# Advanced Quantitative Research Methodology, Lecture Notes: Matching Methods for Causal Inference[1]

Gary King

GaryKing.org

March 31, 2013

# Overview

- Problem: Model dependence (review)

# Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)

# Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications

# Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us choose

# Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us choose
- Problem: The most commonly used method can increase imbalance!

# Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us choose
- Problem: The most commonly used method can increase imbalance!
- Solution: Other methods do not share this problem

# Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us choose
- Problem: The most commonly used method can increase imbalance!
- Solution: Other methods do not share this problem
- (Coarsened Exact Matching is simple, easy, and powerful)

# Overview

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us choose
- Problem: The most commonly used method can increase imbalance!
- Solution: Other methods do not share this problem
- (Coarsened Exact Matching is simple, easy, and powerful)
- $\rightsquigarrow$ Lots of insights revealed in the process

# Model Dependence Example

# Model Dependence Example

## Replication: Doyle and Sambanis, APSR 2000

# Model Dependence Example

## Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars

# Model Dependence Example

## Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success

# Model Dependence Example

## Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)

# Model Dependence Example

## Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status; etc.

# Model Dependence Example

## Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status; etc.
- Counterfactual question: UN intervention switched for each war

# Model Dependence Example

## Replication: Doyle and Sambanis, APSR 2000

- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status; etc.
- Counterfactual question: UN intervention switched for each war
- Data analysis: Logit model

# Model Dependence Example

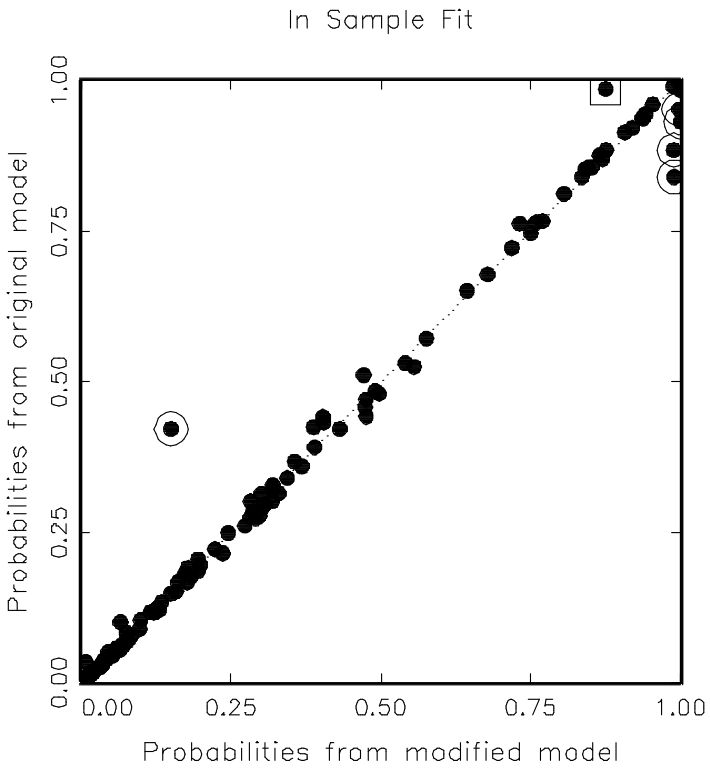
## Replication: Doyle and Sambanis, APSR 2000

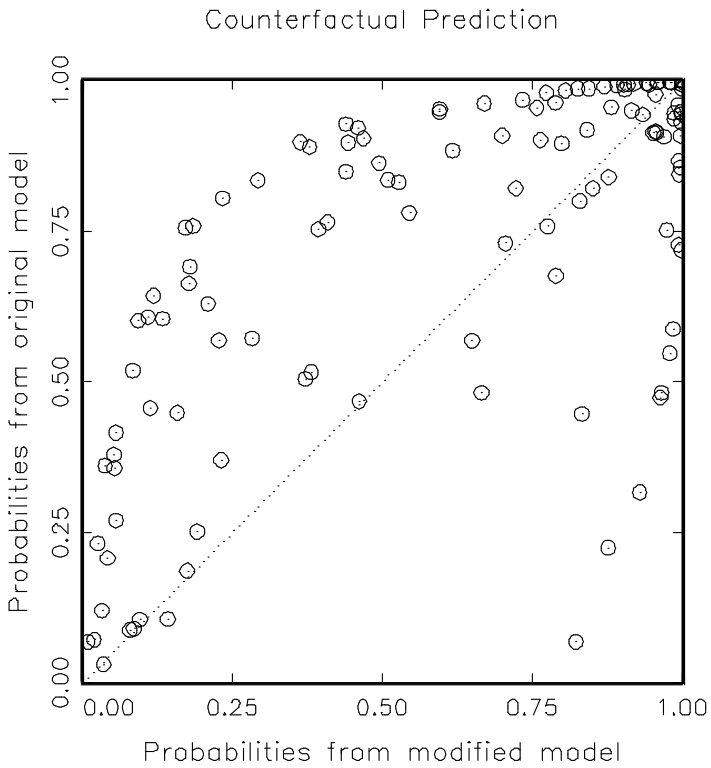
- Data: 124 Post-World War II civil wars
- Dependent variable: peacebuilding success
- Treatment variable: multilateral UN peacekeeping intervention (0/1)
- Control vars: war type, severity, duration; development status; etc.
- Counterfactual question: UN intervention switched for each war
- Data analysis: Logit model
- The question: How *model dependent* are the results?

# Two Logit Models, Apparently Similar Results

| | Original "Interactive" Model | | | Modified Model | | |
|---|---|---|---|---|---|---|
| Variables | Coeff | SE | P-val | Coeff | SE | P-val |
| Wartype | −1.742 | .609 | .004 | −1.666 | .606 | .006 |
| Logdead | −.445 | .126 | .000 | −.437 | .125 | .000 |
| Wardur | .006 | .006 | .258 | .006 | .006 | .342 |
| Factnum | −1.259 | .703 | .073 | −1.045 | .899 | .245 |
| Factnum2 | .062 | .065 | .346 | .032 | .104 | .756 |
| Trnsfcap | .004 | .002 | .010 | .004 | .002 | .017 |
| Develop | .001 | .000 | .065 | .001 | .000 | .068 |
| Exp | −6.016 | 3.071 | .050 | −6.215 | 3.065 | .043 |
| Decade | −.299 | .169 | .077 | −0.284 | .169 | .093 |
| Treaty | 2.124 | .821 | .010 | 2.126 | .802 | .008 |
| UNOP4 | 3.135 | 1.091 | .004 | .262 | 1.392 | .851 |
| Wardur*UNOP4 | — | — | — | .037 | .011 | .001 |
| Constant | 8.609 | 2.157 | 0.000 | 7.978 | 2.350 | .000 |
| N | | 122 | | | 122 | |
| Log-likelihood | | -45.649 | | | -44.902 | |
| Pseudo $R^2$ | | .423 | | | .433 | |

# Doyle and Sambanis: Model Dependence



# Overview of Matching for Causal Inference

# Overview of Matching for Causal Inference

- Goal: reduce model dependence

# Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach

# Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)

# Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of $T$ and $X$, but not $Y$)

# Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of $T$ and $X$, but not $Y$)
- Apply model to preprocessed (pruned) rather than raw data

# Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of $T$ and $X$, but not $Y$)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP

# Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of $T$ and $X$, but not $Y$)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP
- Overall idea:

# Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of $T$ and $X$, but not $Y$)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP
- Overall idea:
  - If each treated unit exactly matches a control unit w.r.t. $X$, then: (1) treated and control groups are identical, (2) $X$ is no longer a confounder, (3) no need to worry about the functional form ($\bar{X}_T - \bar{X}_C$ is good enough).

# Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of $T$ and $X$, but not $Y$)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP
- Overall idea:
  - If each treated unit exactly matches a control unit w.r.t. $X$, then: (1) treated and control groups are identical, (2) $X$ is no longer a confounder, (3) no need to worry about the functional form ($\bar{X}_T - \bar{X}_C$ is good enough).
  - If treated and control groups are better balanced than when you started, due to pruning, model dependence is reduced

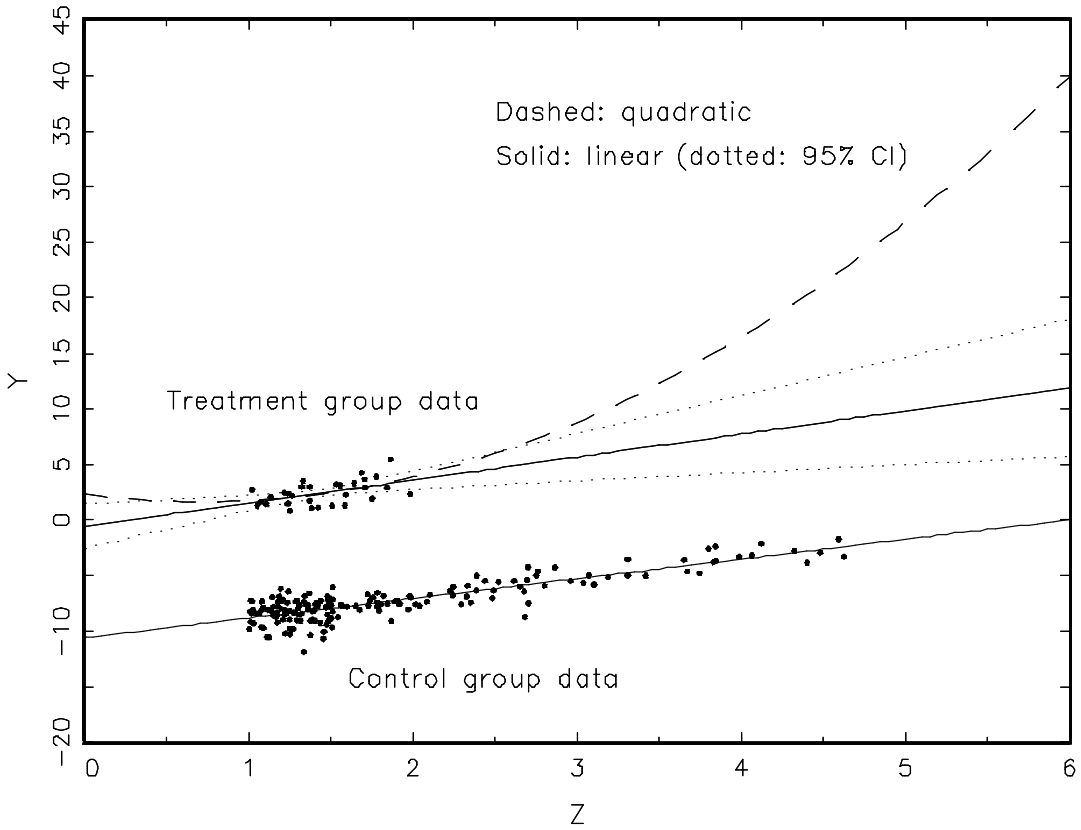
# Model Dependence: A Simpler Example

# Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

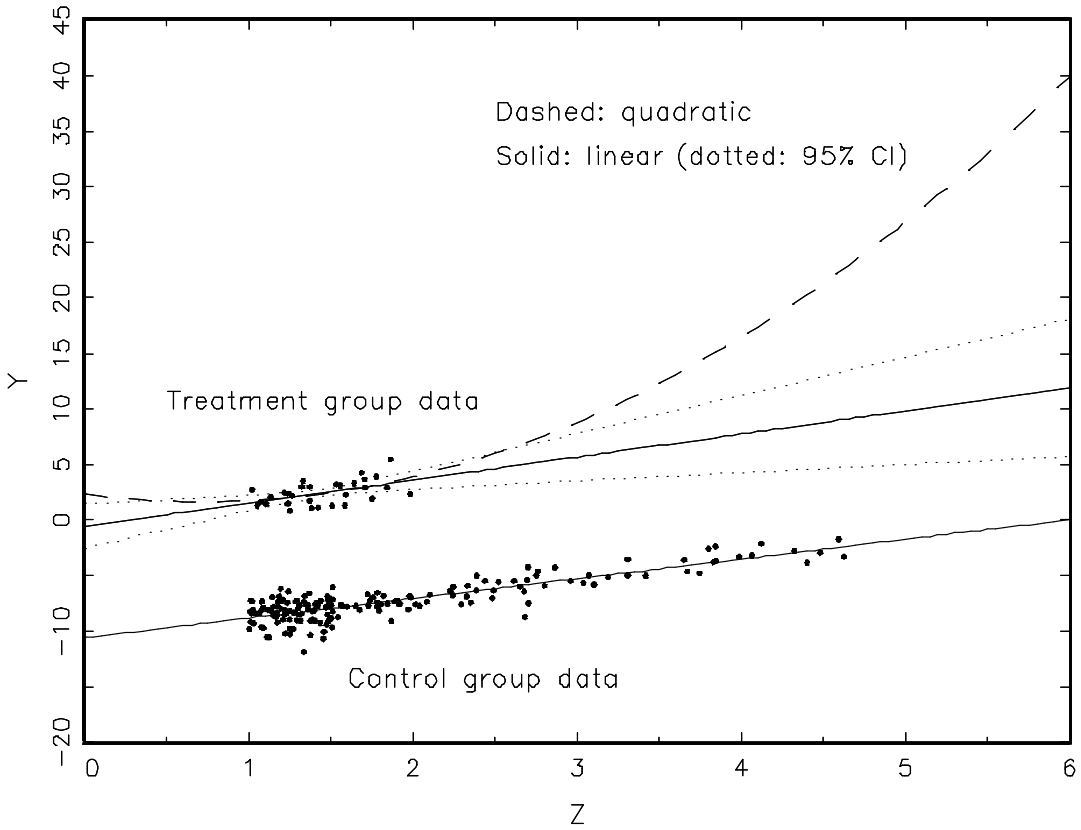
# Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)



