Policy Evaluation - Quasi-Experimental Research Designs

April 22, 2025

Estimating Treatment Effects Review

- $ATE = Avg_n[Y^1 - Y^0]$
- $ATE_{est} = Avg_n[Y^1|D=1] - Avg_n[Y^0|D=0]$
- When $(Y^1, Y^0) \not\perp\!\!\!\perp D$:

$$ATE_{est} = ATE + \underbrace{\{Avg_n[Y^0|D=1] - Avg_n[Y^0|D=0]\}}_{\text{Selection Bias}} + \underbrace{(1-\pi)(ATT - ATU)}_{\text{Heterogeneous Treatment Effect Bias}}$$

- $ATE_{est} = \beta_0 + \beta_1 D + \beta_2 X_1 + \beta_3 X_2 + ...\beta_k X_{k-1} + \varepsilon$
- Natural experiment w/randomization (Oregon): $(Y^1, Y^0) \perp\!\!\!\perp D$

Quasi-Experimental Research Design

- But we rarely encounter natural experiments w/ randomized assignment to treatment, so what can we do?

Quasi-Experimental Research Design

- But we rarely encounter natural experiments w/ randomized assignment to treatment, so what can we do?
- Method 1: Difference-in-differences
  - Intuition: Compare units exposed to treatment before and after exposure to unexposed units.

Quasi-Experimental Research Design

- But we rarely encounter natural experiments w/ randomized assignment to treatment, so what can we do?
- Method 1: Difference-in-differences
  - Intuition: Compare units exposed to treatment before and after exposure to unexposed units.
  - ATE $= (Treat_{post} - Treat_{pre}) - (Control_{post} - Control_{pre})$

Quasi-Experimental Research Design

- But we rarely encounter natural experiments w/ randomized assignment to treatment, so what can we do?
- Method 1: Difference-in-differences
  - Intuition: Compare units exposed to treatment before and after exposure to unexposed units.
  - ATE $= (Treat_{post} - Treat_{pre}) - (Control_{post} - Control_{pre})$
  - ATE $=$ $(E[Y^1_{treat}|Post] - E[Y^0_{treat}|Pre]) - (E[Y^0_{cont}|Post] - E[Y^0_{cont}|Pre])$

Quasi-Experimental Research Design - Difference-in-Differences

- Card & Kruegar (1994) - Does raising the minimum wage reduce employment?

Quasi-Experimental Research Design - Difference-in-Differences

- Card & Kruegar (1994) - Does raising the minimum wage reduce employment?
  - Economic theory suggests that higher employment costs will reduce demand for labor

- Card & Kruegar (1994) - Does raising the minimum wage reduce employment?
  - Economic theory suggests that higher employment costs will reduce demand for labor
  - April 1, 1992: New Jersey raises minimum wage from \$4.25 to \$5.05 per hour

- Card & Kruegar (1994) - Does raising the minimum wage reduce employment?
  - Economic theory suggests that higher employment costs will reduce demand for labor
  - April 1, 1992: New Jersey raises minimum wage from \$4.25 to \$5.05 per hour
  - Collected data on employment in 400 fast food restaurants in the Philadelphia area in February 1992 (pre-NJ increase) and again in November 1992 (post-increase).

Quasi-Experimental Research Design - Difference-in-Differences

- Card & Kruegar (1994) - Minimum Wages and Employment

Quasi-Experimental Research Design - Difference-in-Differences

- ATE $= (E[Y^1_{NJ}|Nov] - E[Y^0_{NJ}|Feb]) - (E[Y^0_{PA}|Nov] - E[Y^0_{PA}|Feb])$
  - Where Y is the average number of FTE employees.

Quasi-Experimental Research Design - Difference-in-Differences

- ATE $= (E[Y^1_{NJ}|Nov] - E[Y^0_{NJ}|Feb]) - (E[Y^0_{PA}|Nov] - E[Y^0_{PA}|Feb])$
  - Where Y is the average number of FTE employees.

