Advanced Quantitative Research Methodology, Lecture Notes: Research Designs for Causal Inference[1]

Gary King

GaryKing.org

April 10, 2016

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 Vol. 171, Part 2 (2008): Pp. 1-22 `http://gking.harvard.edu/files/abs/matchse-abs.shtml`

Notation

Notation

- Sample of n units from a finite population of N units (typically $N >> n$)

Notation

- Sample of n units from a finite population of N units (typically $N >> n$)
- Sample selection: I_i is 1 for units selected, 0 otherwise

Notation

- Sample of n units from a finite population of N units (typically $N >> n$)
- Sample selection: I_i is 1 for units selected, 0 otherwise
- Treatment assignment: T_i is 1 for treated group, 0 control group

Notation

- Sample of n units from a finite population of N units (typically $N >> n$)
- Sample selection: I_i is 1 for units selected, 0 otherwise
- Treatment assignment: T_i is 1 for treated group, 0 control group
- (Assume: treated and control groups are each of size $n/2$)

Notation

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

Notation

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

Notation

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

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

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

Notation

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

  If $T_i = 0$, $\quad Y_i(0) = Y_i \quad Y_i(1) =$ **?**

  If $T_i = 1$, $\quad Y_i(0) =$ **?** $\quad Y_i(1) = Y_i$
- (I_i, T_i, Y_i) are random; $Y_i(1)$ and $Y_i(0)$ are fixed.

Quantities of Interest

Quantities of Interest

- Treatment Effect (for unit i):

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

Quantities of Interest

- Treatment Effect (for unit i):

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

- Population Average Treatment Effect:

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

Quantities of Interest

- Treatment Effect (for unit i):

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

- Population Average Treatment Effect:

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

- Sample Average Treatment Effect:

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

Decomposition of Causal Effect Estimation Error

Decomposition of Causal Effect Estimation Error

- Difference in means estimator:

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

Decomposition of Causal Effect Estimation Error

- Difference in means estimator:

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

- Estimation Error:

$$\Delta \equiv \text{PATE} - D$$

Decomposition of Causal Effect Estimation Error

- Difference in means estimator:

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

- Estimation Error:

$$\Delta \equiv \text{PATE} - D$$

- Pretreatment confounders: X are observed and U are unobserved

Decomposition of Causal Effect Estimation Error

- Difference in means estimator:

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

- Estimation Error:

$$\Delta \equiv \text{PATE} - D$$

- Pretreatment confounders: X are observed and U are unobserved
- Decomposition:

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

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

Selection Error

Selection Error

- Definition:

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

where NATE is the nonsample average treatment effect.

Selection Error

- Definition:

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

  where NATE is the nonsample average treatment effect.
- Δ_S vanishes if:

Selection Error

- Definition:

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

  where NATE is the nonsample average treatment effect.

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

Selection Error

- Definition:

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

  where NATE is the nonsample average treatment effect.

- Δ_S vanishes if:
  1. The sample is a census ($I_i = 1$ for all observations and $n = N$);
  2. SATE = NATE; or

Selection Error

- Definition:

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

  where NATE is the nonsample average treatment effect.

- Δ_S vanishes if:
  1. The sample is a census ($I_i = 1$ for all observations and $n = N$);
  2. SATE $=$ NATE; or
  3. Switch quantity of interest from PATE to SATE (recommended!)

Decomposing Selection Error

Decomposing Selection Error

- Decomposition:

$$\Delta_S = \Delta_{S_X} + \Delta_{S_U}$$

Decomposing Selection Error

- Decomposition:

$$\Delta_S = \Delta_{S_X} + \Delta_{S_U}$$

- $\Delta_{S_X} = 0$ 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

- Decomposition:

$$\Delta_S = \Delta_{S_X} + \Delta_{S_U}$$

- $\Delta_{S_X} = 0$ 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

- Decomposition:

$$\Delta_S = \Delta_{S_X} + \Delta_{S_U}$$

- $\Delta_{S_X} = 0$ 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: $\widetilde{F}(U \mid I = 0) = \widetilde{F}(U \mid I = 1)$.
- conditions are unverifiable: X is observed only in sample and U is not observed at all.

