Incremental Causal Effects for Time to Treatment

Zhichen Zhao

Department of Mathematics
University of California, San Diego
zhz147@ucsd.edu

Collaborators: Andrew Ying, Google Inc; Ronghui (Lily) Xu, UCSD

Time to treatment

Examples

- cancer screening
- vaccine during pregnancy
- organ transplant
- anti-rheumatic therapy (eg. Methotrexate)

Traditional causal effects might be ‘if treated’, or ‘when to treat’.

Incremental interventions and effects

This is a type of *stochastic* intervention (as opposed to 'deterministic'):

- instead of intervening on the value of a treatment (Y/N, when), it intervenes on the distribution of the treatment assignment (process);
- may have interpretation as (health) policy effect.

Ref: Kennedy (2019), Bonvini *et al.* (2023) for binary treatment and later developments for discrete time longitudinal treatments.

Setup and notation

Let

- T - time to treatment (initialization)
- Y - outcome of interest measured at time τ
- L - baseline covariates

We observe $\{Y, U = T \wedge \tau, L\}$, denote also
$\Delta = 1(T < \tau) = 1(U < \tau)$.

- Assume that T is absolutely continuous, with
  $\lambda(t|l) = \lim_{h \to 0} P(t \leq T < t + h | T \geq t, L)/h$ and
  $\Lambda(t|l) = \int_0^t \lambda(s|L)ds$ as its conditional hazard function and cumulative hazard function, respectively.

Potential outcome

Denote potential outcome

- Y_t if the treatment is initialized at time t $(t < \tau)$;
- $Y_\infty = Y_\tau = Y_t$ $(t \geq \tau)$ for someone who is never treated.

Q: what would be the expected outcome Y if the hazard function $\lambda(t|l)$ were to be modified, eg. doubled?

In general, we may consider changing the hazard of T by a multiplicative factor $\theta(t, l) > 0$.

Stochastic intervention

Consider $T(\theta)$ a random draw $\sim$ the hazard function $\theta(t, I) \cdot \lambda(t|I)$.

- The incremental causal effect is

$$\psi(\theta) = E(Y_{T(\theta)}).$$

When $\theta \equiv 1$, $\psi(1) = E(Y_{T(1)}) = E(Y)$ corresponds to the factual ('*natural*') distribution of T that we have observed.

Assumptions

1. Consistency: $Y_{T \wedge \tau} = Y$.

2. No Unmeasured Confounding: $T \perp Y_t \mid L, \quad \forall t \in [0, \tau]$.

Identification

- Likelihood for observed data:

$$P(Y|U,L)\cdot\lambda(U|L)^{\Delta}e^{-\int_0^U d\Lambda(t|L)}\cdot P(L)$$

- The target likelihood under intervention:

$$P(Y|U,L)\cdot\{\theta(U,L)\lambda(U|L)\}^{\Delta}e^{-\int_0^U \theta(t,L)d\Lambda(t|L)}\cdot P(L)$$

The likelihood ratio, i.e. Radon-Nikodym derivative, is

$$\theta(U,L)^{\Delta}e^{-\int_0^U\{\theta(t,L)-1\}d\Lambda(t|L)}.$$

Lemma (Identification, ICLR 2025)

Under Assumptions 1 and 2,

$$\psi(\theta)=E(Y_{T(\theta)})=E\left[Y\theta(U,L)^{\Delta}e^{-\int_0^U\{\theta(t,L)-1\}d\Lambda(t|L)}\right].$$

Efficient Influence Function (AIPW)

Under Assumptions 1 and 2, the EIF for the incremental causal effect curve $\psi(\theta)$ is $\phi(\theta; \Lambda, \mu) - \psi(\theta)$, where

$$\begin{aligned}\phi(\theta; \Lambda, \mu) = {} & Y\theta(U, L)^{\Delta} e^{-\int_0^U \{\theta(v,L)-1\} d\Lambda(v|L)} \\ & + \int_0^\tau \mu(u, L)\theta(u, L)^{\delta} e^{-\int_0^u \{\theta(v,L)-1\} d\Lambda(v|L)} \left[\int_0^{U \wedge u} \frac{\theta(v, L) - 1}{S(v|L)} d\Lambda(v|L)\right] dF(u|L) \\ & - \frac{\theta(U, L) - 1}{S(U|L)} \int_{U+}^\tau \mu(u, L)\theta(u, L)^{\delta} e^{-\int_0^u \{\theta(v,L)-1\} d\Lambda(v|L)} dF(u|L),\end{aligned}$$

$\delta = 1(u < \tau)$, $\mu(u, l) = E(Y|U = u, L = l)$, and $F(u|l) = 1 - S(u|l) = 1 - e^{-\Lambda(u|l)}$.

See Hines *et al.* (2022) for tutorial on how to derive EIF.