State	Time	Outcome	Diff 1	Diff 2
New Jersey	Pre	$Y = FTE_{NJ}$		
	Post	$Y = FTE_{NJ} + T + D$		
Pennsylvania	Pre	$Y = FTE_{PA}$		
	Post	$Y = FTE_{PA} + T$		

- ATE $= (E[Y^1_{NJ}|Nov] - E[Y^0_{NJ}|Feb]) - (E[Y^0_{PA}|Nov] - E[Y^0_{PA}|Feb])$
  - Where Y is the average number of FTE employees.

State	Time	Outcome	Diff 1	Diff 2
New Jersey	Pre	$Y = FTE_{NJ}$		
	Post	$Y = FTE_{NJ} + T + D$	T+D	
Pennsylvania	Pre	$Y = FTE_{PA}$		
	Post	$Y = FTE_{PA} + T$	T	

Quasi-Experimental Research Design - Difference-in-Differences

- $\text{ATE} = (E[Y^1_{NJ}|Nov] - E[Y^0_{NJ}|Feb]) - (E[Y^0_{PA}|Nov] - E[Y^0_{PA}|Feb])$
  - Where Y is the average number of FTE employees.

State	Time	Outcome	Diff 1	Diff 2
New Jersey	Pre	$Y = FTE_{NJ}$		
	Post	$Y = FTE_{NJ} + T + D$	T+D	
				D
Pennsylvania	Pre	$Y = FTE_{PA}$		
	Post	$Y = FTE_{PA} + T$	T	

Quasi-Experimental Research Design - Difference-in-Differences

- ATE $= (E[Y^1_{NJ}|Nov] - E[Y^0_{NJ}|Feb]) - (E[Y^0_{PA}|Nov] - E[Y^0_{PA}|Feb])$
  - Where Y is the average number of FTE employees.

State	Time	Outcome	Diff 1	Diff 2
New Jersey	Pre	20.4		
	Post	21.0		
Pennsylvania	Pre	23.3		
	Post	21.1		

Quasi-Experimental Research Design - Difference-in-Differences

- ATE $= (E[Y^1_{NJ}|Nov] - E[Y^0_{NJ}|Feb]) - (E[Y^0_{PA}|Nov] - E[Y^0_{PA}|Feb])$
  - Where Y is the average number of FTE employees.

State	**Time**	**Outcome**	**Diff 1**	**Diff 2**
New Jersey	Pre	20.4		
	Post	21.0	0.6	
Pennsylvania	Pre	23.3		
	Post	21.1	-2.2	

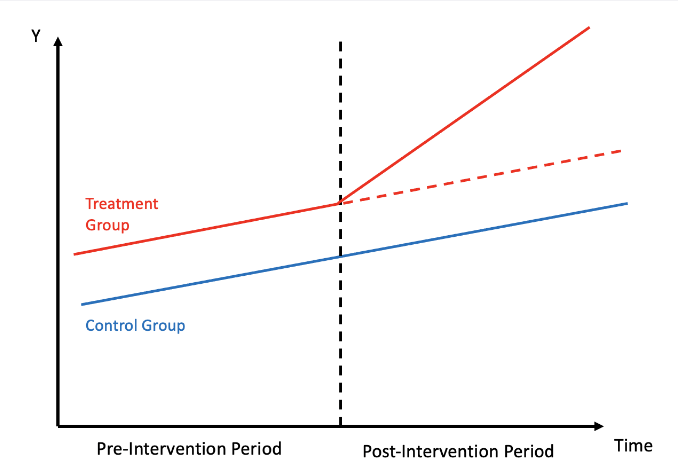
Quasi-Experimental Research Design - Difference-in-Differences

- ATE $= (E[Y^1_{NJ}|Nov] - E[Y^0_{NJ}|Feb]) - (E[Y^0_{PA}|Nov] - E[Y^0_{PA}|Feb])$
  - Where Y is the average number of FTE employees.

