FLS 6441 - Methods III: Explanation and Causation

Week 10 - Matching

Jonathan Phillips

May 2020

Classification of Research Designs

		Independence of Treatment Assignment	**Researcher Controls Treatment Assignment?**
Controlled Experiments	Field Experiments	✓	✓
	Survey and Lab Experiments	✓	✓
Natural Experiments	Natural Experiments	✓	
	Instrumental Variables	✓	
	Discontinuities	✓	
Observational Studies	Difference-in-Differences		
	Controlling for Confounding		
	Matching		
	Comparative Cases and Process Tracing		

Section 1

Matching

The Weakness of Controlling

- Controlling for confounding with regression has three weak spots:

The Weakness of Controlling

- Controlling for confounding with regression has three weak spots:
  1. **Lack of overlap** - Extreme treated outliers alter our results, even when there are no comparable control units in the data

The Weakness of Controlling

- Controlling for confounding with regression has three weak spots:
    1. **Lack of overlap** - Extreme treated outliers alter our results, even when there are no comparable control units in the data
    2. **Model-dependence** - Variable X is a confounder, but is it linear, quadratic, cubic or what? The wrong model of the real relationship with the outcome biases our results

The Weakness of Controlling

- Controlling for confounding with regression has three weak spots:
  1. **Lack of overlap** - Extreme treated outliers alter our results, even when there are no comparable control units in the data
  2. **Model-dependence** - Variable X is a confounder, but is it linear, quadratic, cubic or what? The wrong model of the real relationship with the outcome biases our results
  3. **Researcher/publication bias** - Lots of freedom to tweak the regression to get positive results

The Weakness of Controlling

- Controlling for confounding with regression has three weak spots:
  1. **Lack of overlap** - Extreme treated outliers alter our results, even when there are no comparable control units in the data
  2. **Model-dependence** - Variable X is a confounder, but is it linear, quadratic, cubic or what? The wrong model of the real relationship with the outcome biases our results
  3. **Researcher/publication bias** - Lots of freedom to tweak the regression to get positive results
- All of these reflect the fact that regression is **parametric**

The Weakness of Controlling

- Controlling for confounding with regression has three weak spots:
  1. **Lack of overlap** - Extreme treated outliers alter our results, even when there are no comparable control units in the data
  2. **Model-dependence** - Variable X is a confounder, but is it linear, quadratic, cubic or what? The wrong model of the real relationship with the outcome biases our results
  3. **Researcher/publication bias** - Lots of freedom to tweak the regression to get positive results
- All of these reflect the fact that regression is **parametric**
  1. It uses *ALL* of the data
  2. It requires us to specify the parameters of a model

The Weakness of Controlling

- The solution?

The Weakness of Controlling

- The solution? **Non-parametric** methods for controlling for confounding

The Weakness of Controlling

- The solution? **Non-parametric** methods for controlling for confounding
  1. We use *ONLY SOME* of the data
  2. We do not specify the parameters of any model

The Weakness of Controlling

- The solution? **Non-parametric** methods for controlling for confounding
  1. We use *ONLY SOME* of the data
  2. We do not specify the parameters of any model
- **Matching** is a non-parametric method

The Weakness of Controlling

- The solution? **Non-parametric** methods for controlling for confounding
  1. We use *ONLY SOME* of the data
  2. We do not specify the parameters of any model
- **Matching** is a non-parametric method
  - A **pre-processing** stage
  - Analysis of the results is separate and comes later

Matching

- If treated and control groups have the **same values** of **all** of the confounding variables, we know that treatment is (conditionally) independent of potential outcomes

Matching

- If treated and control groups have the **same values** of **all** of the confounding variables, we know that treatment is (conditionally) independent of potential outcomes
- There is no variation in the confounders that could possibly explain the difference between the outcomes in treated and control groups

Matching

- If treated and control groups have the **same values** of **all** of the confounding variables, we know that treatment is (conditionally) independent of potential outcomes
- There is no variation in the confounders that could possibly explain the difference between the outcomes in treated and control groups
- So how do we force balance on multiple variables?

Matching

- If treated and control groups have the **same values** of **all** of the confounding variables, we know that treatment is (conditionally) independent of potential outcomes
- There is no variation in the confounders that could possibly explain the difference between the outcomes in treated and control groups
- So how do we force balance on multiple variables?
    1. One way is by **adjusting/extrapolating** each treated observation to predict what it would 'look like' if it were identical to a control observation - a regression model

Matching

- If treated and control groups have the **same values** of **all** of the confounding variables, we know that treatment is (conditionally) independent of potential outcomes
- There is no variation in the confounders that could possibly explain the difference between the outcomes in treated and control groups
- So how do we force balance on multiple variables?
  1. One way is by **adjusting/extrapolating** each treated observation to predict what it would 'look like' if it were identical to a control observation - a regression model
  2. An alternative is just to **throw out** all of the treated observations that do not have a comparable control observation - this is matching

Matching

- Matching should really be called **trimming** or **pruning**
  - Dropping units that don’t have good counterfactuals in the data

Matching

- Matching should really be called **trimming** or **pruning**
  - Dropping units that don't have good counterfactuals in the data
- It succeeds only where we can measure and create balance on all confounding variables

Matching

- Matching should really be called **trimming** or **pruning**
  - Dropping units that don't have good counterfactuals in the data
- It succeeds only where we can measure and create balance on all confounding variables
- Matching is **NOT** an experimental method

Matching

1. For each treated unit, find a control unit with very close values of all confounding variables, and keep both

Matching

1. For each treated unit, find a control unit with very close values of all confounding variables, and keep both
2. Repeat for every treated unit

Matching

1. For each treated unit, find a control unit with very close values of all confounding variables, and keep both
2. Repeat for every treated unit
3. Drop all the unmatched units (eg. ’extra’ control units that are ’far away’ from any treated units)

Matching

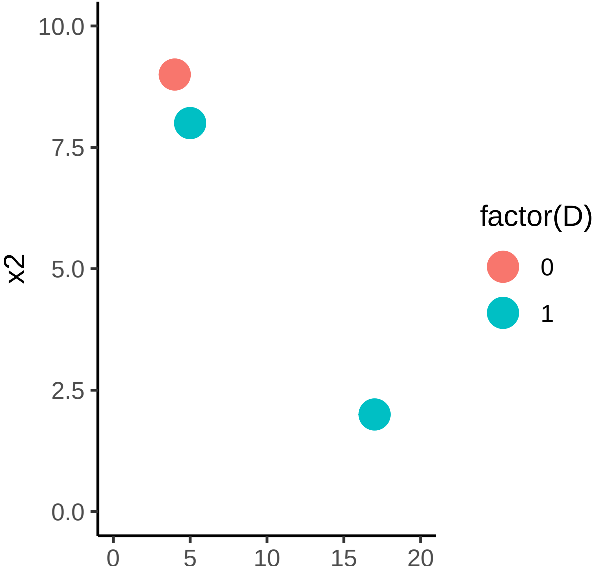
1. For each treated unit, find a control unit with very close values of all confounding variables, and keep both
2. Repeat for every treated unit
3. Drop all the unmatched units (eg. ’extra’ control units that are ’far away’ from any treated units)
4. Assess balance

Matching

1. For each treated unit, find a control unit with very close values of all confounding variables, and keep both
2. Repeat for every treated unit
3. Drop all the unmatched units (eg. ’extra’ control units that are ’far away’ from any treated units)
4. Assess balance
5. If balance is low, re-run the matching process as many times as you can to maximize balance!

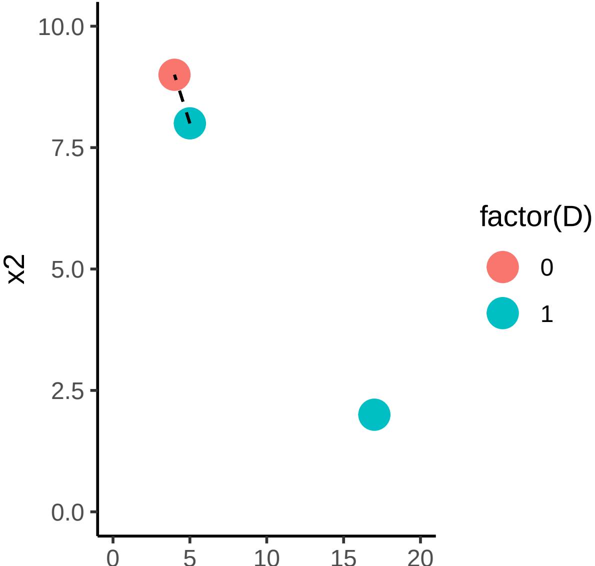
Matching

1. For example:



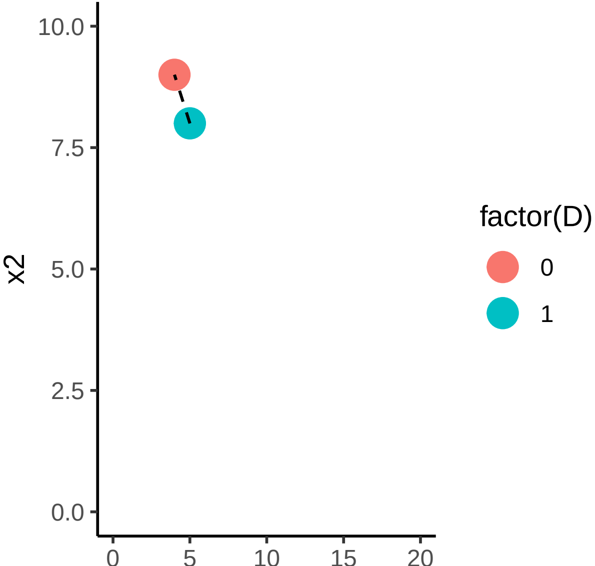
Matching

1. For example:



Matching

1. For example:



Matching

- Matching *always* produces a smaller dataset

Matching

- Matching *always* produces a smaller dataset
  - So there is a trade-off between improving balance and retaining a large sample

Matching

- Matching *always* produces a smaller dataset
  - So there is a trade-off between improving balance and retaining a large sample
- After matching, for the analysis we can either:
  1. Calculate the difference in means between treated and control groups

Matching

- Matching *always* produces a smaller dataset
  - So there is a trade-off between improving balance and retaining a large sample
- After matching, for the analysis we can either:
  1. Calculate the difference in means between treated and control groups
  2. Conduct the normal regression: $Y \sim D$
     - Option to include all our matching variables as controls
     - This will help control for any **residual imbalance** (esp. for continuous variables)

Matching

- Which variables to match on?