Estimation

Let $\hat{\psi}(\theta) = \sum_{i=1}^{n} \phi_i(\theta; \hat{\Lambda}, \hat{\mu})/n$, where $\hat{\mu}(u, l)$ and $\hat{\Lambda}(u|l)$ are estimated (semi)parametrically or nonparametrically (with cross-fitting).

We have

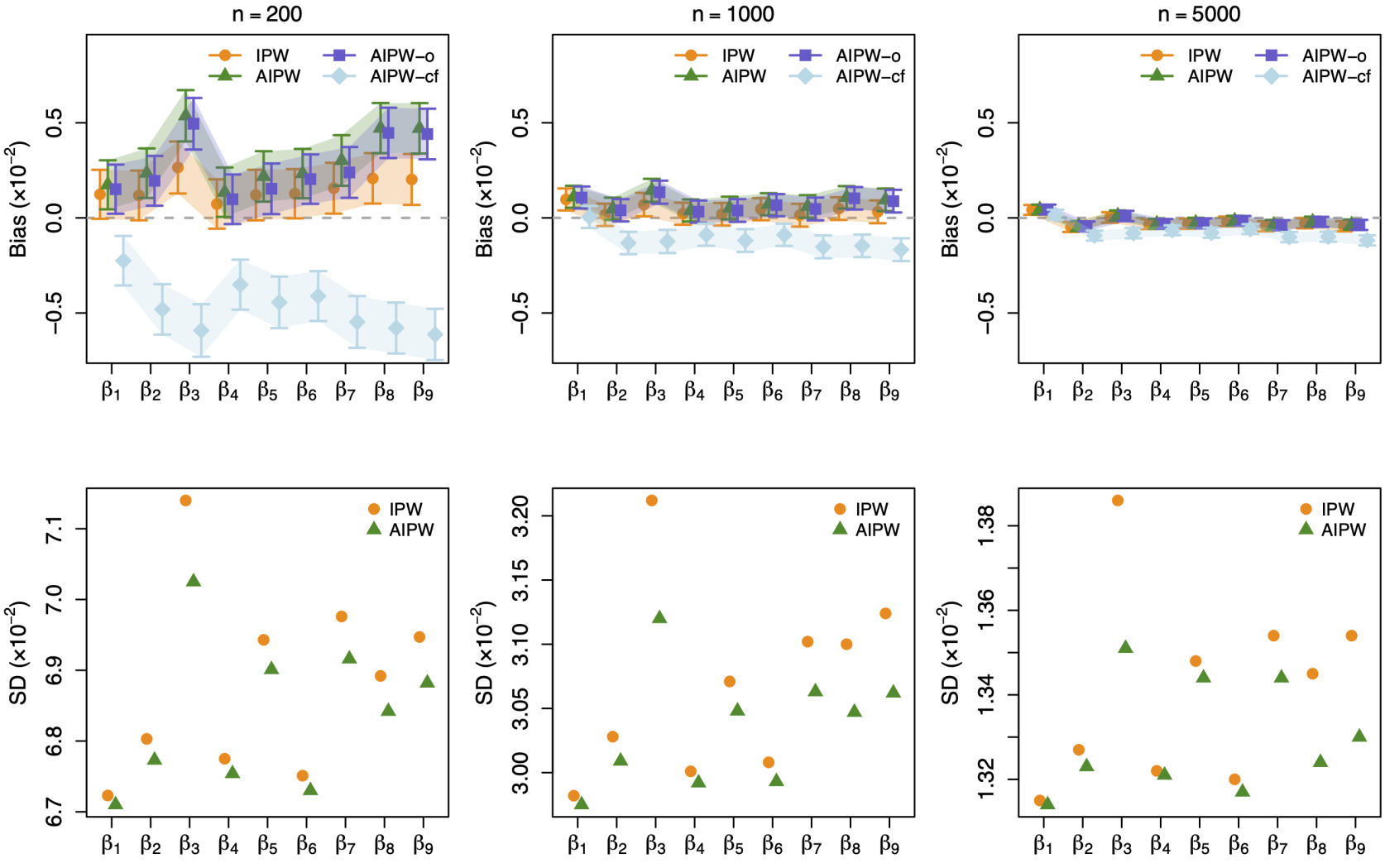
- consistency and AN under suitable conditions;
- uniform inference over θ (work in progress)