State	Time	Outcome	Diff 1	Diff 2
New Jersey	Pre	20.4		
	Post	21.0	0.6	
				2.8
Pennsylvania	Pre	23.3		
	Post	21.1	-2.2	

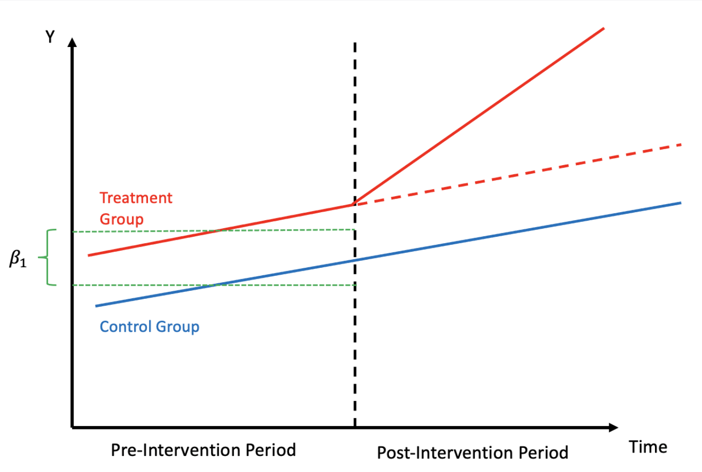
Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$



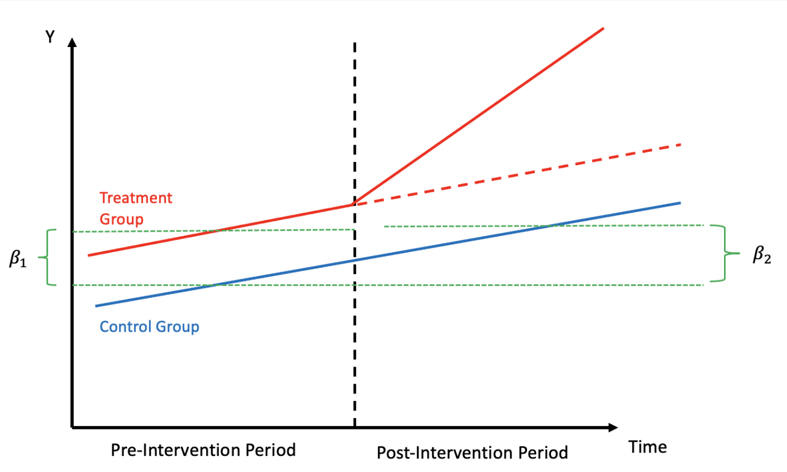
Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $$Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$$



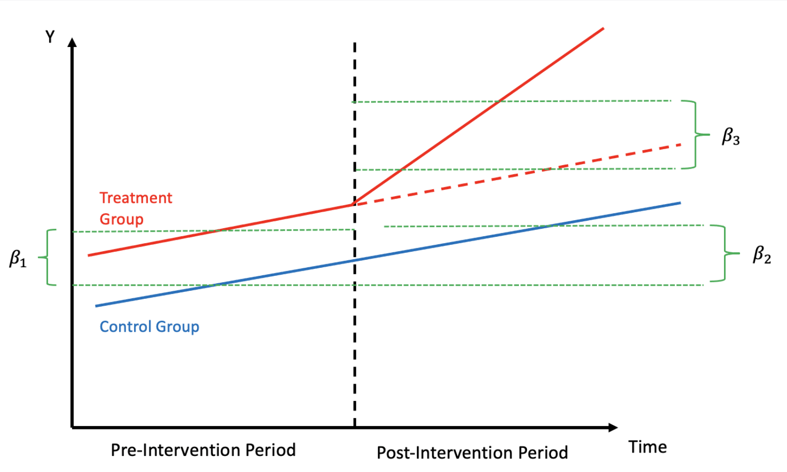
Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$



Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$



- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$

- Why estimate a regression? Why not just compare means?

Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$
- Why estimate a regression? Why not just compare means?
  - We can, but a regression allows us to control for observable differences in treatment and control units.

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$

- Why estimate a regression? Why not just compare means?
  - We can, but a regression allows us to control for observable differences in treatment and control units.

- What ATE are we estimating with DD?

Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$

- Why estimate a regression? Why not just compare means?
  - We can, but a regression allows us to control for observable differences in treatment and control units.

- What ATE are we estimating with DD?
  - Depends. Could be ITT. Could be TOT. Are all units in the treatment group exposed to treatment?

Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$

- Why estimate a regression? Why not just compare means?
  - We can, but a regression allows us to control for observable differences in treatment and control units.

- What ATE are we estimating with DD?
  - Depends. Could be ITT. Could be TOT. Are all units in the treatment group exposed to treatment?

- Requirements for causal interpretation:

Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$

- Why estimate a regression? Why not just compare means?
  - We can, but a regression allows us to control for observable differences in treatment and control units.

- What ATE are we estimating with DD?
  - Depends. Could be ITT. Could be TOT. Are all units in the treatment group exposed to treatment?

- Requirements for causal interpretation:
  - Policy exogeneity
    - Policy enactment is unrelated to outcome of interest
    - Policy is "unanticipated"

Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$

- Why estimate a regression? Why not just compare means?
  - We can, but a regression allows us to control for observable differences in treatment and control units.

- What ATE are we estimating with DD?
  - Depends. Could be ITT. Could be TOT. Are all units in the treatment group exposed to treatment?

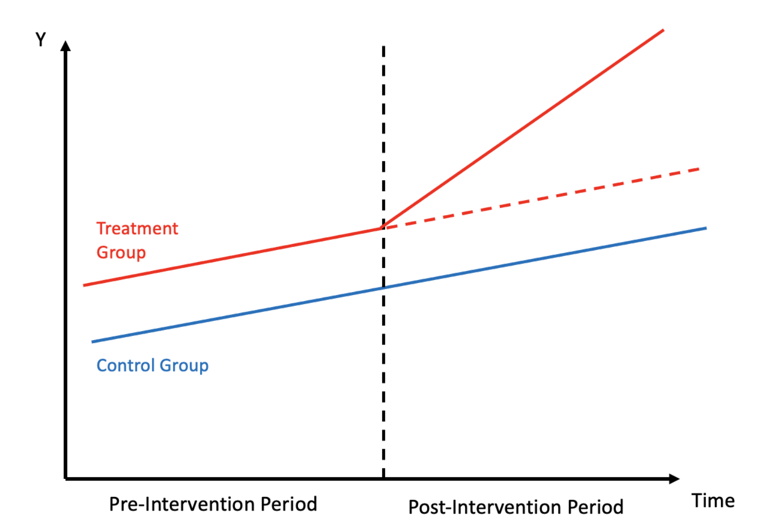
- Requirements for causal interpretation:
  - Policy exogeneity
    - Policy enactment is unrelated to outcome of interest
    - Policy is “unanticipated”
  - Parallel trends assumption

