Econometrics II

Lecture 5: Matching Estimators

Konrad Burchardi

Stockholm University

16th of April 2024

Literature

1. **"Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction"**, Imbens and Rubin
   Chapters 12.1-12.3
2. **"Mostly Harmless Econometrics"**, Angrist and Pischke
   Chapter 3.3.1-3.3.3

These notes draw on those books. All mistakes are mine.

Plan for Today

Conditioncal Independence

In Lecture 1 we discussed that we require

$$\mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] = 0$$

for causal inference about ATT. This is fundamentally not testable.

Maybe plausible: $\{Y_i(0), Y_i(1)\} \perp D_i|X_i$, and X_i are observable (CIA).[1]

Conditional Independence Assumption implies:

$$\begin{aligned}\mathbb{E}[Y|D = 1, X = x] - \mathbb{E}[Y|D = 0, X = x] \\ = \mathbb{E}[Y(1) - Y(0)|X = x] \equiv \tau_x\end{aligned}$$

Suggests one could condition analysis on X. But also...

[1] Fundamentally not testable either! Requires substantial information.

Why Covariate Balance?

...when does a simple difference in means comparison give $\mathbb{E}_X[\tau_x]$?

$$
\begin{aligned}
& \mathbb{E}[Y|D=1] - \mathbb{E}[Y|D=0] \\
= & \int \mathbb{E}[Y|D=1, X=x] f_{X|D=1} dx - \int \mathbb{E}[Y|D=0, X=x] f_{X|D=0} dx && \text{[LIE]} \\
= & \int \left(\mathbb{E}[Y|D=1, X=x] - \mathbb{E}[Y|D=0, X=x]\right) f_X dx && \text{[If } f_{X|D=1} = f_{X|D=0}\text{!]} \\
= & \mathbb{E}_X[\tau_x]
\end{aligned}
$$

Balanced covariate distribution is important!

Why Covariate Balance?

Today:

Check whether (confounding) covariates X_i are balanced.

Exactly analogous to balance test of randomised experiment.

Make sure the (confounding) covariates X_i are balanced.

Exactly analogous to stratification in randomised experiment.

Plan for Today

Propensity Score

Building Intuition, Step 1:
Why might imbalances in covariates be problematic?

Imagine for some $x_1 \neq x_2$

$$\{Y_i(0), Y_i(1)\}|X_i = x_1 \not\sim \{Y_i(0), Y_i(1)\}|X_i = x_2,$$

and we are simply comparing outcomes of treated and control observations.

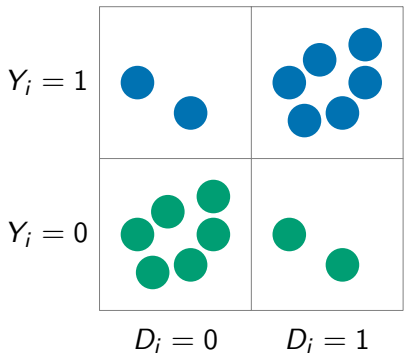
When will this yield an unbiased estimate of $\mathbb{E}_X[\tau_x]$*?*

Propensity Score

A simple example:

- $x_1 = 0$, $x_2 = 1$
- $(Y_i(0)|X_i = 0) = (Y_i(1)|X_i = 0) = 0$
- $(Y_i(0)|X_i = 1) = (Y_i(1)|X_i = 1) = 1$

Suppose share treated at $X_i = 0$ is 0.25 and at $X_i = 1$ it is 0.75.



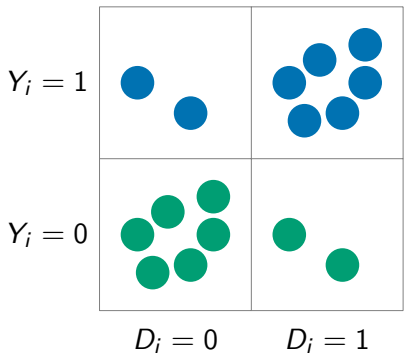
Propensity Score

A simple example:

- $x_1 = 0$, $x_2 = 1$
- $(Y_i(0)|X_i = 0) = (Y_i(1)|X_i = 0) = 0$
- $(Y_i(0)|X_i = 1) = (Y_i(1)|X_i = 1) = 1$

Suppose share treated at $X_i = 0$ is 0.25 and at $X_i = 1$ it is 0.75.

