

# Simplifying Matching Methods for Causal Inference

Gary King<sup>1</sup>

Institute for Quantitative Social Science  
Harvard University

Hebrew University of Jerusalem, 1/1/2020

---

<sup>1</sup>GaryKing.org

## 3 Problems, 3 Solutions

## 3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 179,000 articles!) sounds magical:

## 3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 179,000 articles!) sounds magical:
  - ~→ “Why Propensity Scores Should Not Be Used for Matching” (*PA*, 2019. Gary King, Richard Nielsen)

## 3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 179,000 articles!) sounds magical:
  - ~→ “Why Propensity Scores Should Not Be Used for Matching” (*PA*, 2019. Gary King, Richard Nielsen)
2. Do powerful methods have to be complicated?

## 3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 179,000 articles!) sounds magical:
  - ~> “Why Propensity Scores Should Not Be Used for Matching” (PA, 2019. Gary King, Richard Nielsen)
2. Do powerful methods have to be complicated?
  - ~> “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)

## 3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 179,000 articles!) sounds magical:
  - ~> “Why Propensity Scores Should Not Be Used for Matching” (PA, 2019. Gary King, Richard Nielsen)
2. Do powerful methods have to be complicated?
  - ~> “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)
3. Matching methods optimize either imbalance ( $\approx$  bias) or # units pruned ( $\approx$  variance); users need both simultaneously:

### 3 Problems, 3 Solutions

1. The most popular method (propensity score matching, used in 179,000 articles!) sounds magical:
  - ~> “Why Propensity Scores Should Not Be Used for Matching” (PA, 2019. Gary King, Richard Nielsen)
2. Do powerful methods have to be complicated?
  - ~> “Causal Inference Without Balance Checking: Coarsened Exact Matching” (PA, 2011. Stefano Iacus, Gary King, and Giuseppe Porro)
3. Matching methods optimize either imbalance ( $\approx$  bias) or # units pruned ( $\approx$  variance); users need both simultaneously:
  - ~> “The Balance-Sample Size Frontier in Matching Methods for Causal Inference” (In press, *AJPS*; Gary King, Christopher Lucas and Richard Nielsen)



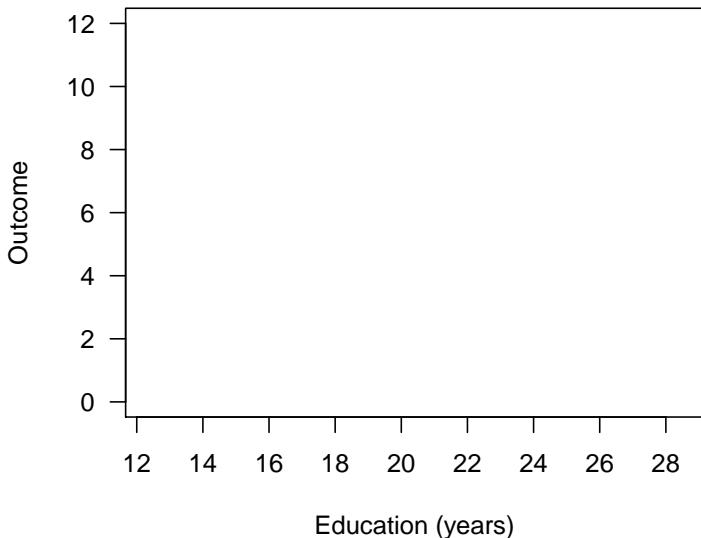
# Matching to Reduce Model Dependence

# Matching to Reduce Model Dependence

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

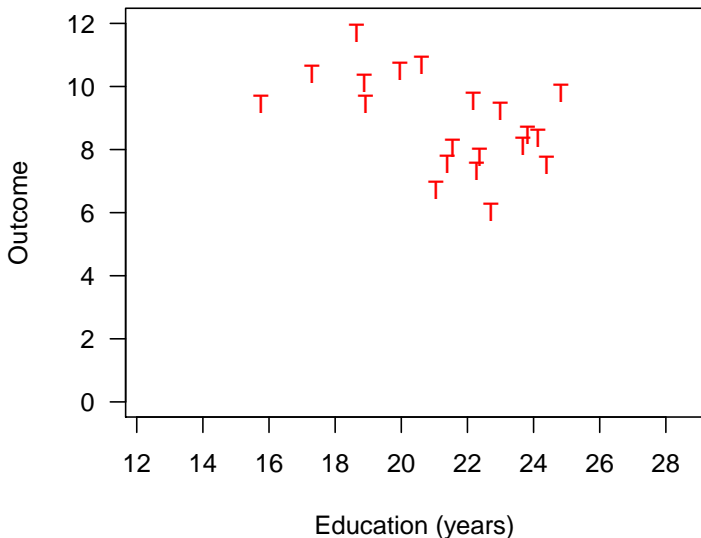
# Matching to Reduce Model Dependence

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



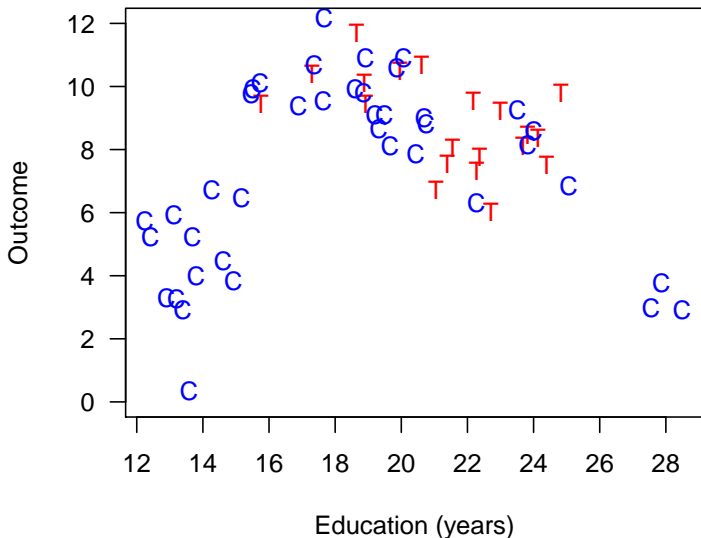
# Matching to Reduce Model Dependence

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



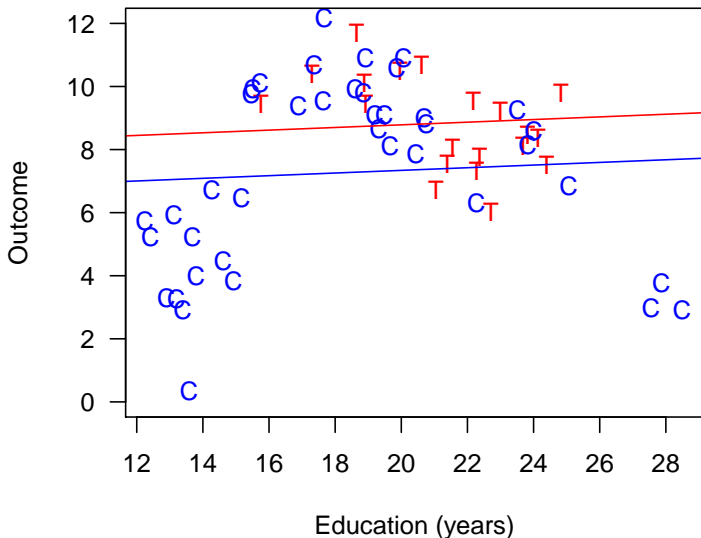
# Matching to Reduce Model Dependence

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



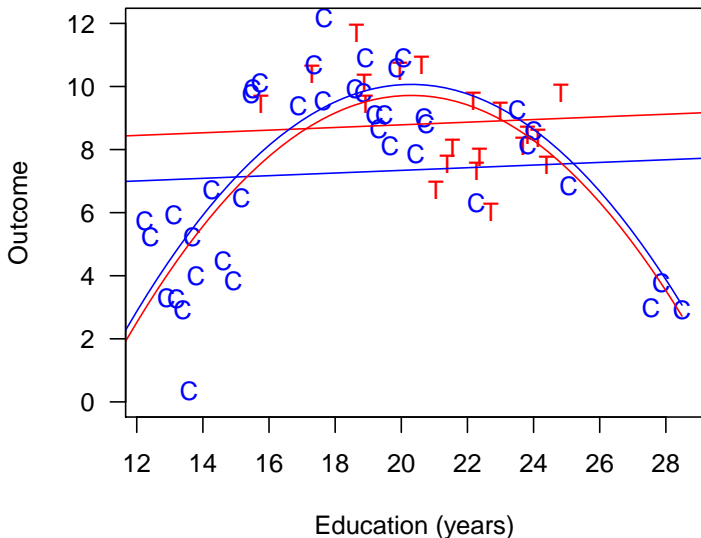
# Matching to Reduce Model Dependence

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



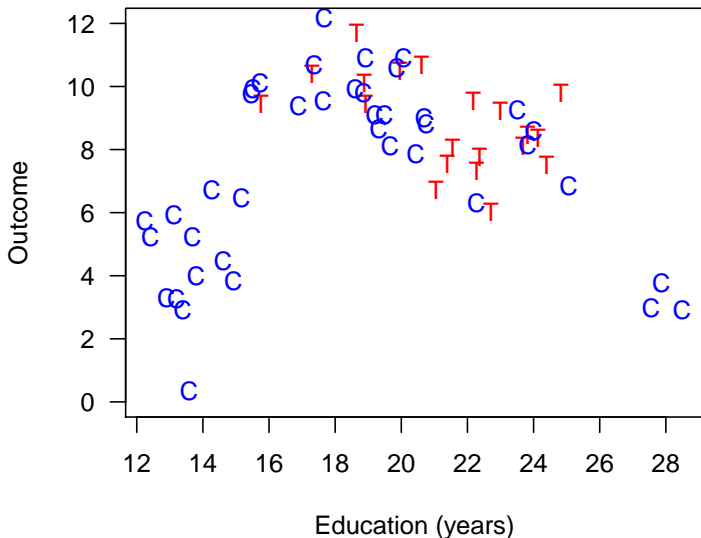
# Matching to Reduce Model Dependence

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



# Matching to Reduce Model Dependence

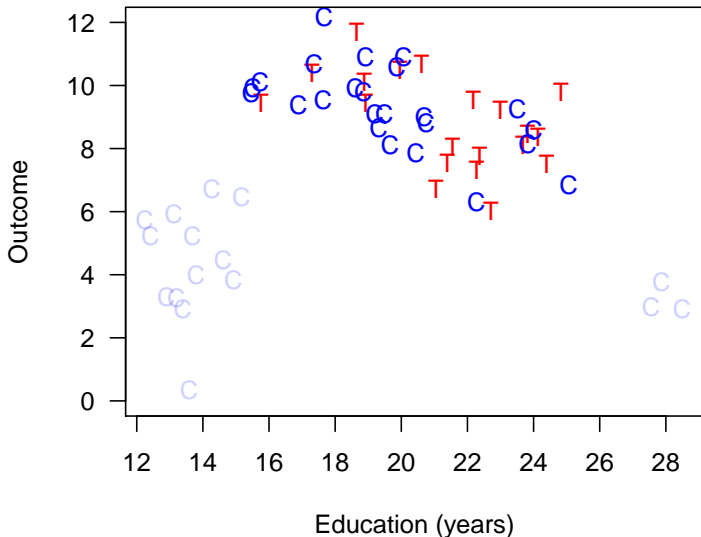
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)





