Estimating treatment effects with competing intercurrent events in randomized controlled trials

Sizhu Lu

UC Berkeley, Statistics

September 23, 2025
Online Causal Inference Seminar

Joint work with Yanyao Yi, Yongming Qu, Karen Liu (Eli Lilly),
Ting Ye (University of Washington, Biostatistics),
and Peng Ding (UC Berkeley, Statistics)

Overview

Motivating randomized controlled trials (RCTs)

Causal parameter of interest

Assumption and identification

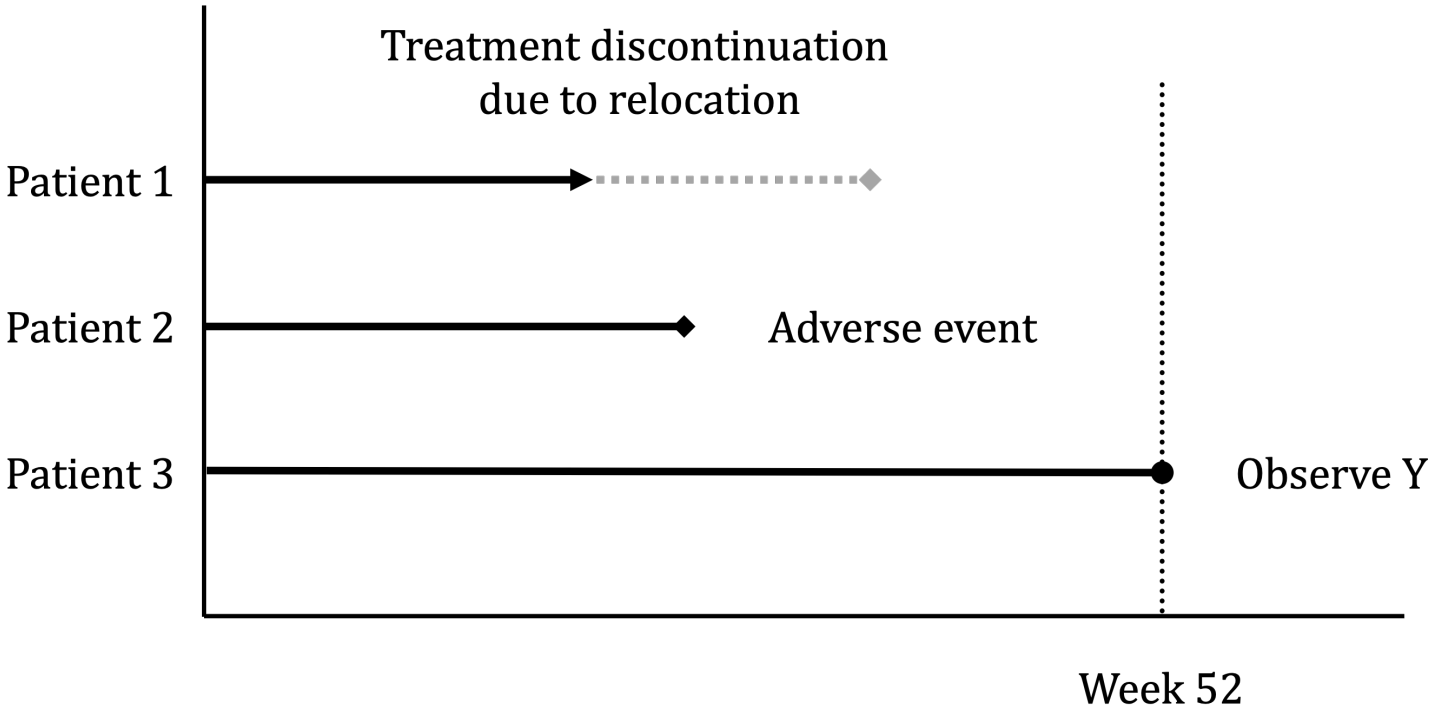
From basic estimators to augmented, robust estimators

Real-world application

Motivating example: two phase-3 immunology trials

- Morand et al. (2023) and Petri et al. (2023)
- Causal effect of baricitinib versus placebo on Systemic Lupus Erythematosus
- Primary endpoint: an immune response index measured *52 weeks after* treatment initiation
- Ideally, comparisons between two groups
- Outcomes not measured: 218/760 and 211/775 in two trials