Find positive treatment effect!



Propensity Score

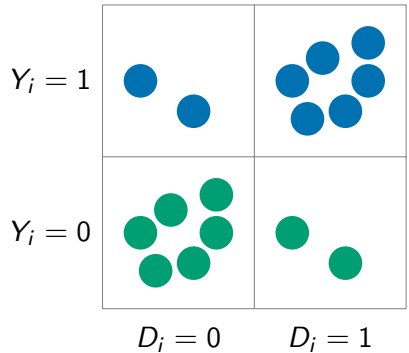
A simple example:

- $x_1 = 0$, $x_2 = 1$
- $(Y_i(0)|X_i = 0) = (Y_i(1)|X_i = 0) = 0$
- $(Y_i(0)|X_i = 1) = (Y_i(1)|X_i = 1) = 1$

Suppose share treated at $X_i = 0$ is 0.25 and at $X_i = 1$ it is 0.75.

Find positive treatment effect!

The problem, somehow:
Observations with different $(Y_i(0), Y_i(1))$ are assigned to treatment at different frequencies.



Propensity Score

A simple example:

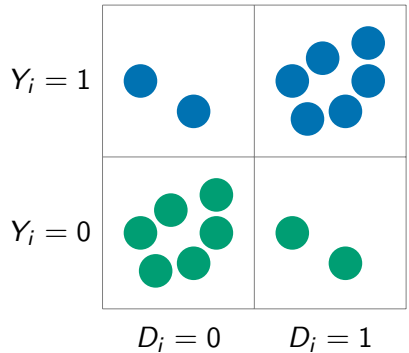
- $x_1 = 0$, $x_2 = 1$
- $(Y_i(0)|X_i = 0) = (Y_i(1)|X_i = 0) = 0$
- $(Y_i(0)|X_i = 1) = (Y_i(1)|X_i = 1) = 1$

Suppose share treated at $X_i = 0$ is 0.25 and at $X_i = 1$ it is 0.75.

Find positive treatment effect!

The problem, somehow:
Observations with different $(Y_i(0), Y_i(1))$ are assigned to treatment at different frequencies.

Note: Find imbalanced covariate distributions.



Propensity Score

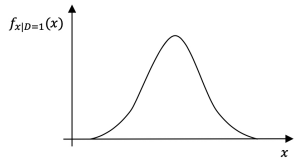
Building Intuition, Step 2:
How might imbalances in covariates occur?

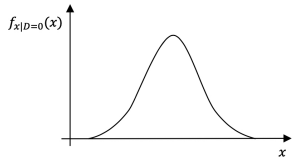
Denote with $e(x) \equiv Pr(D = 1|X = x)$ the probability of treatment given x.
We call this the '**propensity score**'.

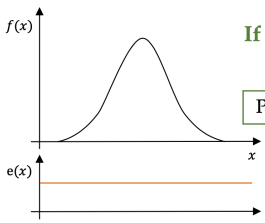
Can it be that $e(x_1) \neq e(x_2)$ for some x_1 and x_2, yet $f_{X|D=1}(x) = f_{X|D=0}(x)$ for all x?

Propensity Score

When will the covariate distribution be balanced across t/c?







If and only if $e(X)$ is a constant!

Probably most important insight today.

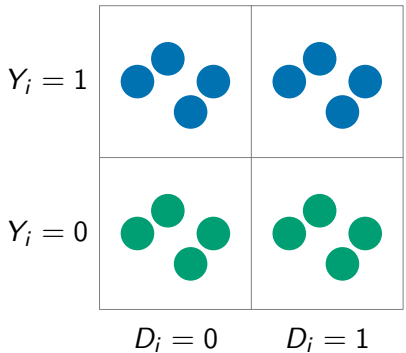
Proof?
See Question 3 of Problem Set 2.