Matching

- Which variables to match on?
  - Treatment variable?

Matching

- Which variables to match on?
  - Treatment variable? **No!** We need treated and control units who are both male

Matching

- Which variables to match on?
  - Treatment variable? **No!** We need treated and control units who are both male
  - Outcome variable?

Matching

- Which variables to match on?
  - Treatment variable? **No!** We need treated and control units who are both male
  - Outcome variable? **No!** That's selecting on the dependent variable - biased!

Matching

- Which variables to match on?
  - Treatment variable? **No!** We need treated and control units who are both male
  - Outcome variable? **No!** That's selecting on the dependent variable - biased!
  - Post-treatment variables?

Matching

- Which variables to match on?
  - Treatment variable? **No!** We need treated and control units who are both male
  - Outcome variable? **No!** That's selecting on the dependent variable - biased!
  - Post-treatment variables? **No!** This will bias our causal effect, just as in regression

Matching

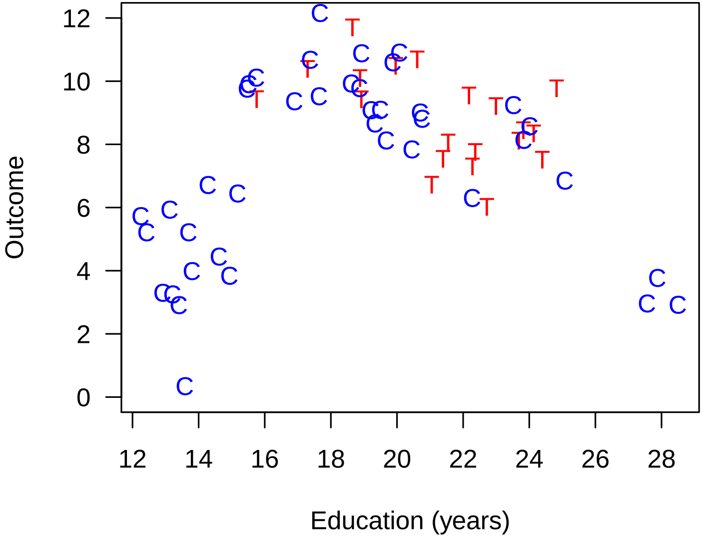
- Which variables to match on?
  - Treatment variable? **No!** We need treated and control units who are both male
  - Outcome variable? **No!** That's selecting on the dependent variable - biased!
  - Post-treatment variables? **No!** This will bias our causal effect, just as in regression
  - Pre-treatment Confounders?

Matching

- Which variables to match on?
  - Treatment variable? **No!** We need treated and control units who are both male
  - Outcome variable? **No!** That's selecting on the dependent variable - biased!
  - Post-treatment variables? **No!** This will bias our causal effect, just as in regression
  - Pre-treatment Confounders? **Yes!** We want to remove imbalance due to confounders

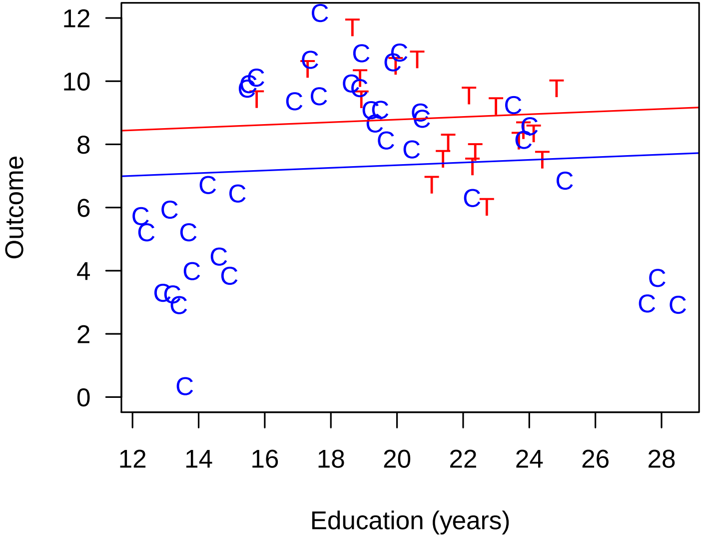
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



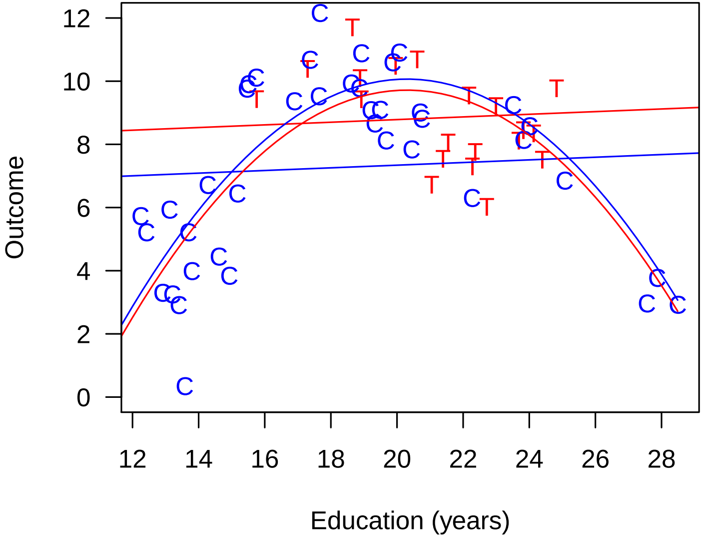
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



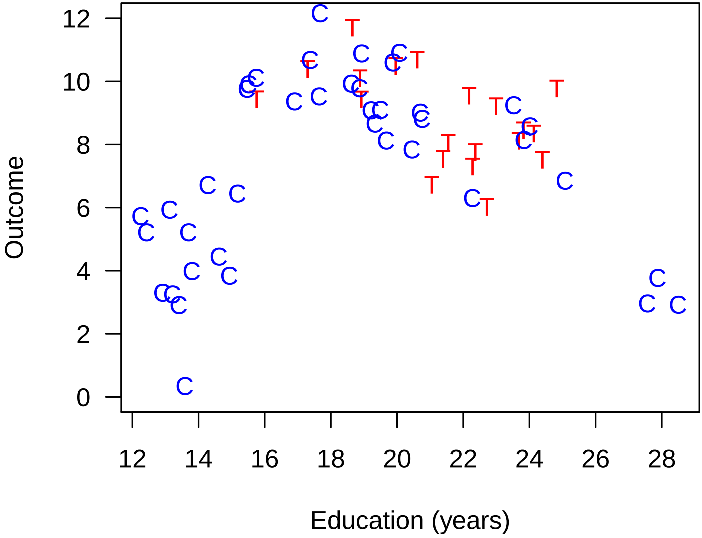
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



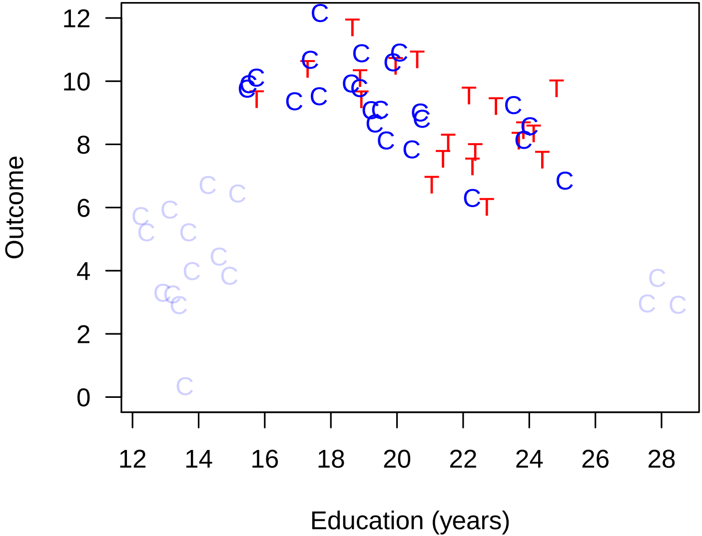
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