Decomposing Selection Error

- Decomposition:

$$\Delta_S = \Delta_{S_X} + \Delta_{S_U}$$

- $\Delta_{S_X} = 0$ 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: $\widetilde{F}(U \mid I = 0) = \widetilde{F}(U \mid I = 1)$.
- conditions are unverifiable: X is observed only in sample and U is not observed at all.
- Δ_{S_X} vanishes if weighting on X

Decomposing Selection Error

- Decomposition:

$$\Delta_S = \Delta_{S_X} + \Delta_{S_U}$$

- $\Delta_{S_X} = 0$ 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: $\widetilde{F}(U \mid I = 0) = \widetilde{F}(U \mid I = 1)$.
- conditions are unverifiable: X is observed only in sample and U is not observed at all.
- Δ_{S_X} vanishes if weighting on X
- Δ_{S_U} cannot be corrected after the fact

Decomposing Treatment Imbalance

Decomposing Treatment Imbalance

- Decomposition:

$$\Delta_T = \Delta_{T_X} + \Delta_{T_U}$$

Decomposing Treatment Imbalance

- Decomposition:

$$\Delta_T = \Delta_{T_X} + \Delta_{T_U}$$

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

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

  Verifiable from data; can be generated ex ante by blocking or enforced ex post via matching or parametric adjustment

Decomposing Treatment Imbalance

- Decomposition:

$$\Delta_T = \Delta_{T_X} + \Delta_{T_U}$$

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

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

  Verifiable from data; can be generated ex ante by blocking or enforced ex post via matching or parametric adjustment

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

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

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

Alternative Quantities of Interest

Alternative Quantities of Interest

- 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

- Sample average treatment effect on the treated:

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

- Population average treatment effect on the treated:

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

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

Alternative Quantities of Interest

- Sample average treatment effect on the treated:

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

- Population average treatment effect on the treated:

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

  (where $N^* = \sum_{i=1}^{N} T_i$ is the number of treated units in the population)
- When are these of more interest than PATE and SATE? Why never for randomized experiments? Why usually for matching?

Alternative Quantities of Interest

- Sample average treatment effect on the treated:

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

- Population average treatment effect on the treated:

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

  (where $N^* = \sum_{i=1}^{N} T_i$ is the number of treated units in the population)
- When are these of more interest than PATE and SATE? Why never for randomized experiments? Why usually for matching?
- Analogous estimation error decomposition: $\Delta' = \text{PATT} - D$, holds:

$$\Delta' \quad = \quad (\Delta'_{S_X} + \Delta'_{S_U}) + (\Delta'_{T_X} + \Delta'_{T_U})$$

Effects of Design Components on Estimation Error

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

Effects of Design Components on Estimation Error

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

Effects of Design Components on Estimation Error

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

Effects of Design Components on Estimation Error

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

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

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	$= 0$	$=?$		

Effects of Design Components on Estimation Error

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	$\to ?$	$\to ?$	$\to ?$	$\to ?$

Effects of Design Components on Estimation Error

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	$\to ?$	$\to ?$	$\to ?$	$\to ?$
Random treatment assignment			$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$

Effects of Design Components on Estimation Error

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	$\to ?$	$\to ?$	$\to ?$	$\to ?$
Random treatment assignment			$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$
Complete blocking			$= 0$	$=?$

Effects of Design Components on Estimation Error

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	$\to ?$	$\to ?$	$\to ?$	$\to ?$
Random treatment assignment			$\overset{\text{avg}}{=} 0$	$\overset{\text{avg}}{=} 0$
Complete blocking			$= 0$	$=?$
Exact matching			$= 0$	$=?$

Effects of Design Components on Estimation Error

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	$\to?$	$\to?$	$\to?$	$\to?$
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

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	$\to ?$	$\to ?$	$\to ?$	$\to ?$
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

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	$\to ?$	$\to ?$	$\to ?$	$\to ?$
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

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	$\rightarrow ?$	$\rightarrow ?$	$\rightarrow ?$	$\rightarrow ?$
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 and Matching