# Model Dependence: A Simpler Example

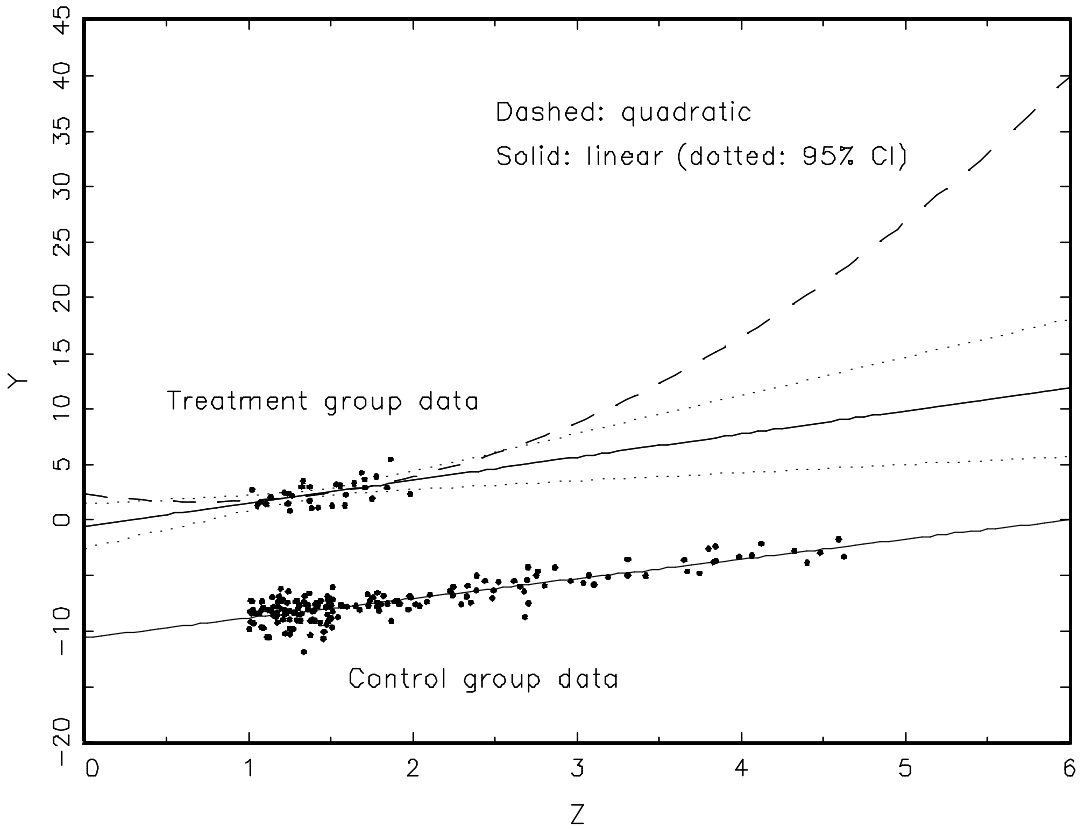
(King and Zeng, 2006: fig.4 *Political Analysis*)




What to do?

# Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

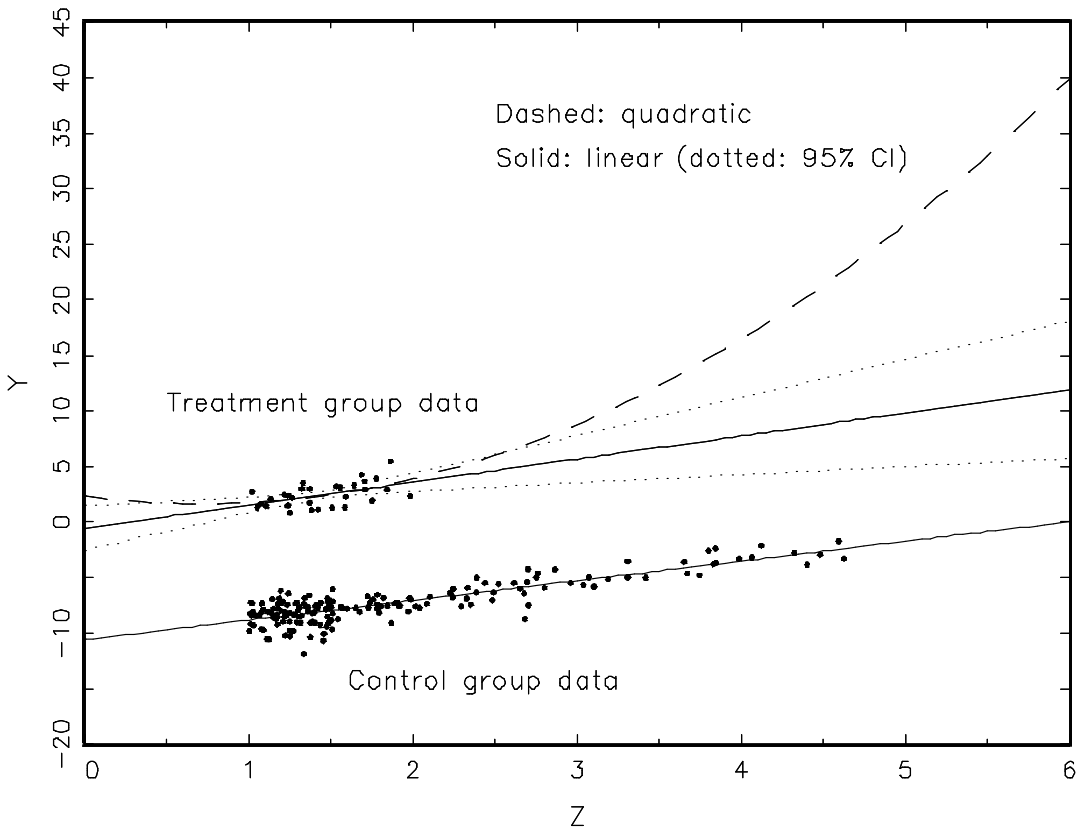



What to do?

- Preprocess I: Eliminate extrapolation region

# Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

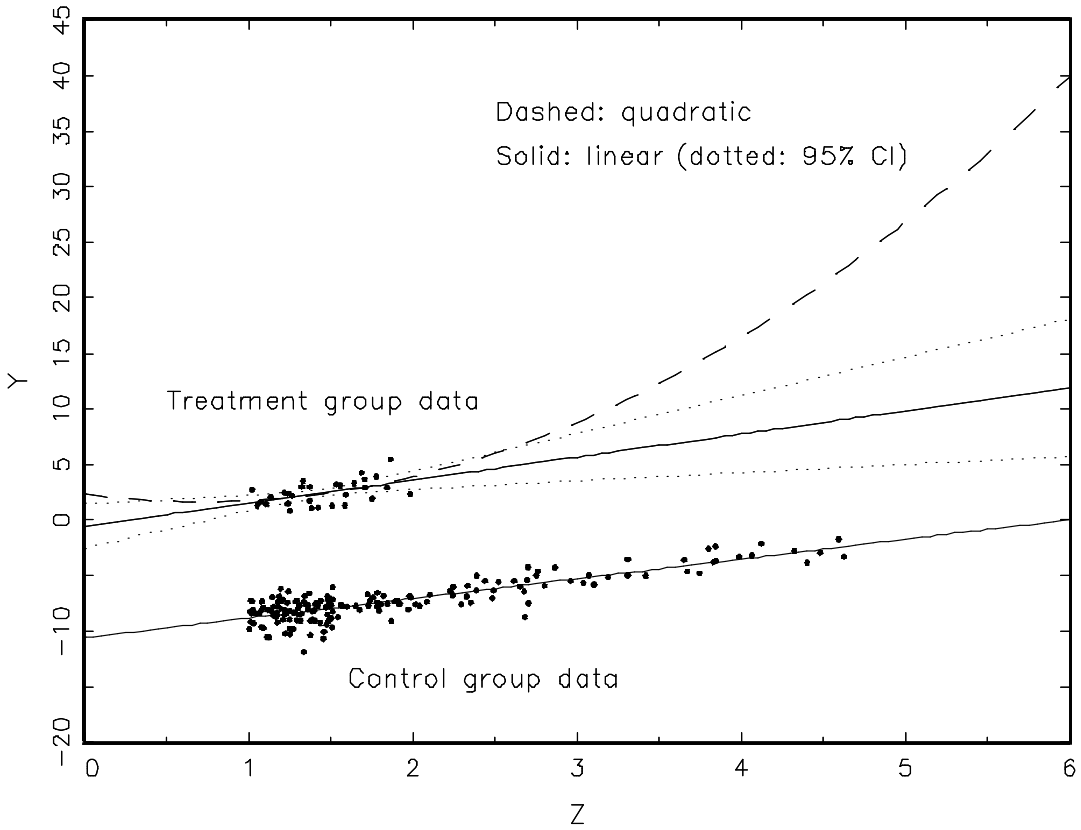



What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune bad matches) within interpolation region

# Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

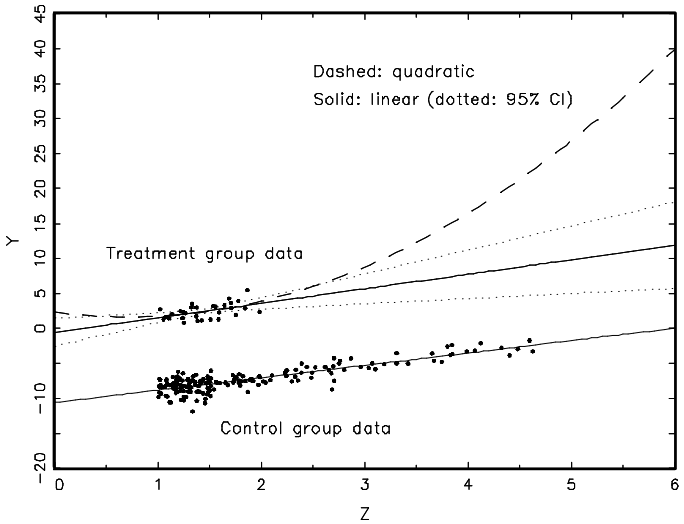



What to do?

