Incremental Causal Effect for Time to Treatment Initialization

Andrew Ying [1], Zhichen Zhao [2], Ronghui Xu [2,3,4]

[1]Irvine [2]Department of Mathematics, University of California San Diego [3]Herbert Wertheim School of Public Health & Human Longevity Science, University of California San Diego

[4]Halıcıoğlu Data Science Institute, University of California San Diego

Background

- Causal questions often involve **time to treatment initialization**, such as when individuals begin vaccination or undergo screening. These decisions are not always deterministic.
- A motivating example from (Bonvini et al., 2023) is about behavioral health services for probationers, in order to reduce their chances of re-arrest. This can be achieved via affecting probationers' *likelihood* of attending services by, for example, providing transportation stipends.
- Standard causal effects (like ATE) often require the **positivity assumption**: everyone must have a chance to receive every treatment level.
- Positivity is often violated in practice (ineligibility, mandatory treatment). **Many traditional causal quantities of interest might be vetoed by absence of positivity alone after checking the data**.
- On the other hand, we introduced the incremental causal effect, bridging this gap by **not requiring positivity**.

Incremental Intervention

- We are interested in answering the question like, "What would happen to outcomes if the **intensity of treatment initialization** were increased or decreased?"
- This leads to the idea of an **incremental intervention**, which shifts the *intensity* (hazard function) of treatment initialization.
- Let $\lambda(t|L)$ be the baseline hazard. Intervention shifts it to $\theta(t, L)\lambda(t|L)$ ($\theta(t, L) > 0$).
- Let $T(\theta)$ be the time-to-treatment under the shifted hazard.
- **Estimand:** Incremental causal effect $\psi(\theta) = \mathbb{E}[Y_{T(\theta)}]$ (mean potential outcome Y under shifted treatment hazard).
- **Key Advantage:** Identification does NOT require positivity.

Identification

Assumptions:

- Consistency: $Y = Y_{T\wedge\tau}$.
- No Unmeasured Confounding (NUC): $T \perp Y_t|L$.

Theorem 1 (IPW Identification): Under Consistency and NUC,

$$\psi(\theta) = \mathbb{E}\left[W(\theta, T, \Delta, L)Y\right],$$

where the weight $W(\theta, T, \Delta, L)$ is:

$$\theta(T, L)^{\Delta} \exp\left\{-\int_0^{T\wedge\tau} (\theta(t, L) - 1)d\Lambda(t|L)\right\}$$

($\Delta = 1(T < \tau)$, $\Lambda(t|L)$ is cumulative hazard).

- **Avoids the need for positivity.**

Estimation

1. Estimate cumulative hazard $\Lambda(t|L)$ using standard methods (e.g., Cox model), get $\hat{\Lambda}(t|L)$.
2. Compute plug-in IPW estimator:

$$\hat{\psi}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \theta(T_i, L_i)^{\Delta_i} \times e^{-\int_0^{T_i\wedge\tau}\{\theta(t,L_i)-1\}d\hat{\Lambda}(t|L)} Y_i.$$

3. Use bootstrap (e.g., multiplier) for confidence intervals.

Theorem 2 (Consistency and Asymptotic Normality): Under Assumptions Consistency, NUC, and some regularity conditions, we have that $\hat{\psi}(\theta)$ converges to $\psi(\theta)$ in probability, that is, for any constant $\varepsilon > 0$,

$$\mathbb{P}(|\hat{\psi}(\theta) - \psi(\theta)| > \varepsilon) \to 0,$$

when $n \to \infty$, and the root-n scaled centered difference $\sqrt{n}\{\hat{\psi}(\theta) - \psi(\theta)\}$ is asymtotically linear and thus converges to a normal variable weakly, that is,

$$\sqrt{n}\{\hat{\psi}(\theta) - \psi(\theta)\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Xi_i + o_P(1) \to \mathcal{N}(0, \mathrm{Var}(\Xi_1)),$$

for some random variable Ξ_1, in distribution.