Motivating example: two phase-3 immunology trials



- Treatment discontinuation due to relocation and adverse event in the example are called intercurrent events (ICEs)
- ICEs: events that occur *after* the treatment initiation and *affect* either the interpretation or existence of outcome measurements

ICEs in the motivating RCTs

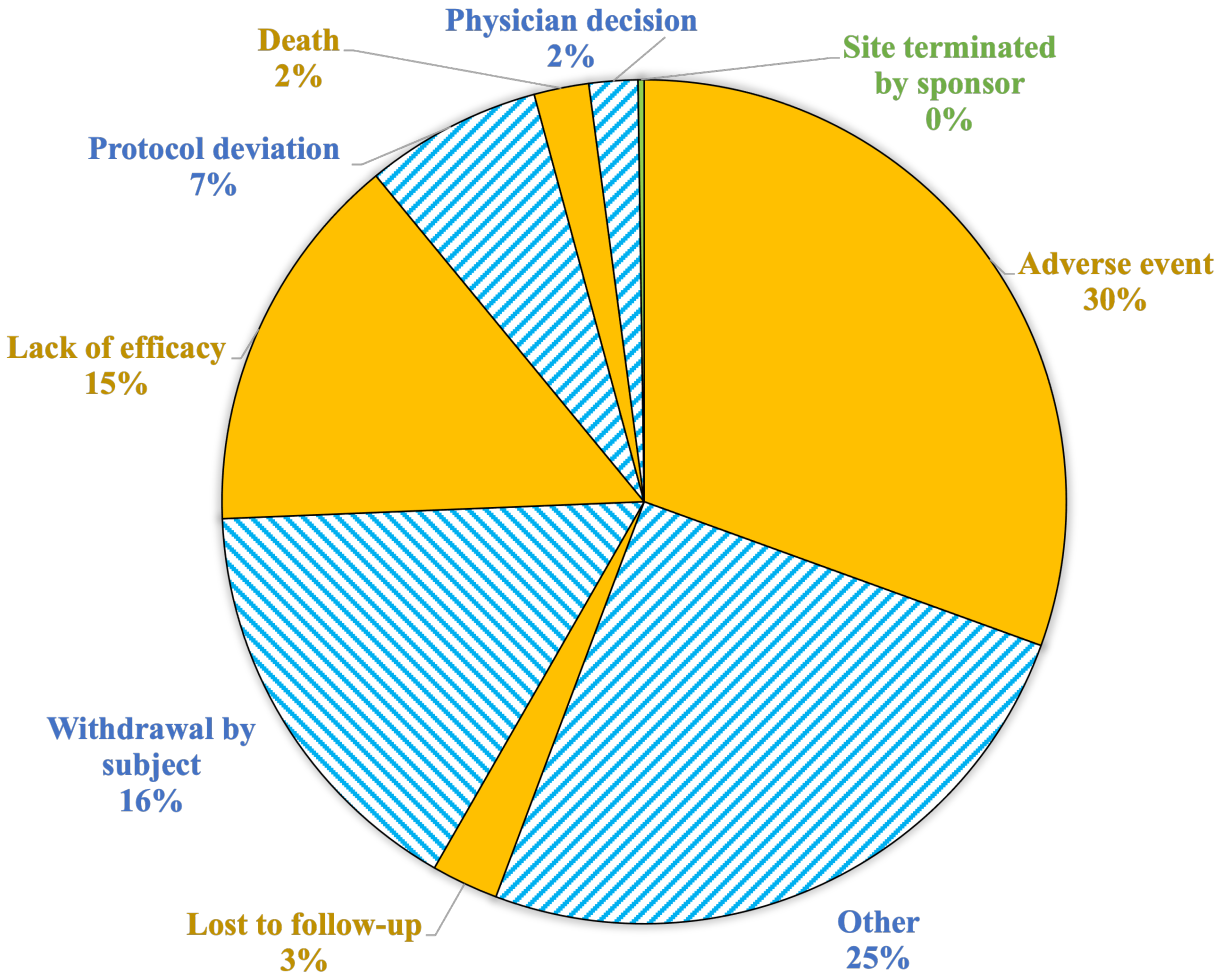


Figure: Pie chart showing the ICE types and proportions

Five strategies to address ICEs

- ICH E9 (R1): a guideline published in 2019 to address ICEs by the International Council for Harmonisation (ICH)
- Treatment policy strategy: intention-to-treat-type principle
- Hypothetical strategy: what if hypothetically the ICE would not occur
- Composite outcome strategy: modify the causal parameter of interest
  - an ICE is itself informative about the patients' outcome of interest
  - e.g., when the outcome is success or failure, the occurrence of ICE can be treated as another mode of failure
- While-on-treatment strategy: compare outcomes before ICEs
- Principal stratification strategy: causal effects on subgroups

