Quantitative Social Science Methods, I, Lecture Notes: Research Designs for Causal Inference

Gary King[1]
Institute for Quantitative Social Science
Harvard University

August 17, 2020

[1] GaryKing.org

Components of Causal Estimation Error

Research Designs

Issues in Ideal Designs

Reference

- Kosuke Imai, Gary King, and Elizabeth Stuart. Misunderstandings among Experimentalists and Observationalists: Balance Test Fallacies in Causal Inference *Journal of the Royal Statistical Society*, Series A, 171, Part 2 (2008): 1–22.
- `http://j.mp/MisExpObs`

Notation

Notation

- Sample n units from finite population size N (typically $N \gg n$)

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i
- Sample selection: $I_i = 1$ if selected, 0 otherwise

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i
- Sample selection: $I_i = 1$ if selected, 0 otherwise
- Treatment assignment: $T_i = 1$ if treated group, 0 if control

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i
- Sample selection: $I_i = 1$ if selected, 0 otherwise
- Treatment assignment: $T_i = 1$ if treated group, 0 if control
- (Assume: treated and control groups are each of size $n/2$)

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i
- Sample selection: $I_i = 1$ if selected, 0 otherwise
- Treatment assignment: $T_i = 1$ if treated group, 0 if control
- (Assume: treated and control groups are each of size $n/2$)
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$, Y_i when T_i is 1 or 0

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i
- Sample selection: $I_i = 1$ if selected, 0 otherwise
- Treatment assignment: $T_i = 1$ if treated group, 0 if control
- (Assume: treated and control groups are each of size $n/2$)
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$, Y_i when T_i is 1 or 0
- Fundamental problem of causal inference. Only one potential outcome is ever observed:

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i
- Sample selection: $I_i = 1$ if selected, 0 otherwise
- Treatment assignment: $T_i = 1$ if treated group, 0 if control
- (Assume: treated and control groups are each of size $n/2$)
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$, Y_i when T_i is 1 or 0
- Fundamental problem of causal inference. Only one potential outcome is ever observed:

$$\text{If } T_i = 0, \quad Y_i(0) = Y_i \quad Y_i(1) = ?$$

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i
- Sample selection: $I_i = 1$ if selected, 0 otherwise
- Treatment assignment: $T_i = 1$ if treated group, 0 if control
- (Assume: treated and control groups are each of size $n/2$)
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$, Y_i when T_i is 1 or 0
- Fundamental problem of causal inference. Only one potential outcome is ever observed:

$$\begin{array}{lll} \text{If } T_i = 0, & Y_i(0) = Y_i & Y_i(1) = ? \\ \text{If } T_i = 1, & Y_i(0) = ? & Y_i(1) = Y_i \end{array}$$

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i
- Sample selection: $I_i = 1$ if selected, 0 otherwise
- Treatment assignment: $T_i = 1$ if treated group, 0 if control
- (Assume: treated and control groups are each of size $n/2$)
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$, Y_i when T_i is 1 or 0
- Fundamental problem of causal inference. Only one potential outcome is ever observed:

$$\begin{array}{lll} \text{If } T_i = 0, & Y_i(0) = Y_i & Y_i(1) = ? \\ \text{If } T_i = 1, & Y_i(0) = ? & Y_i(1) = Y_i \end{array}$$

- (I_i, T_i, Y_i) are random; $Y_i(1)$ and $Y_i(0)$ are fixed.

Notation

- Sample n units from finite population size N (typically $N \gg n$)
- Observed outcome variable: Y_i
- Sample selection: $I_i = 1$ if selected, 0 otherwise
- Treatment assignment: $T_i = 1$ if treated group, 0 if control
- (Assume: treated and control groups are each of size $n/2$)
- Potential outcomes: $Y_i(1)$ and $Y_i(0)$, Y_i when T_i is 1 or 0
- Fundamental problem of causal inference. Only one potential outcome is ever observed:

$$\begin{array}{lll} \text{If } T_i = 0, & Y_i(0) = Y_i & Y_i(1) = ? \\ \text{If } T_i = 1, & Y_i(0) = ? & Y_i(1) = Y_i \end{array}$$

- (I_i, T_i, Y_i) are random; $Y_i(1)$ and $Y_i(0)$ are fixed.
- Quiz: How can Y_i be random when $Y_i(0)$ and $Y_i(1)$ are fixed?

Quantities of Interest

Quantities of Interest

- Treatment Effect (for unit i):

$$\mathsf{TE}_i \quad \equiv \quad Y_i(1) - Y_i(0)$$

Quantities of Interest

- Treatment Effect (for unit i):

$$\mathsf{TE}_i \quad \equiv \quad Y_i(1) - Y_i(0)$$

- Population Average Treatment Effect

$$\mathsf{PATE} \quad \equiv \quad \frac{1}{N}\sum_{i=1}^{N} \mathsf{TE}_i$$

Quantities of Interest

- Treatment Effect (for unit i):

$$\mathrm{TE}_i \equiv Y_i(1) - Y_i(0)$$

- Population Average Treatment Effect

$$\mathrm{PATE} \equiv \frac{1}{N}\sum_{i=1}^{N} \mathrm{TE}_i$$

- Sample Average Treatment Effect

$$\mathrm{SATE} \equiv \frac{1}{n}\sum_{i\in\{I_i=1\}} \mathrm{TE}_i$$

Decomposition of Causal Effect Estimation Error

Decomposition of Causal Effect Estimation Error

- Difference in means estimator

$$D \quad \equiv \quad \bar{Y}_1 - \bar{Y}_0 = \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 1\}} Y_i \right) - \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 0\}} Y_i \right).$$

Decomposition of Causal Effect Estimation Error

- Difference in means estimator

$$D \quad \equiv \quad \bar{Y}_1 - \bar{Y}_0 = \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 1\}} Y_i \right) - \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 0\}} Y_i \right).$$

- Pretreatment confounders: observed X; unobserved U

Decomposition of Causal Effect Estimation Error

- Difference in means estimator

$$D \equiv \bar{Y}_1 - \bar{Y}_0 = \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 1\}} Y_i \right) - \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 0\}} Y_i \right).$$