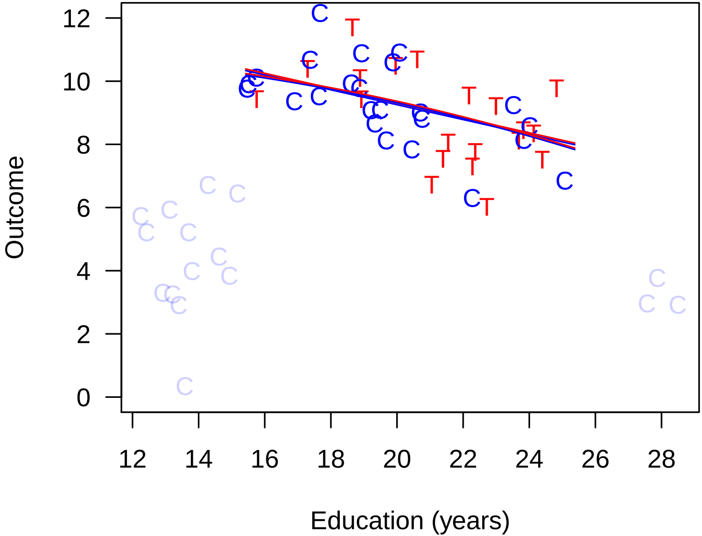
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching

- To identify 'close' matches we need some measure of distance between units' covariates

Matching

- To identify ’close’ matches we need some measure of distance between units’ covariates

1. Matching on few categorical variables: **Exact Matching**

Matching

- To identify 'close' matches we need some measure of distance between units' covariates

1. Matching on few categorical variables: **Exact Matching**
2. Matching on continuous variables (sequentially): **Nearest-Neighbour Matching**

Matching

- To identify ’close’ matches we need some measure of distance between units’ covariates

1. Matching on few categorical variables: **Exact Matching**
2. Matching on continuous variables (sequentially): **Nearest-Neighbour Matching**
3. Matching to maximize balance: **Optimal/Genetic Matching**

Matching

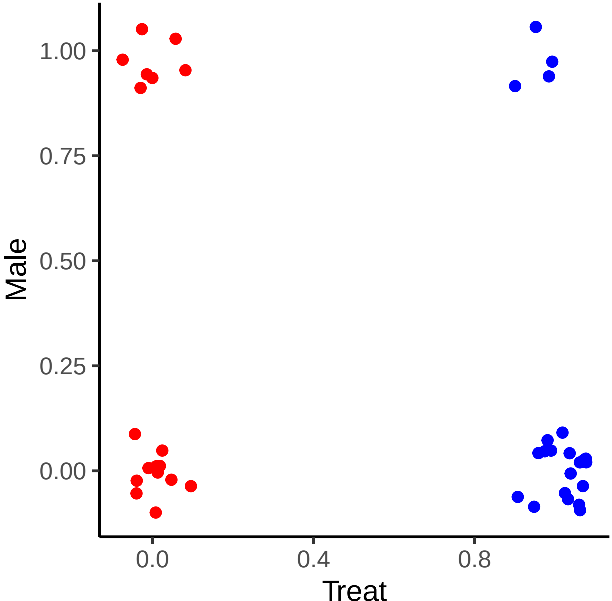
- To identify ’close’ matches we need some measure of distance between units’ covariates

1. Matching on few categorical variables: **Exact Matching**
2. Matching on continuous variables (sequentially): **Nearest-Neighbour Matching**
3. Matching to maximize balance: **Optimal/Genetic Matching**
4. Matching to balance the probability of treatment: **Propensity Score Matching**

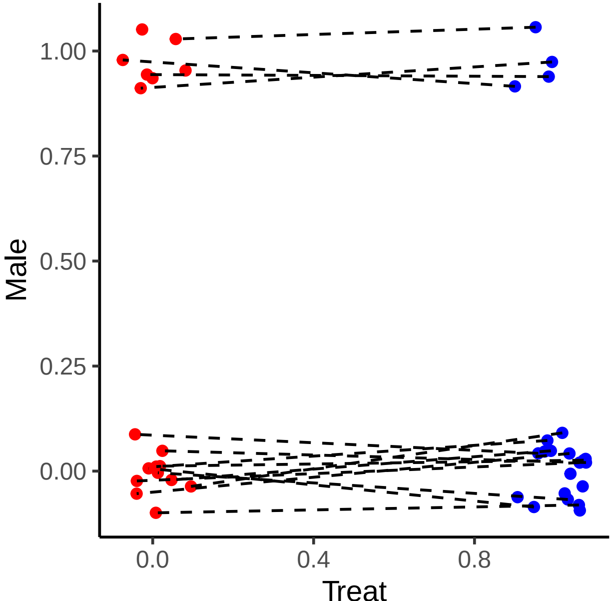
Section 2

Alternative Matching Methods

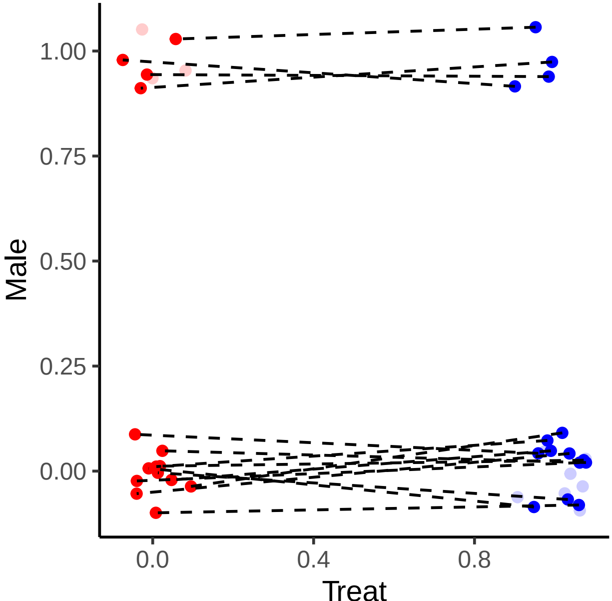
Exact Matching



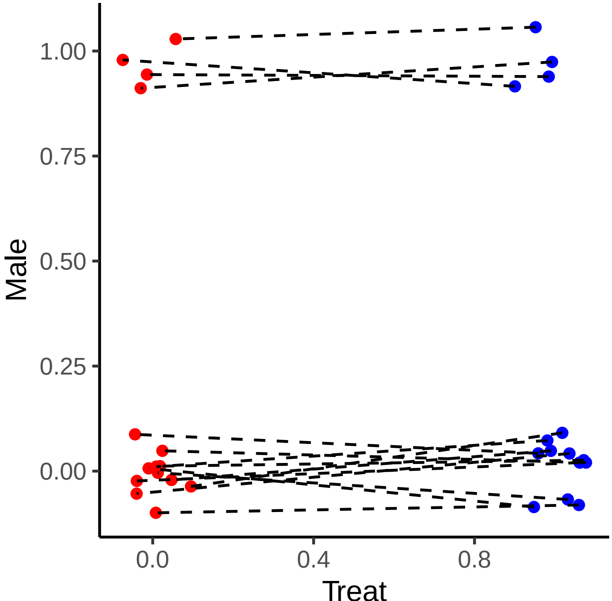
Exact Matching



Exact Matching



Exact Matching



Exact Matching

- Exact matching defines clear counterfactuals:
  - What is the difference in the outcome between treated and control units **for units of the same gender**

Exact Matching

- Exact matching defines clear counterfactuals:
  - What is the difference in the outcome between treated and control units **for units of the same gender**
- After matching, we **prune/remove** unmatched units

Exact Matching

- Exact matching defines clear counterfactuals:
  - What is the difference in the outcome between treated and control units **for units of the same gender**
- After matching, we **prune/remove** unmatched units
- **Then delete the link between the paired units, we don't need it any more**

Exact Matching

- Exact matching defines clear counterfactuals:
  - What is the difference in the outcome between treated and control units **for units of the same gender**
- After matching, we **prune/remove** unmatched units
- **Then delete the link between the paired units, we don't need it any more**
- Then compare the outcome of the **remaining** treated and control units

Exact Matching

- Exact matching defines clear counterfactuals:
  - What is the difference in the outcome between treated and control units **for units of the same gender**
- After matching, we **prune/remove** unmatched units
- **Then delete the link between the paired units, we don't need it any more**
- Then compare the outcome of the **remaining** treated and control units
  - Difference in means

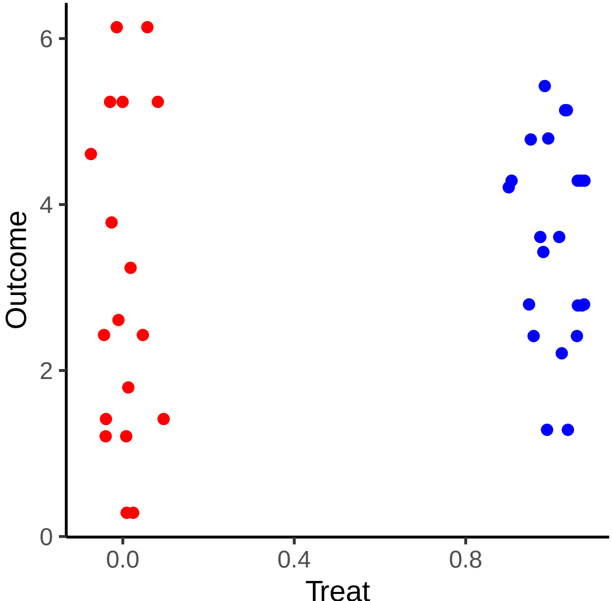
Exact Matching

- Exact matching defines clear counterfactuals:
  - What is the difference in the outcome between treated and control units **for units of the same gender**
- After matching, we **prune/remove** unmatched units
- **Then delete the link between the paired units, we don't need it any more**
- Then compare the outcome of the **remaining** treated and control units
  - Difference in means
  - Or regression of outcome on treatment