Our proposal: combine two strategies

- Classify ICEs into two broad types:
  - effect-informative ICEs, e.g., adverse effect, lack of efficacy
  - effect-uninformative ICEs, e.g., treatment discontinuation due to relocation or COVID-19 lockdown
- Combining composite outcome and hypothetical strategies
  - effect-informative ICEs: composite outcome strategy
  - effect-uninformative ICEs: hypothetical strategy
- Key challenges:
  - combining composite outcome and hypothetical strategies needs new theory and method
  - need to deal with competing ICEs

Challenge in combining the two strategies

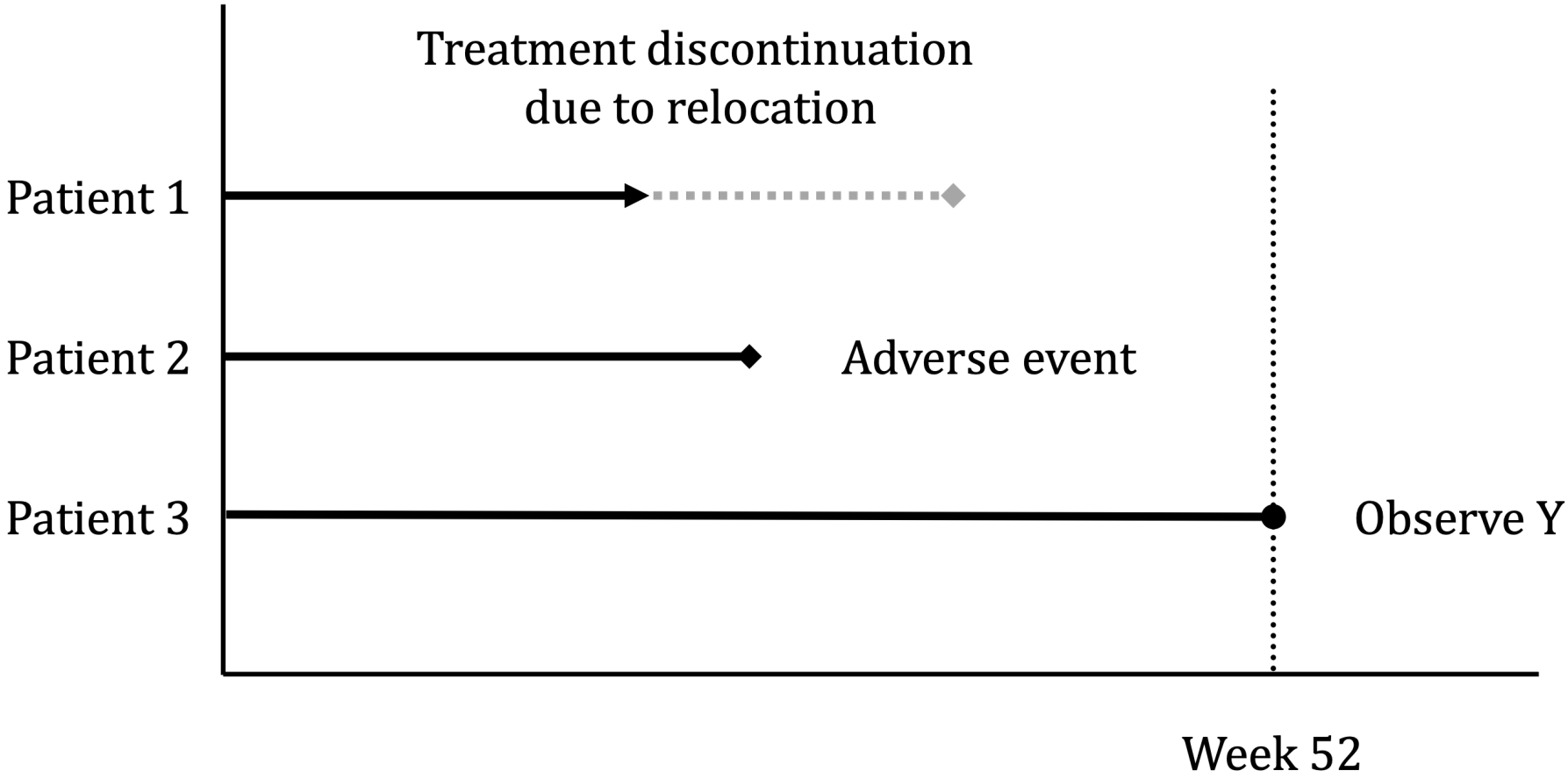


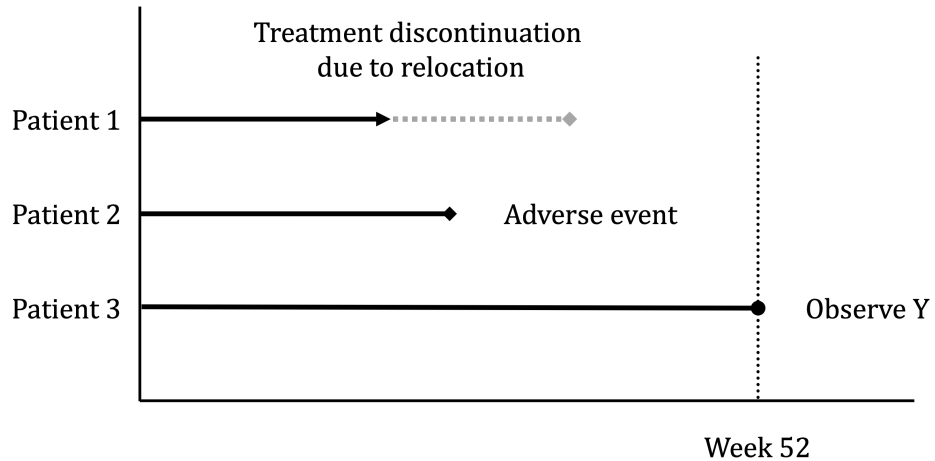
Figure: Illustration of the motivating immunology trial example

Notation and potential outcomes

- Binary treatment: $A = 1, 0$ for treatment and control
- Primary endpoint: Y, measured at a pre-specified time point k
- Two types of ICEs:
  - effect-informative ICEs: event time T
  - effect-uninformative ICEs: event time C
- Define $T = \infty$ if $T > k$, and $C = \infty$ if $C > k$