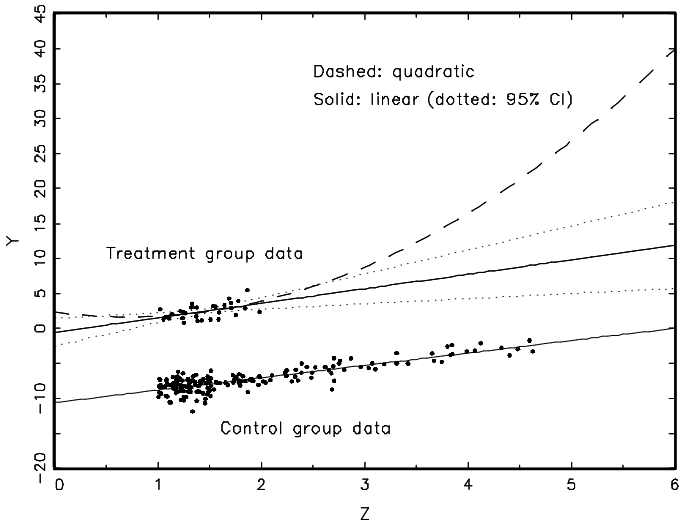
- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune bad matches) within interpolation region
- Model remaining imbalance

# Remove Extrapolation Region, then Match

# Remove Extrapolation Region, then Match

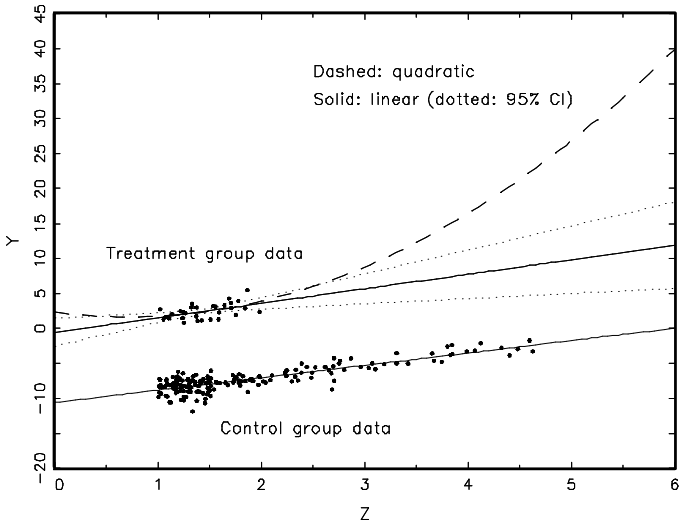


# Remove Extrapolation Region, then Match



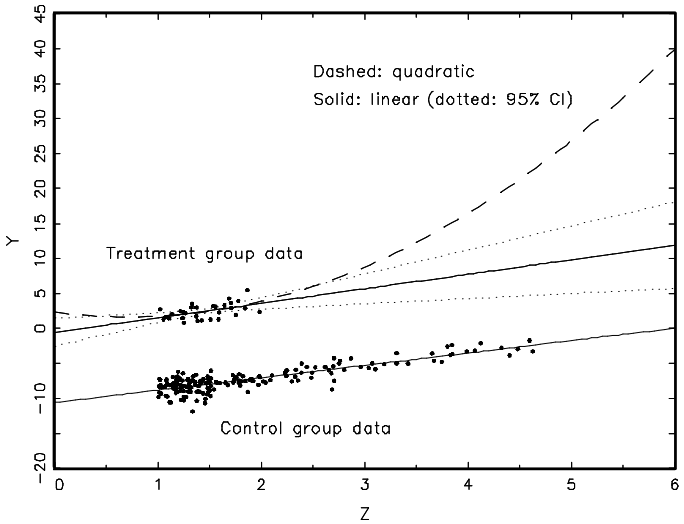

- Must remove data (selecting on $X$) to avoid extrapolation.

# Remove Extrapolation Region, then Match



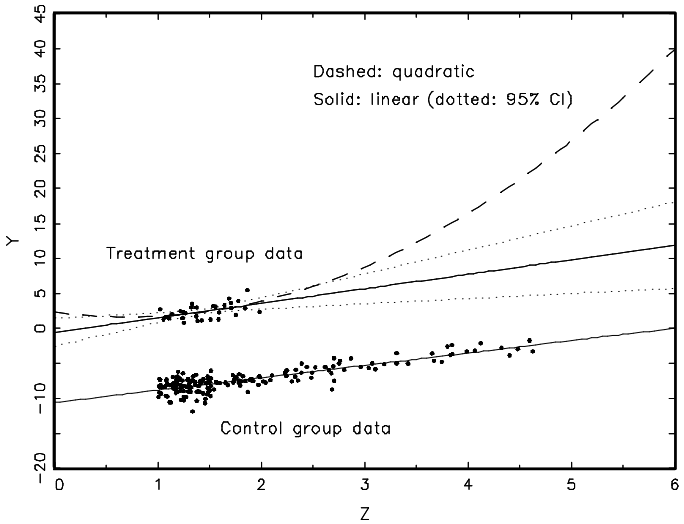

- Must remove data (selecting on $X$) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$

# Remove Extrapolation Region, then Match



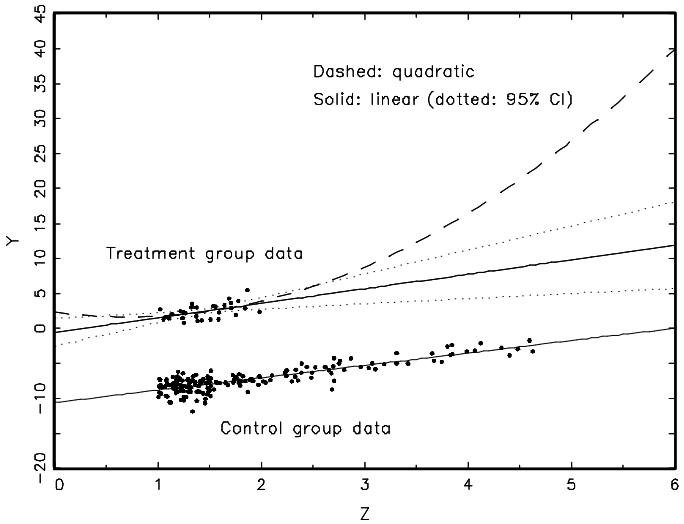

- Must remove data (selecting on $X$) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points

# Remove Extrapolation Region, then Match



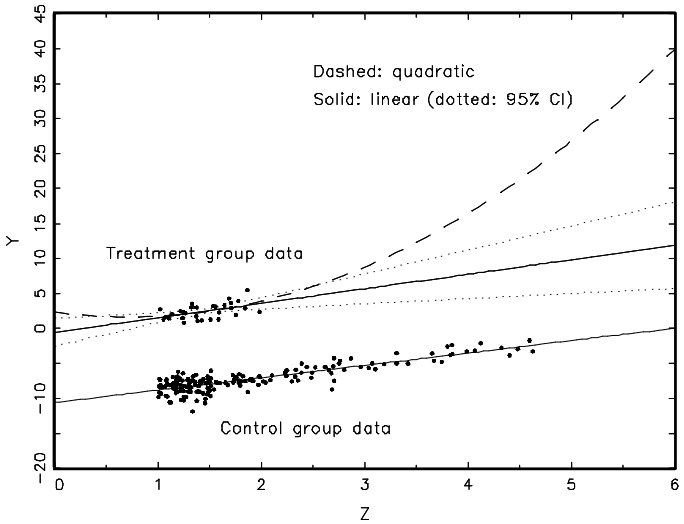

- Must remove data (selecting on $X$) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach

# Remove Extrapolation Region, then Match



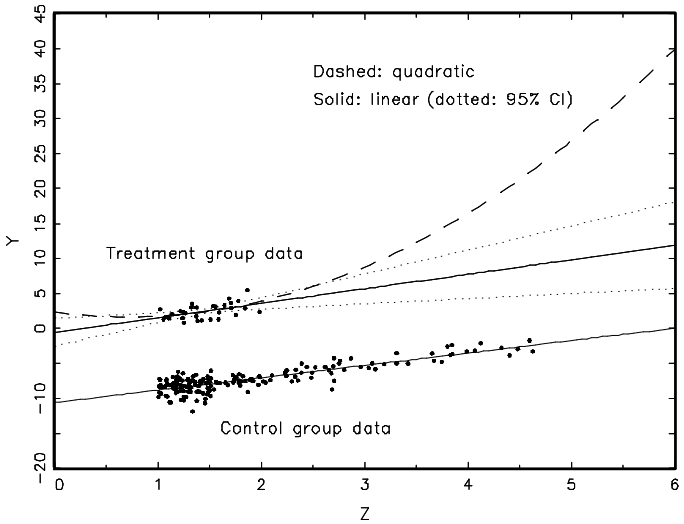

- Must remove data (selecting on $X$) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let $T^*$ and $X^*$ denote subsets of $T$ and $X$ s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$

# Remove Extrapolation Region, then Match



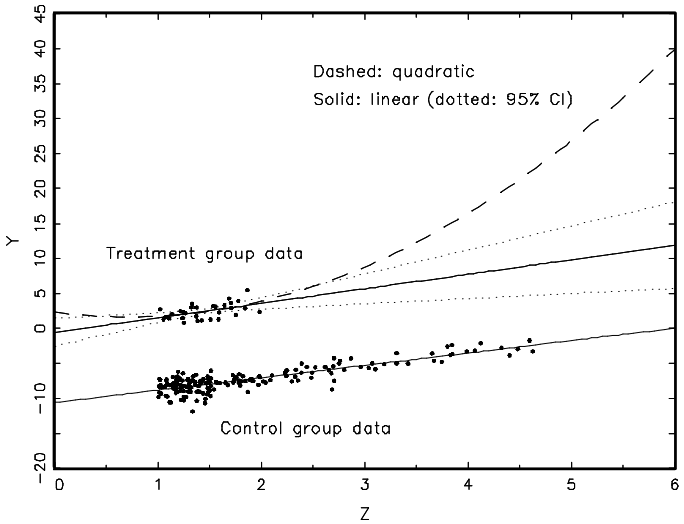

- Must remove data (selecting on $X$) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let $T^*$ and $X^*$ denote subsets of $T$ and $X$ s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$
     - use $X^*$ as estimate of common support (deleting remaining observations)

# Remove Extrapolation Region, then Match




- Must remove data (selecting on $X$) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let $T^*$ and $X^*$ denote subsets of $T$ and $X$ s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$
     - use $X^*$ as estimate of common support (deleting remaining observations)
  3. Other approaches, based on distance metrics, pscores, etc.