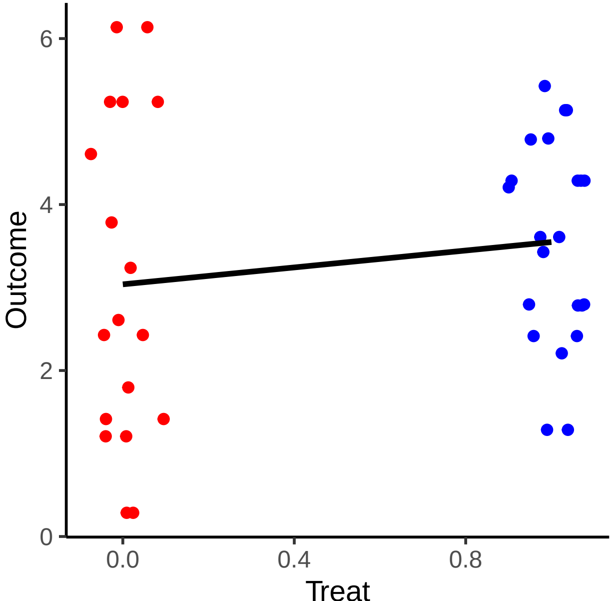
Exact Matching

	Units	Means Treated	Means Control	Mean Diff
1	All	0.18	0.39	-0.21
2	Matched	0.27	0.27	0.00