Propensity Score

A simple example:

- $x_1 = 0$, $x_2 = 1$
- $(Y_i(0)|X_i = 0) = (Y_i(1)|X_i = 0) = 0$
- $(Y_i(0)|X_i = 1) = (Y_i(1)|X_i = 1) = 1$

Suppose $e(0) = 0.5$ and $e(1) = 0.5$.



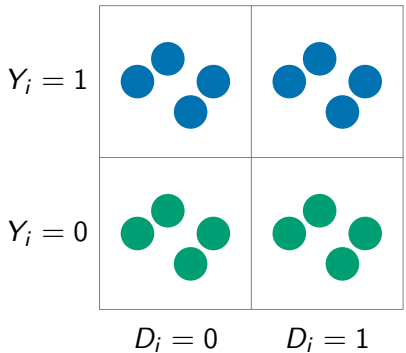
Propensity Score

A simple example:

- $x_1 = 0$, $x_2 = 1$
- $(Y_i(0)|X_i = 0) = (Y_i(1)|X_i = 0) = 0$
- $(Y_i(0)|X_i = 1) = (Y_i(1)|X_i = 1) = 1$

Suppose $e(0) = 0.5$ and $e(1) = 0.5$.

Find no treatment effect!



Propensity Score

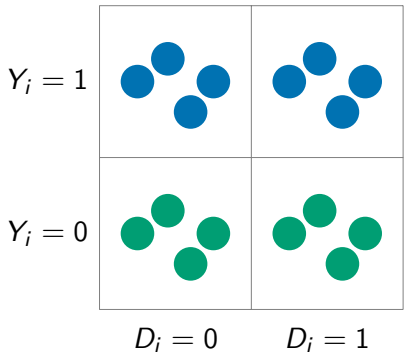
A simple example:

- $x_1 = 0$, $x_2 = 1$
- $(Y_i(0)|X_i = 0) = (Y_i(1)|X_i = 0) = 0$
- $(Y_i(0)|X_i = 1) = (Y_i(1)|X_i = 1) = 1$

Suppose $e(0) = 0.5$ and $e(1) = 0.5$.

Find no treatment effect!

Find no imbalances in covariate distributions!



Propensity Score

Note: the source of the problem is *not*...

- ... that distribution of outcomes is different at different X.
- ... that $e(X_i) \neq 0.5$.

Two diagnostics:

- $X_i = x$ is differently frequent in different treatment arms.

  Motivation for standard balance checks.

- $e(X_i)$ not independent of x.

Result: In fact, these two diagnostics are equivalent![2]

[2] Probably most important insight today.

Plan for Today

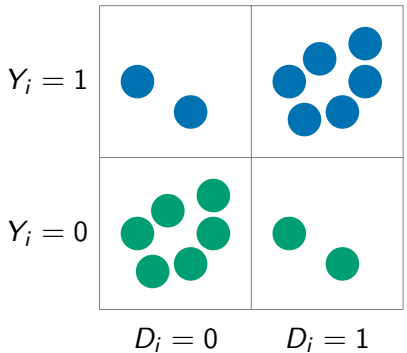
1. Conditional Independence Assumption and Balance
2. Propensity Score
3. Achieving Balance
   - Matching
   - Propensity Score Matching
   - Reweighting
4. Matching and Regressions
5. Practical Issues

Achieving Balance

Solution 1, Matching:

Control for influence of X by conditioning analysis on X, find τ_X.

You see why that solves the problem?



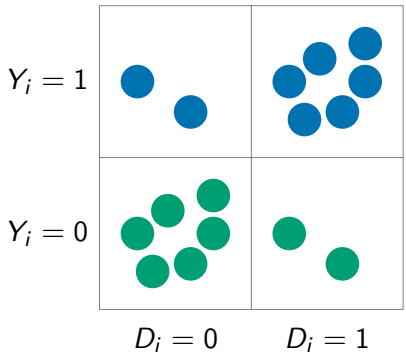
Achieving Balance

Solution 1, Matching:

Control for influence of X by conditioning analysis on X, find τ_X.