Quasi-Experimental Research Design - Difference-in-Differences

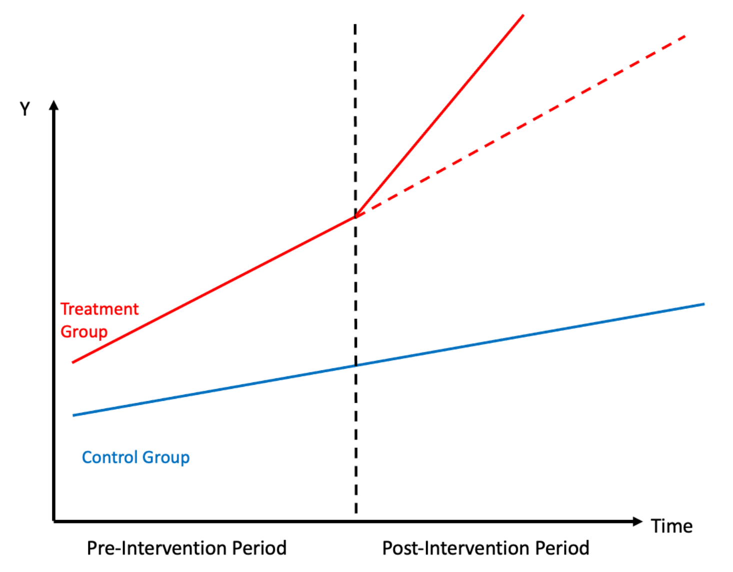
State	Time	Outcome	Diff 1	Diff 2
New Jersey	Pre	Y = FTE_{NJ}		
	Post	Y = FTE_{NJ} + T + D	T+D	
				D
Pennsylvania	Pre	Y = FTE_{PA}		
	Post	Y = FTE_{PA} + T	T	

- Note that Diff 1 assumes that $T_{NJ} = T_{PA}$

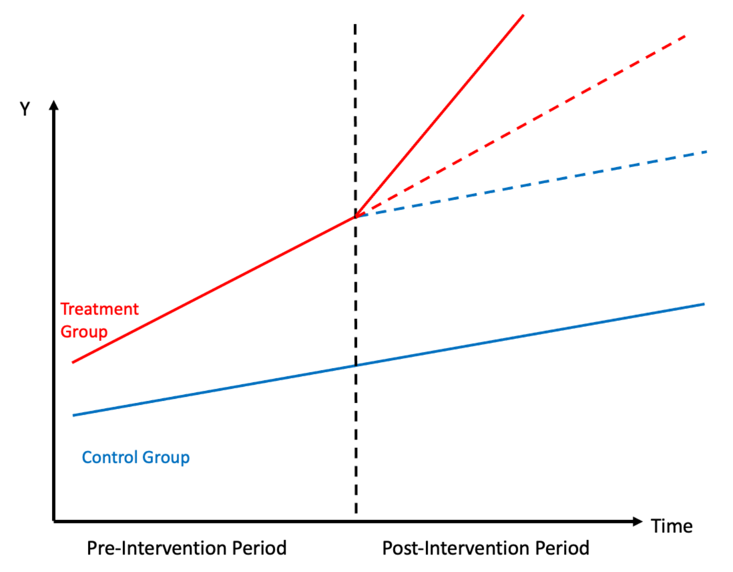
Quasi-Experimental Research Design - Difference-in-Differences



Quasi-Experimental Research Design - Difference-in-Differences



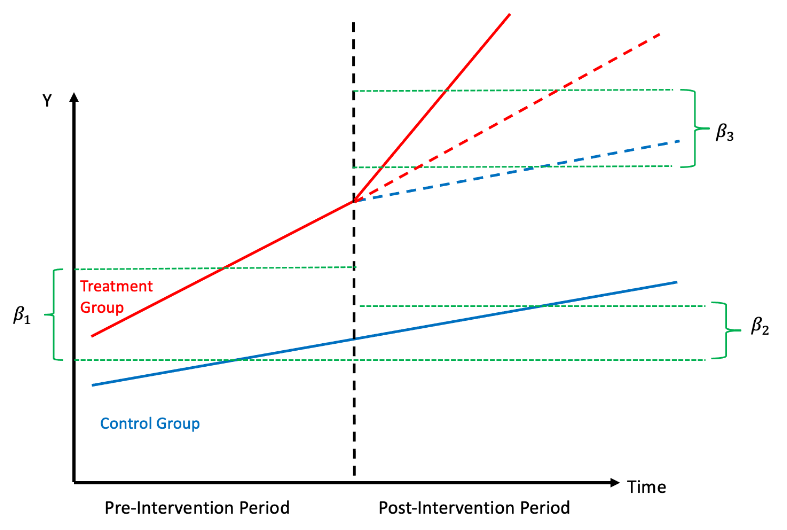
Quasi-Experimental Research Design - Difference-in-Differences



