The Roles of Estimands and Assumptions in Causal Inference

Discussion on *"Chasing Shadows: How Implausible Assumptions Skew Our Understanding of Causal Estimands"*

Sizhu Lu and Peng Ding

Department of Statistics, UC Berkeley

ASA BIOP Distance Learning Webinar
November 7, 2025

Distinction between causal estimands and assumptions

- Estimands: correspond to the scientific question of interest
- Identification assumptions: guided by our understanding of the data-generating processes
- Point identification versus bounds on the estimand
- Additional assumptions for analytical convenience, not fully justified by our understanding of the data-generating process: sensitivity analyses

Problem setup

- RCT: randomized treatment $T \in \{0, 1\}$; covariates X; outcome Y
- Potential outcomes $Y(1), Y(0)$
- Observed outcome $Y = TY(1) + (1 - T)Y(0)$
- Target estimand: average causal effect of T on Y

$$\tau = E\{Y(1) - Y(0)\} = E(Y \mid T = 1) - E(Y \mid T = 0)$$

Problem setup

- RCT: randomized treatment $T \in \{0, 1\}$; covariates X; outcome Y
- Potential outcomes $Y(1), Y(0)$
- Observed outcome $Y = TY(1) + (1 - T)Y(0)$
- Target estimand: average causal effect of T on Y

$$\tau = E\{Y(1) - Y(0)\} = E(Y \mid T = 1) - E(Y \mid T = 0)$$

- **Challenge**: non-adherence can happen after the treatment initiation
- Adherence indicator $A \in \{0, 1\}$, with $A = 1$ if adhered to the assigned treatment until the measure of Y; $A = 0$ otherwise

Intention-to-treat

$$\tau^{\mathrm{ITT}} = E\{Y(1) - Y(0)\} = E(Y \mid T = 1) - E(Y \mid T = 0)$$

- Causal effect of *being assigned* to treatment versus control, view non-adherence as a reflection of the clinical practice
- Same identification if Y measured for all patients regardless of their adherence

Intention-to-treat

$$\tau^{\text{ITT}} = E\{Y(1) - Y(0)\} = E(Y \mid T = 1) - E(Y \mid T = 0)$$

- Causal effect of *being assigned* to treatment versus control, view non-adherence as a reflection of the clinical practice
- Same identification if Y measured for all patients regardless of their adherence

- Non-adherence often creates missing Y with missing not at random
- Dilution under low adherence
- Poor generalizability, if adherence proportion changes in the future
- Terminating events, e.g., death, leading to not well-defined outcome

Principal stratification

- Principal stratum-specific causal effect:

$$\tau_{11}^{\mathrm{PS}} = E\{Y(1) - Y(0) \mid A(1) = 1, A(0) = 1\}$$

- Subgroup effect: among patients who would always adhere
- Separates biologic efficacy from behavioral adherence
- Clinically relevant when tolerability/adherence is a stable patient trait

Principal stratification

- Principal stratum-specific causal effect:

$$\tau_{11}^{\mathrm{PS}} = E\{Y(1) - Y(0) \mid A(1) = 1, A(0) = 1\}$$

- Subgroup effect: among patients who would always adhere
- Separates biologic efficacy from behavioral adherence
- Clinically relevant when tolerability/adherence is a stable patient trait
- Identification assumptions:
  - Monotonicity: $A(1) \geq A(0)$
  - Principal ignorability: $Y(t) \perp\!\!\!\perp A(t-1) \mid A(t), X$ for $t = 0, 1$

Principal stratification

- Principal stratum-specific causal effect:

$$\tau_{11}^{\mathrm{PS}} = E\{Y(1) - Y(0) \mid A(1) = 1, A(0) = 1\}$$

- Subgroup effect: among patients who would always adhere
- Separates biologic efficacy from behavioral adherence
- Clinically relevant when tolerability/adherence is a stable patient trait
- Identification assumptions:
  - Monotonicity: $A(1) \geq A(0)$
  - Principal ignorability: $Y(t) \perp\!\!\!\perp A(t-1) \mid A(t), X$ for $t = 0, 1$
- Strong and untestable
- Sensitivity analysis methodologies available for both (Jiang et al., 2022)

Hypothetical strategy

- *What if* the non-adherence had not occurred
- Indicates that A is manipulatable and on the causal pathway from T to Y (Figure 1 in Vansteelandt and Van Lancker, 2025)
- An augmented notation: $Y(t, a)$
- Hypothetical strategy estimand: $\tau^{\text{HYP}} = E\{Y(1,1) - Y(0,1)\}$, controlled direct effect
- What would the treatment effect be in the hypothetical scenario where all patients adhere
- Strong and untestable identification assumptions (more details in the paper)

Choosing the right estimand

	ITT	Principal stratification	Hypothetical
Estimand	$E\{Y(1, A(1)) - Y(0, A(0))\}$	$E\{Y(1,1) - Y(0,1) \mid a_1, a_0\}$ or $E\{Y(1, A(1)) - Y(0, A(0)) \mid a_1, a_0\}$	$E\{Y(1,1) - Y(0,1)\}$
Clinical	What happens in practice regardless of adherence	Among people who would adhere no matter which arm, what is the biological treatment effect under full adherence	What would the treatment effect be if everyone could adhere fully in both arms
Use when	Pragmatic; adherence part of intervention	Adherence/tolerability is a stable trait	Adherence modifiable and can be improved/enforced
ID	Weaker	Strong and untestable	Strong and untestable
Pros	Preserves randomization; policy-relevant; simple	Separates biology vs adherence; can transport across settings with different adherence proportions	Relatively transparent identification; clinically relevant when adherence modifiable
Cons	Dilution under low adherence; not terminating events	Strong untestable assumptions; latent strata	Change when adherence patterns change

Thank you very much!

sizhu_lu@berkeley.edu
pengdingpku@berkeley.edu

Identification formulas for τ_{11}

Let $\pi(X) = \Pr(T = 1 \mid X)$, $p_t(X) = \Pr(A = 1 \mid T = t, X)$, $p_t = \Pr(A = 1 \mid T = t)$. Then

$$\begin{aligned}\tau_{11} &= E\Big[\frac{p_0(X)}{p_0}\big\{E(Y \mid A = 1, T = 1, X) - E(Y \mid A = 1, T = 0, X)\big\}\Big] \\ &= E\left\{\frac{p_0(X)}{p_0}\frac{A}{p_1(X)}\frac{T}{\pi(X)}Y\right\} - E\left\{\frac{A}{p_0}\frac{1-T}{1-\pi(X)}Y\right\}.\end{aligned}$$

Remark: Efficient implementations can combine outcome and propensity models; sensitivity analysis recommended.