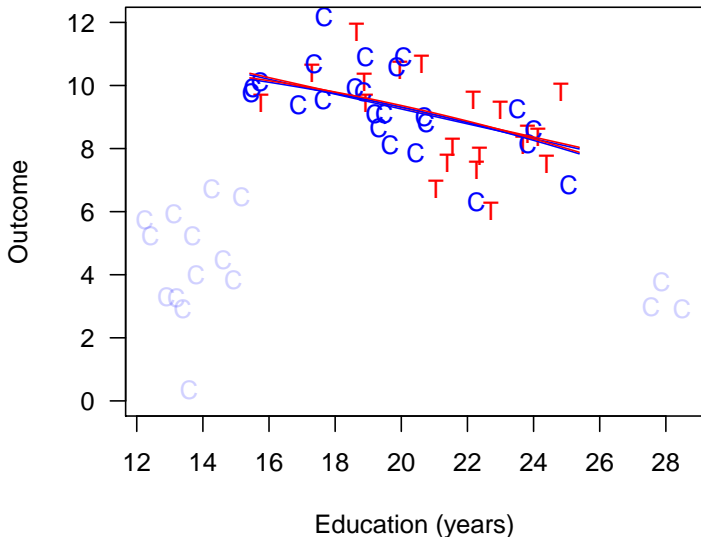
# Matching to Reduce Model Dependence

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



# Matching to Reduce Model Dependence

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



# The Problems Matching Solves

# The Problems Matching Solves

Without Matching:

# The Problems Matching Solves

Without Matching:

Imbalance

# The Problems Matching Solves

Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence

# The Problems Matching Solves

Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion

# The Problems Matching Solves

Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias



# The Problems Matching Solves

## Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias

- Qualitative choice from unbiased estimates = biased estimator

# The Problems Matching Solves

## Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments

# The Problems Matching Solves

## Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse<sub>[effrt]</sub>

# The Problems Matching Solves

## Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse<sub>[effrt]</sub>
- conscientious effort doesn't avoid biases (Banaji 2013)<sub>[acc]</sub>

# The Problems Matching Solves

## Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse<sub>[effrt]</sub>
- conscientious effort doesn't avoid biases (Banaji 2013)<sub>[acc]</sub>
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)<sub>[exprt]</sub>

# The Problems Matching Solves

## Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse<sub>[effrt]</sub>
- conscientious effort doesn't avoid biases (Banaji 2013)<sub>[acc]</sub>
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)<sub>[exprt]</sub>
- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)<sub>[tch]</sub>

# The Problems Matching Solves

## Without Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse<sub>[effrt]</sub>
- conscientious effort doesn't avoid biases (Banaji 2013)<sub>[acc]</sub>
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)<sub>[exppt]</sub>
- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)<sub>[tch]</sub>
- “Teaching psychology is mostly a waste of time” (Kahneman 2011)

# The Problems Matching Solves

~~Without~~ Matching:

Imbalance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias



# The Problems Matching Solves

~~Without~~ Matching:

~~Im~~balance  $\rightsquigarrow$  Model Dependence  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias

# The Problems Matching Solves

~~Without~~ Matching:

~~Im~~balance  $\rightsquigarrow$  ~~Model Dependence~~  $\rightsquigarrow$  Researcher discretion  $\rightsquigarrow$  Bias

# The Problems Matching Solves

Without Matching:

~~Imbalance~~  $\rightsquigarrow$  ~~Model Dependence~~  $\rightsquigarrow$  ~~Researcher discretion~~  $\rightsquigarrow$  Bias

# The Problems Matching Solves

~~Without~~ Matching:

~~Im~~balance  $\rightsquigarrow$  ~~Model Dependence~~  $\rightsquigarrow$  ~~Researcher discretion~~  $\rightsquigarrow$  ~~Bias~~

# The Problems Matching Solves

Without Matching:

~~Imbalance~~  $\rightsquigarrow$  ~~Model Dependence~~  $\rightsquigarrow$  ~~Researcher discretion~~  $\rightsquigarrow$  ~~Bias~~

A central project of statistics: Automating away human discretion

# What's Matching?

## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders

## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :



## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :

$$TE_i = Y_i(1) - Y_i(0)$$

## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :

$$\begin{aligned} TE_i &= Y_i(1) - Y_i(0) \\ &= \text{observed} - \text{unobserved} \end{aligned}$$

## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :

$$\begin{aligned} TE_i &= Y_i - Y_i(0) \\ &= \text{observed} - \text{unobserved} \end{aligned}$$

## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :

$$\begin{aligned} TE_i &= Y_i - Y_i(0) \\ &= \text{observed} - \text{unobserved} \end{aligned}$$

- Estimate  $Y_i(0)$  with  $Y_j$  with a matched ( $X_i \approx X_j$ ) control

## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :

$$\begin{aligned} TE_i &= Y_i - Y_i(0) \\ &= \text{observed} - \text{unobserved} \end{aligned}$$

- Estimate  $Y_i(0)$  with  $Y_j$  with a matched ( $X_i \approx X_j$ ) control
- Quantities of Interest:

## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :

$$\begin{aligned} TE_i &= Y_i - Y_i(0) \\ &= \text{observed} - \text{unobserved} \end{aligned}$$

- Estimate  $Y_i(0)$  with  $Y_j$  with a matched ( $X_i \approx X_j$ ) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$SATT = \text{Mean}_{i \in \{T_i=1\}} (TE_i)$$

## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :

$$\begin{aligned} TE_i &= Y_i - Y_i(0) \\ &= \text{observed} - \text{unobserved} \end{aligned}$$

- Estimate  $Y_i(0)$  with  $Y_j$  with a matched ( $X_i \approx X_j$ ) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$SATT = \text{Mean}_{i \in \{T_i=1\}} (TE_i)$$

2. FSATT: Feasible SATT (prune badly matched treateds too)

## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :

$$\begin{aligned} TE_i &= Y_i - Y_i(0) \\ &= \text{observed} - \text{unobserved} \end{aligned}$$

- Estimate  $Y_i(0)$  with  $Y_j$  with a matched ( $X_i \approx X_j$ ) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$SATT = \text{Mean}_{i \in \{T_i=1\}} (TE_i)$$

2. FSATT: Feasible SATT (prune badly matched treateds too)
- **Big convenience:** Follow preprocessing with whatever statistical method you'd have used without matching



## What's Matching?

- $Y_i$  dep var,  $T_i$  (1=treated, 0=control),  $X_i$  confounders
- Treatment Effect for treated observation  $i$ :

$$\begin{aligned} TE_i &= Y_i - Y_i(0) \\ &= \text{observed} - \text{unobserved} \end{aligned}$$

- Estimate  $Y_i(0)$  with  $Y_j$  with a matched ( $X_i \approx X_j$ ) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$SATT = \text{Mean}_{i \in \{T_i=1\}} (TE_i)$$

2. FSATT: Feasible SATT (prune badly matched treateds too)
- **Big convenience:** Follow preprocessing with whatever statistical method you'd have used without matching
  - **Pruning nonmatches makes control vars matter less:** reduces imbalance, model dependence, researcher discretion, & bias

# Matching: Finding Hidden Randomized Experiments

# Matching: Finding Hidden Randomized Experiments

# Matching: Finding Hidden Randomized Experiments


## Types of Experiments



# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

*Complete  
Randomization*



# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

<i>Complete Randomization</i>	<i>Fully Blocked</i>

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	<hr/>	
<i>Unobserved</i>		

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	
<i>Unobserved</i>		



# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	
<i>Unobserved</i>	On average	

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	Exact
<i>Unobserved</i>	On average	

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	Exact
<i>Unobserved</i>	On average	On average

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

~> *Fully blocked dominates complete randomization*

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

~> *Fully blocked* dominates *complete randomization* for:

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	Exact
<i>Unobserved</i>	On average	On average

⇒ *Fully blocked* dominates *complete randomization* for:  
imbalance,

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

⇒ *Fully blocked* dominates *complete randomization* for:  
imbalance, model dependence,

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

⇒ *Fully blocked* dominates *complete randomization* for:  
imbalance, model dependence, power,



# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

⇒ *Fully blocked* dominates *complete randomization* for:  
imbalance, model dependence, power, efficiency,

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

⇒ *Fully blocked* dominates *complete randomization* for:  
imbalance, model dependence, power, efficiency, bias,

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

~> *Fully blocked* dominates *complete randomization* for:  
imbalance, model dependence, power, efficiency, bias, research costs,

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

~> *Fully blocked* dominates *complete randomization* for:  
imbalance, model dependence, power, efficiency, bias, research costs, robustness.

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

~> *Fully blocked* dominates *complete randomization* for:  
imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

Balance	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	Exact
<i>Unobserved</i>	On average	On average

~> *Fully blocked* dominates *complete randomization* for:  
imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

## Goal of Each Matching Method (in Observational Data)

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

⇒ *Fully blocked* dominates *complete randomization* for:  
imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

## Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

	<i>Complete</i>	<i>Fully</i>
Balance		
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

⇒ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

## Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*
- Other methods: *fully blocked*



# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

	<i>Complete</i>	<i>Fully</i>
Balance		
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

⇒ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

## Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*
- Other methods: *fully blocked*
- **Other matching methods dominate PSM**

# Matching: Finding Hidden Randomized Experiments

## Types of Experiments

	<i>Complete</i>	<i>Fully</i>
Covariates:	<i>Randomization</i>	<i>Blocked</i>
<i>Observed</i>	On average	<b>Exact</b>
<i>Unobserved</i>	On average	On average

⇒ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

## Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*
- Other methods: *fully blocked*
- **Other matching methods dominate PSM** (wait, it gets worse)

# Method 1: Mahalanobis Distance Matching

# Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

# Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
2. **Estimation** Difference in means or a model

# Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

## 1. Preprocess (Matching)

- $\text{Distance}(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$

## 2. Estimation Difference in means or a model

# Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

## 1. Preprocess (Matching)

- $\text{Distance}(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
- (Mahalanobis is for methodologists; in applications, use Euclidean!)