- Both ICEs are *post-treatment* variables, thus having potential values $T(a)$ and $C(a)$
- Potential outcomes: $Y(a, t, c)$
- Consistency: the observed outcome $Y = Y(A, T(A), C(A))$

Causal parameter of interest

$$\tau = E[\underbrace{Y(1, c = \infty)}_{\text{hypothetical}} \underbrace{1\{T(1) = \infty\}}_{\text{composite}}] - E[Y(0, c = \infty)\, 1\{T(0) = \infty\}]$$



observed ICE types	(T, C, k)-relationship	$Y(A, c = \infty)$	$1(T = \infty)$	composite outcome
TD	$T \wedge k > C$	?	?	?
AE	$C \wedge k > T$	?	0	0
no AE/TD	$C \wedge T > k$	Y	1	Y

Identification assumptions

Assumption 1 (Randomization)

$A \perp\!\!\!\perp \{Y(a, c = \infty), T(a), C(a)\} \mid X$ *for* $a = 0, 1$.

- Guaranteed by the experimental design in a randomized trial

Identification assumptions

Assumption 1 (Randomization)

$A \perp\!\!\!\perp \{Y(a, c = \infty), T(a), C(a)\} \mid X$ *for* $a = 0, 1$.

- Guaranteed by the experimental design in a randomized trial

Assumption 2 (Effect-uninformative ICE time)

$C(a) \perp\!\!\!\perp \{Y(a, c = \infty), T(a)\} \mid X$ *for* $a = 0, 1$.