Comparing Blocking and Matching

- Adding blocking to random assignment is always as or more efficient, and never biased (blocking on pretreatment variables related to the outcome is more efficient)

Comparing Blocking and Matching

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

Comparing Blocking and Matching

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

Comparing Blocking and Matching

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

Comparing Blocking and Matching

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

Comparing Blocking and Matching

- Adding blocking to random assignment is always as or more efficient, and never biased (blocking on pretreatment variables related to the outcome is more efficient)
- Blocking is like regression adjustment, where the functional form and the parameter values are known
- Matching is like blocking, except:
  1. to avoid selection error, must change quantity of interest from PATE to PATT or SATT,
  2. random treatment assignment following matching is impossible
  3. Exact matching, unlike blocking, is dependent on the already-collected data happening to contain sufficiently good matches.

Comparing Blocking and Matching

- Adding blocking to random assignment is always as or more efficient, and never biased (blocking on pretreatment variables related to the outcome is more efficient)
- Blocking is like regression adjustment, where the functional form and the parameter values are known
- Matching is like blocking, except:
    1. to avoid selection error, must change quantity of interest from PATE to PATT or SATT,
    2. random treatment assignment following matching is impossible
    3. Exact matching, unlike blocking, is dependent on the already-collected data happening to contain sufficiently good matches.
    4. In the worst case scenerio, matching (just like regression adjustment) can increase bias (this cannot occur with blocking plus random assignment)

Comparing Blocking and Matching

- Adding blocking to random assignment is always as or more efficient, and never biased (blocking on pretreatment variables related to the outcome is more efficient)
- Blocking is like regression adjustment, where the functional form and the parameter values are known
- Matching is like blocking, except:
  1. to avoid selection error, must change quantity of interest from PATE to PATT or SATT,
  2. random treatment assignment following matching is impossible
  3. Exact matching, unlike blocking, is dependent on the already-collected data happening to contain sufficiently good matches.
  4. In the worst case scenerio, matching (just like regression adjustment) can increase bias (this cannot occur with blocking plus random assignment)
- Adding matching to a parametric model almost always reduces model dependence and bias, and sometimes variance too

The Benefits of Major Research Designs: Overview

	Δ_{S_X}	Δ_{S_U}	Δ_{T_X}	Δ_{T_U}
Ideal experiment	$\to 0$	$\to 0$	$= 0$	$\to 0$
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)	$\neq 0$	$\neq 0$	$\to 0$	$\to 0$
Survey Experiment (Limited or no blocking)	$\to 0$	$\to 0$	$\to 0$	$\to 0$
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	$\neq 0$

For column Q, "$\to 0$" denotes $E(Q) = 0$ and $\lim_{n\to\infty} \text{Var}(Q) = 0$, whereas "$\overset{\text{avg}}{=} 0$" indicates zero on average, or $E(Q) = 0$, for a design with a small n. Δ_S can be set to zero if we switch from PATE to SATE.

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
- $E(\Delta_{S_X}) = 0$, $\lim_{n\to\infty} V(\Delta_{S_X}) = 0$

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
- $E(\Delta_{S_X}) = 0$, $\lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0$, $\lim_{n\to\infty} V(\Delta_{S_U}) = 0$

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
- $E(\Delta_{S_X}) = 0$, $\lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0$, $\lim_{n\to\infty} V(\Delta_{S_U}) = 0$
- $\Delta_{T_X} = 0$

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
- $E(\Delta_{S_X}) = 0$, $\lim_{n\to\infty} V(\Delta_{S_X}) = 0$
- $E(\Delta_{S_U}) = 0$, $\lim_{n\to\infty} V(\Delta_{S_U}) = 0$
- $\Delta_{T_X} = 0$
- $E(\Delta_{T_U}) = 0$, $\lim_{n\to\infty} V(\Delta_{T_U}) = 0$

An Even More Ideal Experiment

An Even More Ideal Experiment

- Begin with a well-defined population

An Even More Ideal Experiment

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

An Even More Ideal Experiment

- Begin with a well-defined population
- Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata (if strata sample is proportional to population fraction, no weights are needed)

An Even More Ideal Experiment

- Begin with a well-defined population
- Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata (if strata sample is proportional to population fraction, no weights are needed)
- large *n*