- Pretreatment confounders: observed X; unobserved U
- Decomposition

$$\Delta \equiv \text{PATE} - D \qquad \text{(Estimation error)}$$

Decomposition of Causal Effect Estimation Error

- Difference in means estimator

$$D \quad \equiv \quad \bar{Y}_1 - \bar{Y}_0 = \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 1\}} Y_i \right) - \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 0\}} Y_i \right).$$

- Pretreatment confounders: observed X; unobserved U
- Decomposition

$$\begin{aligned} \Delta \quad &\equiv \quad \text{PATE} - D \qquad \text{(Estimation error)} \\ &= \quad \Delta_S + \Delta_T \end{aligned}$$

Error due to: Δ_S (sample selection), Δ_T (treatment imbalance),

Decomposition of Causal Effect Estimation Error

- Difference in means estimator

$$D \quad \equiv \quad \bar{Y}_1 - \bar{Y}_0 = \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 1\}} Y_i \right) - \left(\frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 0\}} Y_i \right).$$

- Pretreatment confounders: observed X; unobserved U
- Decomposition

$$\begin{aligned} \Delta \quad &\equiv \quad \text{PATE} - D \qquad \text{(Estimation error)} \\ &= \quad \Delta_S + \Delta_T \\ &= \quad (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U}) \end{aligned}$$

Error due to: Δ_S (sample selection), Δ_T (treatment imbalance), and each due to observed (X_i) and unobserved (U_i) covariates

Decomposing Selection Error

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = \left(\Delta_{S_X} + \Delta_{S_U}\right) + \Delta_T$

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\Delta_S \equiv \text{PATE} - \text{SATE}$$

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$
\begin{aligned}
\Delta_S &\equiv \text{PATE} - \text{SATE} \\
&= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE}
\end{aligned}
$$

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if
  - The sample is a census ($I_i = 1$ for all observations and $n = N$);

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if
  - The sample is a census ($I_i = 1$ for all observations and $n = N$);
  - SATE = NATE (i.e., nothing to correct)

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if
  - The sample is a census ($I_i = 1$ for all observations and $n = N$);
  - SATE = NATE (i.e., nothing to correct)
  - Switch quantity of interest from PATE to SATE (recommended!)

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if
  - The sample is a census ($I_i = 1$ for all observations and $n = N$);
  - SATE = NATE (i.e., nothing to correct)
  - Switch quantity of interest from PATE to SATE (recommended!)
- $\Delta_{S_X} = 0$ when empirical distribution of (observed) X is identical in population and sample: $\widetilde{F}(X \mid I = 0) = \widetilde{F}(X \mid I = 1)$.

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if
    - The sample is a census ($I_i = 1$ for all observations and $n = N$);
    - SATE = NATE (i.e., nothing to correct)
    - Switch quantity of interest from PATE to SATE (recommended!)
- $\Delta_{S_X} = 0$ when empirical distribution of (observed) X is identical in population and sample: $\widetilde{F}(X \mid I = 0) = \widetilde{F}(X \mid I = 1)$.
- $\Delta_{S_U} = 0$ when empirical distribution of (unobserved) U is identical in population and sample: $\widetilde{F}(U \mid I = 0) = \widetilde{F}(U \mid I = 1)$.

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if
  - The sample is a census ($I_i = 1$ for all observations and $n = N$);
  - SATE = NATE (i.e., nothing to correct)
  - Switch quantity of interest from PATE to SATE (recommended!)
- $\Delta_{S_X} = 0$ when empirical distribution of (observed) X is identical in population and sample: $\widetilde{F}(X \mid I = 0) = \widetilde{F}(X \mid I = 1)$.
- $\Delta_{S_U} = 0$ when empirical distribution of (unobserved) U is identical in population and sample: $\widetilde{F}(U \mid I = 0) = \widetilde{F}(U \mid I = 1)$.
- Unverifiable:

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if
  - The sample is a census ($I_i = 1$ for all observations and $n = N$);
  - SATE = NATE (i.e., nothing to correct)
  - Switch quantity of interest from PATE to SATE (recommended!)
- $\Delta_{S_X} = 0$ when empirical distribution of (observed) X is identical in population and sample: $\widetilde{F}(X \mid I = 0) = \widetilde{F}(X \mid I = 1)$.
- $\Delta_{S_U} = 0$ when empirical distribution of (unobserved) U is identical in population and sample: $\widetilde{F}(U \mid I = 0) = \widetilde{F}(U \mid I = 1)$.
- Unverifiable: X unobserved out of sample;

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if
  - The sample is a census ($I_i = 1$ for all observations and $n = N$);
  - SATE = NATE (i.e., nothing to correct)
  - Switch quantity of interest from PATE to SATE (recommended!)
- $\Delta_{S_X} = 0$ when empirical distribution of (observed) X is identical in population and sample: $\widetilde{F}(X \mid I = 0) = \widetilde{F}(X \mid I = 1)$.
- $\Delta_{S_U} = 0$ when empirical distribution of (unobserved) U is identical in population and sample: $\widetilde{F}(U \mid I = 0) = \widetilde{F}(U \mid I = 1)$.
- Unverifiable: X unobserved out of sample; U unobserved

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned} \Delta_S &\equiv \text{PATE} - \text{SATE} \\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE} \end{aligned}$$

- Δ_S vanishes if
  - The sample is a census ($I_i = 1$ for all observations and $n = N$);
  - SATE = NATE (i.e., nothing to correct)
  - Switch quantity of interest from PATE to SATE (recommended!)
- $\Delta_{S_X} = 0$ when empirical distribution of (observed) X is identical in population and sample: $\widetilde{F}(X \mid I = 0) = \widetilde{F}(X \mid I = 1)$.
- $\Delta_{S_U} = 0$ when empirical distribution of (unobserved) U is identical in population and sample: $\widetilde{F}(U \mid I = 0) = \widetilde{F}(U \mid I = 1)$.
- Unverifiable: X unobserved out of sample; U unobserved
- Δ_{S_X}: vanishes if weighting on X (and examples exist in sample)

Decomposing Selection Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + \Delta_T$