Quasi-Experimental Research Design - Difference-in-Differences

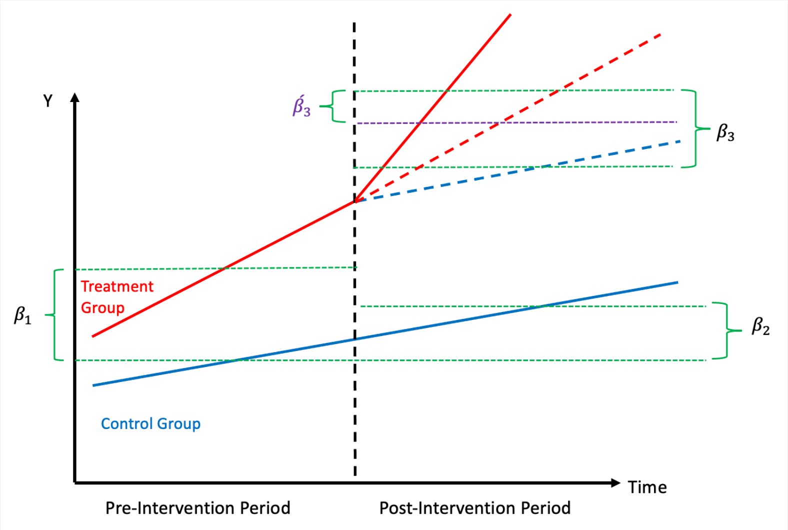
- DD Estimating Equation:

$$Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$$



Quasi-Experimental Research Design - Difference-in-Differences

- DD Estimating Equation:
  $$Y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post_t + \beta_3 Treat_s \times Post_t + \varepsilon_{st}$$



Quasi-Experimental Research Design

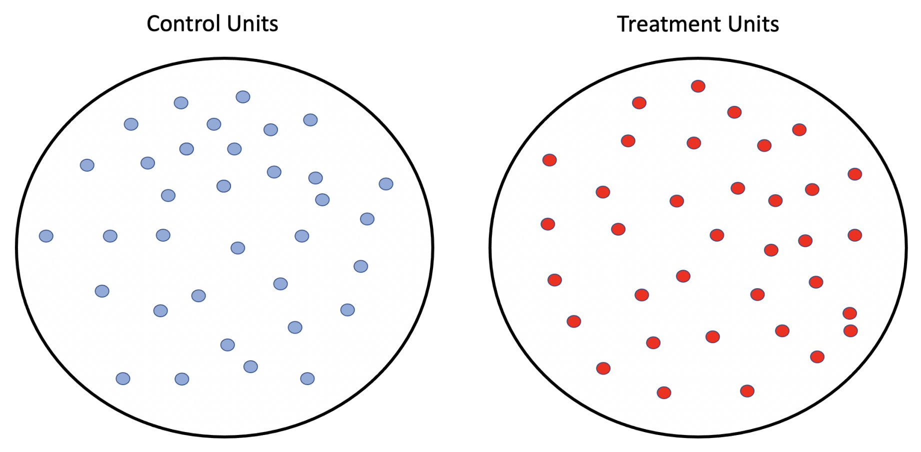
- Can we improve the likelihood that estimating the ATE with non-random assignment returns causal estimates?