## 2. Estimation Difference in means or a model

# Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

## 1. Preprocess (Matching)

- $\text{Distance}(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
- (Mahalanobis is for methodologists; in applications, use Euclidean!)
- Match each treated unit to the nearest control unit

## 2. Estimation Difference in means or a model



# Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

## 1. Preprocess (Matching)

- $\text{Distance}(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
- (Mahalanobis is for methodologists; in applications, use Euclidean!)
- Match each treated unit to the nearest control unit
- Control units: not reused; pruned if unused

## 2. Estimation Difference in means or a model

# Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

## 1. Preprocess (Matching)

- $\text{Distance}(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
- (Mahalanobis is for methodologists; in applications, use Euclidean!)
- Match each treated unit to the nearest control unit
- Control units: not reused; pruned if unused
- Prune matches if  $\text{Distance} > \text{caliper}$

## 2. Estimation Difference in means or a model

# Method 1: Mahalanobis Distance Matching

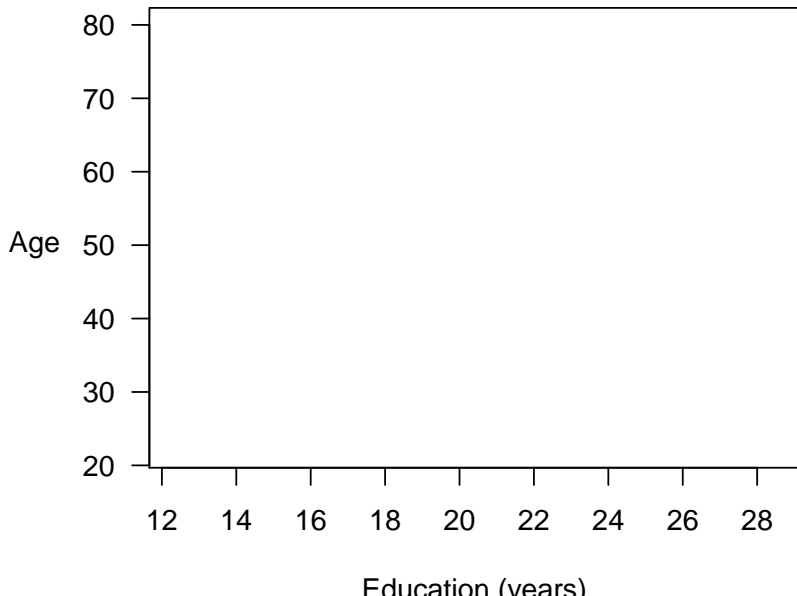
(Approximates Fully Blocked Experiment)

## 1. Preprocess (Matching)

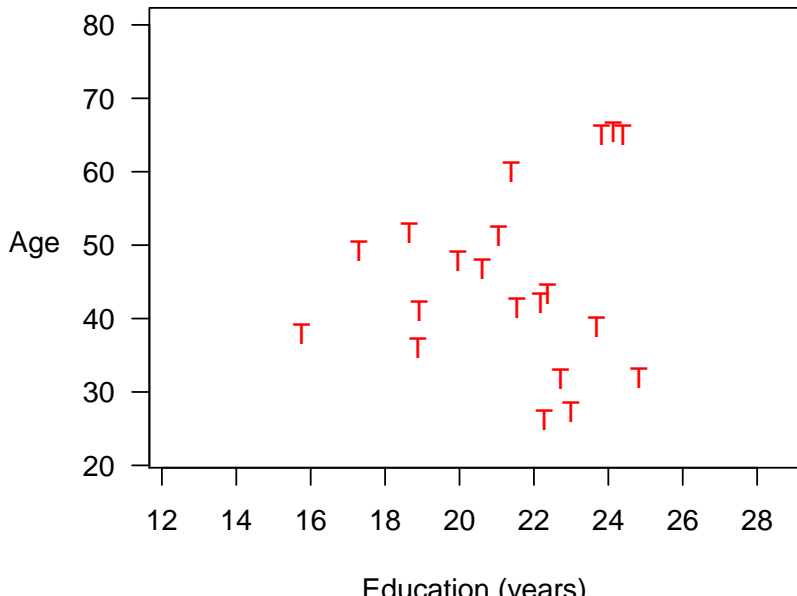
- $\text{Distance}(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
- (Mahalanobis is for methodologists; in applications, use Euclidean!)
- Match each treated unit to the nearest control unit
- Control units: not reused; pruned if unused
- Prune matches if  $\text{Distance} > \text{caliper}$
- (Many adjustments available to this basic method)