- Definition

$$\begin{aligned}\Delta_S &\equiv \text{PATE} - \text{SATE}\\ &= \frac{N-n}{N}(\text{NATE} - \text{SATE}), \quad \text{NATE: nonsample ATE}\end{aligned}$$

- Δ_S vanishes if
  - The sample is a census ($I_i = 1$ for all observations and $n = N$);
  - SATE = NATE (i.e., nothing to correct)
  - Switch quantity of interest from PATE to SATE (recommended!)
- $\Delta_{S_X} = 0$ when empirical distribution of (observed) X is identical in population and sample: $\widetilde{F}(X \mid I = 0) = \widetilde{F}(X \mid I = 1)$.
- $\Delta_{S_U} = 0$ when empirical distribution of (unobserved) U is identical in population and sample: $\widetilde{F}(U \mid I = 0) = \widetilde{F}(U \mid I = 1)$.
- Unverifiable: X unobserved out of sample; U unobserved
- Δ_{S_X}: vanishes if weighting on X (and examples exist in sample)

Decomposing Treatment Imbalance

Decomposing Treatment Imbalance

$\Delta = \Delta_S + \Delta_T = \Delta_S + (\Delta_{T_X} + \Delta_{T_U})$

Decomposing Treatment Imbalance

$\Delta = \Delta_S + \Delta_T = \Delta_S + (\Delta_{T_X} + \Delta_{T_U})$

- $\Delta_{T_X} = 0$: when X balanced between treateds and controls:

$$\widetilde{F}(X \mid T = 1, I = 1) \quad = \quad \widetilde{F}(X \mid T = 0, I = 1).$$

Decomposing Treatment Imbalance

$\Delta = \Delta_S + \Delta_T = \Delta_S + (\Delta_{T_X} + \Delta_{T_U})$

- $\Delta_{T_X} = 0$: when X balanced between treateds and controls:

$$\widetilde{F}(X \mid T = 1, I = 1) \quad = \quad \widetilde{F}(X \mid T = 0, I = 1).$$

Verifiable;

Decomposing Treatment Imbalance

$\Delta = \Delta_S + \Delta_T = \Delta_S + (\Delta_{T_X} + \Delta_{T_U})$

- $\Delta_{T_X} = 0$: when X balanced between treateds and controls:

$$\widetilde{F}(X \mid T = 1, I = 1) \quad = \quad \widetilde{F}(X \mid T = 0, I = 1).$$

Verifiable; generated ex ante by blocking or ex post via matching or modeling

Decomposing Treatment Imbalance

$\Delta = \Delta_S + \Delta_T = \Delta_S + (\Delta_{T_X} + \Delta_{T_U})$

- $\Delta_{T_X} = 0$: when X balanced between treateds and controls:

$$\widetilde{F}(X \mid T = 1, I = 1) \quad = \quad \widetilde{F}(X \mid T = 0, I = 1).$$

  Verifiable; generated ex ante by blocking or ex post via matching or modeling

- $\Delta_{T_U} = 0$: when U balanced between treateds and controls:

$$\widetilde{F}(U \mid T = 1, I = 1) \quad = \quad \widetilde{F}(U \mid T = 0, I = 1).$$

Decomposing Treatment Imbalance

$\Delta = \Delta_S + \Delta_T = \Delta_S + (\Delta_{T_X} + \Delta_{T_U})$

- $\Delta_{T_X} = 0$: when X balanced between treateds and controls:

$$\widetilde{F}(X \mid T = 1, I = 1) \quad = \quad \widetilde{F}(X \mid T = 0, I = 1).$$

  Verifiable; generated ex ante by blocking or ex post via matching or modeling

- $\Delta_{T_U} = 0$: when U balanced between treateds and controls:

$$\widetilde{F}(U \mid T = 1, I = 1) \quad = \quad \widetilde{F}(U \mid T = 0, I = 1).$$

  Unverifiable;

Decomposing Treatment Imbalance

$\Delta = \Delta_S + \Delta_T = \Delta_S + (\Delta_{T_X} + \Delta_{T_U})$

- $\Delta_{T_X} = 0$: when X balanced between treateds and controls:

  $$\widetilde{F}(X \mid T = 1, I = 1) \quad = \quad \widetilde{F}(X \mid T = 0, I = 1).$$

  Verifiable; generated ex ante by blocking or ex post via matching or modeling
- $\Delta_{T_U} = 0$: when U balanced between treateds and controls:

  $$\widetilde{F}(U \mid T = 1, I = 1) \quad = \quad \widetilde{F}(U \mid T = 0, I = 1).$$

  Unverifiable; Achieved only by assumption or, on average, by random treatment assignment

Alternative Quantities of Interest: For Matching

Alternative Quantities of Interest: For Matching

- Population average treatment effect on the treated

$$\mathsf{PATT} \equiv \frac{1}{N^*} \sum_{i \in \{T_i = 1\}} \mathsf{TE}_i$$

($N^* = \sum_{i=1}^{N} T_i$: number of treated units in population)

Alternative Quantities of Interest: For Matching

- Population average treatment effect on the treated

$$\text{PATT} \equiv \frac{1}{N^*} \sum_{i \in \{T_i = 1\}} \text{TE}_i$$

($N^* = \sum_{i=1}^{N} T_i$: number of treated units in population)

- Sample average treatment effect on the treated

$$\text{SATT} \equiv \frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 1\}} \text{TE}_i$$

Alternative Quantities of Interest: For Matching

- Population average treatment effect on the treated

$$\text{PATT} \equiv \frac{1}{N^*} \sum_{i \in \{T_i = 1\}} \text{TE}_i$$

($N^* = \sum_{i=1}^{N} T_i$: number of treated units in population)

- Sample average treatment effect on the treated

$$\text{SATT} \equiv \frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 1\}} \text{TE}_i$$

- Analogous estimation error decomposition holds:

$$\Delta' = \text{PATT} - D = (\Delta'_{S_X} + \Delta'_{S_U}) + (\Delta'_{T_X} + \Delta'_{T_U})$$

Alternative Quantities of Interest: For Matching

- Population average treatment effect on the treated

$$\text{PATT} \equiv \frac{1}{N^*} \sum_{i \in \{T_i = 1\}} \text{TE}_i$$

($N^* = \sum_{i=1}^{N} T_i$: number of treated units in population)

- Sample average treatment effect on the treated

$$\text{SATT} \equiv \frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 1\}} \text{TE}_i$$

- Analogous estimation error decomposition holds:

$$\Delta' = \text{PATT} - D = (\Delta'_{S_X} + \Delta'_{S_U}) + (\Delta'_{T_X} + \Delta'_{T_U})$$

- Quiz: Why PATT and SATT rather than PATE and SATE for matching?