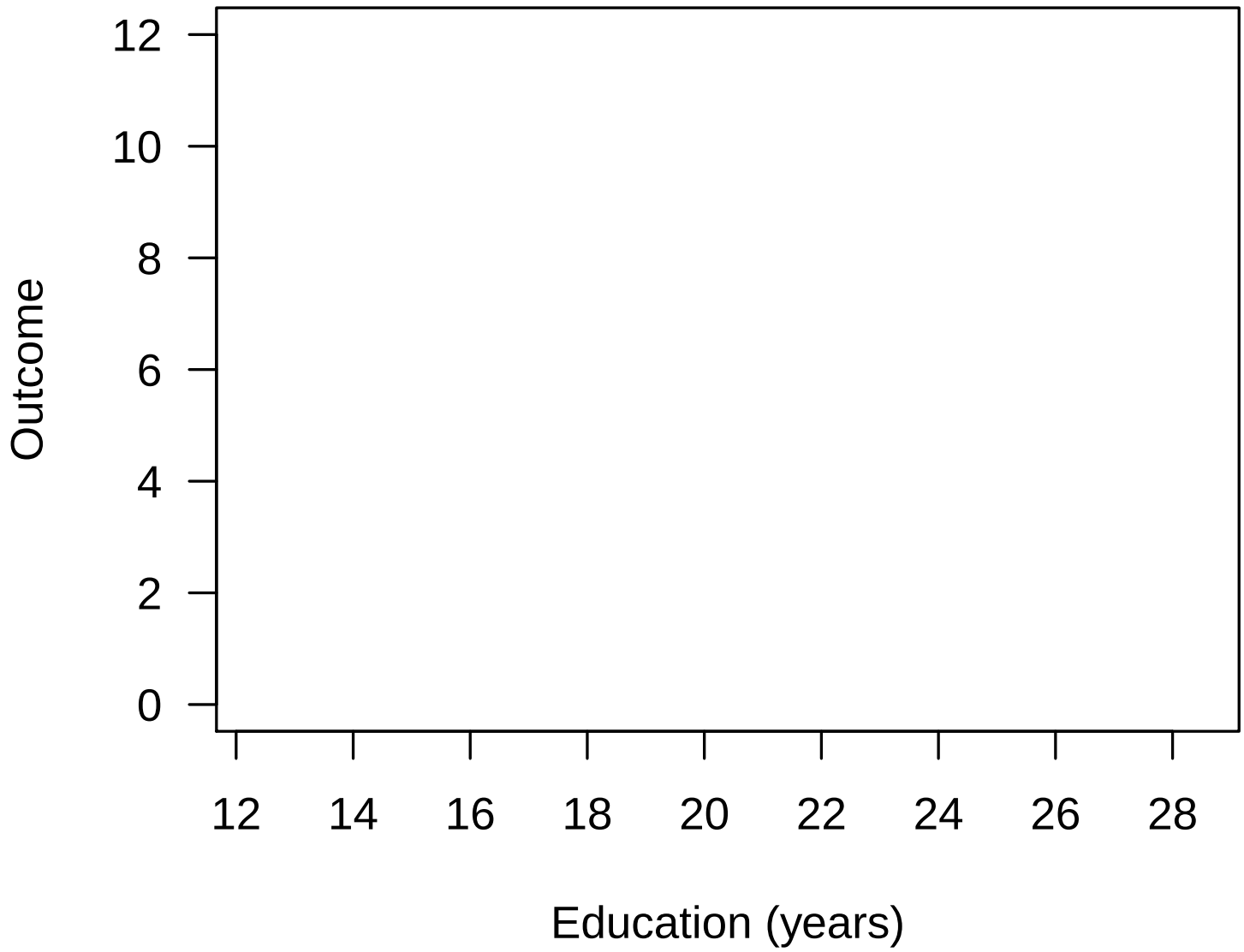
# Remove Extrapolation Region, then Match




- Must remove data (selecting on $X$) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let $T^*$ and $X^*$ denote subsets of $T$ and $X$ s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$
     - use $X^*$ as estimate of common support (deleting remaining observations)
  3. Other approaches, based on distance metrics, pscores, etc.
  4. Easiest: Coarsened Exact Matching, no separate step needed

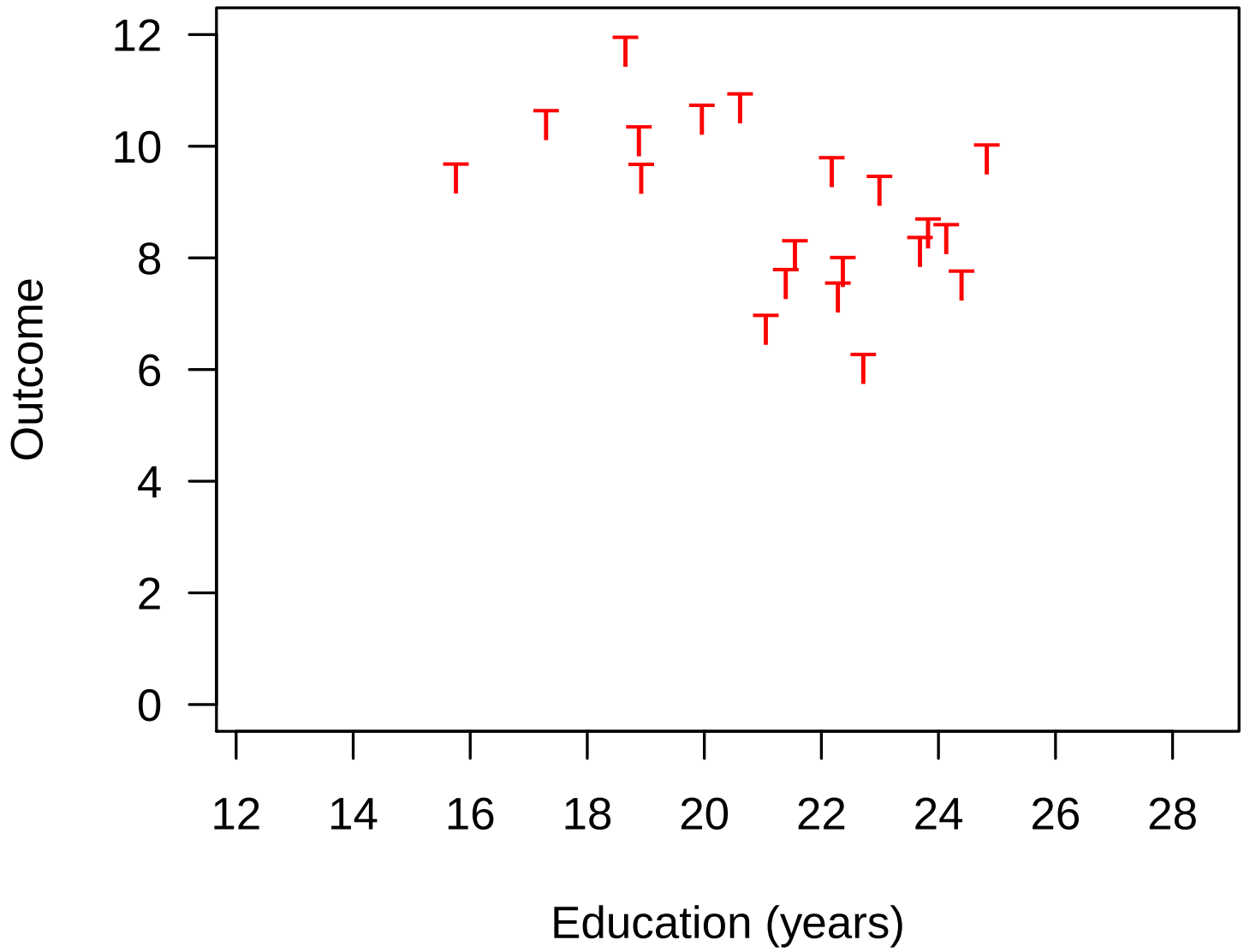
# Matching within the Interpolation Region

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



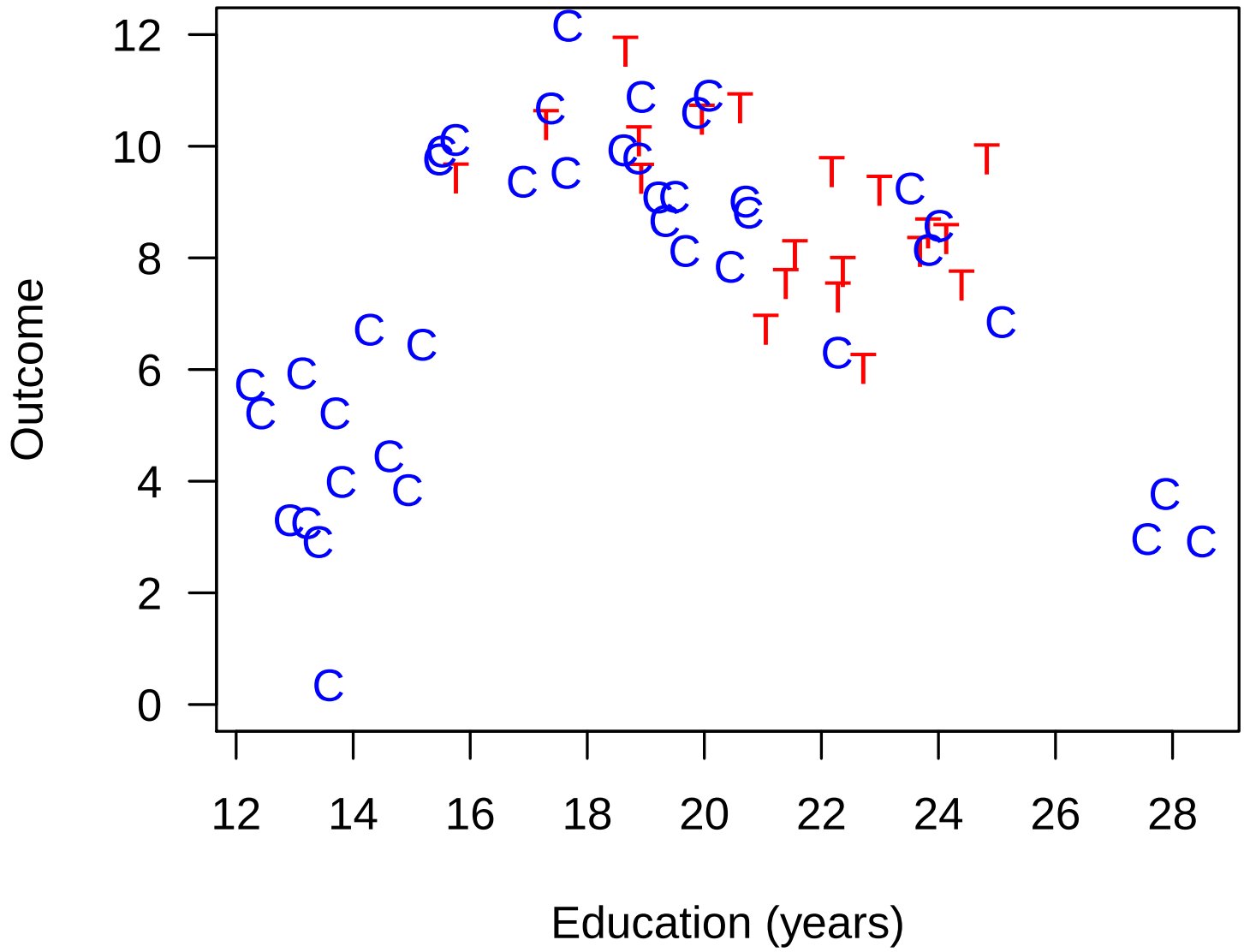
# Matching within the Interpolation Region

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



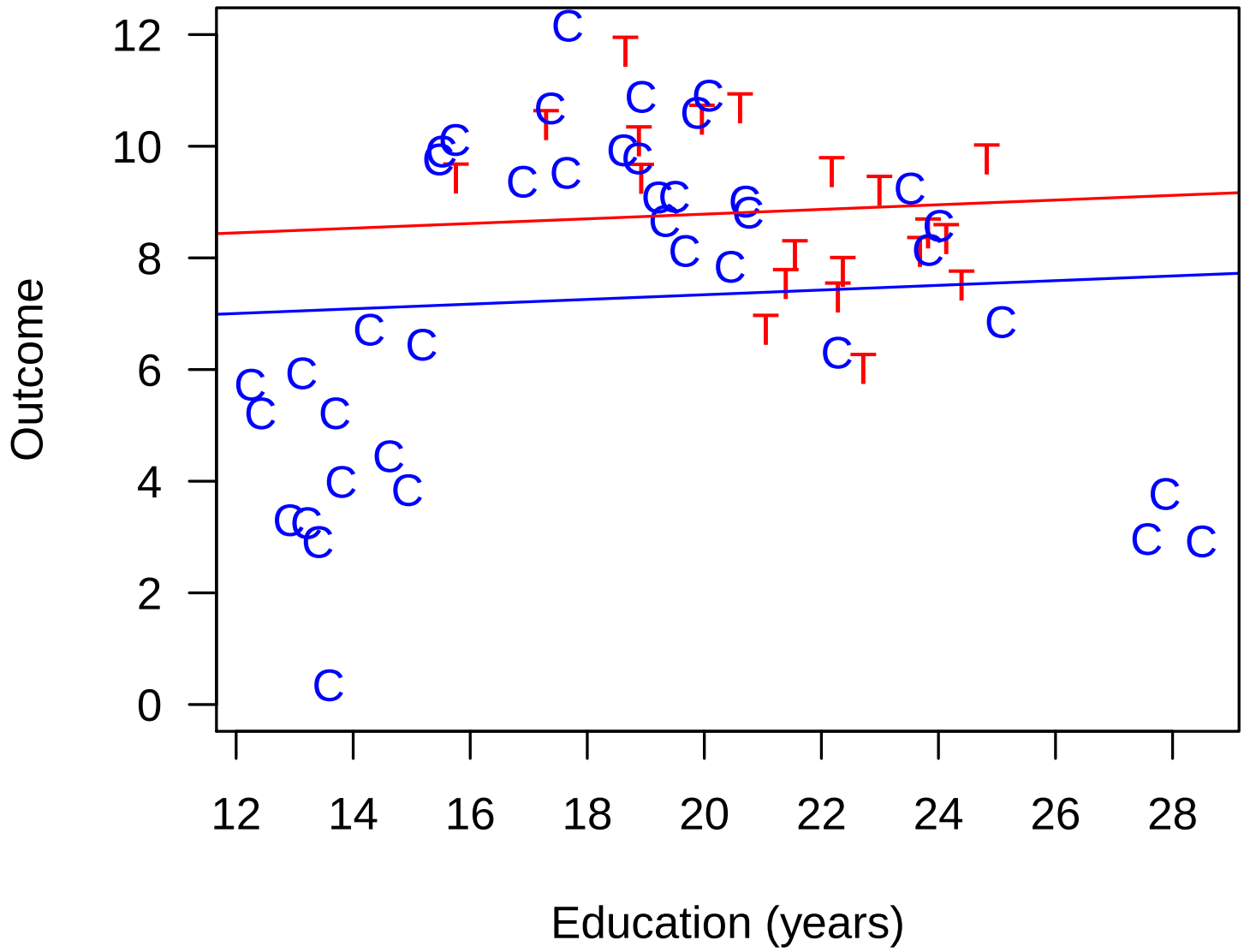
# Matching within the Interpolation Region