## 2. Estimation Difference in means or a model

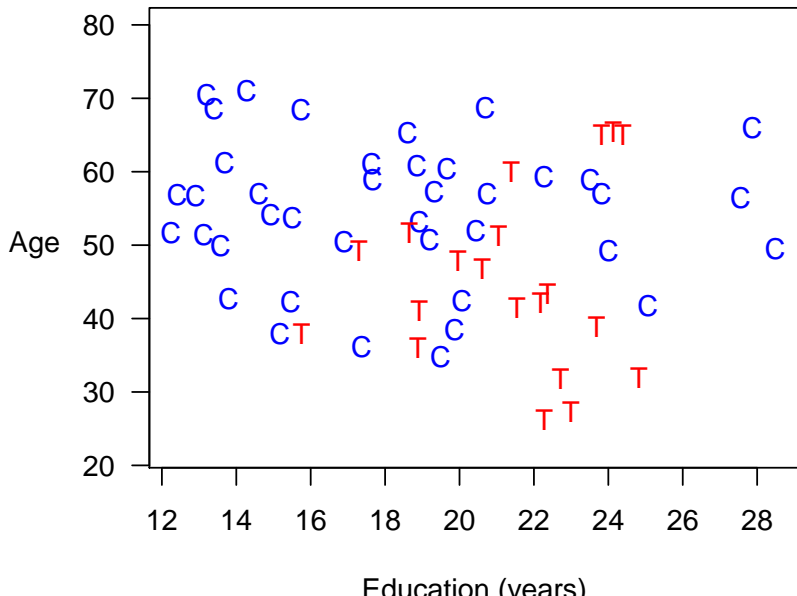
## Mahalanobis Distance Matching



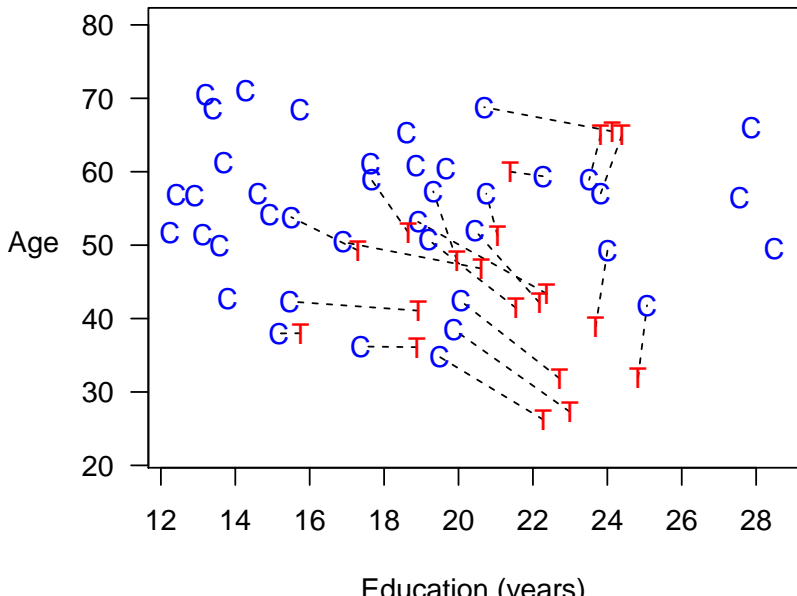
## Mahalanobis Distance Matching



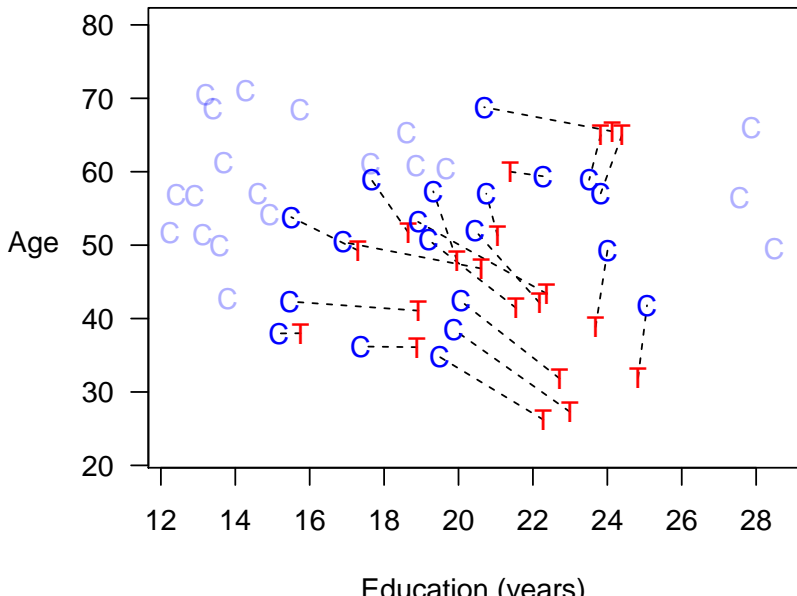
## Mahalanobis Distance Matching



## Mahalanobis Distance Matching

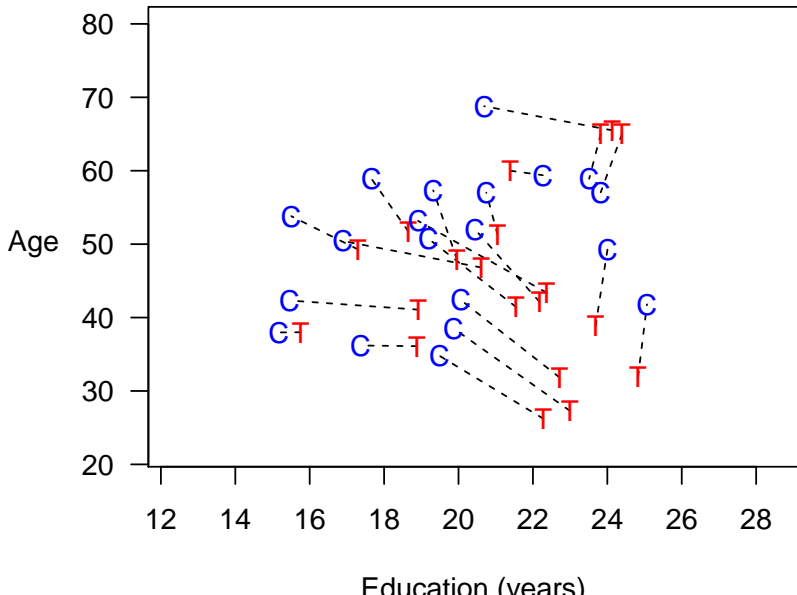


## Mahalanobis Distance Matching

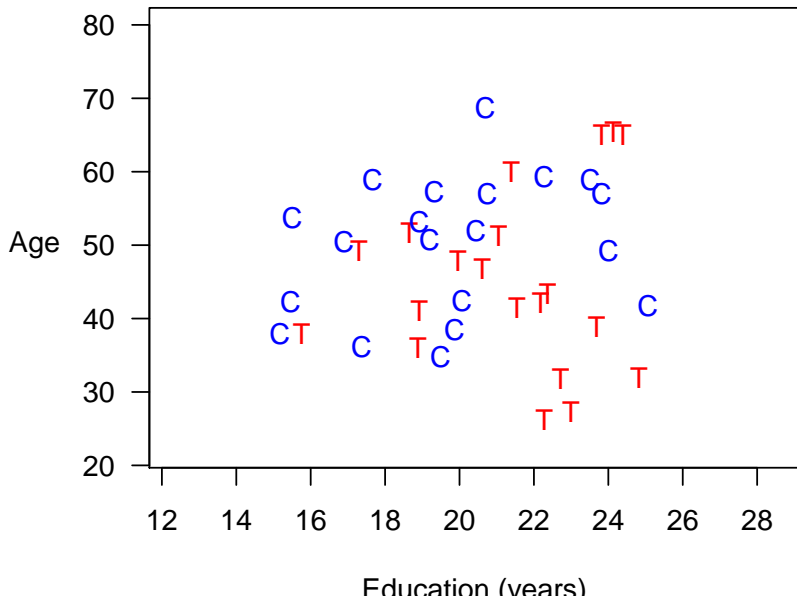




## Mahalanobis Distance Matching

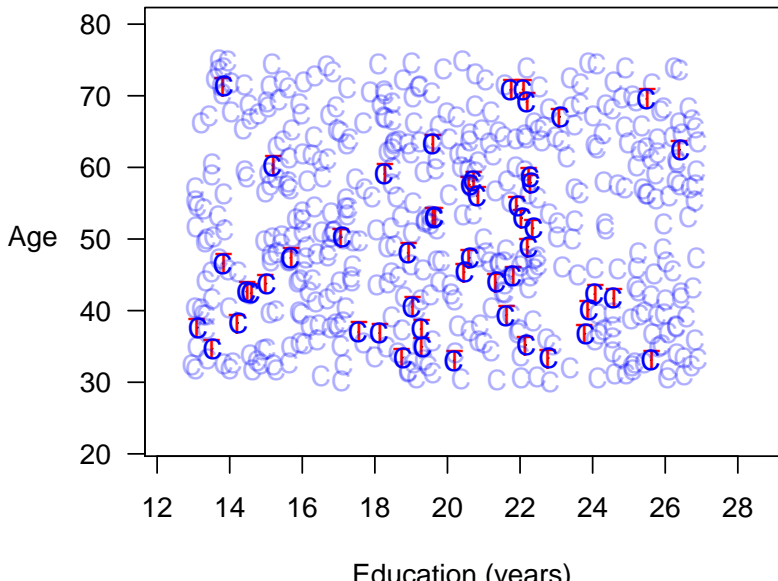


## Mahalanobis Distance Matching

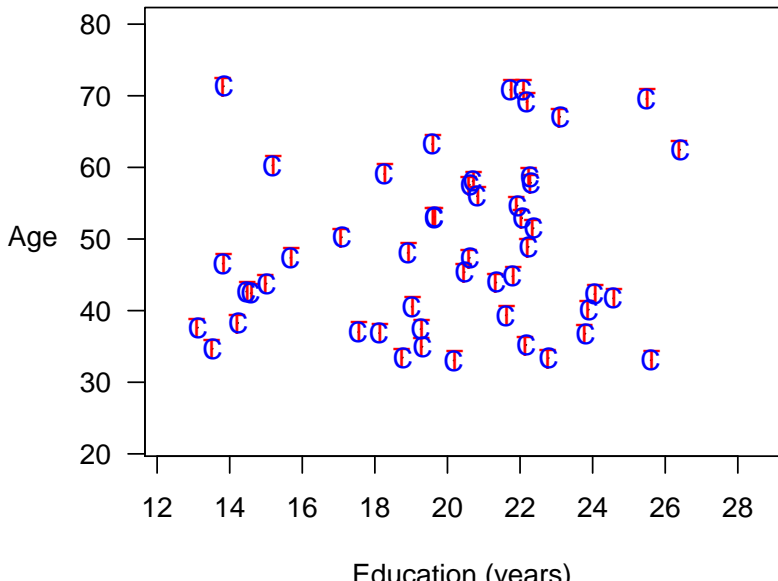


## Best Case: Mahalanobis Distance Matching

## Best Case: Mahalanobis Distance Matching



## Best Case: Mahalanobis Distance Matching



## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach) (Approximates Fully Blocked Experiment)

## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach) (Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
2. **Estimation** Difference in means or a model



## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach) (Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
  - Temporarily coarsen  $X$  as much as you're willing
2. **Estimation** Difference in means or a model

## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach) (Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
  - Temporarily coarsen  $X$  as much as you're willing
    - e.g., Education (grade school, high school, college, graduate)
2. **Estimation** Difference in means or a model

## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach) (Approximates Fully Blocked Experiment)

### 1. **Preprocess** (Matching)

- Temporarily coarsen  $X$  as much as you're willing
  - e.g., Education (grade school, high school, college, graduate)
- Apply exact matching to the coarsened  $X$ ,  $C(X)$

### 2. **Estimation** Difference in means or a model

## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach) (Approximates Fully Blocked Experiment)

### 1. **Preprocess** (Matching)

- Temporarily coarsen  $X$  as much as you're willing
  - e.g., Education (grade school, high school, college, graduate)
- Apply exact matching to the coarsened  $X$ ,  $C(X)$ 
  - Sort observations into strata, each with unique values of  $C(X)$

### 2. **Estimation** Difference in means or a model

## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

### 1. **Preprocess** (Matching)

- Temporarily coarsen  $X$  as much as you're willing
  - e.g., Education (grade school, high school, college, graduate)
- Apply exact matching to the coarsened  $X$ ,  $C(X)$ 
  - Sort observations into strata, each with unique values of  $C(X)$
  - Prune any stratum with 0 treated or 0 control units

### 2. **Estimation** Difference in means or a model

## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

### 1. Preprocess (Matching)

- Temporarily coarsen  $X$  as much as you're willing
  - e.g., Education (grade school, high school, college, graduate)
- Apply exact matching to the coarsened  $X$ ,  $C(X)$ 
  - Sort observations into strata, each with unique values of  $C(X)$
  - Prune any stratum with 0 treated or 0 control units
- Pass on original (uncoarsened) units except those pruned

### 2. Estimation Difference in means or a model

## Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

### 1. Preprocess (Matching)

- Temporarily coarsen  $X$  as much as you're willing
  - e.g., Education (grade school, high school, college, graduate)
- Apply exact matching to the coarsened  $X$ ,  $C(X)$ 
  - Sort observations into strata, each with unique values of  $C(X)$
  - Prune any stratum with 0 treated or 0 control units
- Pass on original (uncoarsened) units except those pruned

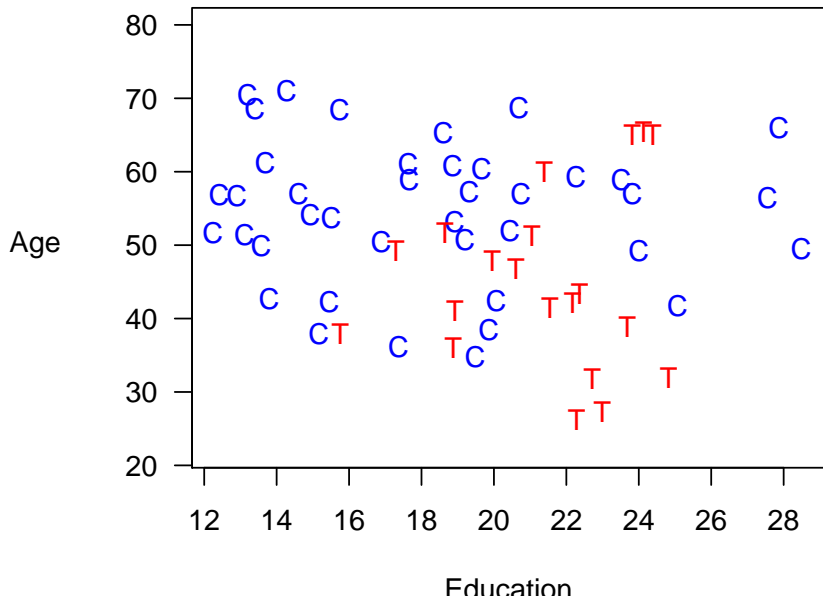
### 2. Estimation Difference in means or a model