Quasi-Experimental Research Design

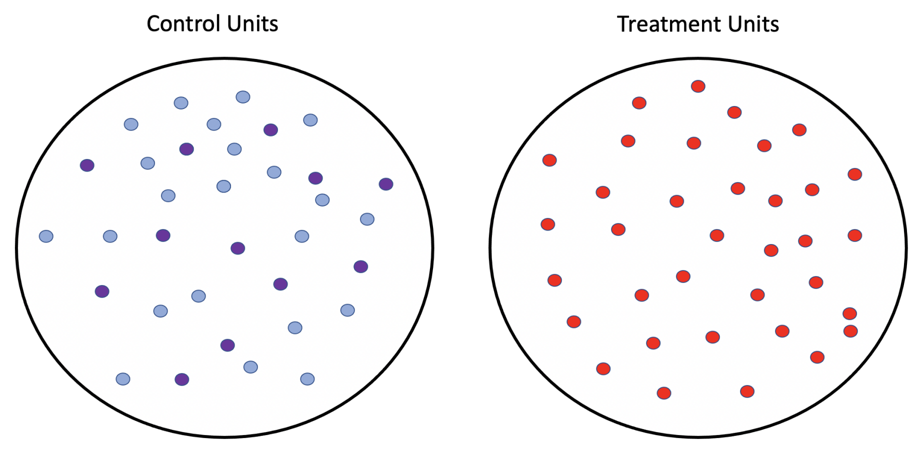
- Can we improve the likelihood that estimating the ATE with non-random assignment returns causal estimates?

- Method 2: Propensity Score Matching
  - Intuition: control units that more closely resemble treatment units on *observables* will also more closely resemble treatment units on *unobservables*.

Quasi-Experimental Research Design - Propensity Score Matching



Quasi-Experimental Research Design - Propensity Score Matching



Quasi-Experimental Research Design - Propensity Score Matching

- PSM Steps:

Quasi-Experimental Research Design - Propensity Score Matching

- PSM Steps:
  1. Select covariates
     - Match on characteristics related to treatment status.

Quasi-Experimental Research Design - Propensity Score Matching

- PSM Steps:
  1. Select covariates
     - Match on characteristics related to treatment status.
  2. Specify regression model for matching
     - Predict probability of treatment.
     - Assign a “propensity score” to each control unit.

Quasi-Experimental Research Design - Propensity Score Matching

- PSM Steps:
  1. Select covariates
     - Match on characteristics related to treatment status.
  2. Specify regression model for matching
     - Predict probability of treatment.
     - Assign a “propensity score” to each control unit.
  3. Select a matching method (several options).

Quasi-Experimental Research Design - Propensity Score Matching

- PSM Steps:
  1. Select covariates
     - Match on characteristics related to treatment status.
  2. Specify regression model for matching
     - Predict probability of treatment.
     - Assign a "propensity score" to each control unit.
  3. Select a matching method (several options).
  4. Create matches.

Quasi-Experimental Research Design - Propensity Score Matching

- PSM Steps:
  1. Select covariates
     - Match on characteristics related to treatment status.
  2. Specify regression model for matching
     - Predict probability of treatment.
     - Assign a "propensity score" to each control unit.
  3. Select a matching method (several options).
  4. Create matches.
  5. Compare balance.
     - How similar are observables between the treatment and matched controls?

Quasi-Experimental Research Design - Propensity Score Matching

- PSM Steps:
  1. Select covariates
     - Match on characteristics related to treatment status.
  2. Specify regression model for matching
     - Predict probability of treatment.
     - Assign a "propensity score" to each control unit.
  3. Select a matching method (several options).
  4. Create matches.
  5. Compare balance.
     - How similar are observables between the treatment and matched controls?
  6. Estimate the ATE.

Quasi-Experimental Research Design - Propensity Score Matching

- Matching Methods

1. Nearest neighbor matching

- Matching Methods

1. Nearest neighbor matching
    - Sequentially move through the sample of treated units matching each unit with the closest control unit.

- Matching Methods

1. Nearest neighbor matching
    - Sequentially move through the sample of treated units matching each unit with the closest control unit.
    - "Closest" is determined by the propensity score.