Alternative Quantities of Interest: For Matching

- Population average treatment effect on the treated

$$\text{PATT} \equiv \frac{1}{N^*} \sum_{i \in \{T_i = 1\}} \text{TE}_i$$

($N^* = \sum_{i=1}^{N} T_i$: number of treated units in population)

- Sample average treatment effect on the treated

$$\text{SATT} \equiv \frac{1}{n/2} \sum_{i \in \{I_i = 1, T_i = 1\}} \text{TE}_i$$

- Analogous estimation error decomposition holds:

$$\Delta' = \text{PATT} - D = (\Delta'_{S_X} + \Delta'_{S_U}) + (\Delta'_{T_X} + \Delta'_{T_U})$$

- Quiz: Why PATT and SATT rather than PATE and SATE for matching?
- Quiz: How do they differ in randomized experiments?

Effects of Design Components on Estimation Error

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\stackrel{\text{avg}}{=} 0$	$\stackrel{\text{avg}}{=} 0$		

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\overset{\text{avg}}{=} 0$		

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\stackrel{\text{avg}}{=} 0$	$\stackrel{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\stackrel{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\overset{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		
Weighting for nonrandom sampling	$= 0$	$= ?$		

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\overset{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		
Weighting for nonrandom sampling	$= 0$	$= ?$		
Large sample size	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\stackrel{\text{avg}}{=} 0$	$\stackrel{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\stackrel{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		
Weighting for nonrandom sampling	$= 0$	$= ?$		
Large sample size	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$
Random treatment assignment			$\stackrel{\text{avg}}{=} 0$	$\stackrel{\text{avg}}{=} 0$

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\overset{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		
Weighting for nonrandom sampling	$= 0$	$= ?$		
Large sample size	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$
Random treatment assignment			$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$
Complete blocking			$= 0$	$= ?$

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\overset{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		
Weighting for nonrandom sampling	$= 0$	$= ?$		
Large sample size	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$
Random treatment assignment			$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$
Complete blocking			$= 0$	$= ?$
Exact matching			$= 0$	$= ?$

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\overset{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		
Weighting for nonrandom sampling	$= 0$	$= ?$		
Large sample size	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$
Random treatment assignment			$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$
Complete blocking			$= 0$	$= ?$
Exact matching			$= 0$	$= ?$
Assumption				

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\overset{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		
Weighting for nonrandom sampling	$= 0$	$= ?$		
Large sample size	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$
Random treatment assignment			$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$
Complete blocking			$= 0$	$= ?$
Exact matching			$= 0$	$= ?$
Assumption				
No selection bias	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Design Choice				
Random sampling	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\overset{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		
Weighting for nonrandom sampling	$= 0$	$= ?$		
Large sample size	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$
Random treatment assignment			$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$
Complete blocking			$= 0$	$= ?$
Exact matching			$= 0$	$= ?$
Assumption				
No selection bias	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$		
Ignorability				$\overset{\text{avg}}{=} 0$

Effects of Design Components on Estimation Error

$\Delta = \Delta_S + \Delta_T = (\Delta_{S_X} + \Delta_{S_U}) + (\Delta_{T_X} + \Delta_{T_U})$

Design Choice	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Random sampling	$\stackrel{\text{avg}}{=} 0$	$\stackrel{\text{avg}}{=} 0$		
Complete stratified random sampling	$= 0$	$\stackrel{\text{avg}}{=} 0$		
Focus on SATE rather than PATE	$= 0$	$= 0$		
Weighting for nonrandom sampling	$= 0$	$= ?$		
Large sample size	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$	$\longrightarrow ?$
Random treatment assignment			$\stackrel{\text{avg}}{=} 0$	$\stackrel{\text{avg}}{=} 0$
Complete blocking			$= 0$	$= ?$
Exact matching			$= 0$	$= ?$
Assumption				
No selection bias	$\stackrel{\text{avg}}{=} 0$	$\stackrel{\text{avg}}{=} 0$		
Ignorability				$\stackrel{\text{avg}}{=} 0$
No omitted variables				$= 0$

Comparing Blocking (i.e., before) and Matching (i.e., after)

Comparing Blocking (i.e., before) and Matching (i.e., after)

- Adding blocking (on pretreatment vars related to outcome) to random assignment: as or more efficient, and never biased

Comparing Blocking (i.e., before) and Matching (i.e., after)

- Adding blocking (on pretreatment vars related to outcome) to random assignment: as or more efficient, and never biased
- Blocking: like regression adjustment, where functional form and the parameter values are known

Comparing Blocking (i.e., before) and Matching (i.e., after)

- Adding blocking (on pretreatment vars related to outcome) to random assignment: as or more efficient, and never biased
- Blocking: like regression adjustment, where functional form and the parameter values are known
- Matching is like blocking, except:

Comparing Blocking (i.e., before) and Matching (i.e., after)

- Adding blocking (on pretreatment vars related to outcome) to random assignment: as or more efficient, and never biased
- Blocking: like regression adjustment, where functional form and the parameter values are known
- Matching is like blocking, except:
    - to avoid selection error: change QOI from PATE to PATT/SATT

Comparing Blocking (i.e., before) and Matching (i.e., after)

- Adding blocking (on pretreatment vars related to outcome) to random assignment: as or more efficient, and never biased
- Blocking: like regression adjustment, where functional form and the parameter values are known
- Matching is like blocking, except:
  - to avoid selection error: change QOI from PATE to PATT/SATT
  - random treatment assignment following matching: impossible

Comparing Blocking (i.e., before) and Matching (i.e., after)