Simulation Results

Setup:

1. $L_i \sim \mathrm{Unif}(0, 1)$.
2. $\mathbb{P}(T_i > t|L_i) = \exp\{-\exp(0.25L_i)t\}$.
3. $Y_i \sim \mathcal{N}(\exp(1 - 1.5L_i - (2 - T_i \wedge 2)), 0.5^2)$.

$\hat{\Lambda}(t|L)$ from Cox model.
Results for $n = 1000$, $R = 1000$:

Table 1: Simulation results of the IPW estimator. We report bias, percent bias (%Bias), standard error of estimate (SEE), average estimated standard deviation (SD) and coverage probability of Wald type 95% confidence intervals (95% CP) of $\hat{\psi}(\theta)$ by $B = 200$ multiplier bootstrap, for sample size $n = 1000$ and $R = 1000$ Monte Carlo samples.

$\theta(t\|l) \equiv$	1/2.5	1/2	1/1.5	1.5	2	2.5
$\psi(\theta)$	0.893	0.808	0.694	0.404	0.336	0.297
Bias ($\times 10^{-2}$)	-0.388	-0.253	-0.139	0.031	0.043	0.039
%Bias	-0.434	-0.313	-0.201	0.077	0.128	0.132
SEE ($\times 10^{-2}$)	3.035	2.778	2.508	2.055	2.033	2.096
SD ($\times 10^{-2}$)	2.982	2.745	2.495	2.059	2.032	2.088
95% CP (%)	93.9	93.7	94.0	94.5	95.0	95.2

Application: Methotrexate for Rheumatoid Arthritis

Goal: Evaluate effect of changing Methotrexate (MTX) initialization rate on joint pain (Y) at 1 year. Such an analysis can illustrate how varying levels of aggressiveness or conservatism in prescribing MTX might influence average disease progression.

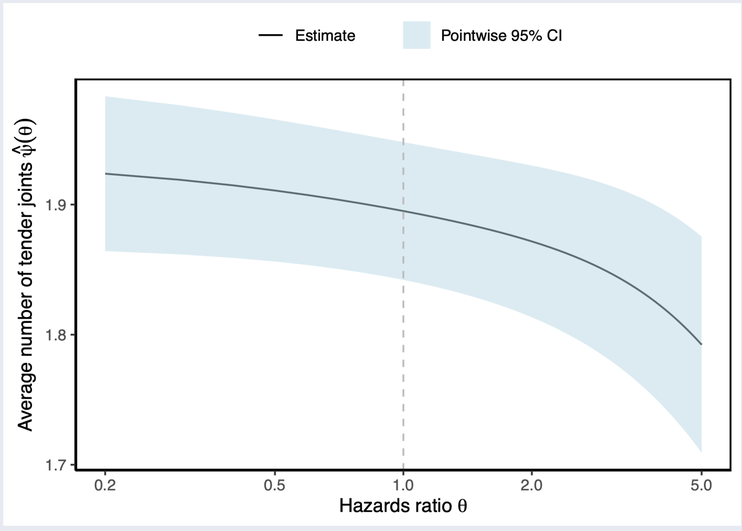

Figure 1: Estimated effect $\hat{\psi}(\theta)$ vs. hazard multiplier θ (constant). Increasing the hazard of starting MTX (higher θ) reduces average joint pain at 1 year.

Findings:

- Doubling the hazard ($\theta = 2$) reduces avg. joint pain by 1.23%.
- Multiplying hazard by 5 ($\theta = 5$) reduces avg. joint pain by 5.42%.
- Consistent with known protective effect of MTX.

Conclusion

- Defined and identified **incremental causal effect** for time-to-treatment initialization.
- Key contribution: Identification **without positivity assumption**.
- Provided practical IPW estimation framework using standard survival analysis tools.
- Useful for evaluating policy-style interventions affecting treatment timing.
- Future work: Efficiency improvements (IPW is known to be inefficient), extensions to complex longitudinal settings, relaxing NUC.