- Weight controls in each stratum to equal treated

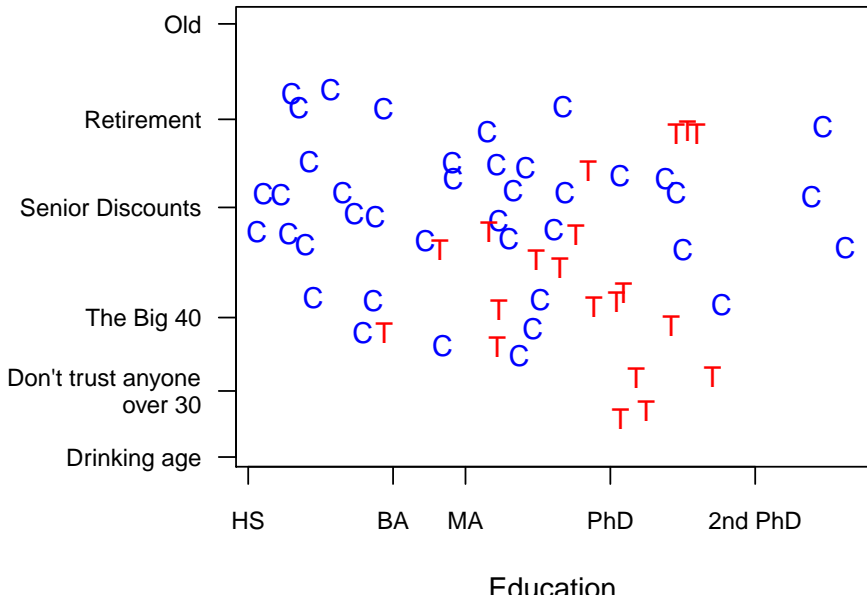
# Coarsened Exact Matching



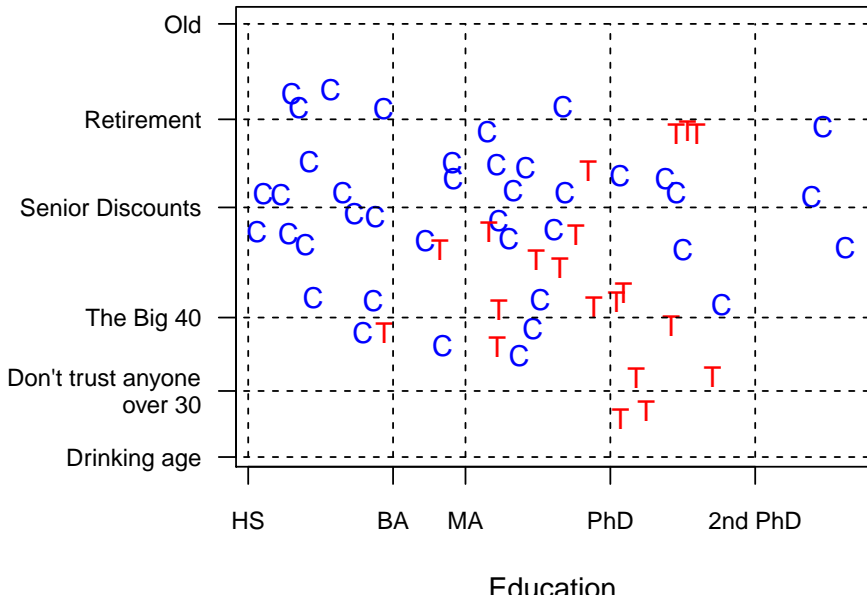
## Coarsened Exact Matching



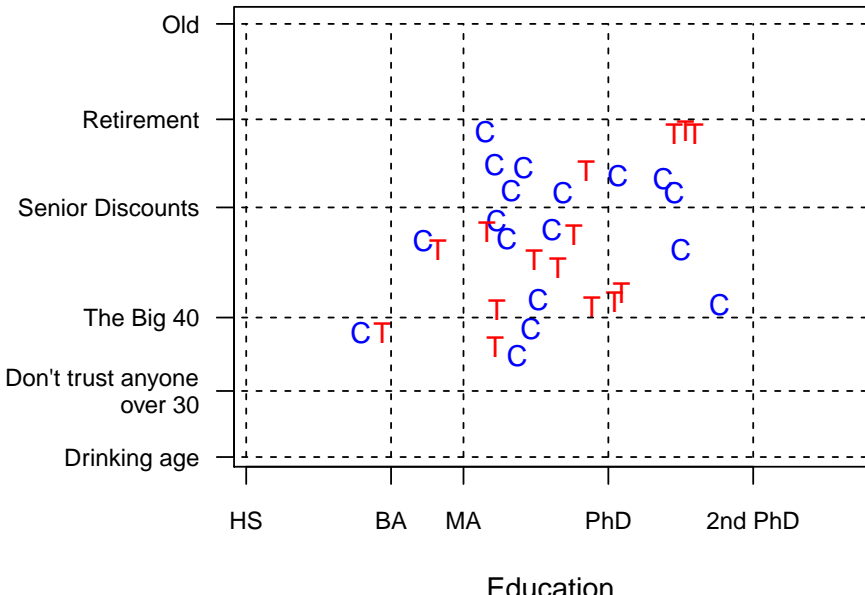
## Coarsened Exact Matching



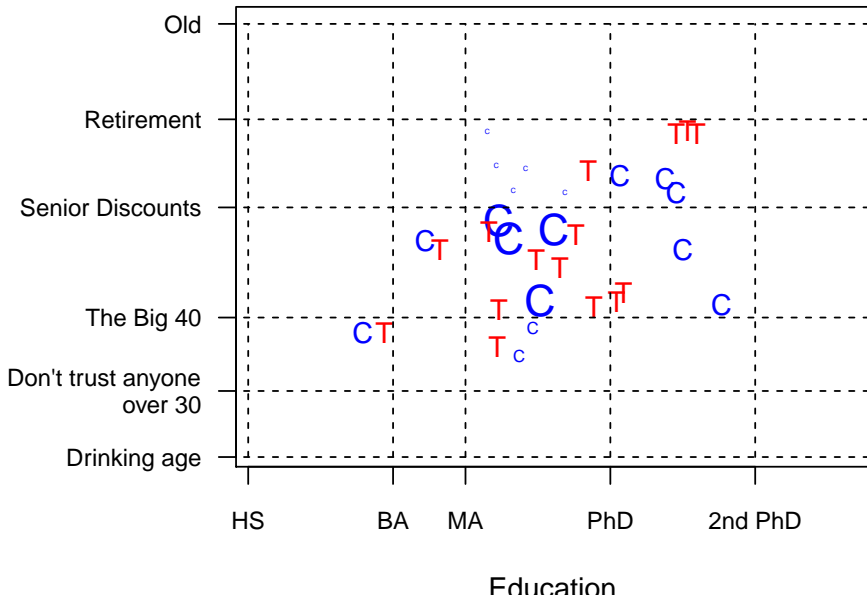
## Coarsened Exact Matching



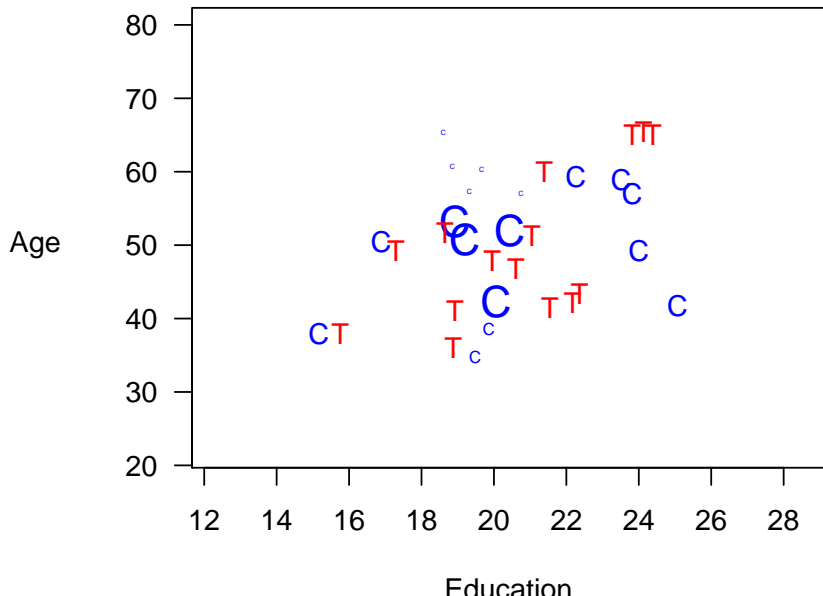
## Coarsened Exact Matching



## Coarsened Exact Matching

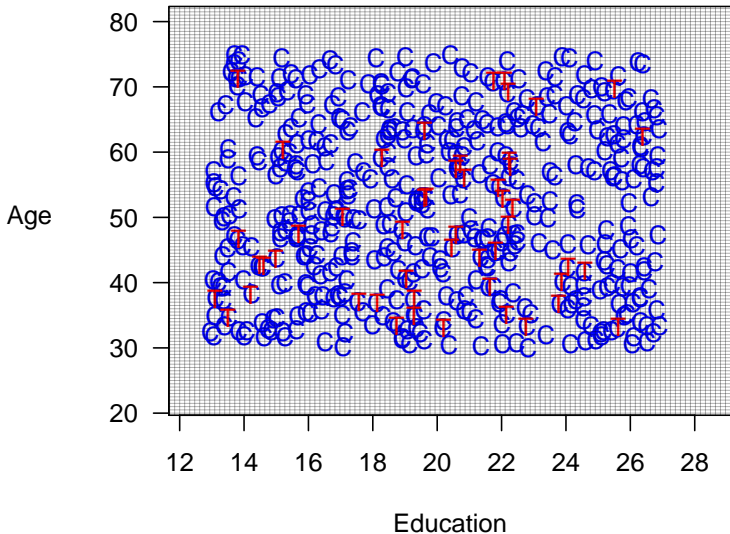


## Coarsened Exact Matching



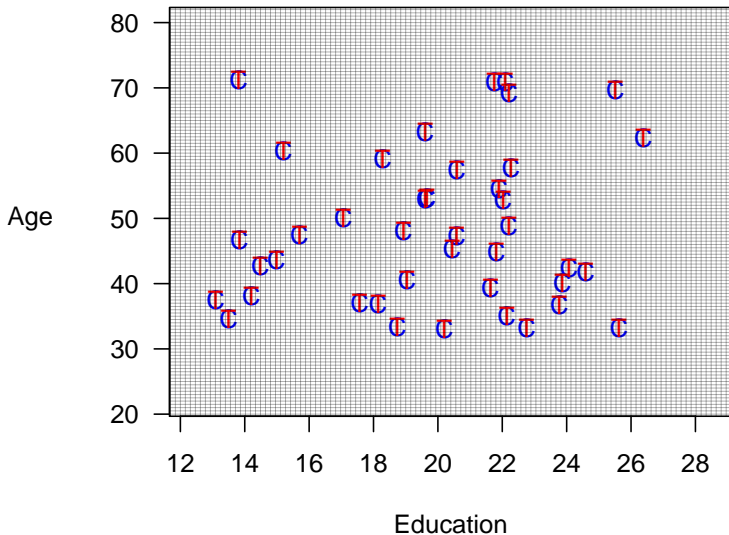
## Best Case: Coarsened Exact Matching

## Best Case: Coarsened Exact Matching

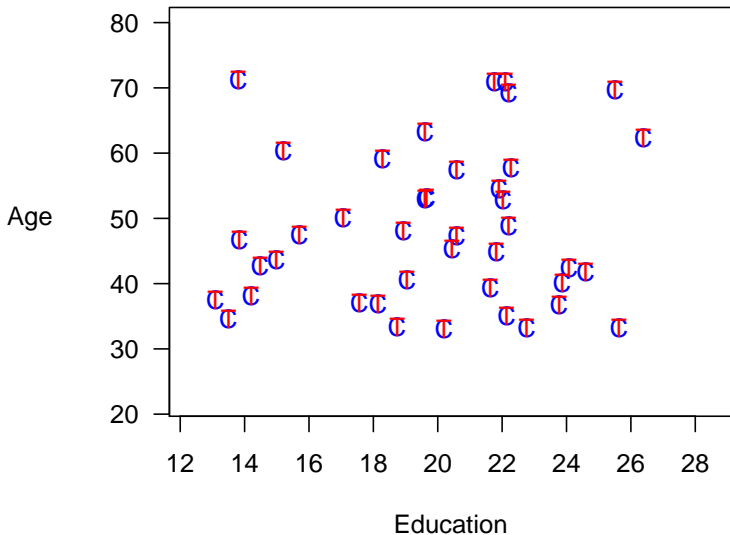




## Best Case: Coarsened Exact Matching



## Best Case: Coarsened Exact Matching



## Method 3: Propensity Score Matching

## Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

## Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

1. **Preprocess** (Matching)
2. **Estimation** Difference in means or a model

# Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

## 1. Preprocess (Matching)

- Reduce  $k$  elements of  $X$  to scalar

$$\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$$

## 2. Estimation Difference in means or a model

# Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

## 1. Preprocess (Matching)

- Reduce  $k$  elements of  $X$  to scalar

$$\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$$

- $\text{Distance}(X_c, X_t) = |\pi_c - \pi_t|$

## 2. Estimation Difference in means or a model

# Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

## 1. Preprocess (Matching)

- Reduce  $k$  elements of  $X$  to scalar
$$\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$$
- Distance( $X_c, X_t$ ) =  $|\pi_c - \pi_t|$
- Match each treated unit to the nearest control unit

## 2. Estimation Difference in means or a model



# Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

## 1. Preprocess (Matching)

- Reduce  $k$  elements of  $X$  to scalar
$$\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$$
- Distance( $X_c, X_t$ ) =  $|\pi_c - \pi_t|$
- Match each treated unit to the nearest control unit
- Control units: not reused; pruned if unused

## 2. Estimation Difference in means or a model

# Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

## 1. Preprocess (Matching)

- Reduce  $k$  elements of  $X$  to scalar  
$$\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$$
- Distance( $X_c, X_t$ ) =  $|\pi_c - \pi_t|$
- Match each treated unit to the nearest control unit
- Control units: not reused; pruned if unused
- Prune matches if Distance > *caliper*