- Adding blocking (on pretreatment vars related to outcome) to random assignment: as or more efficient, and never biased
- Blocking: like regression adjustment, where functional form and the parameter values are known
- Matching is like blocking, except:
    - to avoid selection error: change QOI from PATE to PATT/SATT
    - random treatment assignment following matching: impossible
    - Exact matching, unlike blocking: dependent on good matches in already-collected data

Comparing Blocking (i.e., before) and Matching (i.e., after)

- Adding blocking (on pretreatment vars related to outcome) to random assignment: as or more efficient, and never biased
- Blocking: like regression adjustment, where functional form and the parameter values are known
- Matching is like blocking, except:
  - to avoid selection error: change QOI from PATE to PATT/SATT
  - random treatment assignment following matching: impossible
  - Exact matching, unlike blocking: dependent on good matches in already-collected data
  - Worst case scenario: matching on wrong vars (like regression adjustment) can increase bias

Comparing Blocking (i.e., before) and Matching (i.e., after)

- Adding blocking (on pretreatment vars related to outcome) to random assignment: as or more efficient, and never biased
- Blocking: like regression adjustment, where functional form and the parameter values are known
- Matching is like blocking, except:
  - to avoid selection error: change QOI from PATE to PATT/SATT
  - random treatment assignment following matching: impossible
  - Exact matching, unlike blocking: dependent on good matches in already-collected data
  - Worst case scenario: matching on wrong vars (like regression adjustment) can increase bias
- Adding matching to a parametric model: reduces model dependence and bias, and sometimes variance too

Comparing Blocking (i.e., before) and Matching (i.e., after)

- Adding blocking (on pretreatment vars related to outcome) to random assignment: as or more efficient, and never biased
- Blocking: like regression adjustment, where functional form and the parameter values are known
- Matching is like blocking, except:
  - to avoid selection error: change QOI from PATE to PATT/SATT
  - random treatment assignment following matching: impossible
  - Exact matching, unlike blocking: dependent on good matches in already-collected data
  - Worst case scenario: matching on wrong vars (like regression adjustment) can increase bias
- Adding matching to a parametric model: reduces model dependence and bias, and sometimes variance too
- Quiz: Which is preferable: Matching or Blocking?

Components of Causal Estimation Error

Research Designs

Issues in Ideal Designs

The Benefits of Major Research Designs: Overview

	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Ideal experiment	$\rightarrow 0$	$\rightarrow 0$	$= 0$	$\rightarrow 0$
Randomized clinicial trials (Limited or no blocking)	$\neq 0$	$\neq 0$	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$
Randomized clinicial trials (Full blocking)	$\neq 0$	$\neq 0$	$= 0$	$\overset{\text{avg}}{=} 0$
Social Science Field Experiment (Limited or no blocking)	$\neq 0$	$\neq 0$	$\rightarrow 0$	$\rightarrow 0$
Survey Experiment (Limited or no blocking)	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$
Observational Study (Representative data set, Well-matched)	≈ 0	≈ 0	≈ 0	$\neq 0$
Observational Study (Unrepresentative but partially, correctable data, well-matched)	≈ 0	$\neq 0$	≈ 0	$\neq 0$
Observational Study (Unrepresentative data set, Well-matched)	$\neq 0$	$\neq 0$	≈ 0	$\neq 0$

The Benefits of Major Research Designs: Overview

	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Ideal experiment	$\rightarrow 0$	$\rightarrow 0$	$= 0$	$\rightarrow 0$
Randomized clinicial trials (Limited or no blocking)	$\neq 0$	$\neq 0$	$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$
Randomized clinicial trials (Full blocking)	$\neq 0$	$\neq 0$	$= 0$	$\overset{\text{avg}}{=} 0$
Social Science Field Experiment (Limited or no blocking)	$\neq 0$	$\neq 0$	$\rightarrow 0$	$\rightarrow 0$
Survey Experiment (Limited or no blocking)	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$
Observational Study (Representative data set, Well-matched)	≈ 0	≈ 0	≈ 0	$\neq 0$
Observational Study (Unrepresentative but partially, correctable data, well-matched)	≈ 0	$\neq 0$	≈ 0	$\neq 0$
Observational Study (Unrepresentative data set, Well-matched)	$\neq 0$	$\neq 0$	≈ 0	$\neq 0$

- $\rightarrow 0$: $E(Q) = 0$ & $\lim_{n \to \infty} \text{Var}(Q) = 0$

The Benefits of Major Research Designs: Overview

	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Ideal experiment	$\longrightarrow 0$	$\longrightarrow 0$	$= 0$	$\longrightarrow 0$
Randomized clinicial trials (Limited or no blocking)	$\neq 0$	$\neq 0$	$\stackrel{\text{avg}}{=} 0$	$\stackrel{\text{avg}}{=} 0$
Randomized clinicial trials (Full blocking)	$\neq 0$	$\neq 0$	$= 0$	$\stackrel{\text{avg}}{=} 0$
Social Science Field Experiment (Limited or no blocking)	$\neq 0$	$\neq 0$	$\longrightarrow 0$	$\longrightarrow 0$
Survey Experiment (Limited or no blocking)	$\longrightarrow 0$	$\longrightarrow 0$	$\longrightarrow 0$	$\longrightarrow 0$
Observational Study (Representative data set, Well-matched)	≈ 0	≈ 0	≈ 0	$\neq 0$
Observational Study (Unrepresentative but partially, correctable data, well-matched)	≈ 0	$\neq 0$	≈ 0	$\neq 0$
Observational Study (Unrepresentative data set, Well-matched)	$\neq 0$	$\neq 0$	≈ 0	$\neq 0$

- $\longrightarrow 0$: $E(Q) = 0$ & $\lim_{n \to \infty} \text{Var}(Q) = 0$
- $\stackrel{\text{avg}}{=} 0$: $E(Q) = 0$

The Ideal Experiment (according to the paper)

The Ideal Experiment (according to the paper)

- Random selection from well-defined population

The Ideal Experiment (according to the paper)

- Random selection from well-defined population
- large n

The Ideal Experiment (according to the paper)

- Random selection from well-defined population
- large n
- blocking on all known confounders

The Ideal Experiment (according to the paper)

- Random selection from well-defined population
- large n
- blocking on all known confounders
- random treatment assignment within blocks

The Ideal Experiment (according to the paper)

- Random selection from well-defined population
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_X}) = 0$

The Ideal Experiment (according to the paper)

- Random selection from well-defined population
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_U}) = 0$

The Ideal Experiment (according to the paper)

- Random selection from well-defined population
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_U}) = 0$
- $\Delta_{T_X} = 0$

The Ideal Experiment (according to the paper)

- Random selection from well-defined population
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_U}) = 0$
- $\Delta_{T_X} = 0$
- $E(\Delta_{T_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{T_U}) = 0$

The Ideal Experiment (according to the paper)

- Random selection from well-defined population
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_U}) = 0$
- $\Delta_{T_X} = 0$
- $E(\Delta_{T_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{T_U}) = 0$
- Quiz: Is there an even more ideal experiment?