You see why that solves the problem?

Fundamental reason why this works:
Within each X cell, $e(X)$ is constant.



Achieving Balance

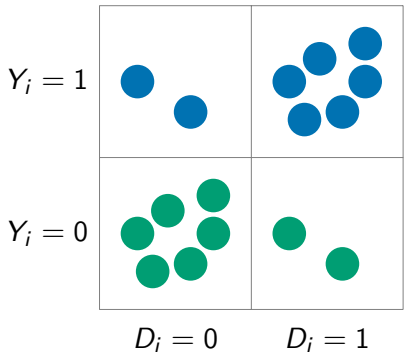
Solution 1, Matching:

Control for influence of X by conditioning analysis on X, find τ_X.

You see why that solves the problem?

Fundamental reason why this works:
Within each X cell, $e(X)$ is constant.

(Typically: Average τ_X's by weights of interest.)



Achieving Balance

Take a just slightly more complex example:

- X has support $\{x_1, x_2, x_3, x_4\}$;
- Distribution of potential outcomes is different;
- $e(x_1) = 0.5$, $e(x_2) = 0.25$, $e(x_3) = 0.5$, $e(x_4) = 0.5$.

Solution 2, Balancing Score Matching: $D_i \perp X_i | b(X_i)$.

Achieving Balance

Solution 3, Propensity Score Matching: Match on $e(X)$.[3]

Recall from before...

- **If $e(X)$ is constant, distribution of X in the treatment arms is the same.**
- And if distribution of X is the same, simple difference in outcomes estimates the average treatment effect, by CIA!

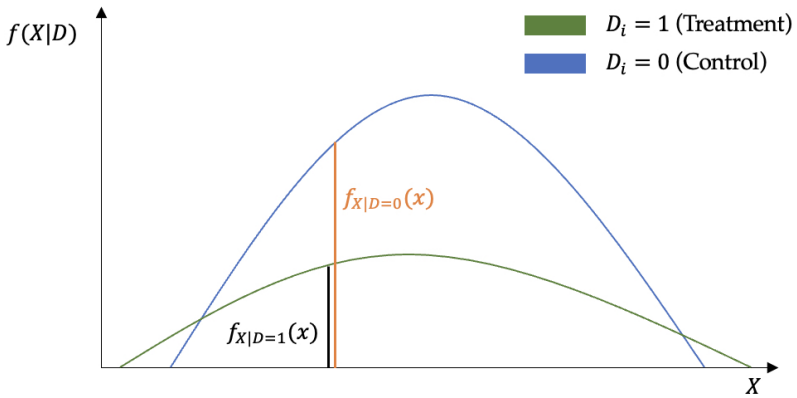
Note:

- This is precisely why randomised trials are useful, they achieve constant $e(X)$.
- Stratification is about forcing realised $\hat{e}(X)$ to be constant.
  Like with RCTs, realised $\hat{e}(X)$ is what matters. *You see why?*

[3] Propensity Score is the coarsest balancing score.

Achieving Balance

Solution 4, Reweighting (Horvitz and Thomson):