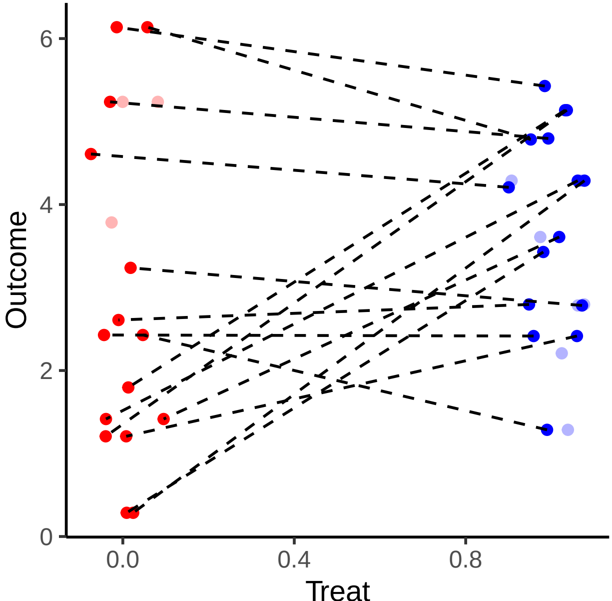
Exact Matching Analysis



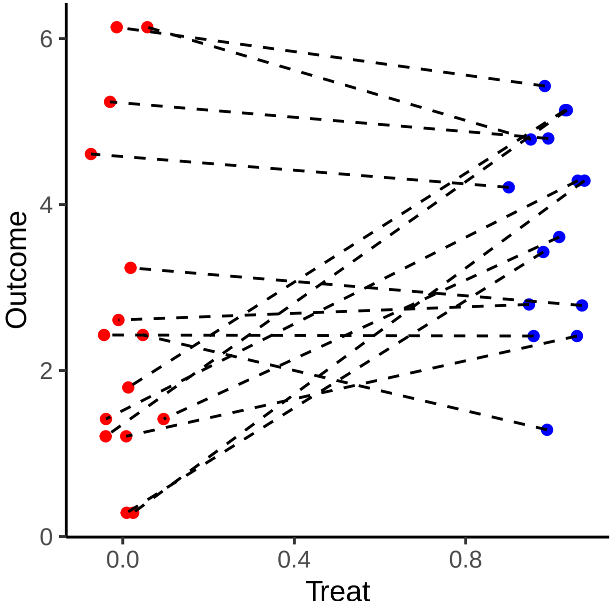
Exact Matching Analysis



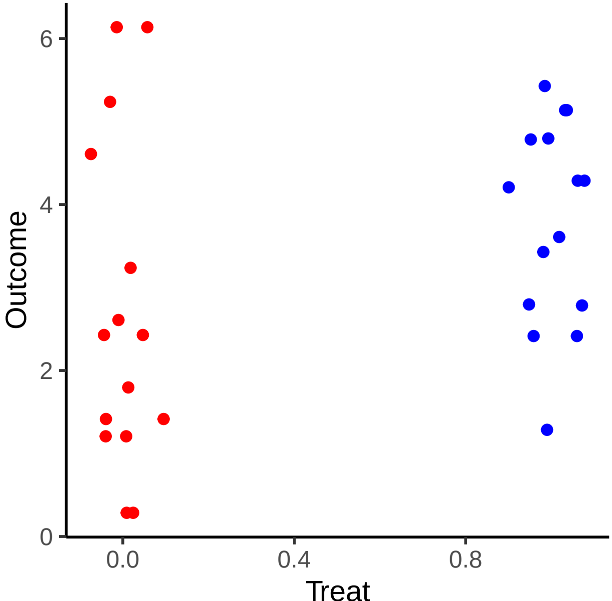
Exact Matching Analysis



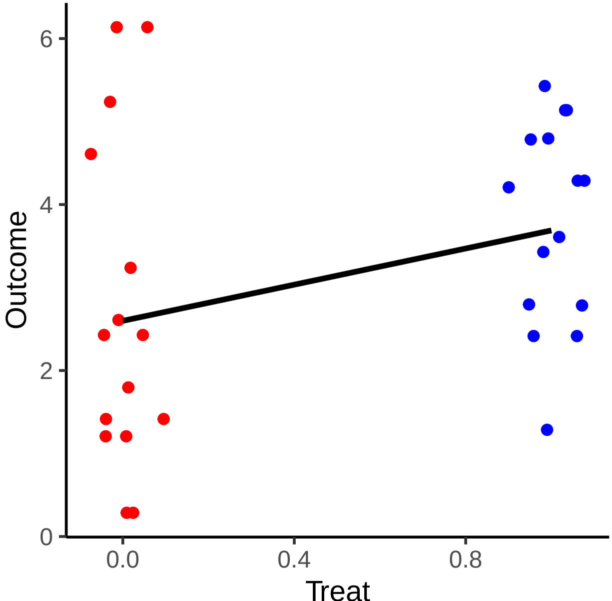
Exact Matching Analysis



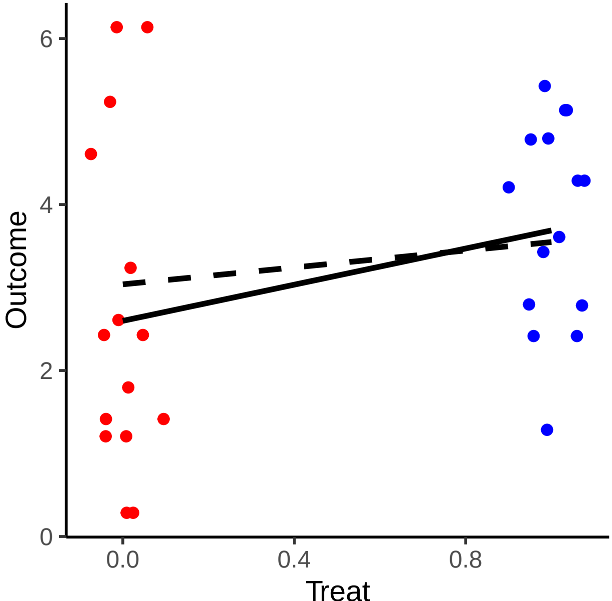
Exact Matching Analysis



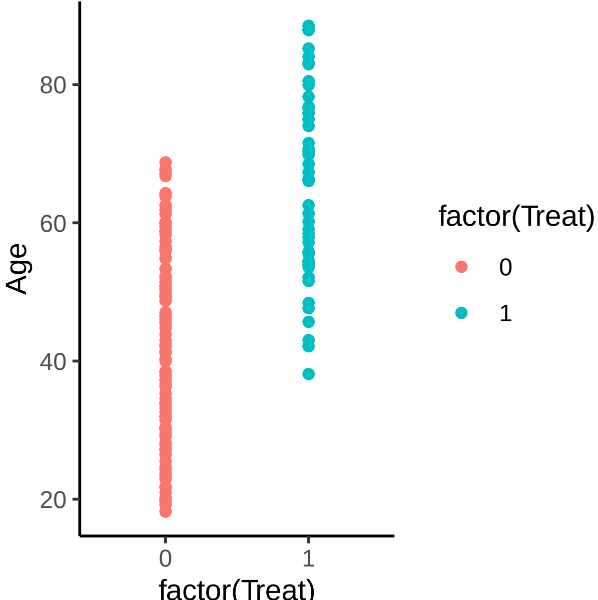
Exact Matching Analysis



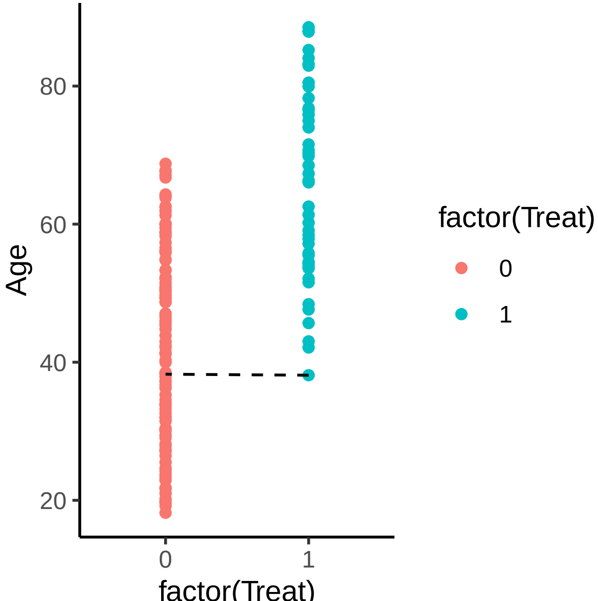
Exact Matching Analysis



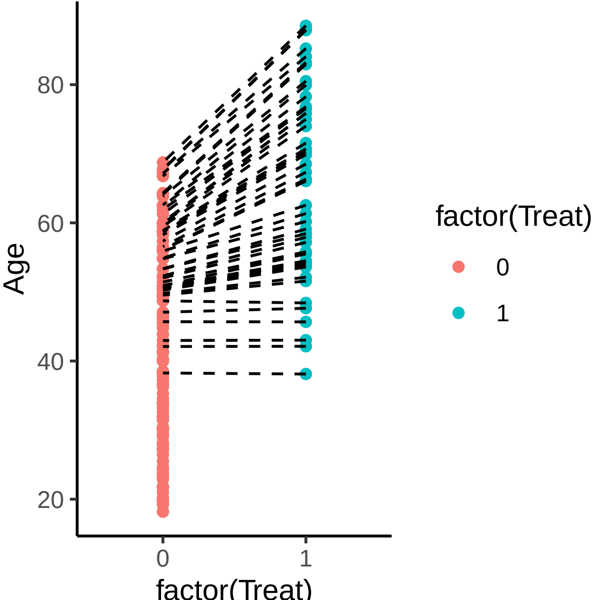
Nearest Neighbour Matching



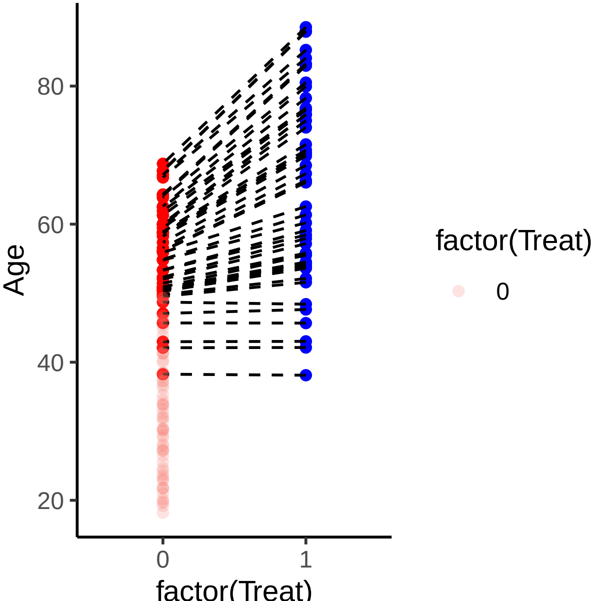
Nearest Neighbour Matching



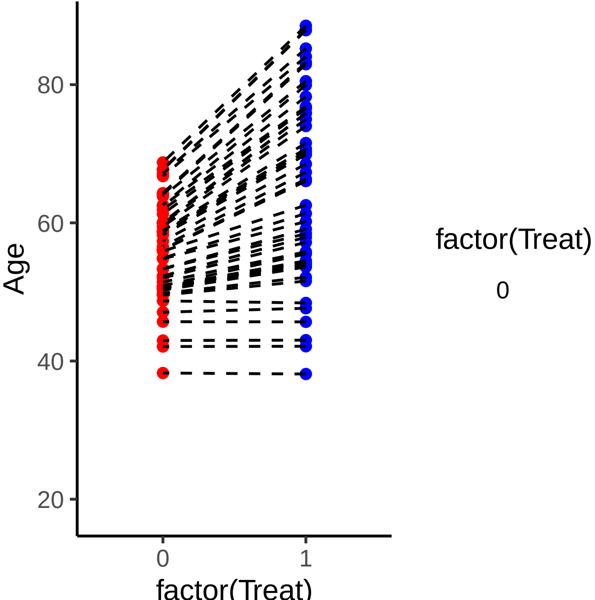
Nearest Neighbour Matching



Nearest Neighbour Matching



Nearest Neighbour Matching



Nearest Neighbour Matching

	Units	Means Treated	Means Control	Mean Diff
1	All	65.70	42.67	23.03
2	Matched	65.70	56.09	9.61

Nearest Neighbour Matching

- Two potential problems with nearest neighbour matching:

Nearest Neighbour Matching

- Two potential problems with nearest neighbour matching:
  1. **Nearest does not mean close:** The oldest treated units are matched with, but very different to, the oldest control units

Nearest Neighbour Matching

- Two potential problems with nearest neighbour matching:
  1. **Nearest does not mean close:** The oldest treated units are matched with, but very different to, the oldest control units
     - We need some **absolute** limits on the distance we can match units within

Nearest Neighbour Matching

- Two potential problems with nearest neighbour matching:
  1. **Nearest does not mean close:** The oldest treated units are matched with, but very different to, the oldest control units
     - We need some **absolute** limits on the distance we can match units within
     - We can add 'calipers' to matching to match only within a fixed range

Nearest Neighbour Matching

- Two potential problems with nearest neighbour matching:
  1. **Nearest does not mean close:** The oldest treated units are matched with, but very different to, the oldest control units
     - We need some **absolute** limits on the distance we can match units within
     - We can add 'calipers' to matching to match only within a fixed range
  2. **The order of matching matters:** The first matches use up units that might make better matches for later treated units

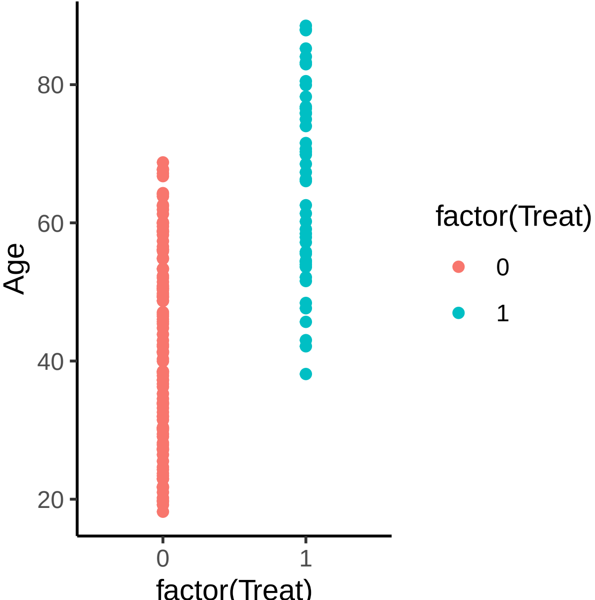
Nearest Neighbour Matching

- Two potential problems with nearest neighbour matching:
  1. **Nearest does not mean close:** The oldest treated units are matched with, but very different to, the oldest control units
     - We need some **absolute** limits on the distance we can match units within
     - We can add 'calipers' to matching to match only within a fixed range
  2. **The order of matching matters:** The first matches use up units that might make better matches for later treated units
     - To maximize balance we need to 'look ahead' and match in the right order

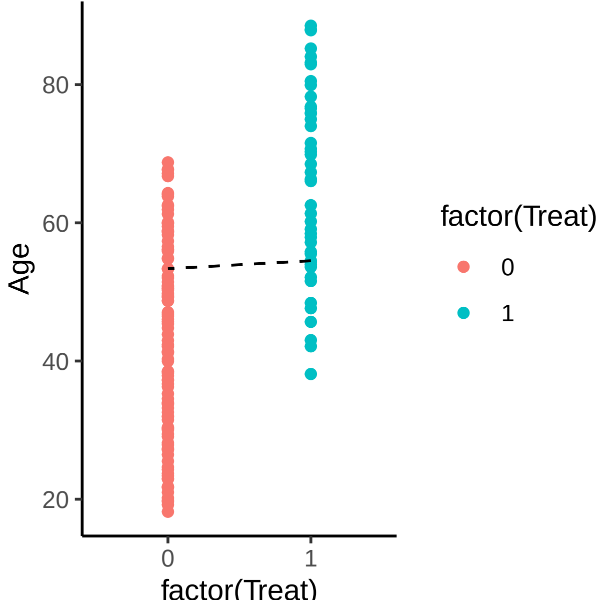
Nearest Neighbour Matching

- Two potential problems with nearest neighbour matching:
  1. **Nearest does not mean close:** The oldest treated units are matched with, but very different to, the oldest control units
     - We need some **absolute** limits on the distance we can match units within
     - We can add 'calipers' to matching to match only within a fixed range
  2. **The order of matching matters:** The first matches use up units that might make better matches for later treated units
     - To maximize balance we need to 'look ahead' and match in the right order
     - For this we can use optimal or genetic matching, which is fully automated