The Ideal Experiment (according to the paper)

- Random selection from well-defined population
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\rightarrow\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0, \quad \lim_{n\rightarrow\infty} V(\Delta_{S_U}) = 0$
- $\Delta_{T_X} = 0$
- $E(\Delta_{T_U}) = 0, \quad \lim_{n\rightarrow\infty} V(\Delta_{T_U}) = 0$
- Quiz: Is there an even more ideal experiment?
- Hint: How can we make $\Delta_{S_X} = 0$?

An Even More Ideal Experiment (not in the paper)

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata
- (if strata sample $\propto$ population size, no weights needed)

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata
- (if strata sample $\propto$ population size, no weights needed)
- large n

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata
- (if strata sample $\propto$ population size, no weights needed)
- large n
- blocking on all known confounders

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata
- (if strata sample $\propto$ population size, no weights needed)
- large n
- blocking on all known confounders
- random treatment assignment within blocks

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata
- (if strata sample $\propto$ population size, no weights needed)
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $\Delta_{S_X} = 0$

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata
- (if strata sample $\propto$ population size, no weights needed)
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $\Delta_{S_X} = 0$
- $E(\Delta_{S_U}) = 0, \lim_{n \longrightarrow \infty} V(\Delta_{S_U}) = 0$

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata
- (if strata sample $\propto$ population size, no weights needed)
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $\Delta_{S_X} = 0$
- $E(\Delta_{S_U}) = 0$, $\lim_{n\to\infty} V(\Delta_{S_U}) = 0$
- $\Delta_{T_X} = 0$

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata
- (if strata sample $\propto$ population size, no weights needed)
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $\Delta_{S_X} = 0$
- $E(\Delta_{S_U}) = 0$, $\lim_{n \to \infty} V(\Delta_{S_U}) = 0$
- $\Delta_{T_X} = 0$
- $E(\Delta_{T_U}) = 0$, $\lim_{n \to \infty} V(\Delta_{T_U}) = 0$

An Even More Ideal Experiment (not in the paper)

- Begin with a well-defined population
- New feature: Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata
- (if strata sample $\propto$ population size, no weights needed)
- large n
- blocking on all known confounders
- random treatment assignment within blocks
- $\Delta_{S_X} = 0$
- $E(\Delta_{S_U}) = 0$, $\lim_{n\to\infty} V(\Delta_{S_U}) = 0$
- $\Delta_{T_X} = 0$
- $E(\Delta_{T_U}) = 0$, $\lim_{n\to\infty} V(\Delta_{T_U}) = 0$
- Wait, why wasn't this in the paper?

Randomized Clinical Trials (Little or no Blocking)

Randomized Clinical Trials (Little or no Blocking)

- nonrandom selection

Randomized Clinical Trials (Little or no Blocking)

- nonrandom selection
- small n

Randomized Clinical Trials (Little or no Blocking)

- nonrandom selection
- small n
- little or no blocking

Randomized Clinical Trials (Little or no Blocking)

- nonrandom selection
- small n
- little or no blocking
- random treatment assignment

Randomized Clinical Trials (Little or no Blocking)

- nonrandom selection
- small n
- little or no blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$

Randomized Clinical Trials (Little or no Blocking)

- nonrandom selection
- small n
- little or no blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$
- $\Delta_{S_U} \neq 0$

Randomized Clinical Trials (Little or no Blocking)

- nonrandom selection
- small n
- little or no blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$
- $\Delta_{S_U} \neq 0$
- $E(\Delta_{T_X}) = 0$

Randomized Clinical Trials (Little or no Blocking)

- nonrandom selection
- small n
- little or no blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$
- $\Delta_{S_U} \neq 0$
- $E(\Delta_{T_X}) = 0$
- $E(\Delta_{T_U}) = 0$

Randomized Clinical Trials (Full Blocking)

Randomized Clinical Trials (Full Blocking)

- nonrandom selection

Randomized Clinical Trials (Full Blocking)

- nonrandom selection
- small n

Randomized Clinical Trials (Full Blocking)

- nonrandom selection
- small n
- Full blocking

Randomized Clinical Trials (Full Blocking)

- nonrandom selection
- small n
- Full blocking
- random treatment assignment

Randomized Clinical Trials (Full Blocking)

- nonrandom selection
- small n
- Full blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$

Randomized Clinical Trials (Full Blocking)

- nonrandom selection
- small n
- Full blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$
- $\Delta_{S_U} \neq 0$

Randomized Clinical Trials (Full Blocking)

- nonrandom selection
- small n
- Full blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$
- $\Delta_{S_U} \neq 0$
- $\Delta_{T_X} = 0$

Randomized Clinical Trials (Full Blocking)

- nonrandom selection
- small n
- Full blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$
- $\Delta_{S_U} \neq 0$
- $\Delta_{T_X} = 0$
- $E(\Delta_{T_U}) = 0$

Social Science Field Experiment

Social Science Field Experiment

- nonrandom selection

Social Science Field Experiment

- nonrandom selection
- large n

Social Science Field Experiment

- nonrandom selection
- large n
- limited or no blocking

Social Science Field Experiment

- nonrandom selection
- large n
- limited or no blocking
- random treatment assignment

Social Science Field Experiment

- nonrandom selection
- large n
- limited or no blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$ or change PATE to SATE and $\Delta_{S_X} = 0$