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



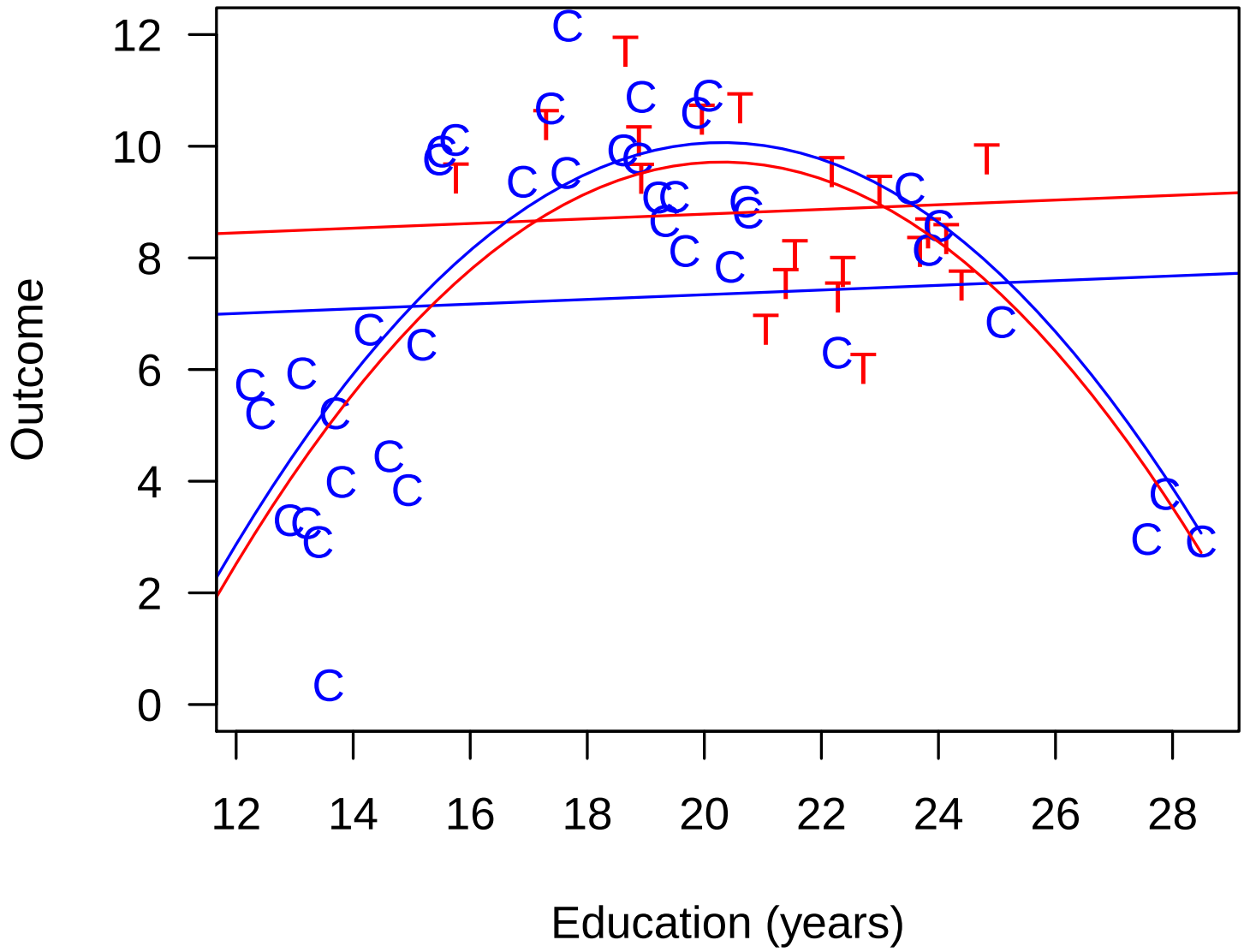
# Matching within the Interpolation Region

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



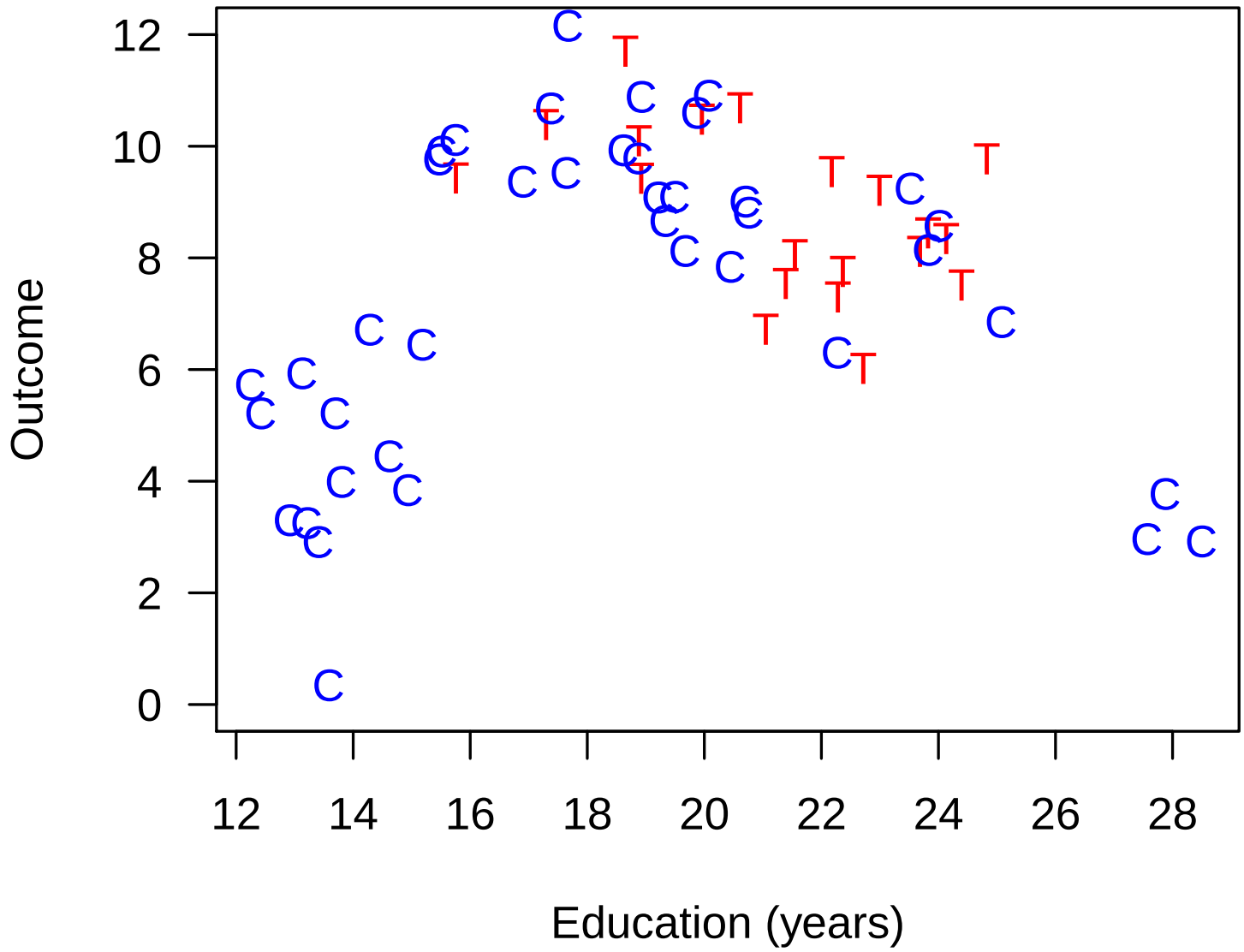
# Matching within the Interpolation Region

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



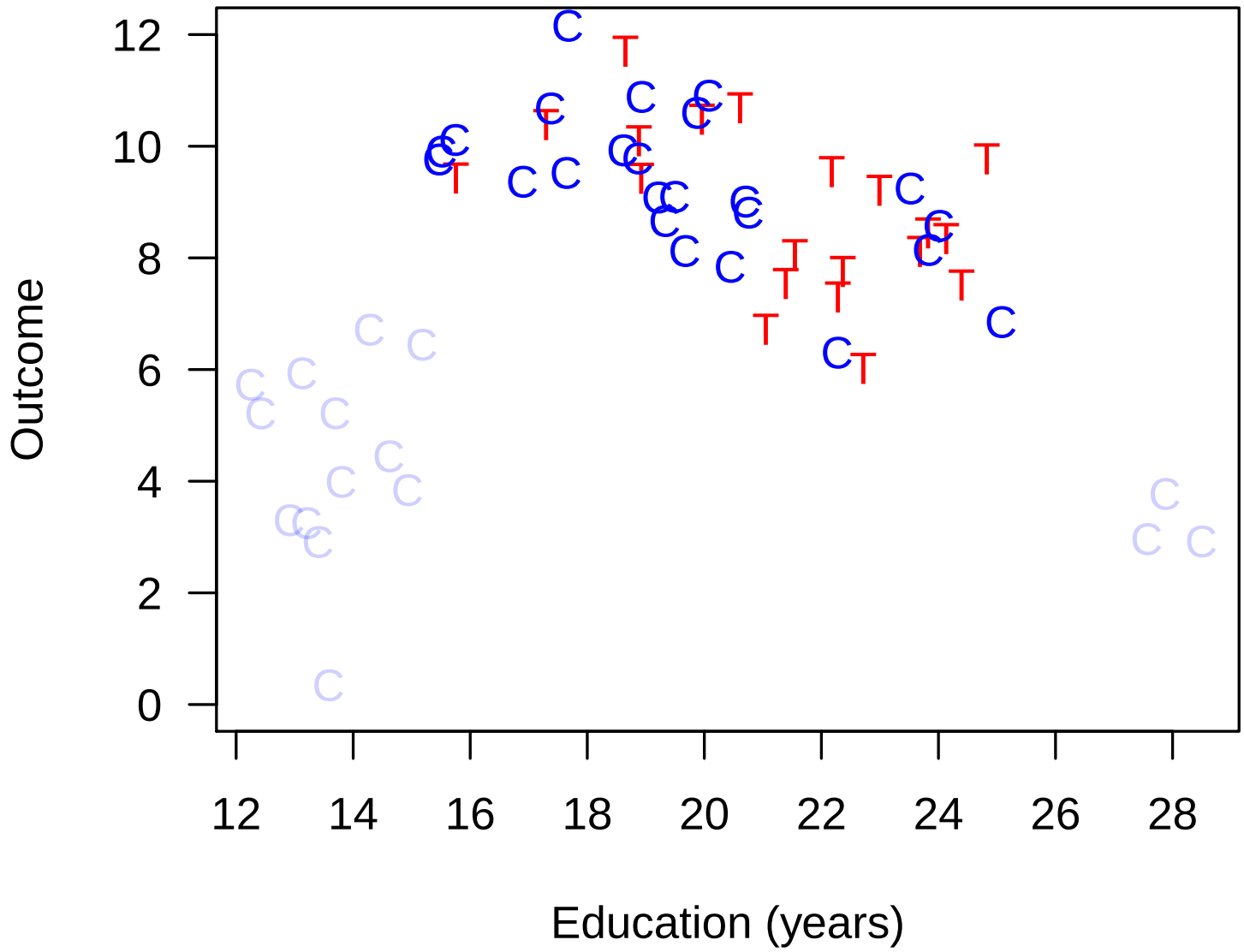
# Matching within the Interpolation Region

