)\|_2+\|K_3(\hat{g},g_0)\|_2\end{aligned}$$

- $\hat{g}=(\hat{F},\hat{G},\hat{S}_D)$; $g_0=(F_0,G_0,S_{D0})$.
- $K_1(\hat{g},g_0)$: integral product between $\hat{F}-F_0$ and $\hat{G}-G_0$;
- $K_2(\hat{g},g_0)$: integral product between $\hat{F}-F_0$ and $\hat{S}_D-S_{D0}$;
- $K_2(\hat{g},g_0)$: higher order integral products.

Simulation – ATE

- Estimand: $\theta = \mathbb{P}^*\{T^*(1) > 3\} - \mathbb{P}^*\{T^*(0) > 3\} = -0.1163$.
- 500 simulated data sets each with sample size 1000; 200 boostrap replications are performed.
- Truncation rate: around 20%; treatment rate: around 50%; censoring rate: around 50%.

Methods	F/G-π-S_D	bias	SD	SE/bootSE	CP/bootCP
Param	Cox1/Cox1-lgs1-Cox1	0.0014	0.0481	0.0494/0.0519	0.960/0.966
	Cox2/Cox1-lgs1-Cox1	0.0010	0.0477	0.0500/0.0515	0.966/0.960
	Cox1/Cox2-lgs1-Cox1	0.0035	0.0450	0.0451/0.0488	0.950/0.964
	Cox1/Cox1-lgs2-Cox1	0.0010	0.0456	0.0456/0.0487	0.952/0.970
	Cox1/Cox1-lgs1-Cox2	0.0031	0.0496	0.0518/0.0532	0.960/0.964
	Cox2/Cox2-lgs1-Cox1	0.0060	0.0451	0.0475/0.0500	0.964/0.964
	Cox2/Cox1-lgs2-Cox1	0.0195	0.0454	0.0460/0.0479	0.930/0.944
	Cox2/Cox1-lgs1-Cox2	-0.0025	0.0495	0.0524/0.0529	0.968/0.954
cf	pCox/pCox-gbm-pCox	-0.0028	0.0631	0.0635/0.0627	0.960/0.951
IPW	- /Cox1-lgs1-Cox1	-0.0013	0.0528	0.0514/0.0515	0.946/0.942
	- /Cox2-lgs1-Cox1	0.0753	0.0477	0.0494/0.0476	0.662/0.648
	- /Cox1-lgs2-Cox1	0.0067	0.0511	0.0474/0.0496	0.928/0.940
	- /Cox1-lgs1-Cox2	-0.0393	0.0521	0.0517/0.0519	0.872/0.866
	- /pCox-gbm-pCox	0.0430	0.0583	0.0567/0.0516	0.866/0.837
full		-0.0009	0.0185	0.0184/0.0184	0.938/0.944

HAAS – CATE

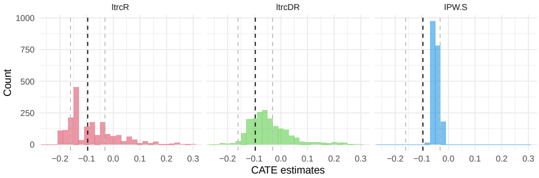


Figure: Histogram of the CATE estimate from different learners; $\nu(t) = I(t > 90)$; the black and grey vertical dashed lines denote the estimate of ATE for the survival probability at 90 years old from the 'cf' estimator and its 95% confidence intervals based on bootstrapped standard errors.