Social Science Field Experiment

- nonrandom selection
- large n
- limited or no blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$ or change PATE to SATE and $\Delta_{S_X} = 0$
- $\Delta_{S_U} \neq 0$ or change PATE to SATE and $\Delta_{S_U} = 0$

Social Science Field Experiment

- nonrandom selection
- large n
- limited or no blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$ or change PATE to SATE and $\Delta_{S_X} = 0$
- $\Delta_{S_U} \neq 0$ or change PATE to SATE and $\Delta_{S_U} = 0$
- $E(\Delta_{T_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{T_X}) = 0$

Social Science Field Experiment

- nonrandom selection
- large n
- limited or no blocking
- random treatment assignment
- $\Delta_{S_X} \neq 0$ or change PATE to SATE and $\Delta_{S_X} = 0$
- $\Delta_{S_U} \neq 0$ or change PATE to SATE and $\Delta_{S_U} = 0$
- $E(\Delta_{T_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{T_X}) = 0$
- $E(\Delta_{T_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{T_U}) = 0$

Survey Experiment

Survey Experiment

- random selection

Survey Experiment

- random selection
- large n

Survey Experiment

- random selection
- large n
- limited or no blocking

Survey Experiment

- random selection
- large n
- limited or no blocking
- random treatment assignment

Survey Experiment

- random selection
- large n
- limited or no blocking
- random treatment assignment
- (only treatments: question wording changes)

Survey Experiment

- random selection
- large n
- limited or no blocking
- random treatment assignment
- (only treatments: question wording changes)
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_X}) = 0$

Survey Experiment

- random selection
- large n
- limited or no blocking
- random treatment assignment
- (only treatments: question wording changes)
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_U}) = 0$

Survey Experiment

- random selection
- large n
- limited or no blocking
- random treatment assignment
- (only treatments: question wording changes)
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_U}) = 0$
- $E(\Delta_{T_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{T_X}) = 0$

Survey Experiment

- random selection
- large n
- limited or no blocking
- random treatment assignment
- (only treatments: question wording changes)
- $E(\Delta_{S_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{S_U}) = 0$
- $E(\Delta_{T_X}) = 0, \quad \lim_{n\to\infty} V(\Delta_{T_X}) = 0$
- $E(\Delta_{T_U}) = 0, \quad \lim_{n\to\infty} V(\Delta_{T_U}) = 0$

Observational Study, well-matched

Observational Study, well-matched

- no stratification, nonrandom selection

Observational Study, well-matched

- no stratification, nonrandom selection
- large n

Observational Study, well-matched

- no stratification, nonrandom selection
- large n
- no blocking, nonrandom treatment assignment

Observational Study, well-matched

- no stratification, nonrandom selection
- large n
- no blocking, nonrandom treatment assignment
- $\Delta_{S_X} \approx 0$ if representative, corrected by weighting, or for estimating SATE; or $\neq 0$ otherwise

Observational Study, well-matched

- no stratification, nonrandom selection
- large n
- no blocking, nonrandom treatment assignment
- $\Delta_{S_X} \approx 0$ if representative, corrected by weighting, or for estimating SATE; or $\neq 0$ otherwise
- $\Delta_{S_U} \neq 0$

Observational Study, well-matched

- no stratification, nonrandom selection
- large n
- no blocking, nonrandom treatment assignment
- $\Delta_{S_X} \approx 0$ if representative, corrected by weighting, or for estimating SATE; or $\neq 0$ otherwise
- $\Delta_{S_U} \neq 0$
- $\Delta_{T_X} \approx 0$ (due to matching well)

Observational Study, well-matched

- no stratification, nonrandom selection
- large n
- no blocking, nonrandom treatment assignment
- $\Delta_{S_X} \approx 0$ if representative, corrected by weighting, or for estimating SATE; or $\neq 0$ otherwise
- $\Delta_{S_U} \neq 0$
- $\Delta_{T_X} \approx 0$ (due to matching well)
- $\Delta_{T_U} \neq 0$ except by assumption

Components of Causal Estimation Error

Research Designs

Issues in Ideal Designs

What is the Best Design?

What is the Best Design?

- Ideal design: rarely feasible

What is the Best Design?

- Ideal design: rarely feasible
- Effort in experimental studies: random assignment

What is the Best Design?

- Ideal design: rarely feasible
- Effort in experimental studies: random assignment
- Effort in observational studies: knowing, measuring, and adjusting for X (via matching or modeling)

What is the Best Design?

- Ideal design: rarely feasible
- Effort in experimental studies: random assignment
- Effort in observational studies: knowing, measuring, and adjusting for X (via matching or modeling)
- Achilles heal of experiments: Δ_S, small n

What is the Best Design?

- Ideal design: rarely feasible
- Effort in experimental studies: random assignment
- Effort in observational studies: knowing, measuring, and adjusting for X (via matching or modeling)
- Achilles heal of experiments: Δ_S, small n
- Achilles heal of observational studies: Δ_T

What is the Best Design?

- Ideal design: rarely feasible
- Effort in experimental studies: random assignment
- Effort in observational studies: knowing, measuring, and adjusting for X (via matching or modeling)
- Achilles heal of experiments: Δ_S, small n
- Achilles heal of observational studies: Δ_T
- Each design: accommodates best to its applications

What is the Best Design?

- Ideal design: rarely feasible
- Effort in experimental studies: random assignment
- Effort in observational studies: knowing, measuring, and adjusting for X (via matching or modeling)
- Achilles heal of experiments: Δ_S, small n
- Achilles heal of observational studies: Δ_T
- Each design: accommodates best to its applications
- Quiz: Astronomers never randomize; is astronomy a science?

Fallacies in Experimental Research

Fallacies in Experimental Research

- Failure to block on all available confounders

Fallacies in Experimental Research

- Failure to block on all available confounders
  - incorrectly seen as requiring fewer assumptions (about what to block on)

Fallacies in Experimental Research

- Failure to block on all available confounders
  - incorrectly seen as requiring fewer assumptions (about what to block on)
  - In fact, blocking helps (except in strange situations)

Fallacies in Experimental Research

- Failure to block on all available confounders
  - incorrectly seen as requiring fewer assumptions (about what to block on)
  - In fact, blocking helps (except in strange situations)
  - Blocking on relevant covariates is better, so choose carefully.