- Time to treatment discontinuation due to relocation is independent of the hypothetical potential outcome and the time to adverse effect given baseline covariates

Nonparametric identification

Theorem 1 (Nonparametric identification)

Under Assumptions 1 and 2, τ is nonparametrically identified by the following identification formulas:

$$\tau = E\left\{\mu_1(X)S_1(k \mid X) - \mu_0(X)S_0(k \mid X)\right\} \quad (1)$$

$$= E\left[\frac{AY1(T \wedge C > k)}{e(X)G_1(k \mid X)} - \frac{(1-A)Y1(T \wedge C > k)}{\{1-e(X)\}\, G_0(k \mid X)}\right]. \quad (2)$$

Nonparametric identification 1: outcome models

$$\tau = E\left\{\mu_1(X)S_1(k \mid X) - \mu_0(X)S_0(k \mid X)\right\}, \tag{1}$$

where

- $\mu_a(X) = E(Y \mid T \wedge C > k, X, A = a)$: conditional mean of observed outcome in the subsample with no ICE and $A = a$, and

Nonparametric identification 1: outcome models

$$\tau = E\left\{\mu_1(X)S_1(k \mid X) - \mu_0(X)S_0(k \mid X)\right\}, \tag{1}$$

where

- $\mu_a(X) = E(Y \mid T \wedge C > k, X, A = a)$, and
- $S_a(k \mid X) = \text{pr}(T > k \mid X, A = a)$: survival probability of AE time larger than k in the subsample $A = a$ conditional on covariates

Nonparametric identification 2: weighting

$$\tau = E\left[\frac{AY1(T \wedge C > k)}{e(X)G_1(k \mid X)} - \frac{(1-A)Y1(T \wedge C > k)}{\{1-e(X)\}G_0(k \mid X)}\right], \qquad (2)$$

where

- $e(X) = \text{pr}(A = 1 \mid X)$: propensity score, and

Nonparametric identification 2: weighting

$$\tau = E\left[\frac{AY1(T \wedge C > k)}{e(X)G_1(k \mid X)} - \frac{(1-A)Y1(T \wedge C > k)}{\{1-e(X)\}\, G_0(k \mid X)}\right], \tag{2}$$

where

- $e(X) = \text{pr}(A = 1 \mid X)$, and
- $G_a(k \mid X) = \text{pr}(C > k \mid X, A = a)$: conditional probability of not censoring up until time k in the subsample $A = a$

Nonparametric identification

$$\tau = E\left\{\mu_1(X)S_1(k \mid X) - \mu_0(X)S_0(k \mid X)\right\} \quad (1)$$

$$= E\left[\frac{AY1(T \wedge C > k)}{e(X)G_1(k \mid X)} - \frac{(1-A)Y1(T \wedge C > k)}{\{1-e(X)\}\, G_0(k \mid X)}\right]. \quad (2)$$

- $\mu_a(X) = E(Y \mid T \wedge C > k, X, A = a)$
- $S_a(k \mid X) = \text{pr}(T > k \mid X, A = a)$
- $e(X) = \text{pr}(A = 1 \mid X)$
- $G_a(k \mid X) = \text{pr}(C > k \mid X, A = a)$
- Identification of $S_a(t \mid X)$ and $G_a(t \mid X)$ for $t \leq k$ (Robin and Rotnitzky, 1992; Robins and Finkelstein, 2000)

Two basic estimators based on two identification formulas

- Outcome regression estimator:

$$\hat{\tau}^{\text{out}} = n^{-1} \sum_{i=1}^{n} \hat{\mu}_1(X_i) \hat{S}_1(k \mid X_i) - n^{-1} \sum_{i=1}^{n} \hat{\mu}_0(X_i) \hat{S}_0(k \mid X_i)$$

- Inverse propensity score weighting estimator:

$$\hat{\tau}^{\text{ipw}} = n^{-1} \sum_{i=1}^{n} \frac{A_i Y_i 1(T_i \wedge C_i > k)}{\hat{e}(X_i) \hat{G}_1(k \mid X_i)} - n^{-1} \sum_{i=1}^{n} \frac{(1 - A_i) Y_i 1(T_i \wedge C_i > k)}{\{1 - \hat{e}(X_i)\} \hat{G}_0(k \mid X_i)}$$

- $\hat{\tau}^{\text{out}}$: consistent if the subsample outcome model and the survival function are correctly specified
- $\hat{\tau}^{\text{ipw}}$: consistent if the propensity score model and the censoring mechanism are correctly specified