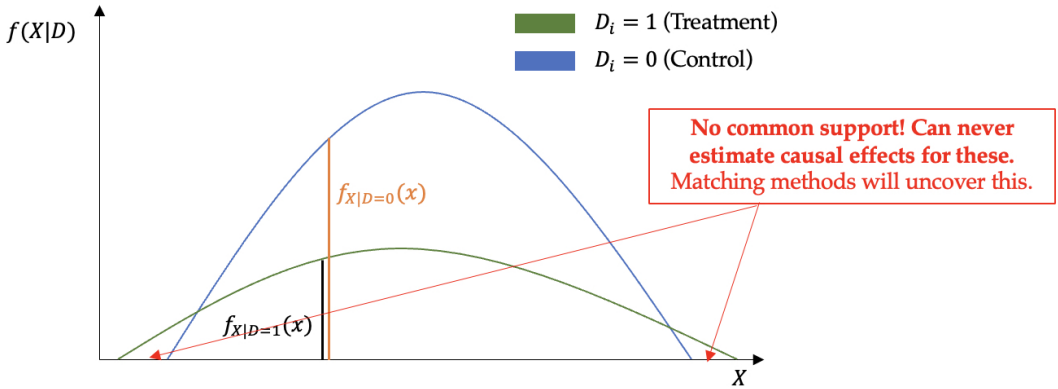


Idea: Find set of weights $\omega(x)$ for all x such that

$$\omega(x) f_{X|D=0}(x) = f_{X|D=1}(x).$$

Achieving Balance

Solution 4, Reweighting (Horvitz and Thomson):



Idea: Find set of weights $\omega(x)$ for all x such that

$$\omega(x) f_{X|D=0}(x) = f_{X|D=1}(x).$$

Achieving Balance

By Bayes' rule (see Angrist and Pischke, ch. 3.3.1):

$$f_{X|D=1}(x) = \frac{Pr(D_i = 1|X_i = x)}{Pr(D_i = 1)} f_X(x) = \frac{e(x)}{e} f_X(x)$$

$$f_{X|D=0}(x) = \frac{Pr(D_i = 0|X_i = x)}{Pr(D_i = 0)} f_X(x) = \frac{1 - e(x)}{1 - e} f_X(x)$$

where $e(x)$ is the propensity score and $e = Pr(D_i = 1)$. Then

$$\omega(x) = \frac{f_{X|D=1}(x)}{f_{X|D=0}(x)} = \frac{e(x)}{1 - e(x)} \frac{1 - e}{e} \tag{1}$$

Then the ATT can be estimated as

$$\hat{\beta}^{ATT} = \frac{1}{N_1} \sum_{i=1}^{N} D_i Y_i - \frac{1}{N_0} \sum_{i=1}^{N} (1 - D_i) Y_i \frac{\hat{e}_i(x)}{1 - \hat{e}_i(x)} \frac{1 - \hat{e}}{\hat{e}} \tag{2}$$

According to Imbens and Rubin, sensitive to the estimation of weights.

Plan for Today

Matching and Regressions

Solution 5, Regression Control:

Upside (see Angrist and Pischke):
Running regression on treatment, fully saturated in X, have

$$\beta_R = \frac{\mathbb{E}[\sigma_D^2(X_i)\tau_X]}{\mathbb{E}[\sigma_D^2(X_i)]}, \text{ where } \sigma_D^2(X_i) \equiv \mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]$$

and a standard ATT matching estimator would be

$$\beta_M = \mathbb{E}[\tau_X Pr(X_i|D_i = 1)].$$

1. Matching estimator weights high Xs with many treatment observations; OLS gives weight to observations with equal treatment shares (max. variance in D_i).
2. If τ_X varies little across X cells, makes little difference.
3. None gives any weight to X cells with no or all treatment.

Matching and Regressions

Solution 5, Regression Control:

Upside (see Angrist and Pischke):
Running regression on treatment, fully saturated in X, have

$$\beta_R = \frac{\mathbb{E}[\sigma_D^2(X_i)\tau_X]}{\mathbb{E}[\sigma_D^2(X_i)]}, \text{ where } \sigma_D^2(X_i) \equiv \mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]$$

and a standard ATT matching estimator would be

$$\beta_M = \mathbb{E}[\tau_X Pr(X_i|D_i = 1)].$$

1. Matching estimator weights high Xs with many treatment observations; OLS gives weight to observations with equal treatment shares (max. variance in D_i).
2. If τ_X varies little across X cells, makes little difference.
3. None gives any weight to X cells with no or all treatment.