## 2. Estimation Difference in means or a model

# Method 3: Propensity Score Matching

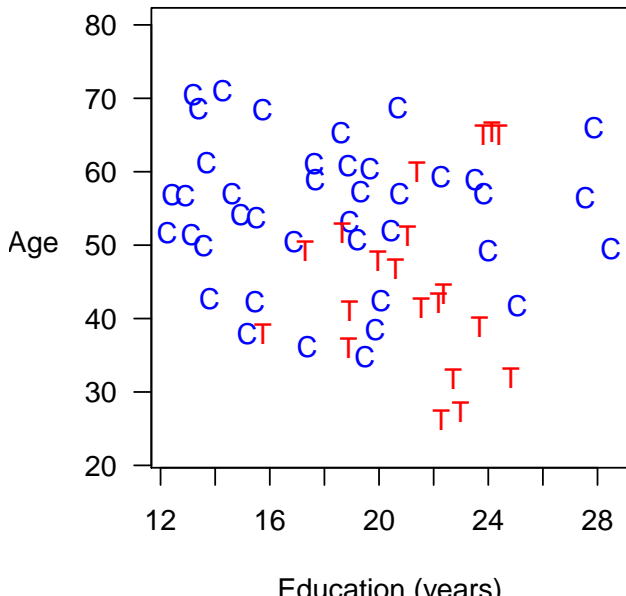
(Approximates Completely Randomized Experiment)

## 1. Preprocess (Matching)

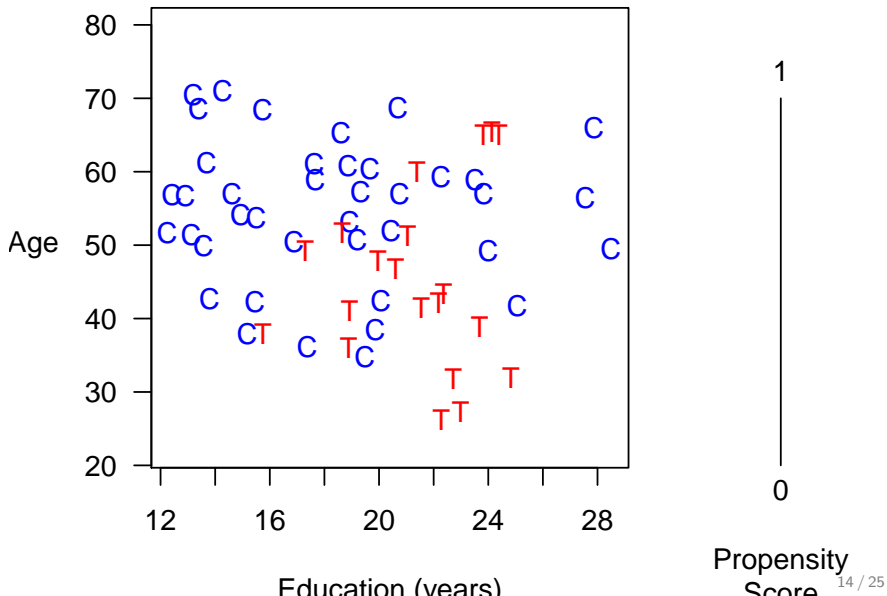
- Reduce  $k$  elements of  $X$  to scalar
$$\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$$
- $\text{Distance}(X_c, X_t) = |\pi_c - \pi_t|$
- Match each treated unit to the nearest control unit
- Control units: not reused; pruned if unused
- Prune matches if  $\text{Distance} > \text{caliper}$
- (Many adjustments available to this basic method)