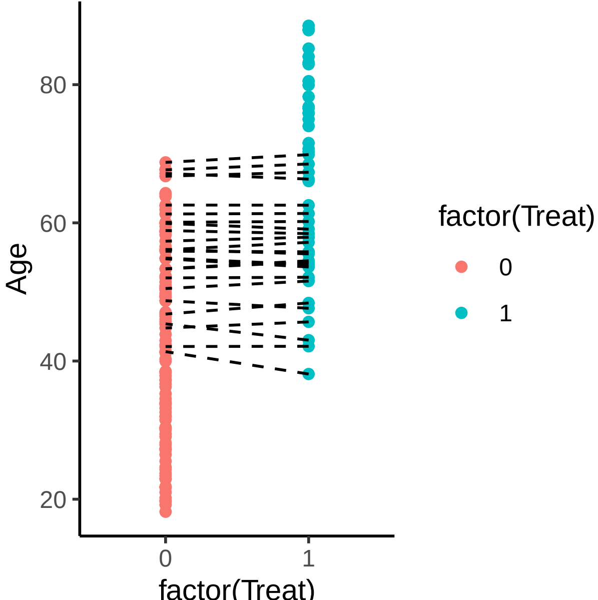
Nearest Neighbour Matching with Caliper



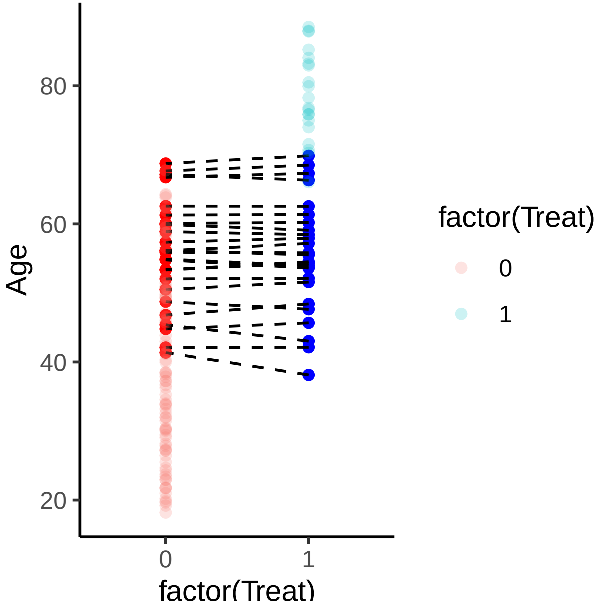
Nearest Neighbour Matching with Caliper



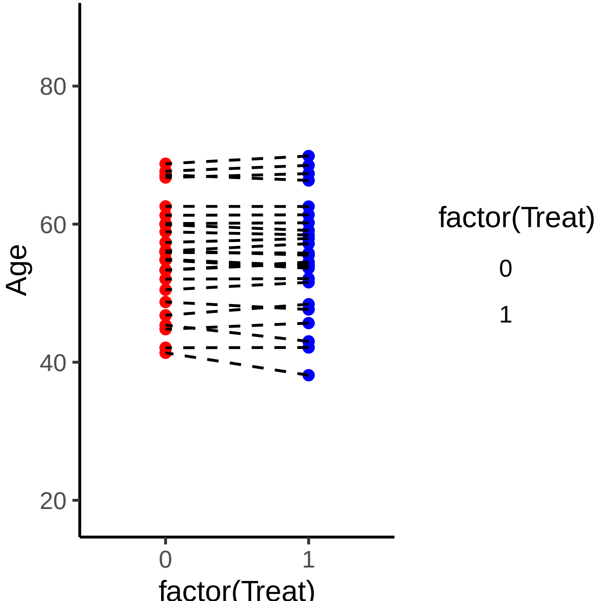
Nearest Neighbour Matching with Caliper



Nearest Neighbour Matching with Caliper



Nearest Neighbour Matching with Caliper



Nearest Neighbour Matching with Caliper

	Units	Means Treated	Means Control	Mean Diff
1	All	65.70	42.67	23.03
2	Matched	55.41	55.46	-0.06

- Note: p-values don't mean so much for balance tests

Nearest Neighbour Matching with Caliper

	Units	Means Treated	Means Control	Mean Diff
1	All	65.70	42.67	23.03
2	Matched	55.41	55.46	-0.06

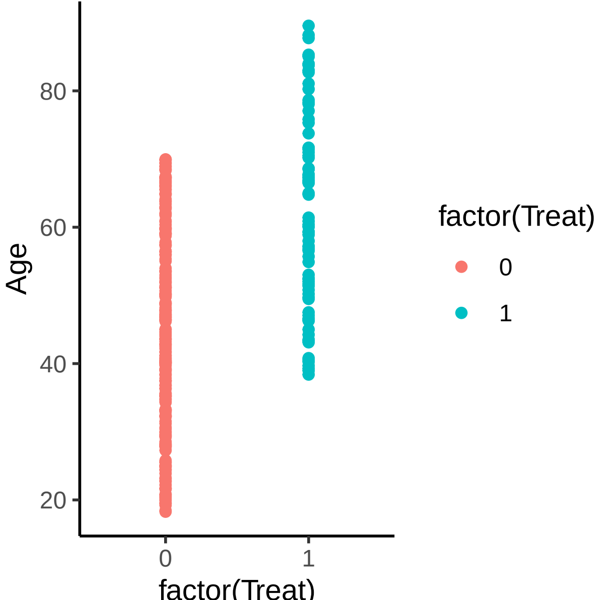
- Note: p-values don't mean so much for balance tests
- We always want to improve balance as much as possible

Nearest Neighbour Matching with Caliper

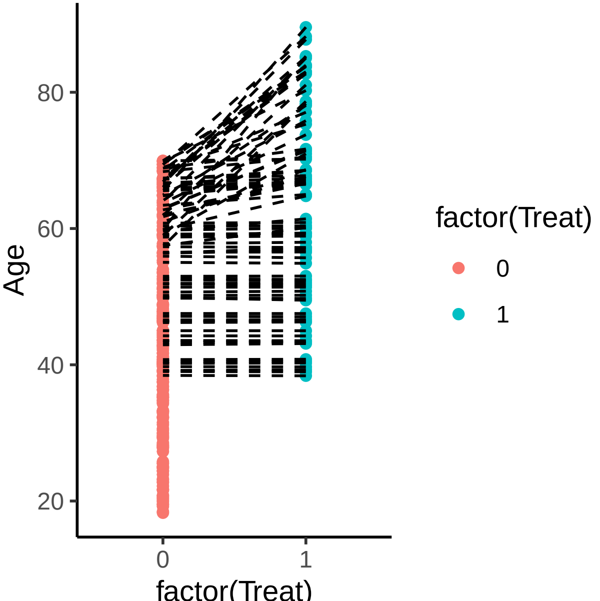
	Units	Means Treated	Means Control	Mean Diff
1	All	65.70	42.67	23.03
2	Matched	55.41	55.46	-0.06

- Note: p-values don't mean so much for balance tests
- We always want to improve balance as much as possible
- Better to compare (standardized) difference in means

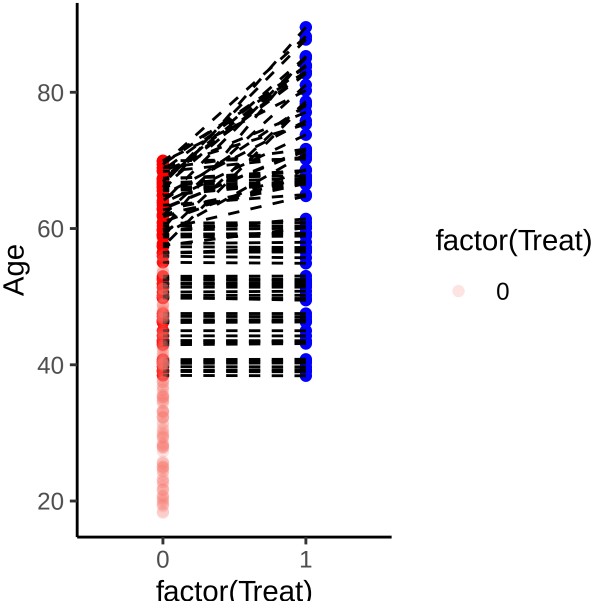
Optimal Matching



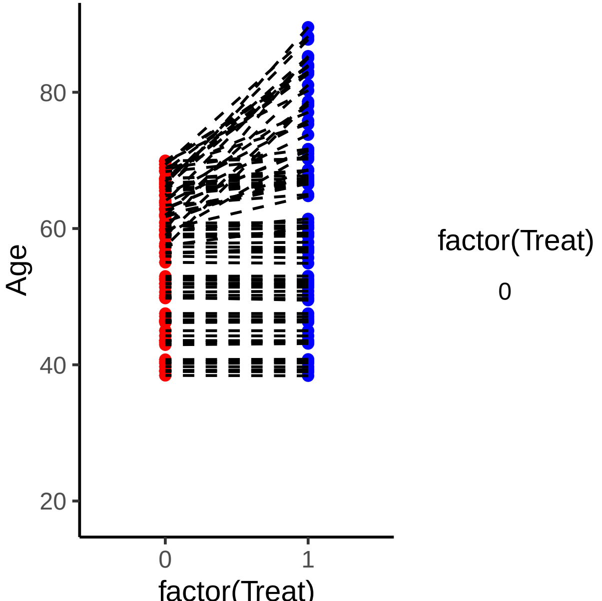
Optimal Matching



Optimal Matching



Optimal Matching



Optimal Matching

	Units	Means Treated	Means Control	Mean Diff
1	All	62.60	44.64	17.96
2	Matched	62.60	57.57	5.03

Propensity Score Matching

- With many covariates we have a ’dimensionality’ challenge

Propensity Score Matching

- With many covariates we have a ’dimensionality’ challenge
  - Overlap is almost zero

Propensity Score Matching

- With many covariates we have a ’dimensionality’ challenge
  - Overlap is almost zero
  - Counterfactuals are impossible to define
- The propensity score simplifies matching to a single dimension

Propensity Score Matching

- With many covariates we have a ’dimensionality’ challenge
  - Overlap is almost zero
  - Counterfactuals are impossible to define
- The propensity score simplifies matching to a single dimension
  - Confounders only matter to the extent they affect the probability of treatment