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



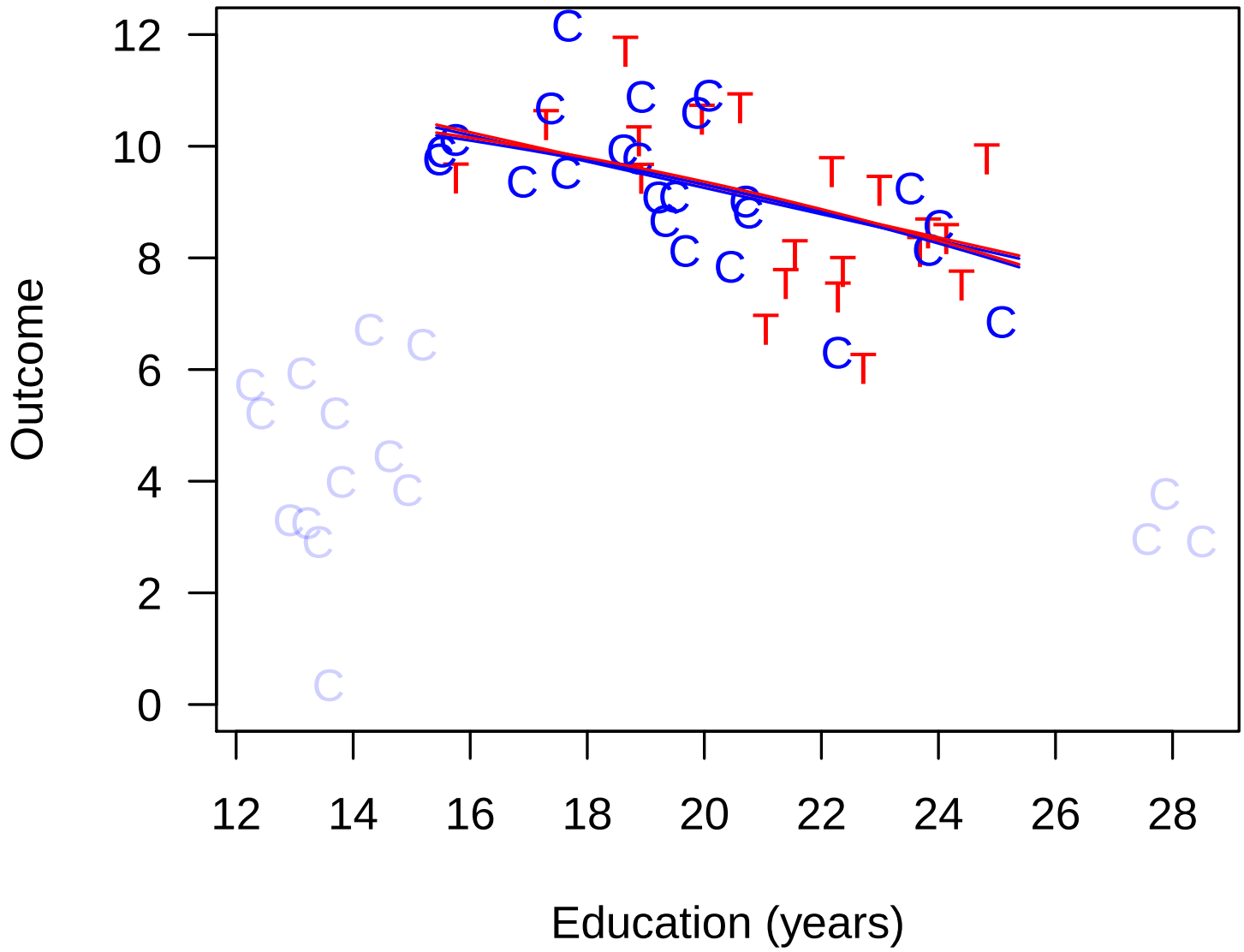
# Matching within the Interpolation Region

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



# Matching within the Interpolation Region

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



# Matching within the Interpolation Region

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

Matching reduces model dependence, bias, and variance

# Empirical Illustration: Carpenter, AJPS, 2002

# Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time

# Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).

# Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.

# Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.
- seven oversight variables (median adjusted ADA scores for House and Senate Committees as well as for House and Senate floors, Democratic Majority in House and Senate, and Democratic Presidency).

# Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.
- seven oversight variables (median adjusted ADA scores for House and Senate Committees as well as for House and Senate floors, Democratic Majority in House and Senate, and Democratic Presidency).
- 18 control variables (clinical factors, firm characteristics, media variables, etc.)

# Evaluating Reduction in Model Dependence

# Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).

# Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.

# Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.

# Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).

# Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).
- run 262,143 possible specifications and calculates ATE for each.

# Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).
- run 262,143 possible specifications and calculates ATE for each.
- Look at *variability* in ATE estimate across specifications.

# Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- omit post-treatment variables.
- use one-to-one nearest neighbor propensity score matching.
- discard 49 units (2 treated and 17 control units).
- run 262,143 possible specifications and calculates ATE for each.
- Look at *variability* in ATE estimate across specifications.
- (Normal applications would only do one or a small number of specifications.)

# Reducing Model Dependence

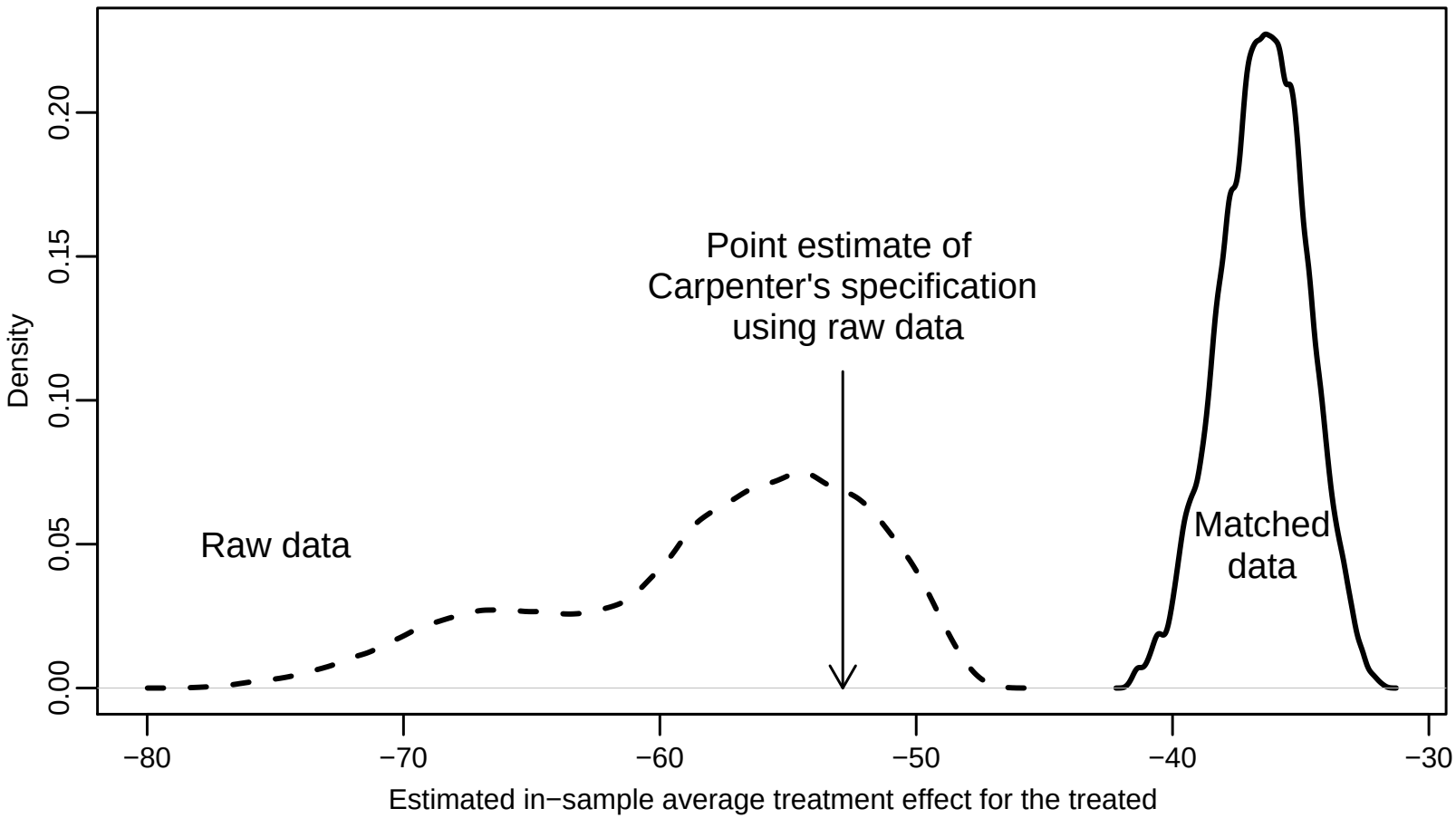



Figure: Histogram of estimated in-sample average treatment effect for the treated (ATT) of the Democratic Senate majority on FDA drug approval time across 262,143 specifications.