Fallacies in Experimental Research

- Failure to block on all available confounders
    - incorrectly seen as requiring fewer assumptions (about what to block on)
    - In fact, blocking helps (except in strange situations)
    - Blocking on relevant covariates is better, so choose carefully.
    - "Block what you can and randomize what you cannot"

Fallacies in Experimental Research

- Failure to block on all available confounders
  - incorrectly seen as requiring fewer assumptions (about what to block on)
  - In fact, blocking helps (except in strange situations)
  - Blocking on relevant covariates is better, so choose carefully.
  - "Block what you can and randomize what you cannot"
- t-test to check balance after random treatment assignment

Fallacies in Experimental Research

- Failure to block on all available confounders
  - incorrectly seen as requiring fewer assumptions (about what to block on)
  - In fact, blocking helps (except in strange situations)
  - Blocking on relevant covariates is better, so choose carefully.
  - "Block what you can and randomize what you cannot"
- t-test to check balance after random treatment assignment
  - blocking vars: balance exactly after treatment assignment; if you're checking, you missed an opportunity to increase efficiency

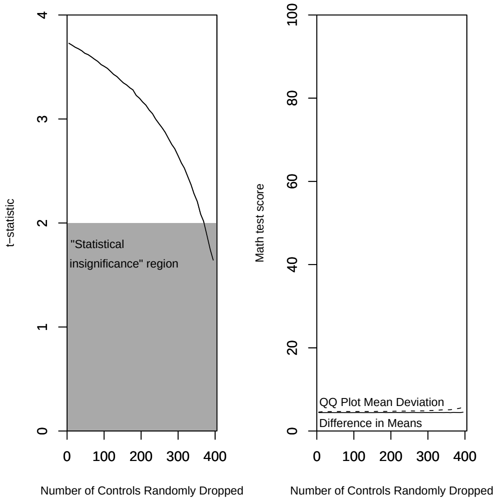
Fallacies in Experimental Research

- Failure to block on all available confounders
  - incorrectly seen as requiring fewer assumptions (about what to block on)
  - In fact, blocking helps (except in strange situations)
  - Blocking on relevant covariates is better, so choose carefully.
  - "Block what you can and randomize what you cannot"
- t-test to check balance after random treatment assignment
  - blocking vars: balance exactly after treatment assignment; if you're checking, you missed an opportunity to increase efficiency
  - if vars become available after treatment assignment: t-test checks if randomization was done appropriately

Fallacies in Experimental Research

- Failure to block on all available confounders
  - incorrectly seen as requiring fewer assumptions (about what to block on)
  - In fact, blocking helps (except in strange situations)
  - Blocking on relevant covariates is better, so choose carefully.
  - "Block what you can and randomize what you cannot"
- t-test to check balance after random treatment assignment
  - blocking vars: balance exactly after treatment assignment; if you're checking, you missed an opportunity to increase efficiency
  - if vars become available after treatment assignment: t-test checks if randomization was done appropriately
  - randomization balances on average: any one random assignment is not balanced exactly (which is why its better to block)

The Balance Test Fallacy in Matching Research



Quiz: randomly dropping observations reduces imbalance??

The Balance Test Fallacy: Explanation

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$
  - Bias in coefficient on T from regressing Y on T (without X): $E(\hat{\beta} - \beta \mid T, X) = G\gamma$ (where G are coefficients from a regression X on a constant and T)

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$
  - Bias in coefficient on T from regressing Y on T (without X): $E(\hat{\beta} - \beta \mid T, X) = G\gamma$ (where G are coefficients from a regression X on a constant and T)
  - Imbalance: G, Importance: γ

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$
  - Bias in coefficient on T from regressing Y on T (without X): $E(\hat{\beta} - \beta \mid T, X) = G\gamma$ (where G are coefficients from a regression X on a constant and T)
  - Imbalance: G, Importance: γ
  - If $G = 0$, bias=0

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$
  - Bias in coefficient on T from regressing Y on T (without X): $E(\hat{\beta} - \beta \mid T, X) = G\gamma$ (where G are coefficients from a regression X on a constant and T)
  - Imbalance: G, Importance: γ
  - If $G = 0$, bias=0
  - If $G \neq 0$, bias can be any size (due to γ)

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$
  - Bias in coefficient on T from regressing Y on T (without X): $E(\hat{\beta} - \beta \mid T, X) = G\gamma$ (where G are coefficients from a regression X on a constant and T)
  - Imbalance: G, Importance: γ
  - If $G = 0$, bias=0
  - If $G \neq 0$, bias can be any size (due to γ)
  - To reduce bias: reduce G without limit

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$
  - Bias in coefficient on T from regressing Y on T (without X): $E(\hat{\beta} - \beta \mid T, X) = G\gamma$ (where G are coefficients from a regression X on a constant and T)
  - Imbalance: G, Importance: γ
  - If $G = 0$, bias=0
  - If $G \neq 0$, bias can be any size (due to γ)
  - To reduce bias: reduce G without limit
- No threshold level is safe

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$
  - Bias in coefficient on T from regressing Y on T (without X): $E(\hat{\beta} - \beta \mid T, X) = G\gamma$ (where G are coefficients from a regression X on a constant and T)
  - Imbalance: G, Importance: γ
  - If $G = 0$, bias=0
  - If $G \neq 0$, bias can be any size (due to γ)
  - To reduce bias: reduce G without limit
- No threshold level is safe
- But prune too much, variance increases

The Balance Test Fallacy: Explanation

- Hypo tests: balance *and* power; only want balance
- Balance is observed: No need for superpopulation or inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$
  - Bias in coefficient on T from regressing Y on T (without X): $E(\hat{\beta} - \beta \mid T, X) = G\gamma$ (where G are coefficients from a regression X on a constant and T)
  - Imbalance: G, Importance: γ
  - If $G = 0$, bias=0
  - If $G \neq 0$, bias can be any size (due to γ)
  - To reduce bias: reduce G without limit
- No threshold level is safe
- But prune too much, variance increases
- Quiz: Should we match on vars that do not influence Y?