## 2. Estimation Difference in means or a model

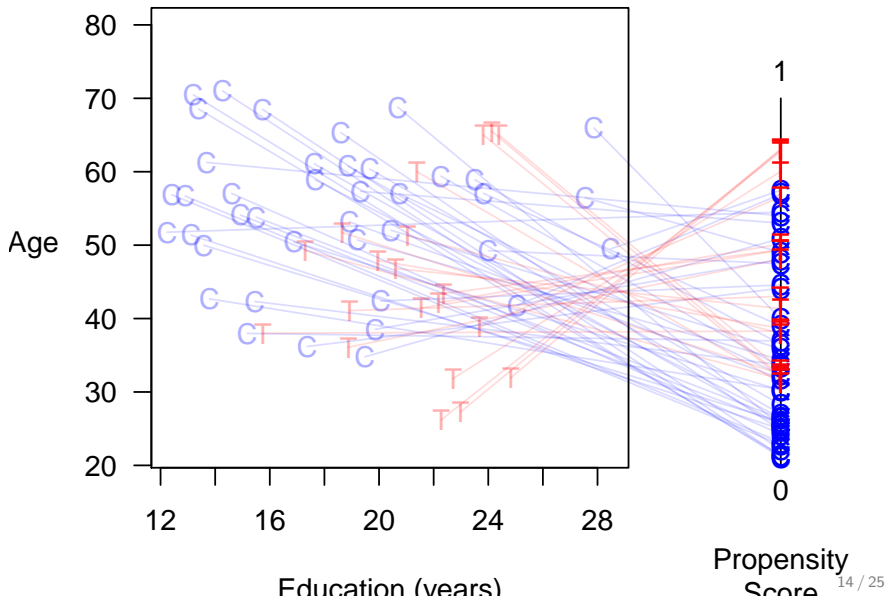
## Propensity Score Matching



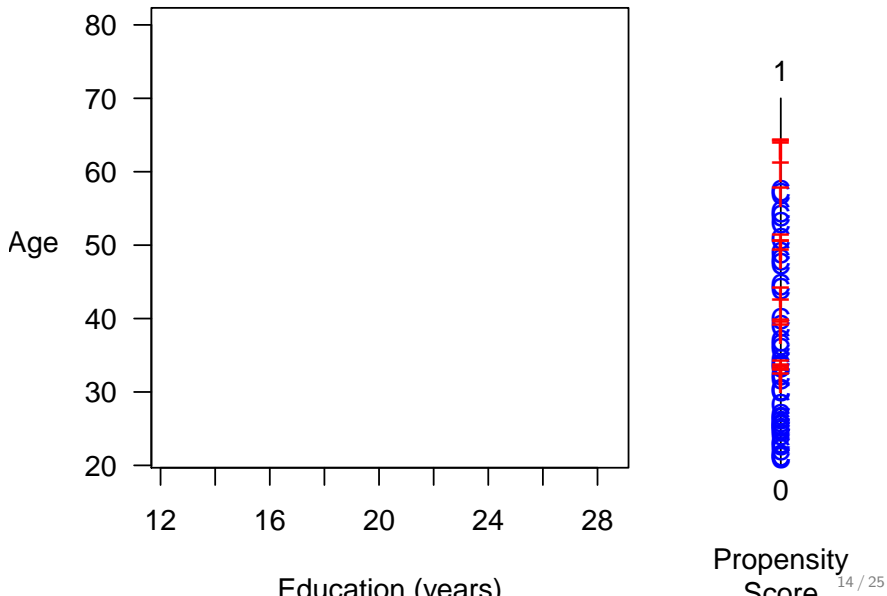
## Propensity Score Matching



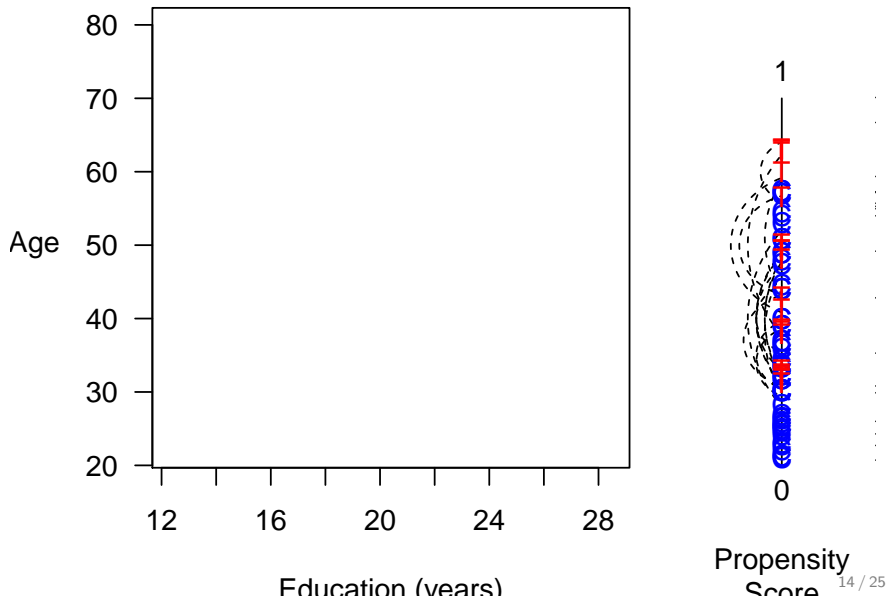
## Propensity Score Matching



## Propensity Score Matching

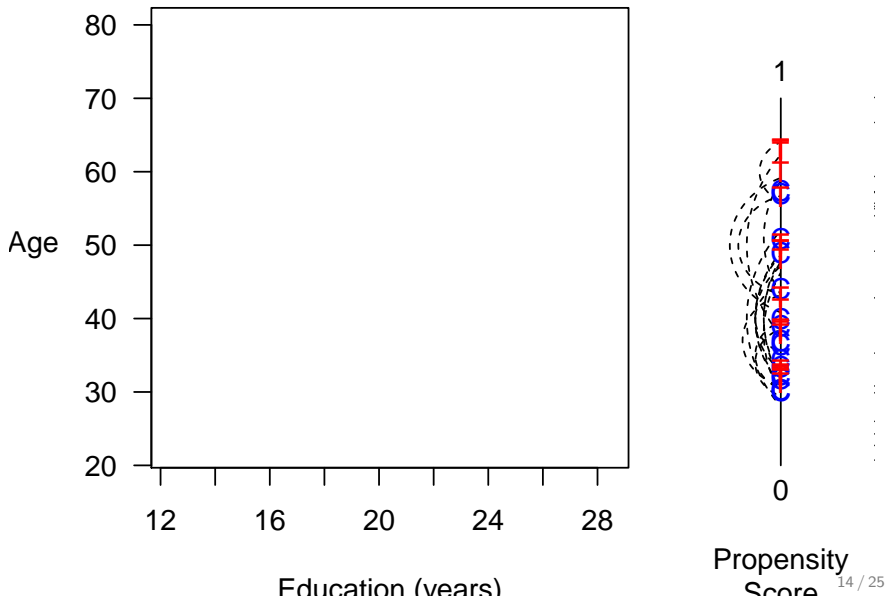


## Propensity Score Matching

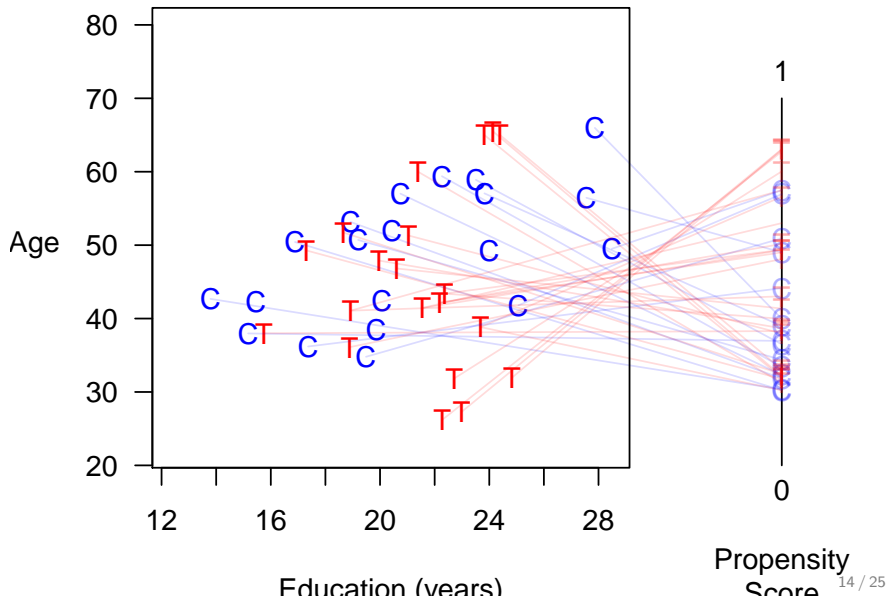




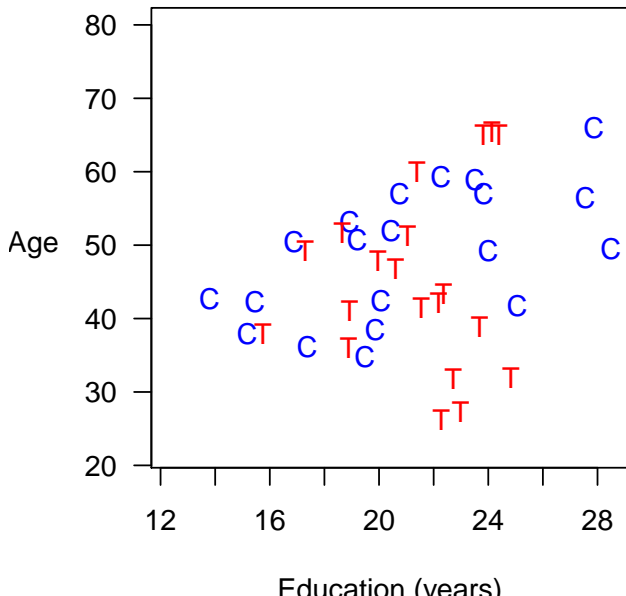
## Propensity Score Matching



## Propensity Score Matching

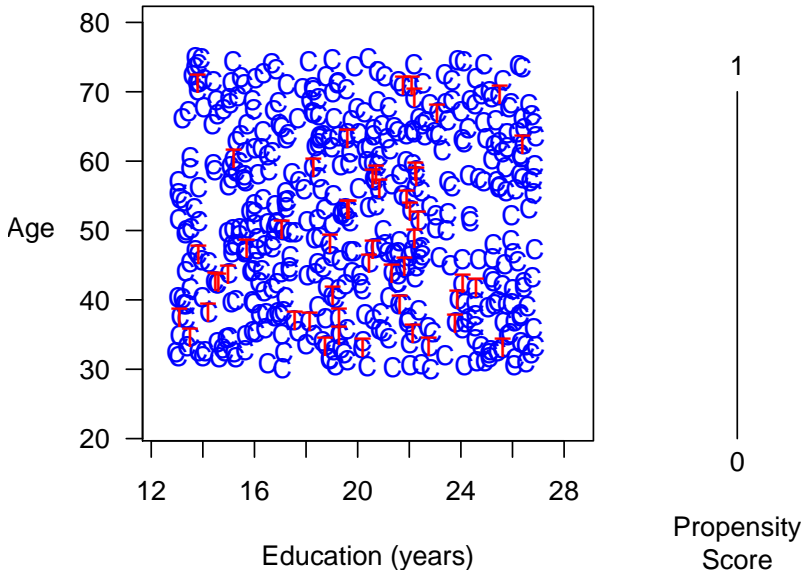


## Propensity Score Matching

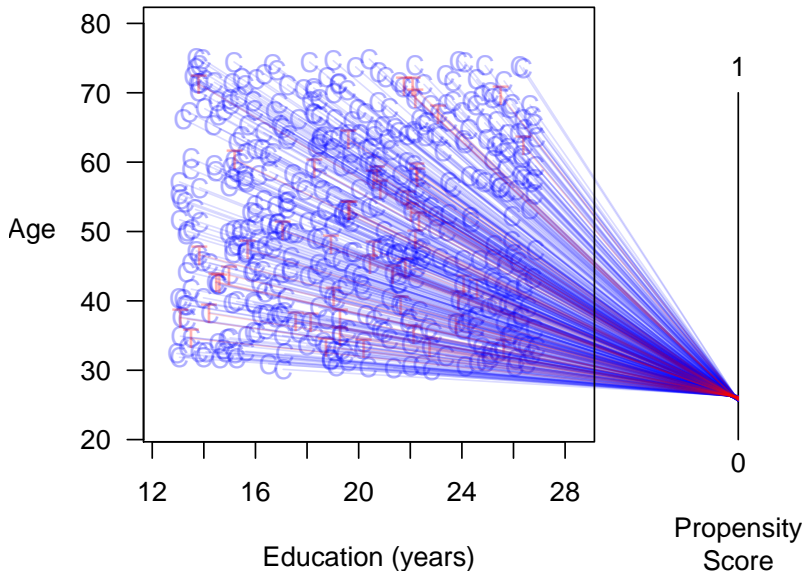


## Best Case: Propensity Score Matching

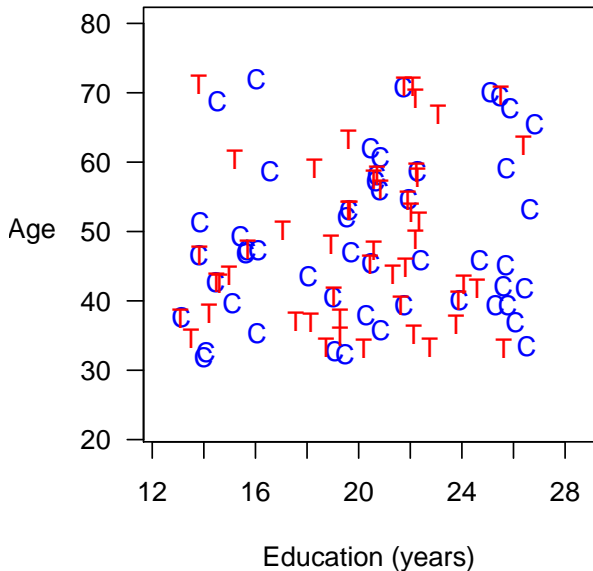
## Best Case: Propensity Score Matching



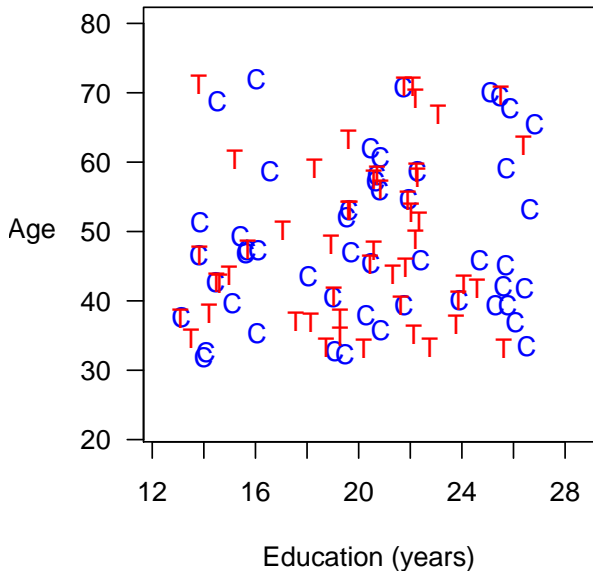
## Best Case: Propensity Score Matching



## Best Case: Propensity Score Matching



## Best Case: Propensity Score Matching is Suboptimal





# PSM's Statistical Properties

# PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes

# PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
  - *Efficient* relative to complete randomization, but

# PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
  - *Efficient* relative to complete randomization, but
  - *Inefficient* relative to (the more powerful) full blocking

# PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
  - *Efficient* relative to complete randomization, but
  - *Inefficient* relative to (the more powerful) full blocking
  - Other methods usually dominate:

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t$$

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\Rightarrow X_c = X_t$$

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\Rightarrow X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse



# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\Rightarrow X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\Rightarrow X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\implies X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\implies X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)  $\rightsquigarrow$  pruning at random

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\implies X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)  $\rightsquigarrow$  pruning at random  $\rightsquigarrow$  Imbalance

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\implies X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)  $\rightsquigarrow$  pruning at random  $\rightsquigarrow$  Imbalance  $\rightsquigarrow$  Inefficiency

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\implies X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)  $\rightsquigarrow$  pruning at random  $\rightsquigarrow$  Imbalance  $\rightsquigarrow$  Inefficiency  $\rightsquigarrow$  Model dependence

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\Rightarrow X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)  $\rightsquigarrow$  pruning at random  $\rightsquigarrow$  Imbalance  $\rightsquigarrow$  Inefficiency  $\rightsquigarrow$  Model dependence  $\rightsquigarrow$  Bias



# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\Rightarrow X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)  $\rightsquigarrow$  pruning at random  $\rightsquigarrow$  Imbalance  $\rightsquigarrow$  Inefficiency  $\rightsquigarrow$  Model dependence  $\rightsquigarrow$  Bias
- If the data have no good matches, the paradox won't be a problem but you're cooked anyway.

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\implies X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)  $\rightsquigarrow$  pruning at random  $\rightsquigarrow$  Imbalance  $\rightsquigarrow$  Inefficiency  $\rightsquigarrow$  Model dependence  $\rightsquigarrow$  Bias
- If the data have no good matches, the paradox won't be a problem but you're cooked anyway.
- Doesn't PSM solve the curse of dimensionality problem?

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

$$\pi_c = \pi_t \not\implies X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)  $\rightsquigarrow$  pruning at random  $\rightsquigarrow$  Imbalance  $\rightsquigarrow$  Inefficiency  $\rightsquigarrow$  Model dependence  $\rightsquigarrow$  Bias
- If the data have no good matches, the paradox won't be a problem but you're cooked anyway.
- Doesn't PSM solve the curse of dimensionality problem? Nope.