# Another Example: Jeffrey Koch, AJPS, 2002

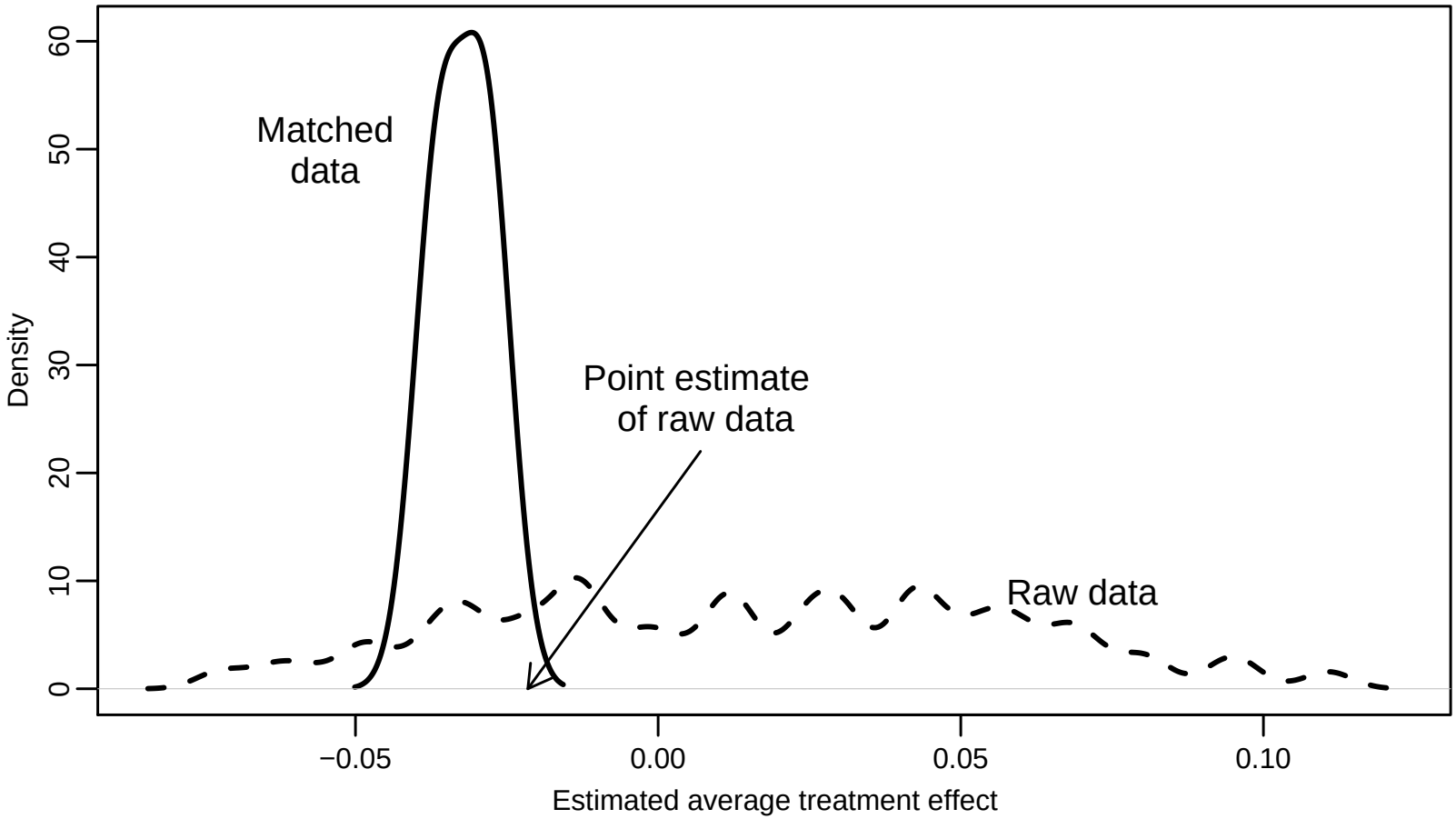



Figure: Estimated effects of being a highly visible female Republican candidate across 63 possible specifications with the Koch data.

# How Matching Works

# How Matching Works

- Notation:

# How Matching Works

- Notation:

  $Y_i$ Dependent variable

# How Matching Works

- Notation:
  - $Y_i$ Dependent variable
  - $T_i$ Treatment variable (0/1)

# How Matching Works

- Notation:
  - $Y_i$ Dependent variable
  - $T_i$ Treatment variable (0/1)
  - $X_i$ Pre-treatment covariates

# How Matching Works

- Notation:
  - $Y_i$ Dependent variable
  - $T_i$ Treatment variable (0/1)
  - $X_i$ Pre-treatment covariates
- Treatment Effect for treated ($T_i = 1$) observation $i$:

# How Matching Works

- Notation:
  - $Y_i$ Dependent variable
  - $T_i$ Treatment variable (0/1)
  - $X_i$ Pre-treatment covariates
- Treatment Effect for treated ($T_i = 1$) observation $i$:

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

# How Matching Works

- Notation:
  - $Y_i$ Dependent variable
  - $T_i$ Treatment variable (0/1)
  - $X_i$ Pre-treatment covariates
- Treatment Effect for treated ($T_i = 1$) observation $i$:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

# How Matching Works

- Notation:
  - $Y_i$ Dependent variable
  - $T_i$ Treatment variable (0/1)
  - $X_i$ Pre-treatment covariates
- Treatment Effect for treated ($T_i = 1$) observation $i$:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls $\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$

# How Matching Works

- Notation:
  - $Y_i$ Dependent variable
  - $T_i$ Treatment variable (0/1)
  - $X_i$ Pre-treatment covariates
- Treatment Effect for treated ($T_i = 1$) observation $i$:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
  $\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$
- Prune unmatched units to improve balance (so $X$ is unimportant)

# How Matching Works

- Notation:
  - $Y_i$ Dependent variable
  - $T_i$ Treatment variable (0/1)
  - $X_i$ Pre-treatment covariates
- Treatment Effect for treated ($T_i = 1$) observation $i$:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
  $\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$
- Prune unmatched units to improve balance (so $X$ is unimportant)
- QoI: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} \text{TE}_i$$

# How Matching Works

- Notation:
  - $Y_i$ Dependent variable
  - $T_i$ Treatment variable (0/1)
  - $X_i$ Pre-treatment covariates
- Treatment Effect for treated ($T_i = 1$) observation $i$:

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} \quad - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with $Y_j$ from matched ($X_i \approx X_j$) controls
  $\hat{Y}_i(0) = Y_j(0)$ or a model $\hat{Y}_i(0) = \hat{g}_0(X_j)$
- Prune unmatched units to improve balance (so $X$ is unimportant)
- QoI: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i = 1\}} \text{TE}_i$$

- or Feasible Average Treatment effect on the Treated: FSATT

# Method 1: Mahalanobis Distance Matching

# Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model

# Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$

2. Estimation Difference in means or a model

# Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit

2. Estimation Difference in means or a model

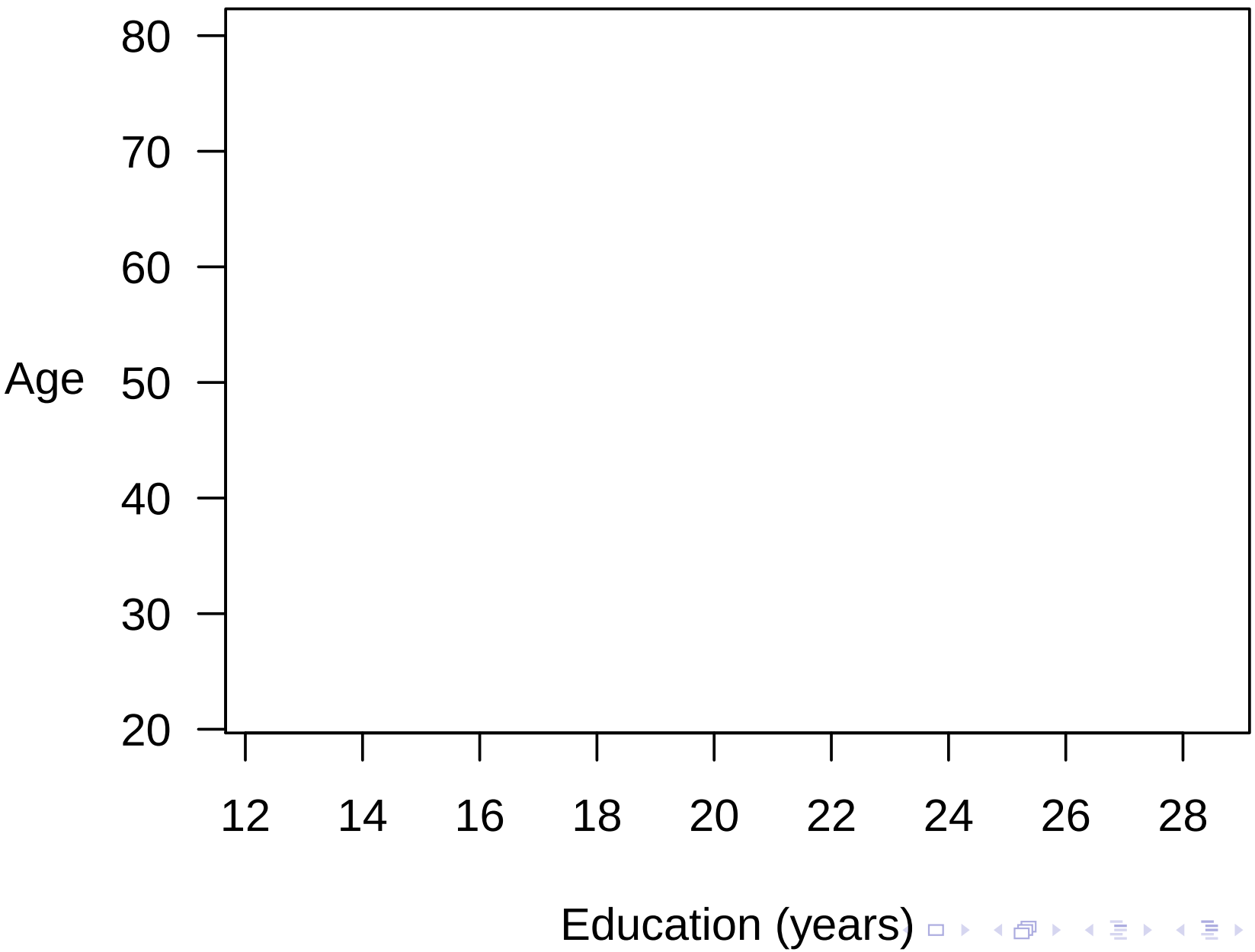
# Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
2. Estimation Difference in means or a model

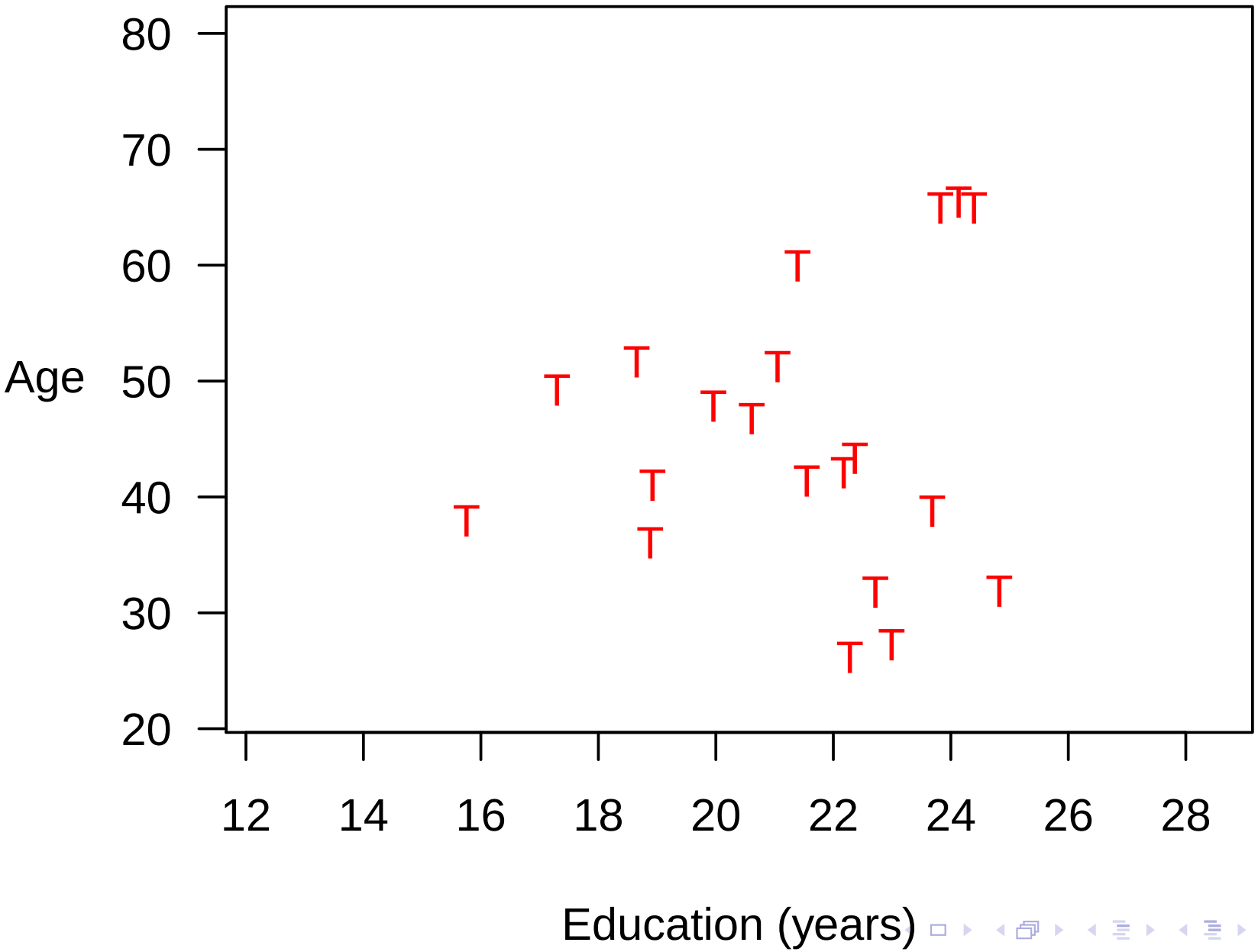
# Method 1: Mahalanobis Distance Matching