An augmented, conditionally doubly robust estimator

- Similar to the classic doubly robust estimator by combining outcome regression and inverse propensity score weighting
- Augment weighting by outcome regression:

$$\hat{\tau}^{\text{aug}} = \hat{\tau}^{\text{ipw}} - n^{-1} \sum_{i=1}^{n} \left\{ \frac{A_i - \hat{e}(X_i)}{\hat{e}(X_i)} \hat{\mu}_1(X_i) \hat{S}_1(k \mid X_i) + \frac{A_i - \hat{e}(X_i)}{1 - \hat{e}(X_i)} \hat{\mu}_0(X_i) \hat{S}_0(k \mid X_i) \right\}.$$

- Conditionally doubly robust: Assume $G_a(k \mid X)$ is correct for $a = 0, 1$. $\hat{\tau}^{\text{aug}}$ is consistent for τ if either $e(X)$ is correct, or both $\mu_a(X)$ and $S_a(k \mid X)$ are correct for $a = 0, 1$
- $\hat{\tau}^{\text{aug}}$ improves $\hat{\tau}^{\text{ipw}}$ but may not improve $\hat{\tau}^{\text{out}}$

Another augmented, doubly robust, and semiparametrically efficient estimator, based on efficient influence function

$$\hat{\tau}^{\text{eif}} = \hat{\tau}^{\text{aug}} + n^{-1} \sum_{i=1}^{n} \frac{A_i}{\hat{e}(X_i)} \hat{\mu}_1(X_i) \hat{S}_1(k \mid X_i) \int_0^{\tilde{T}_i} \frac{dM_{\hat{G}_1}(t)}{\hat{S}_1(t \mid X_i)\, \hat{G}_1(t \mid X_i)}$$
$$- n^{-1} \sum_{i=1}^{n} \frac{1 - A_i}{1 - \hat{e}(X_i)} \hat{\mu}_0(X_i) \hat{S}_0(k \mid X_i) \int_0^{\tilde{T}_i} \frac{dM_{\hat{G}_0}(t)}{\hat{S}_0(t \mid X_i)\, \hat{G}_0(t \mid X_i)} .$$

- $\tilde{T}_i = C_i \wedge T_i \wedge k$ and $\Delta_i = 1(C_i \geq T_i \wedge k)$
- Further augmentation based on martingales: $\mathrm{d}M_{G_a}(t) = 1(C \in (t, t + \mathrm{d}t], \Delta = 0) - 1(\tilde{T} \geq t)\mathrm{d}\Lambda_a(t \mid X)$ with $\Lambda_a(t \mid X)$ denoting the conditional cumulative hazard function for the effect-uninformative ICE C in the treatment group $A = a$ for $a = 0, 1$

Double robustness and semiparametric efficiency

- $\hat{\tau}^{\text{eif}}$ is doubly robust in the sense that it is consistent for τ if either
  - $\mu_a(X)$ and $S_a(t \mid X)$ are correct for $t \leq k$ and $a = 0, 1$; or
  - $e(X)$ and $G_a(t \mid X)$ are correct for $t \leq k$ and $a = 0, 1$
- $\hat{\tau}^{\text{eif}}$ improves the previous three estimators in terms of robustness
- Asymptotically linear and achieves the semiparametric efficiency bound

Real-world application

- Two double-blinded, randomized, placebo-controlled phase-3 immunology trials
- Effect of baricitinib on systemic lupus erythematosus
- Doses: 2mg baricitinib, 4mg baricitinib, and placebo
- Primary outcome: Systemic lupus erythematosus Responder Index 4 (SRI4) at week 52, a binary composite responder index based on:
  - improvement in disease activity, and
  - without worsening of the overall condition or the development of substantial disease activity in new organ systems
- Effect-informative ICEs: 82.6% and 84.4%; and effect-uninformative ICEs: 17.4% and 15.6%
- Covariates: geographic region, corticosteriod use, Physician's Global Assessment score

Data analysis results

		Trial 1 (Petri et al., 2023)				Trial 2 (Morand et al., 2023)			
		$\hat{\tau}^{\text{out}}$	$\hat{\tau}^{\text{ipw}}$	$\hat{\tau}^{\text{aug}}$	$\hat{\tau}^{\text{eif}}$	$\hat{\tau}^{\text{out}}$	$\hat{\tau}^{\text{ipw}}$	$\hat{\tau}^{\text{aug}}$	$\hat{\tau}^{\text{eif}}$
2mg	point	0.030	0.029	0.026	0.026	0.019	0.019	0.022	0.022
	se	0.042	0.043	0.042	0.043	0.042	0.043	0.043	0.042
	p-value	0.479	0.504	0.534	0.540	0.643	0.662	0.602	0.606
4mg	point	0.113	0.120	0.115	0.113	−0.002	−0.002	−0.002	−0.002
	se	0.046	0.046	0.046	0.046	0.042	0.042	0.042	0.042
	p-value	0.013	0.008	0.012	0.013	0.961	0.962	0.966	0.969