An Even More Ideal Experiment

- Begin with a well-defined population
- Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata (if strata sample is proportional to population fraction, no weights are needed)
- large *n*
- blocking on all known confounders

An Even More Ideal Experiment

- Begin with a well-defined population
- Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata (if strata sample is proportional to population fraction, no weights are needed)
- large n
- blocking on all known confounders
- random treatment assignment within blocks

An Even More Ideal Experiment

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

An Even More Ideal Experiment

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

An Even More Ideal Experiment

- Begin with a well-defined population
- Define sampling strata based on cross-classification of all known confounders
- Random sampling within strata (if strata sample is proportional to population fraction, no weights are 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

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

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 $\Delta_{S_U} = 0$
- $E(\Delta_{T_X}) = 0$, $\lim_{n\to\infty} V(\Delta_{T_X}) = 0$

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 $\Delta_{S_U} = 0$
- $E(\Delta_{T_X}) = 0$, $\lim_{n\to\infty} V(\Delta_{T_X}) = 0$
- $E(\Delta_{T_U}) = 0$, $\lim_{n\to\infty} V(\Delta_{T_U}) = 0$

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 (only question wording treatments)

Survey Experiment

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

Survey Experiment

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

Survey Experiment

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

Survey Experiment

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

Observational Study, well-matched

Observational Study, well-matched

- nonrandom selection

Observational Study, well-matched

- nonrandom selection
- large n

Observational Study, well-matched

- nonrandom selection
- large n
- no blocking

Observational Study, well-matched

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

Observational Study, well-matched

- 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

- 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

- 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

- 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

What is the Best Design?

What is the Best Design?

- The ideal design is rarely feasible

What is the Best Design?

- The ideal design is rarely feasible
- Effort in experimental studies: random assignment

What is the Best Design?

- The ideal design is rarely feasible
- Effort in experimental studies: random assignment
- Effort in observational studies: measuring and adjusting for X (via matching or modeling)

What is the Best Design?

- The ideal design is rarely feasible
- Effort in experimental studies: random assignment
- Effort in observational studies: measuring and adjusting for X (via matching or modeling)
- the Achilles heal of experimental studies: Δ_S, small n, definition of T

What is the Best Design?

- The ideal design is rarely feasible
- Effort in experimental studies: random assignment
- Effort in observational studies: measuring and adjusting for X (via matching or modeling)
- the Achilles heal of experimental studies: Δ_S, small n, definition of T
- the Achilles heal of observational studies: Δ_U

What is the Best Design?

- The ideal design is rarely feasible
- Effort in experimental studies: random assignment
- Effort in observational studies: measuring and adjusting for X (via matching or modeling)
- the Achilles heal of experimental studies: Δ_S, small n, definition of T
- the Achilles heal of observational studies: Δ_U
- Each design accomodates best to the applications for which it was designed

Fallacies in Experimental Research

Fallacies in Experimental Research

- Failure to block on all known covariates

Fallacies in Experimental Research

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

Fallacies in Experimental Research

- Failure to block on all known covariates
  - seen incorrectly 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 known covariates
  - seen incorrectly 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 known covariates
  - seen incorrectly 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" (Box et al.)

Fallacies in Experimental Research

- Failure to block on all known covariates
  - seen incorrectly 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" (Box et al.)
- Conducting t-tests to check for balance after random treatment assignment

Fallacies in Experimental Research

- Failure to block on all known covariates
  - seen incorrectly 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” (Box et al.)
- Conducting t-tests to check for balance after random treatment assignment
  - variables blocked on balance exactly after treatment assignment; so if you’re checking, you missed an opportunity to increase efficiency