# PSM's Statistical Properties

## 1. Low Standards: Sometimes helps, never optimizes

- *Efficient* relative to complete randomization, but
- *Inefficient* relative to (the more powerful) full blocking
- Other methods usually dominate:

$$X_c = X_t \implies \pi_c = \pi_t \text{ but}$$

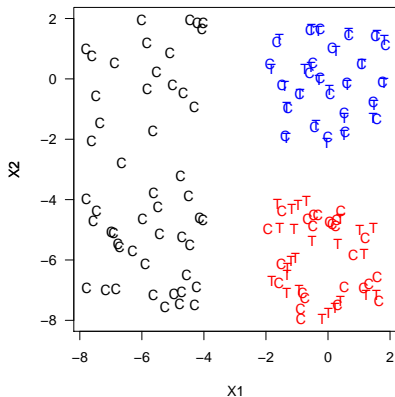
$$\pi_c = \pi_t \not\Rightarrow X_c = X_t$$

## 2. The PSM Paradox: When you do “better,” you do worse

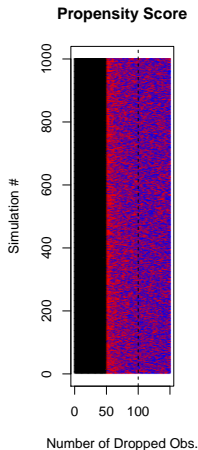
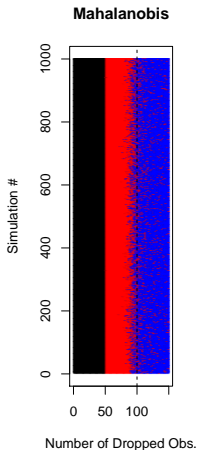
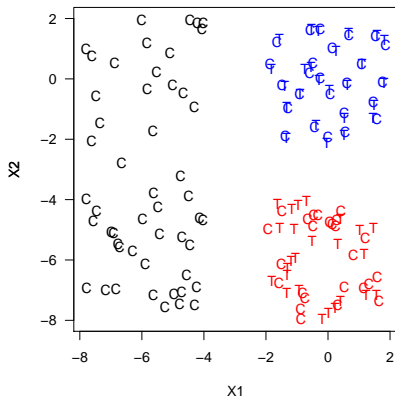
- Background: Random matching increases imbalance
- When PSM approximates complete randomization (to begin with or, after some pruning)  $\rightsquigarrow$  all  $\hat{\pi} \approx 0.5$  (or constant within strata)  $\rightsquigarrow$  pruning at random  $\rightsquigarrow$  Imbalance  $\rightsquigarrow$  Inefficiency  $\rightsquigarrow$  Model dependence  $\rightsquigarrow$  Bias
- If the data have no good matches, the paradox won't be a problem but you're cooked anyway.
- Doesn't PSM solve the curse of dimensionality problem? Nope. The PSM Paradox gets worse with more covariates

# PSM is Blind Where Other Methods Can See

# PSM is Blind Where Other Methods Can See

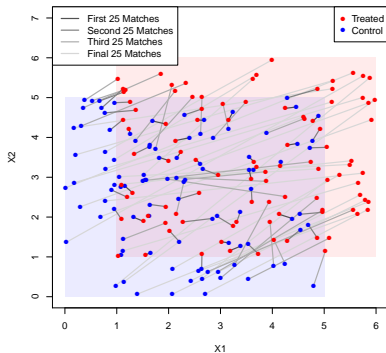


# PSM is Blind Where Other Methods Can See

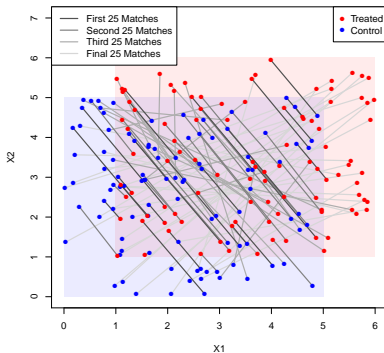


# What Does PSM Match?

## MDM Matches



## PSM Matches



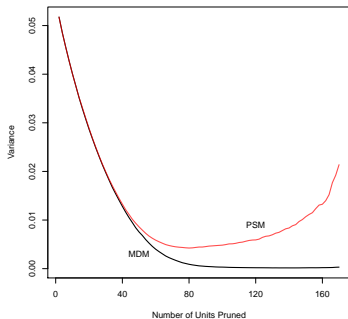
Controls:  $X_1, X_2 \sim \text{Uniform}(0,5)$

Treateds:  $X_1, X_2 \sim \text{Uniform}(1,6)$

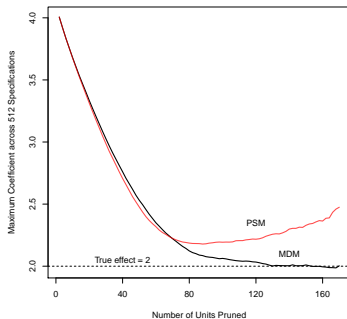


# PSM Increases Model Dependence & Bias

## Model Dependence



## Bias

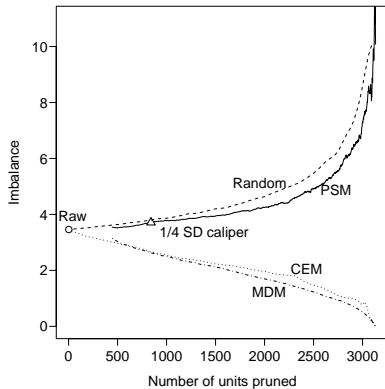


$$Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i$$
$$\epsilon_i \sim N(0, 1)$$

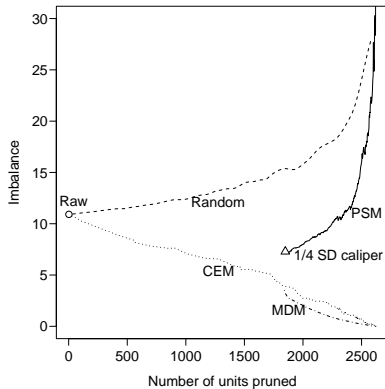
# The Propensity Score Paradox in Real Data

# The Propensity Score Paradox in Real Data

Finkel et al. (JOP, 2012)

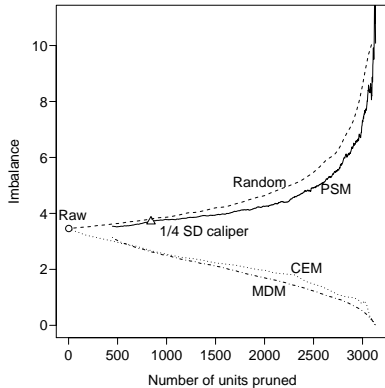


Nielsen et al. (AJPS, 2011)

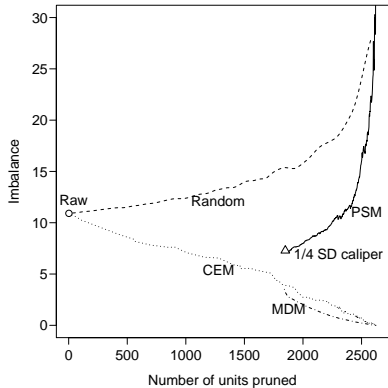


# The Propensity Score Paradox in Real Data

Finkel et al. (JOP, 2012)



Nielsen et al. (AJPS, 2011)



Similar pattern for > 20 other real data sets we checked

# The Matching Frontier

# The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each  $n$

# The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each  $n$
- Bias-Variance trade off  $\rightsquigarrow$  Imbalance- $n$  Trade Off

# The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each  $n$
- Bias-Variance trade off  $\rightsquigarrow$  Imbalance- $n$  Trade Off
- Simple to use



# The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each  $n$
- Bias-Variance trade off  $\rightsquigarrow$  Imbalance- $n$  Trade Off
- Simple to use
- No need to choose or use a matching method

# The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each  $n$
- Bias-Variance trade off  $\rightsquigarrow$  Imbalance- $n$  Trade Off
- Simple to use
- No need to choose or use a matching method
- All solutions are optimal

# The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each  $n$
- Bias-Variance trade off  $\rightsquigarrow$  Imbalance- $n$  Trade Off
- Simple to use
- No need to choose or use a matching method
- All solutions are optimal
- No iteration or diagnostics required

# The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each  $n$
- Bias-Variance trade off  $\rightsquigarrow$  Imbalance- $n$  Trade Off
- Simple to use
- No need to choose or use a matching method
- All solutions are optimal
- No iteration or diagnostics required
- No cherry picking possible; you see everything optimal

# The Matching Frontier

- **Frontier** = matched dataset with lowest imbalance for each  $n$
- Bias-Variance trade off  $\rightsquigarrow$  Imbalance- $n$  Trade Off
- Simple to use
- No need to choose or use a matching method
- All solutions are optimal
- No iteration or diagnostics required
- No cherry picking possible; you see everything optimal
- Choose an imbalance metric, then run.

How hard is the frontier to calculate?

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$



## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N - 1, \dots, 1$

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N - 1, \dots, 1$
  - The combination is the (gargantuan) “power set”

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N - 1, \dots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g.,  $N > 300$  requires more imbalance evaluations than elementary particles in the universe

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N - 1, \dots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g.,  $N > 300$  requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$  It's **hard** to calculate!

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N - 1, \dots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g.,  $N > 300$  requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$  It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:



## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N-1, \dots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g.,  $N > 300$  requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$  It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N - 1, \dots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g.,  $N > 300$  requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$  It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N-1, \dots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g.,  $N > 300$  requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$  It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N-1, \dots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g.,  $N > 300$  requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$  It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets

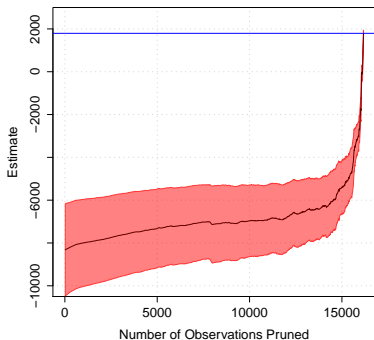
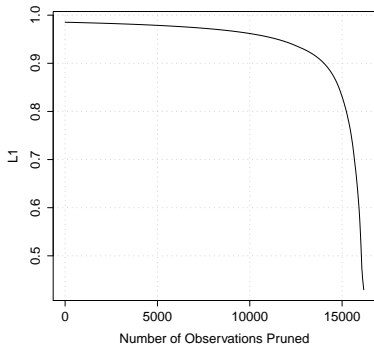
## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N-1, \dots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g.,  $N > 300$  requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$  It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)

## How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of  $N$  control units  $X_0$
  - Calculate imbalance for all  $\binom{N}{n}$  subsets of rows of  $X_0$
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$  evaluations for each sample size  $n = N, N - 1, \dots, 1$
  - The combination is the (gargantuan) “power set”
  - e.g.,  $N > 300$  requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$  It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as “greedy” but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)
  - $\rightsquigarrow$  It's **easy** to calculate!

## Job Training Data: Frontier and Causal Estimates



- 185 Ts; pruning most 16,252 Cs won't increase variance much
- Huge bias-variance trade-off after pruning most Cs
- Estimates converge to experiment after removing bias
- No mysteries: basis of inference clearly revealed

# Conclusions



# Conclusions

- Propensity score matching:

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM:

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates;

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data;

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support;

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score;



# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- A Simple and Powerful Method: CEM

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- A Simple and Powerful Method: CEM
- A New General Approach: The Matching Frontier

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- A Simple and Powerful Method: CEM
- A New General Approach: The Matching Frontier
  - Fast; easy; no iteration; Software: MatchingFrontier

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- A Simple and Powerful Method: CEM
- A New General Approach: The Matching Frontier
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- A Simple and Powerful Method: CEM
- A New General Approach: The Matching Frontier
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

# Conclusions

- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- A Simple and Powerful Method: CEM
- A New General Approach: The Matching Frontier
  - Fast; easy; no iteration; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- $\rightsquigarrow$  Using more information is simpler and more powerful

For more information, articles, & software

GaryKing.org