Matching and Regressions

Solution 5, Regression Control:

Upside (see Angrist and Pischke):
Running regression on treatment, fully saturated in X, have

$$\beta_R = \frac{\mathbb{E}[\sigma_D^2(X_i)\tau_X]}{\mathbb{E}[\sigma_D^2(X_i)]}, \text{ where } \sigma_D^2(X_i) \equiv \mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]$$

and a standard ATT matching estimator would be

$$\beta_M = \mathbb{E}[\tau_X Pr(X_i|D_i = 1)].$$

1. Matching estimator weights high Xs with many treatment observations; OLS gives weight to observations with equal treatment shares (max. variance in D_i).
2. If τ_X varies little across X cells, makes little difference.
3. None gives any weight to X cells with no or all treatment.

Matching and Regressions

Solution 5, Regression Control:

Upside (see Angrist and Pischke):
Running regression on treatment, fully saturated in X, have

$$\beta_R = \frac{\mathbb{E}[\sigma_D^2(X_i)\tau_X]}{\mathbb{E}[\sigma_D^2(X_i)]}, \text{ where } \sigma_D^2(X_i) \equiv \mathbb{E}[(D_i - \mathbb{E}[D_i|X_i])^2|X_i]$$

and a standard ATT matching estimator would be

$$\beta_M = \mathbb{E}[\tau_X Pr(X_i|D_i = 1)].$$

1. Matching estimator weights high Xs with many treatment observations; OLS gives weight to observations with equal treatment shares (max. variance in D_i).
2. If τ_X varies little across X cells, makes little difference.
3. None gives any weight to X cells with no or all treatment.

Matching and Regressions

Downside (see Imbens and Rubin):

Their argument is that regression requires:

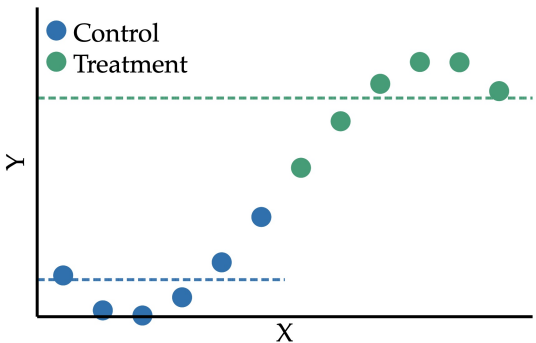
1. unconfoundedness assumption,

and additionally:

2. functional form assumptions for X.

Is strong and unnecessary.

Matching and Regressions

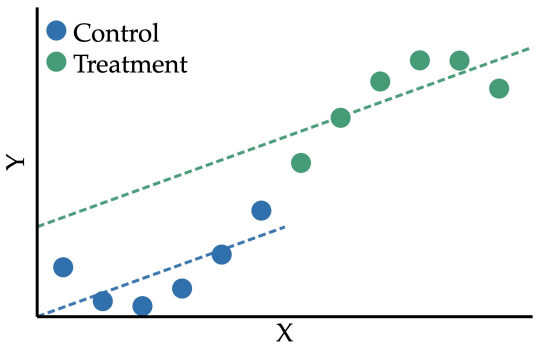


Suppose you run regression of Y on treatment status D and...

- **no control for** X,
- linear control for X, or
- quadratic control for X interacted with D.

What is the coefficient estimate on D?

Matching and Regressions

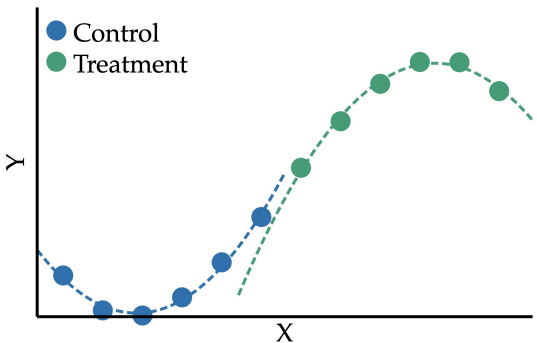


Suppose you run regression of Y on treatment status D and...