- Coherent results across estimators: no severe model misspecification
- Different from ad hoc methods (details in the paper)
- Incoherent results from two trials: negative results for drug approval

Thank you very much!

sizhu_lu@berkeley.edu

Simulation

	$\hat{\tau}^{\text{out}}$			$\hat{\tau}^{\text{ipw}}$			$\hat{\tau}^{\text{aug}}$			$\hat{\tau}^{\text{eif}}$		
	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR
all_correct	0.003	0.100	0.989	0.005	0.154	0.977	0.004	0.152	0.980	0.004	0.110	0.982
e_wrong	−0.002	0.134	0.978	−0.338	0.527	0.916	−0.003	0.479	0.961	0.000	0.234	0.979
e_G_wrong	−0.002	0.130	0.985	−0.194	1.591	0.831	0.141	1.559	0.972	−0.013	0.354	0.969
μ_S_wrong	−0.085	0.170	0.936	−0.009	0.186	0.974	−0.009	0.186	0.974	−0.012	0.176	0.972
all_wrong	0.230	0.233	0.837	0.660	5.799	0.898	0.850	5.782	0.967	0.573	3.022	0.919

Semiparametric efficient influence function (EIF)

Theorem 2 (EIF for μ_1)

Under the nonparametric model with Assumptions 1 and 2, the EIF for μ_1 is

$$
\begin{aligned}
D_{\mu_1} \quad = \quad & \frac{A}{e(X)} \left\{ \frac{Y 1(T \wedge C > k)}{G_1(k \mid X)} + \mu_1(X) S_1(k \mid X) \int_0^{\tilde{T}} \frac{\mathrm{d}M_{G_1}(t)}{S_1(t \mid X) G_1(t \mid X)} \right\} \\
& - \frac{A - e(X)}{e(X)} \mu_1(X) S_1(k \mid X) - \mu_1, \qquad (3)
\end{aligned}
$$

- $\tilde{T} = T \wedge C \wedge k$: the observed event time;
- $\mathrm{d}M_{G_1}(t) = 1(C \in (t, t + \mathrm{d}t], \Delta = 0) - 1(\tilde{T} \geq t)\mathrm{d}\Lambda_1(t \mid X)$: the martingale constructed from the censoring counting process;
- $\Lambda_1(t \mid X)$: the conditional cumulative hazard function for the censoring C in the treatment subgroup.

EIF estimator

With discrete observed time points, the integration part for i:

$$\int_0^{\tilde{T}_i} \frac{\mathrm{d}M_{\hat{G}_1}(t)}{\hat{S}_1(t \mid X_i)\hat{G}_1(t \mid X_i)} = \sum_{t \le \tilde{T}_i} \frac{1(\Delta_i = 0, C_i = t) - \hat{\lambda}_{C_1}(t \mid X_i)}{\hat{S}_1(t \mid X_i)\hat{G}_1(t \mid X_i)}$$

$$= -\sum_{t \le \tilde{T}_i} \frac{\hat{\lambda}_{C_1}(t \mid X_i)}{\hat{S}_1(t \mid X_i)\hat{G}_1(t \mid X_i)} + \frac{1(\Delta_i = 0)}{\hat{S}_1(\tilde{T}_i \mid X_i)\hat{G}_1(\tilde{T}_i \mid X_i)}. \quad (4)$$

- $\hat{\lambda}_{C_1}(t \mid X_i)$: estimated conditional hazard of censoring;
- First term in (4): summation of $-\hat{\lambda}_{C_1}(t \mid X_i)/\{\hat{S}_1(t \mid X_i)\hat{G}_1(t \mid X_i)\}$ over all observed event time points before $\tilde{T}_i$.
- Second term in (4):
  - 0 for observations that are not right-censored by LF;
  - $1/\{\hat{S}_1(\tilde{T}_i \mid X_i)\hat{G}_1(\tilde{T}_i \mid X_i)\}$ for observations with an LF event happened at time $\tilde{T}_i$.