Propensity Score Matching

- With many covariates we have a ’dimensionality’ challenge
  - Overlap is almost zero
  - Counterfactuals are impossible to define
- The propensity score simplifies matching to a single dimension
  - Confounders only matter to the extent they affect the probability of treatment
  - So let’s use the confounders to **predict treatment**

Propensity Score Matching

- With many covariates we have a 'dimensionality' challenge
  - Overlap is almost zero
  - Counterfactuals are impossible to define
- The propensity score simplifies matching to a single dimension
  - Confounders only matter to the extent they affect the probability of treatment
  - So let's use the confounders to **predict treatment**
  - That's different to actual treatment status, with the remainder due to 'random' factors (if we include all confounders)

Propensity Score Matching

- With many covariates we have a ’dimensionality’ challenge
  - Overlap is almost zero
  - Counterfactuals are impossible to define
- The propensity score simplifies matching to a single dimension
  - Confounders only matter to the extent they affect the probability of treatment
  - So let’s use the confounders to **predict treatment**
  - That’s different to actual treatment status, with the remainder due to ’random’ factors (if we include all confounders)
- Then use the propensity score (probability 0-1) to match treated and control units which have the same ex ante probability of treatment

Propensity Score Matching

- But some concerns about drawbacks of propensity score matching

Propensity Score Matching

- But some concerns about drawbacks of propensity score matching
- May have poor balance on individual confounders

Propensity Score Matching

- But some concerns about drawbacks of propensity score matching
- May have poor balance on individual confounders
- Balance may get worse as we remove more units

Propensity Score Matching

- But some concerns about drawbacks of propensity score matching
- May have poor balance on individual confounders
- Balance may get worse as we remove more units
- We have to get the functional form of the treatment explanation right (linear, quadratic etc.) so we remain vulnerable to model dependence!

Propensity Score Matching

- Treatment: 1/0
- Confounder: Age
- Logit model predicting treatment:

$$Treat_i = \alpha + \beta Age_i + \epsilon_i$$

Propensity Score Matching

- Treatment: 1/0
- Confounder: Age
- Logit model predicting treatment:

$$Treat_i = \alpha + \beta Age_i + \epsilon_i$$

$$Predicted\_Treat_i = -7.19 + 0.116 Age_i + \epsilon_i$$

Propensity Score Matching

- Treatment: 1/0
- Confounder: Age
- Logit model predicting treatment:

$$Treat_i = \alpha + \beta Age_i + \epsilon_i$$

$$Predicted\_Treat_i = -7.19 + 0.116 Age_i + \epsilon_i$$

- Match on the values of $Predicted\_Treat_i$ (fitted values of the regression)

Propensity Score Matching

- Treatment: 1/0
- Confounder: Age
- Logit model predicting treatment:

$$Treat_i = \alpha + \beta Age_i + \epsilon_i$$

$$Predicted\_Treat_i = -7.19 + 0.116 Age_i + \epsilon_i$$

- Match on the values of $Predicted\_Treat_i$ (fitted values of the regression)
- I.e. match units with a similar *probability* of treatment

Propensity Score Matching

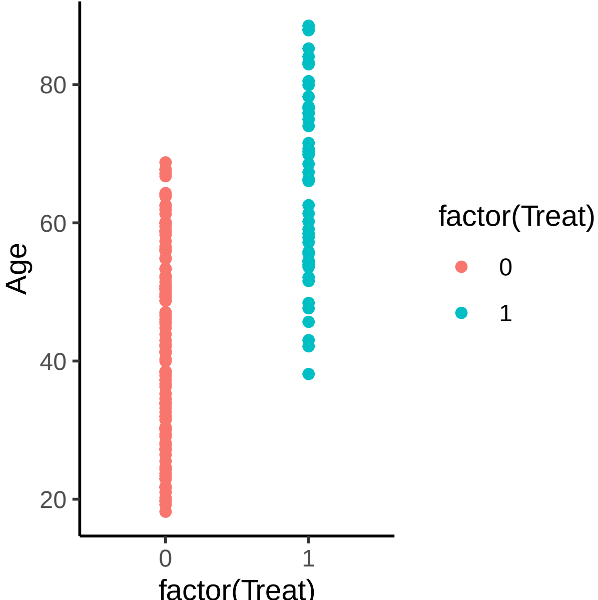
- Treatment: 1/0
- Confounder: Age
- Logit model predicting treatment:

$$Treat_i = \alpha + \beta Age_i + \epsilon_i$$

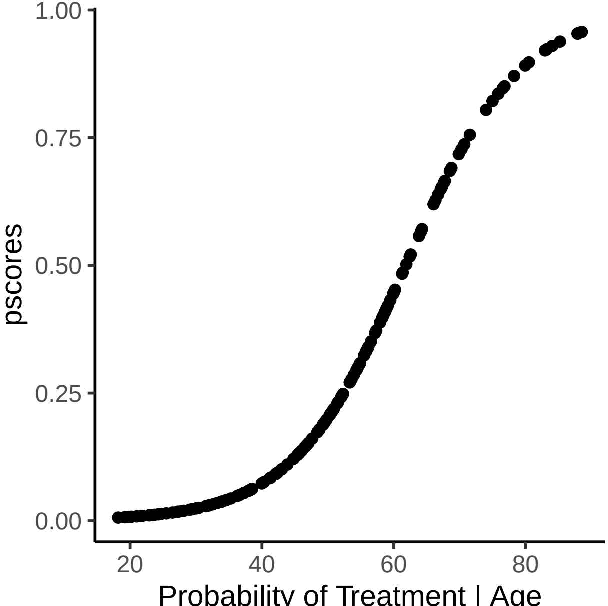
$$Predicted\_Treat_i = -7.19 + 0.116 Age_i + \epsilon_i$$

- Match on the values of $Predicted\_Treat_i$ (fitted values of the regression)
- I.e. match units with a similar *probability* of treatment
- ...Regardless of whether they actually get treated