- Matching Methods

1. Nearest neighbor matching
    - Sequentially move through the sample of treated units matching each unit with the closest control unit.
    - "Closest" is determined by the propensity score.
        - Starting point?
        - Replacement?
        - Caliper adjustment?

Quasi-Experimental Research Design - Propensity Score Matching

- Matching Methods

1. Nearest neighbor matching
    - Sequentially move through the sample of treated units matching each unit with the closest control unit.
    - "Closest" is determined by the propensity score.
        - Starting point?
        - Replacement?
        - Caliper adjustment?

- Downsides:
    - Lots of (ad hoc) decision making required.
    - Discard unmatched control units.

Quasi-Experimental Research Design - Propensity Score Matching

- Matching Methods

2. Inverse probability weighting

Quasi-Experimental Research Design - Propensity Score Matching

- Matching Methods

2. Inverse probability weighting
   - Instead of matching units from the treatment and control groups, IPW re-weights the control group to more closely resemble the treatment group.

Quasi-Experimental Research Design - Propensity Score Matching

- Matching Methods

2. Inverse probability weighting
   - Instead of matching units from the treatment and control groups, IPW re-weights the control group to more closely resemble the treatment group.
   - Use propensity scores as regression weights.

- Matching Methods

2. Inverse probability weighting
   - Instead of matching units from the treatment and control groups, IPW re-weights the control group to more closely resemble the treatment group.
   - Use propensity scores as regression weights.
   - Control units with a high probability of treatment get larger weights and units with a low probability of treatment get smaller weights.

- Matching Methods

2. Inverse probability weighting
    - Instead of matching units from the treatment and control groups, IPW re-weights the control group to more closely resemble the treatment group.
    - Use propensity scores as regression weights.
    - Control units with a high probability of treatment get larger weights and units with a low probability of treatment get smaller weights.

- Remember: regardless of matching technique, PSM assumes that matching on observables removes confounding from unobservables.

- Matching Methods

2. Inverse probability weighting
   - Instead of matching units from the treatment and control groups, IPW re-weights the control group to more closely resemble the treatment group.
   - Use propensity scores as regression weights.
   - Control units with a high probability of treatment get larger weights and units with a low probability of treatment get smaller weights.

- Remember: regardless of matching technique, PSM assumes that matching on observables removes confounding from unobservables.
- We have no definitive way to test this assumption.

Quasi-Experimental Research Design - Propensity Score Matching

- Balance Test: *Shau et al. (2018) - Medicaid is Associated with Increased Readmission and Resource Utilization after Primary Total Knee Arthroplasty*

Table 1
Characteristics of the TKA (ICD-9 code 8154) patients from the 2013 NRD.

Risk factors	Before propensity matching				After propensity matching			
	Medicaid, N = 8372	Other insurance, n = 268,261	*P* Value	Standardized difference	Medicaid, n = 8372	Other insurance, n = 8372	*P* value[a]	Standardized difference[b]
Age, years (mean ± SD)	56.7 ± 9.4	66.7 ± 9.7	<.0001	1.055	56.7 ± 9.4	56.5 ± 9.6	<.001	0.014
Female sex	6072 (72.5%)	165,227 (61.6%)	<.0001	0.234	6071 (72.5%)	6129 (73.2%)	.004	0.016
Severity of illness (major/extreme loss vs other)	549 (6.6%)	13,685 (5.1%)	<.0001	0.062	547 (6.5%)	477 (5.7%)	<.001	0.035
Discharged to skilled facility	2118 (25.3%)	75,028 (28.0%)	<.0001	0.061	2118 (25.3%)	2142 (25.6%)	.261	0.007
Smoking	2711 (32.4%)	59,997 (22.4%)	<.0001	0.226	2710 (32.4%)	2664 (31.8%)	.038	0.012