- no control for X,
- **linear control for** X, or
- quadratic control for X interacted with D.

What is the coefficient estimate on D?

Matching and Regressions



Suppose you run regression of Y on treatment status D and...

- no control for X,
- linear control for X, or
- **quadratic control for X interacted with D.**

What is the coefficient estimate on D?

Plan for Today

Practical Issues

Distinguish:

1. **Design Stage**
   1. Assessing Overlap in Distributions
      - How similar are distributions? [Univariate/Multivariate tests.]
      - Do similar observations with opposite level of treatment exist?
   2. Estimate Propensity Score
      - Estimated propensity score matters; goal is to achieve balance in sample.
      - Machine learning methods can potentially be helpful.
   3. Create Balanced Sample: unbiased/robust inference; power.
      - Match estimation sample: 1 to 1; 1 to many; on PS or X...
      - Trim sample. Two competing forces for power: sample size vs. match quality.

Practical Issues

Distinguish:

1. **Design Stage**
   1. Assessing Overlap in Distributions
      - How similar are distributions? [Univariate/Multivariate tests.]
      - Do similar observations with opposite level of treatment exist?
   2. Estimate Propensity Score
      - Estimated propensity score matters; goal is to achieve balance in sample.
      - Machine learning methods can potentially be helpful.
   3. Create Balanced Sample: unbiased/robust inference; power.
      - Match estimation sample: 1 to 1; 1 to many; on PS or X...
      - Trim sample. Two competing forces for power: sample size vs. match quality.

No need for any Y here!

Can go back and forth, "play around", until balanced sample is found...

... just do not condition this on Y data.

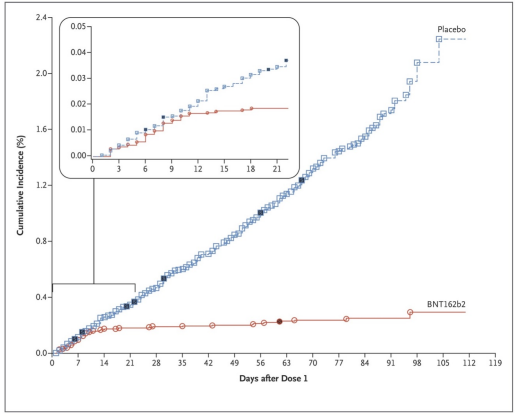
Practical Issues

2. **Assessment Stage**
   - Use pseudo outcomes to validate the approach (see example on next slide).

3. **Analysis Stage**
   - Create PS blocks, or create exact matches, or inexact matches, and estimate mean outcome difference within those.
   - Covariate adjustment might increase precision, see Imbens and Rubin, Part III.
   - Inference easier with exact matches.

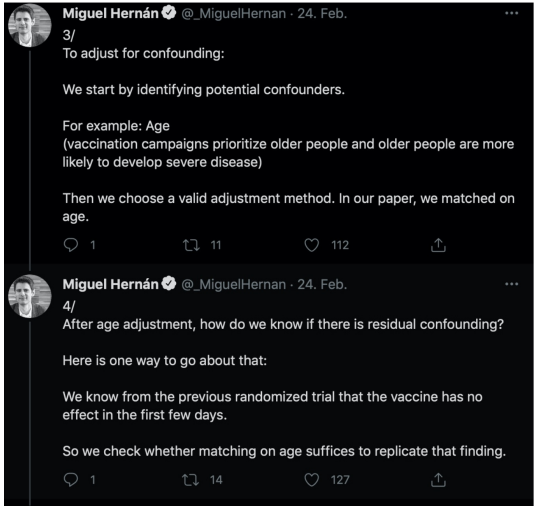
Pseudo Outcomes: Example

Pfizer/BioNTech released results from an RCT in 12/2020 (NEJM, $N = 43548$)