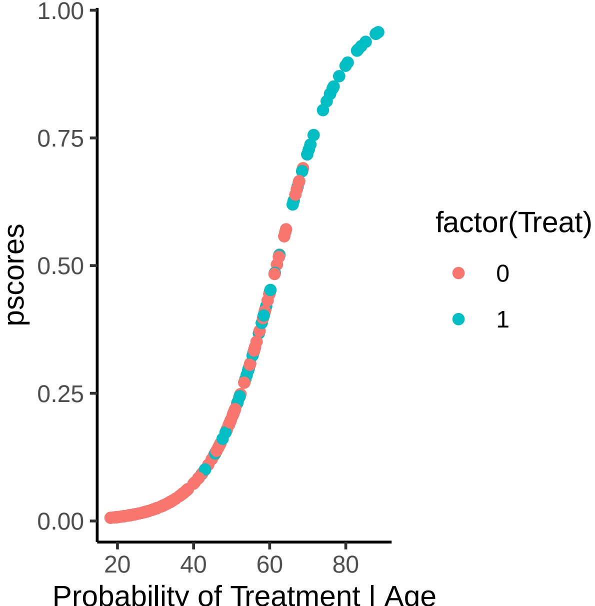
Propensity Score Matching



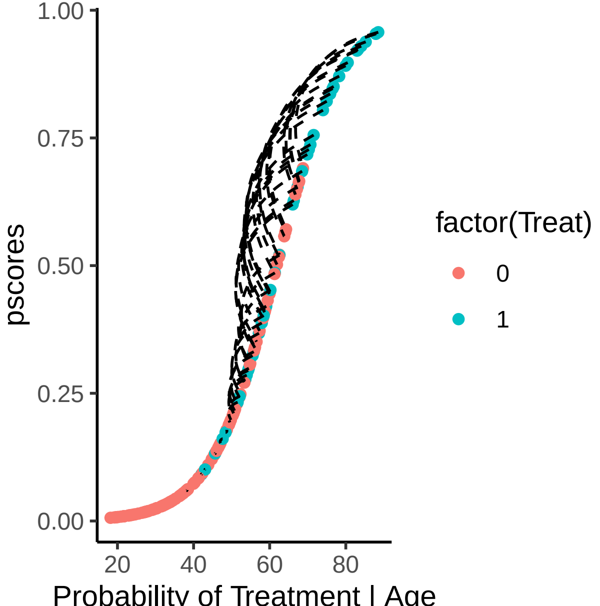
Propensity Score Matching



Propensity Score Matching



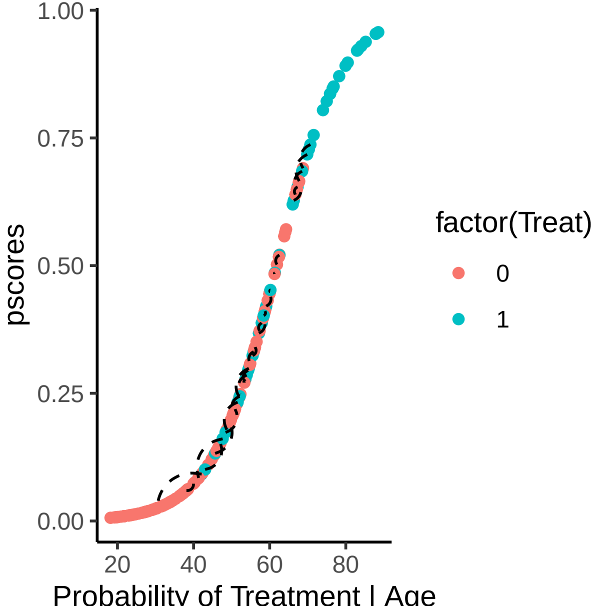
Propensity Score Matching



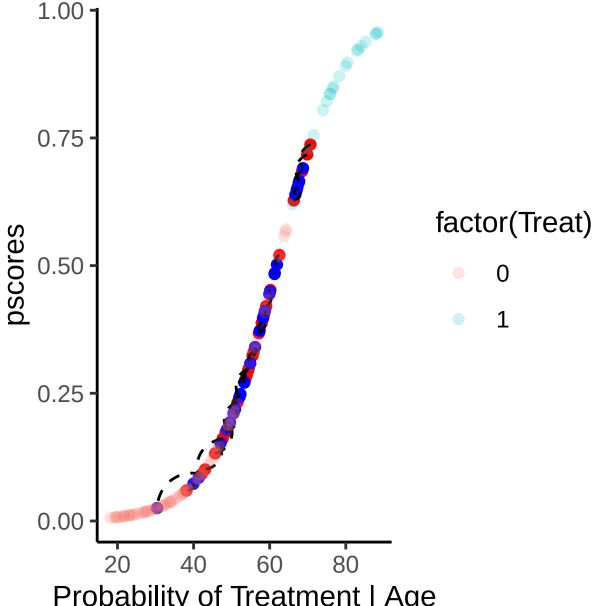
Propensity Score Matching

	Units	Means Treated	Means Control	Mean Diff
1	All	0.57	0.18	0.39
2	Matched	0.57	0.36	0.21

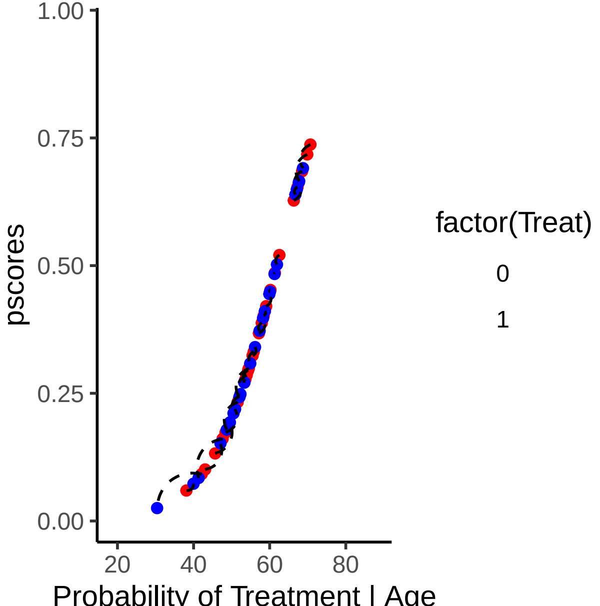
Propensity Score Matching with Caliper



Propensity Score Matching with Caliper



Propensity Score Matching with Caliper



Propensity Score Matching with Caliper

	Units	Means Treated	Means Control	Mean Diff
1	All	0.57	0.18	0.39
2	Matched	0.36	0.35	0.01

Matching

- Matching was supposed to be 'non-parametric' to reduce researcher influence, but there are a lot of options here!

Matching

- Matching was supposed to be 'non-parametric' to reduce researcher influence, but there are a lot of options here!
- That's okay! Regression can be biased if we try to make a p-value significant, but with matching we always want more balance

Matching

- Matching was supposed to be 'non-parametric' to reduce researcher influence, but there are a lot of options here!
- That's okay! Regression can be biased if we try to make a p-value significant, but with matching we always want more balance
  - As long as we do matching **without** looking at the outcome variables

Matching

- Matching was supposed to be 'non-parametric' to reduce researcher influence, but there are a lot of options here!
- That's okay! Regression can be biased if we try to make a p-value significant, but with matching we always want more balance
  - As long as we do matching **without** looking at the outcome variables
- How much trimming/pruning should we undertake?

Matching

- Matching was supposed to be ’non-parametric’ to reduce researcher influence, but there are a lot of options here!
- That’s okay! Regression can be biased if we try to make a p-value significant, but with matching we always want more balance
  - As long as we do matching **without** looking at the outcome variables
- How much trimming/pruning should we undertake?
- We can always enforce **stricter** matching (eg. narrower calipers, more exact matching) to get better balance

Matching

- Matching was supposed to be 'non-parametric' to reduce researcher influence, but there are a lot of options here!
- That's okay! Regression can be biased if we try to make a p-value significant, but with matching we always want more balance
  - As long as we do matching **without** looking at the outcome variables
- How much trimming/pruning should we undertake?
- We can always enforce **stricter** matching (eg. narrower calipers, more exact matching) to get better balance
  - But our N will approach zero, so little statistical power

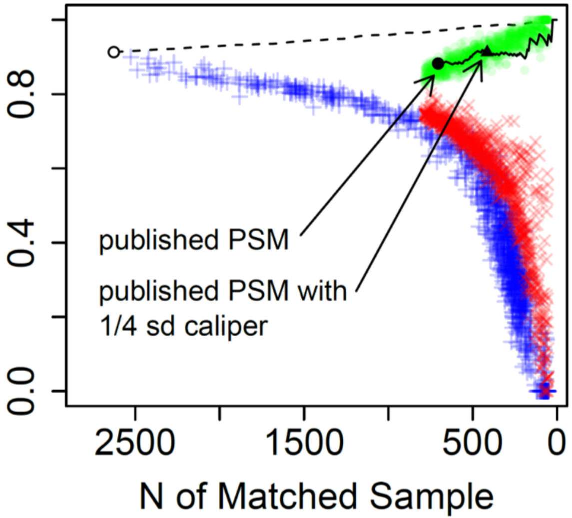
Matching

- Matching was supposed to be 'non-parametric' to reduce researcher influence, but there are a lot of options here!
- That's okay! Regression can be biased if we try to make a p-value significant, but with matching we always want more balance
  - As long as we do matching **without** looking at the outcome variables
- How much trimming/pruning should we undertake?
- We can always enforce **stricter** matching (eg. narrower calipers, more exact matching) to get better balance
  - But our N will approach zero, so little statistical power
  - A Bias-variance trade-off

Matching

- Matching was supposed to be 'non-parametric' to reduce researcher influence, but there are a lot of options here!
- That's okay! Regression can be biased if we try to make a p-value significant, but with matching we always want more balance
  - As long as we do matching **without** looking at the outcome variables
- How much trimming/pruning should we undertake?
- We can always enforce **stricter** matching (eg. narrower calipers, more exact matching) to get better balance
  - But our N will approach zero, so little statistical power
  - A Bias-variance trade-off
  - Try alternatives

Matching



0.0
0.4
0.8
published PSM
published PSM with
1/4 sd caliper
2500
1500
500
0
N of Matched Sample

Matching

- Matching preferred to regression where:

Matching

- Matching preferred to regression where:
  - Never! Do both!
- Matching makes a big contribution where there's poor overlap

Matching

- Matching preferred to regression where:
  - Never! Do both!
- Matching makes a big contribution where there's poor overlap
- Matching + Regression = "Doubly Robust"

Matching

- Matching preferred to regression where:
  - Never! Do both!
- Matching makes a big contribution where there's poor overlap
- Matching + Regression = "Doubly Robust"
  - If **either** matching produces balance **OR** we have the correct functional form for regression, we can make causal inference

Section 3

Matching vs. Experiments

Matching

- Arceneaux, Gerber and Green (2005)

Matching

- Arceneaux, Gerber and Green (2005)
- How does matching work on experimental (IV) data? (eg. for how to get voters to vote)

Matching

- Arceneaux, Gerber and Green (2005)
- How does matching work on experimental (IV) data? (eg. for how to get voters to vote)
- Matching is biased compared to the experimental results

Matching

- Arceneaux, Gerber and Green (2005)
- How does matching work on experimental (IV) data? (eg. for how to get voters to vote)
- Matching is biased compared to the experimental results
- Lots of controls

Matching

- Arceneaux, Gerber and Green (2005)
- How does matching work on experimental (IV) data? (eg. for how to get voters to vote)
- Matching is biased compared to the experimental results
- Lots of controls
- But unobserved confounders mean matching can't recover causal estimates

Matching

- Bias was due to whether people actually answered phone calls

Matching

- Bias was due to whether people actually answered phone calls
- Huge N, **Perfect balance**

Matching

- Bias was due to whether people actually answered phone calls
- Huge N, **Perfect balance** (on what they could measure)

Matching

- Bias was due to whether people actually answered phone calls
- Huge N, **Perfect balance** (on what they could measure)
- Experimental measure: 0.4

Matching

- Bias was due to whether people actually answered phone calls
- Huge N, **Perfect balance** (on what they could measure)
- Experimental measure: 0.4
- OLS estimate: 2.7

Matching

- Bias was due to whether people actually answered phone calls
- Huge N, **Perfect balance** (on what they could measure)
- Experimental measure: 0.4
- OLS estimate: 2.7
- Matching estimate: 2.8

Matching

- Bias was due to whether people actually answered phone calls
- Huge N, **Perfect balance** (on what they could measure)
- Experimental measure: 0.4
- OLS estimate: 2.7
- Matching estimate: 2.8
- We can’t control for likelihood of answering the phone using the (many) covariates they have

Matching

- Bias was due to whether people actually answered phone calls
- Huge N, **Perfect balance** (on what they could measure)
- Experimental measure: 0.4
- OLS estimate: 2.7
- Matching estimate: 2.8
- We can't control for likelihood of answering the phone using the (many) covariates they have
- Matching still relies on **measuring all confounders**