Simulation

Table: Coverage probabilities (%) of 95% confidence intervals

	$\theta(t,l) =$	$\beta_1^\top l$	$\beta_2^\top l$	$\beta_3^\top l$	$\beta_4^\top l$	$\beta_5^\top l$	$\beta_6^\top l$	$\beta_7^\top l$	$\beta_8^\top l$	$\beta_9^\top l$
	$\psi(\theta)$	1.062	0.993	0.820	1.047	1.004	1.008	0.953	0.865	0.853
$n = 200$	IPW	94.3	94.7	94.1	94.3	94.3	94.5	94.5	94.5	94.5
	AIPW	94.3	94.5	94.8	94.1	94.1	94.4	94.2	94.5	94.7
	AIPW-cf	94.6	94.8	94.6	94.8	95.0	94.7	94.7	94.1	94.3
	AIPW-o	94.5	94.6	94.8	94.5	94.4	94.6	94.5	94.3	94.7
$n = 1000$	IPW	94.8	94.6	94.2	94.8	94.5	94.7	94.8	94.5	94.2
	AIPW	94.8	94.7	94.5	95.0	94.9	94.8	94.8	94.7	94.5
	AIPW-cf	94.6	94.6	94.7	94.5	95.0	94.6	94.7	95.1	94.9
	AIPW-o	94.5	94.9	94.3	94.7	94.9	95.0	94.8	94.6	94.6

Simulation

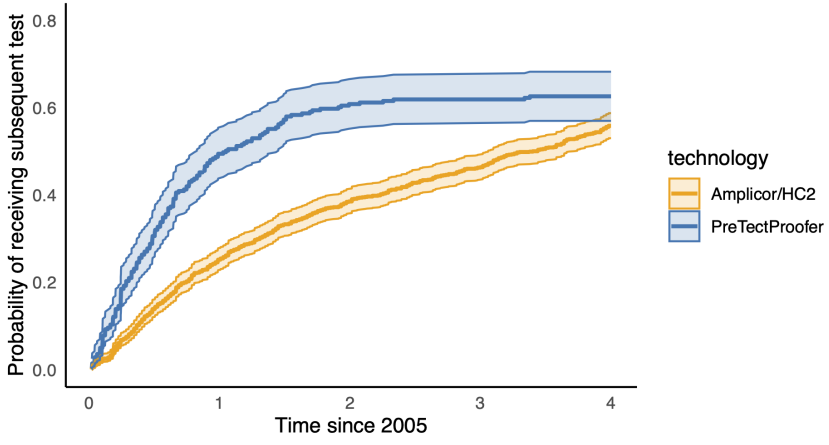

n = 200
n = 1000
n = 5000
IPW
AIPW
AIPW–o
AIPW–cf
Bias (×10⁻²)
SD (×10⁻²)
β1 β2 β3 β4 β5 β6 β7 β8 β9

Data example

PreTectProofer vs. Amplicor/HC2

After inconclusive finding:

- PreTectProofer testing is done more frequently



Y: cervical intraepithelial lesion grade 2 or 3 (CIN2+) by 4 years

- $\bar{Y} = 5.69\%$ for PreTectProofer, 0.87% for Amplicor/HC2

Q: does PreTectProofer have more *false negative*?

Data example

When fitting the Cox PH model to T for Amplicor/HC2 vs. PreTectProofer, we have $\hat{\theta} = 0.677$.

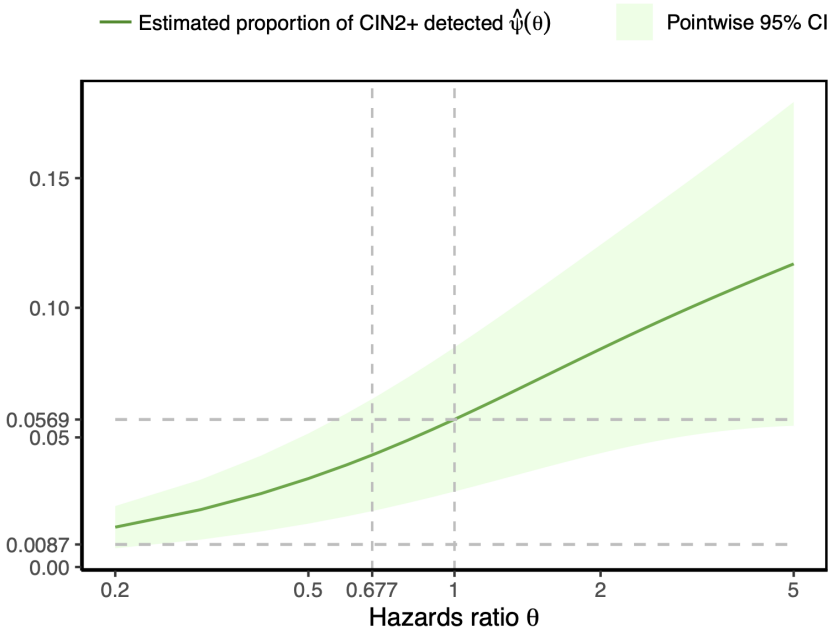


Figure: Estimated incremental causal effects $\hat{\psi}(\theta)$, i.e. proportion of CIN2+ at 4 years, for the PreTectProofer group with pointwise 95% confidence intervals; $\hat{\psi}(0.677) = 0.0432$.