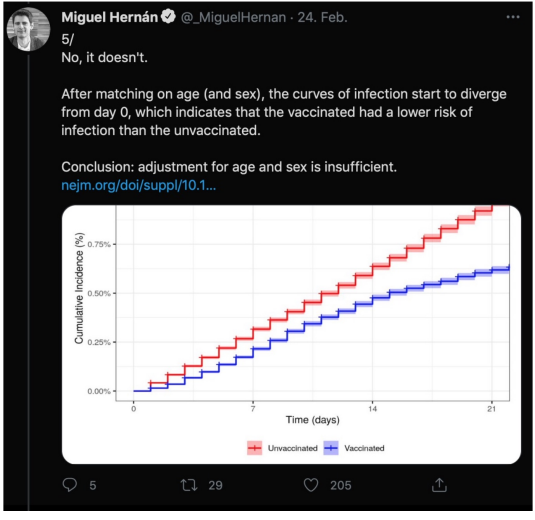
Pseudo Outcomes: Example

Hernán et al. wanted to run such analysis in non-RCT using large-N registry data from Israel. Here is how they tested their matching algorithm.



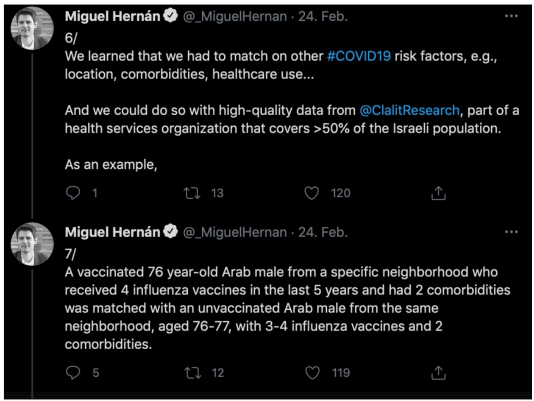
Pseudo Outcomes: Example

Hernán et al. wanted to run such analysis in non-RCT using large-N registry data from Israel. Here is how they tested their matching algorithm.



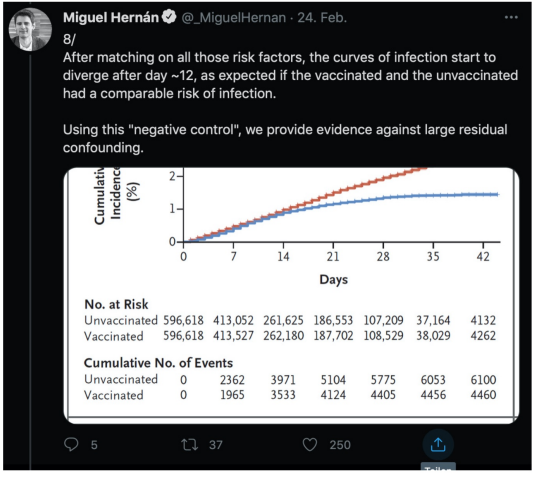
Pseudo Outcomes: Example

Hernán et al. wanted to run such analysis in non-RCT using large-N registry data from Israel. Here is how they tested their matching algorithm.



Pseudo Outcomes: Example

Hernán et al. wanted to run such analysis in non-RCT using large-N registry data from Israel. Here is how they tested their matching algorithm.



Making Matching Persuasive

David McKenzie writes about statistical and ”rhetorical” plausibility. [4]

“Rhetorical” plausibility talks explicitly about why some individuals were treated and others were not.

Examples:

- Separate decision-maker with limited information decides on treatment.
- Capacity limits.
- Treatment consequence of randomization.
- Decision maker caress about different outcome than evaluator.

[4] https://blogs.worldbank.org/impactevaluations/what-do-you-need-do-make-matching-estimator-convincing-rhetorical-vs-statistical

Summary

- Role of balanced covariates distribution.
- Covariates distributions are balanced iff $e(X)$ is constant.
- Obtain balance forcing X to be the same (*matching*),
  or forcing $e(X)$ to be the same (*propensity score matching)*.
- Many practical choices how to implement those ideas.

Questions?