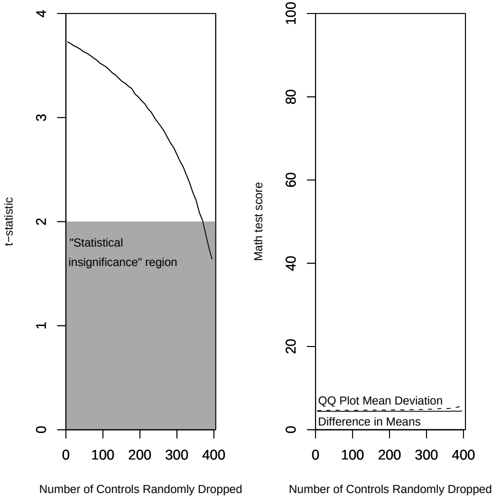
Fallacies in Experimental Research

- Failure to block on all known covariates
  - seen incorrectly 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" (Box et al.)
- Conducting t-tests to check for balance after random treatment assignment
  - variables blocked on balance exactly after treatment assignment; so if you're checking, you missed an opportunity to increase efficiency
  - if variables became available after treatment assignment, a t-test only checks for whether randomization was done appropriately

Fallacies in Experimental Research

- Failure to block on all known covariates
  - seen incorrectly 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" (Box et al.)
- Conducting t-tests to check for balance after random treatment assignment
  - variables blocked on balance exactly after treatment assignment; so if you're checking, you missed an opportunity to increase efficiency
  - if variables became available after treatment assignment, a t-test only checks for whether 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



Randomly dropping observations “reduces” imbalance???

The Balance Test Fallacy: Explanation

The Balance Test Fallacy: Explanation

- hypothesis tests are a function of balance AND power; we only want to measure balance

The Balance Test Fallacy: Explanation

- hypothesis tests are a function of balance AND power; we only want to measure balance
- for SATE, PATE, or SPATE, balance is a characteristic of the sample, not some superpopulation; so no need to make an inference

The Balance Test Fallacy: Explanation

- hypothesis tests are a function of balance AND power; we only want to measure balance
- for SATE, PATE, or SPATE, balance is a characteristic of the sample, not some superpopulation; so no need to make an inference
- Simple linear model (for intution):

The Balance Test Fallacy: Explanation

- hypothesis tests are a function of balance AND power; we only want to measure balance
- for SATE, PATE, or SPATE, balance is a characteristic of the sample, not some superpopulation; so no need to make an inference
- Simple linear model (for intution):
  - Suppose $E(Y \mid T, X) = \theta + T\beta + X\gamma$

The Balance Test Fallacy: Explanation

- hypothesis tests are a function of balance AND power; we only want to measure balance
- for SATE, PATE, or SPATE, balance is a characteristic of the sample, not some superpopulation; so no need to make an 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

- hypothesis tests are a function of balance AND power; we only want to measure balance
- for SATE, PATE, or SPATE, balance is a characteristic of the sample, not some superpopulation; so no need to make an 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 is due to G only.

The Balance Test Fallacy: Explanation

- hypothesis tests are a function of balance AND power; we only want to measure balance
- for SATE, PATE, or SPATE, balance is a characteristic of the sample, not some superpopulation; so no need to make an 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 is due to G only.
  - If $G = 0$, bias=0

The Balance Test Fallacy: Explanation

- hypothesis tests are a function of balance AND power; we only want to measure balance
- for SATE, PATE, or SPATE, balance is a characteristic of the sample, not some superpopulation; so no need to make an 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 is due to G only.
  - If $G = 0$, bias=0
  - If $G \neq 0$, bias can be any size due to γ

The Balance Test Fallacy: Explanation

- hypothesis tests are a function of balance AND power; we only want to measure balance
- for SATE, PATE, or SPATE, balance is a characteristic of the sample, not some superpopulation; so no need to make an 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 is due to G only.
  - If $G = 0$, bias=0
  - If $G \neq 0$, bias can be any size due to γ
  - If you cannot control G (and we don't usually try to minimize γ so we don't cook the books), then G must be reduced without limit.

The Balance Test Fallacy: Explanation

- hypothesis tests are a function of balance AND power; we only want to measure balance
- for SATE, PATE, or SPATE, balance is a characteristic of the sample, not some superpopulation; so no need to make an 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 is due to G only.
  - If $G = 0$, bias=0
  - If $G \neq 0$, bias can be any size due to γ
  - If you cannot control G (and we don't usually try to minimize γ so we don't cook the books), then G must be reduced without limit.
- No threshold level is safe.