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - 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

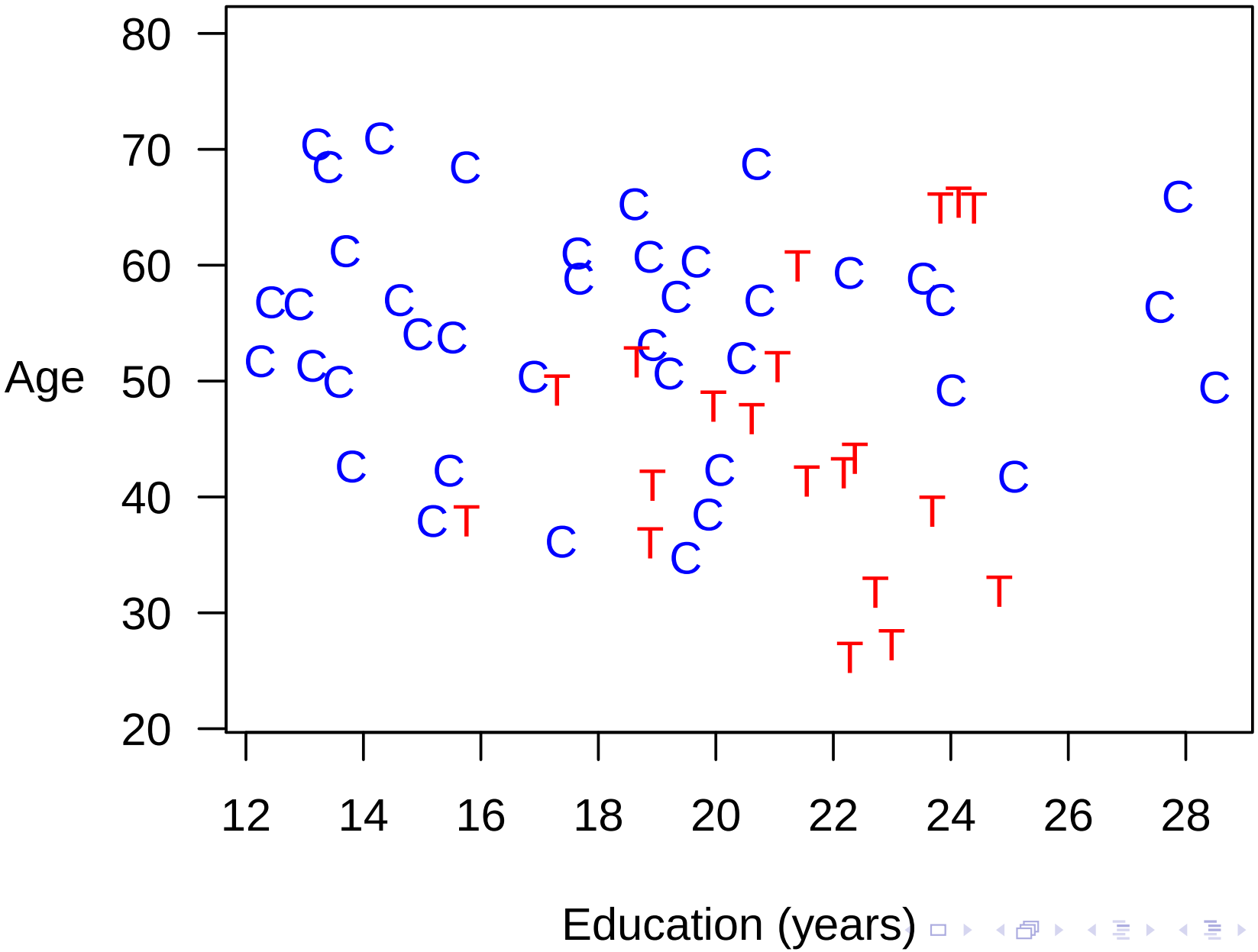
# Mahalanobis Distance Matching



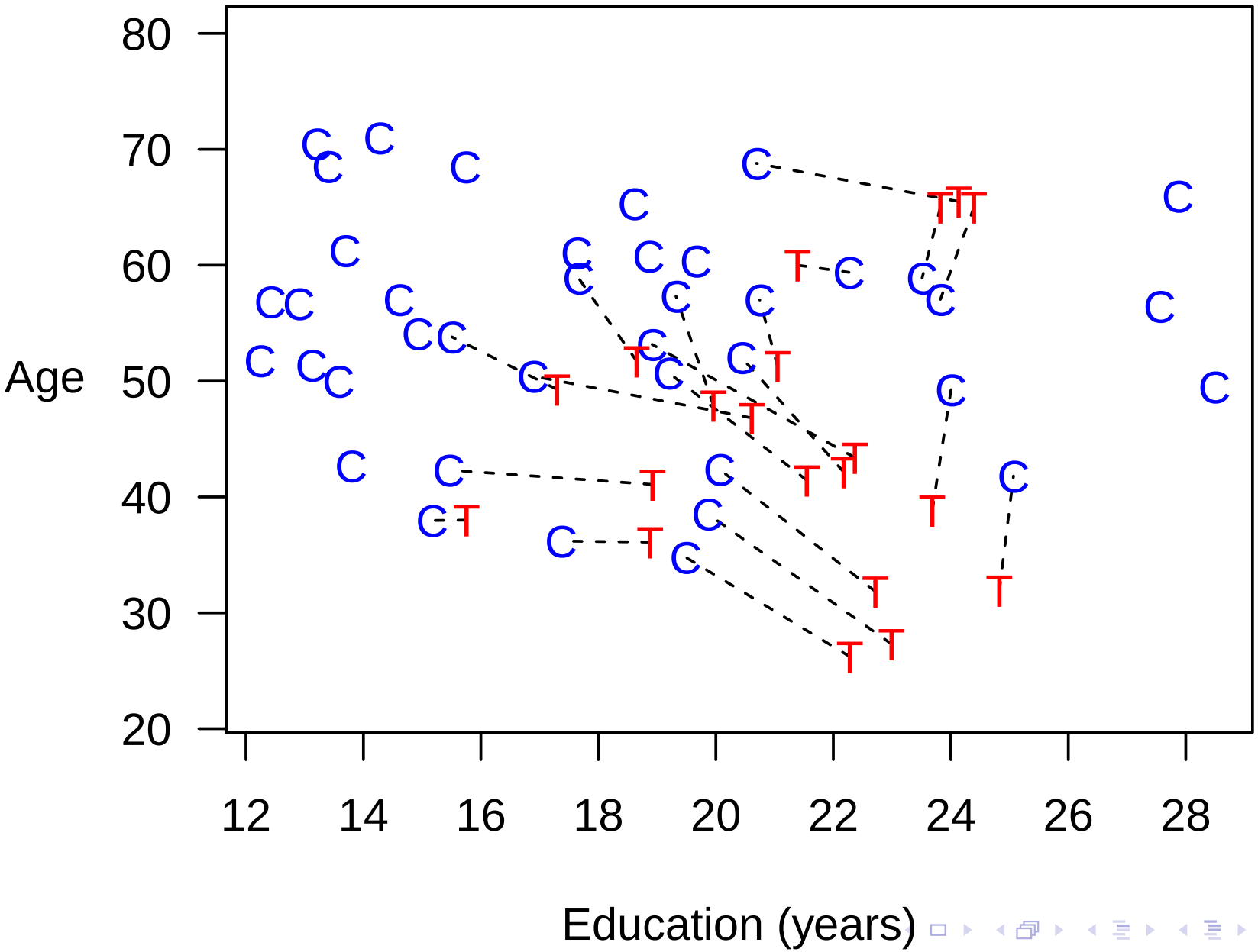
# Mahalanobis Distance Matching



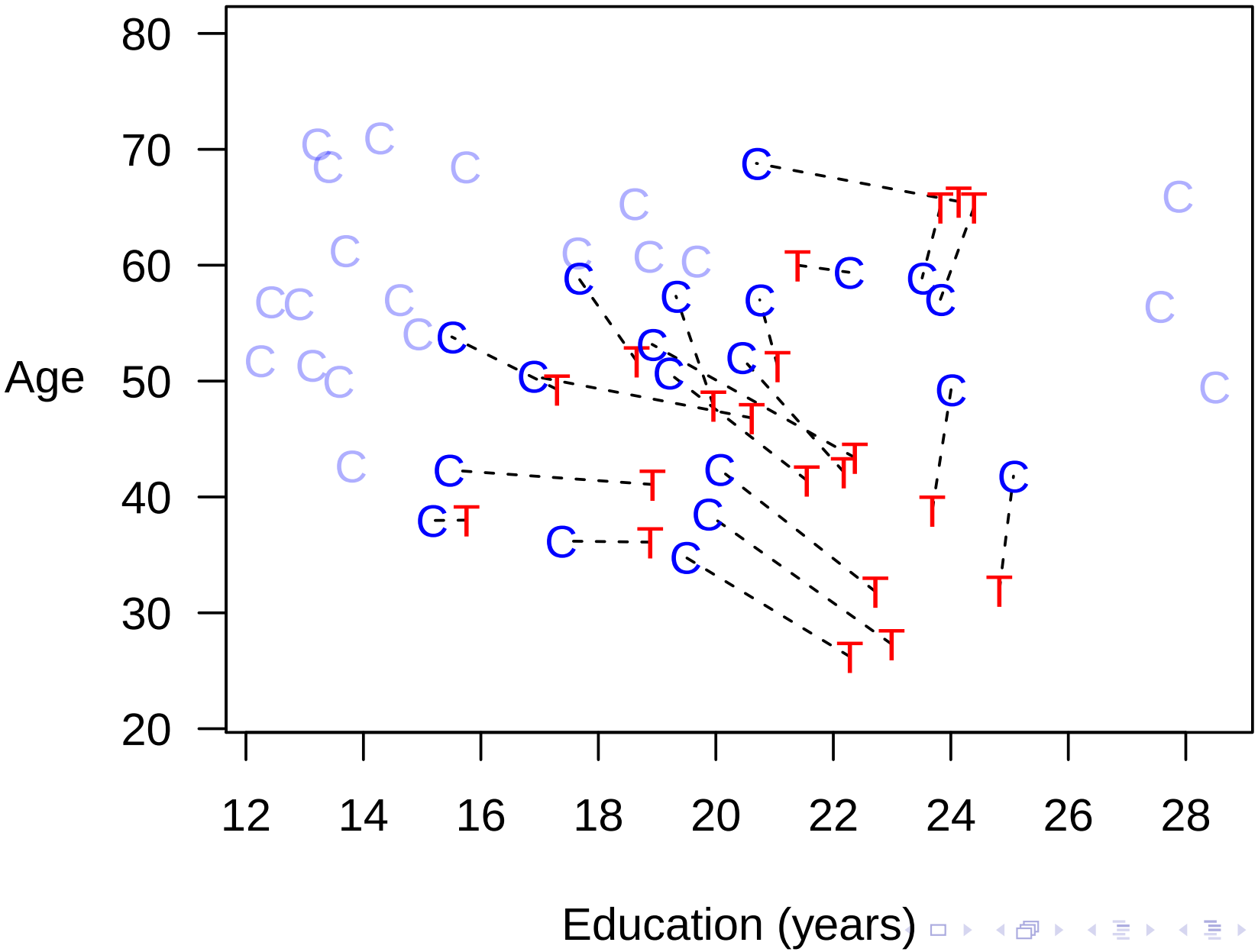
# Mahalanobis Distance Matching



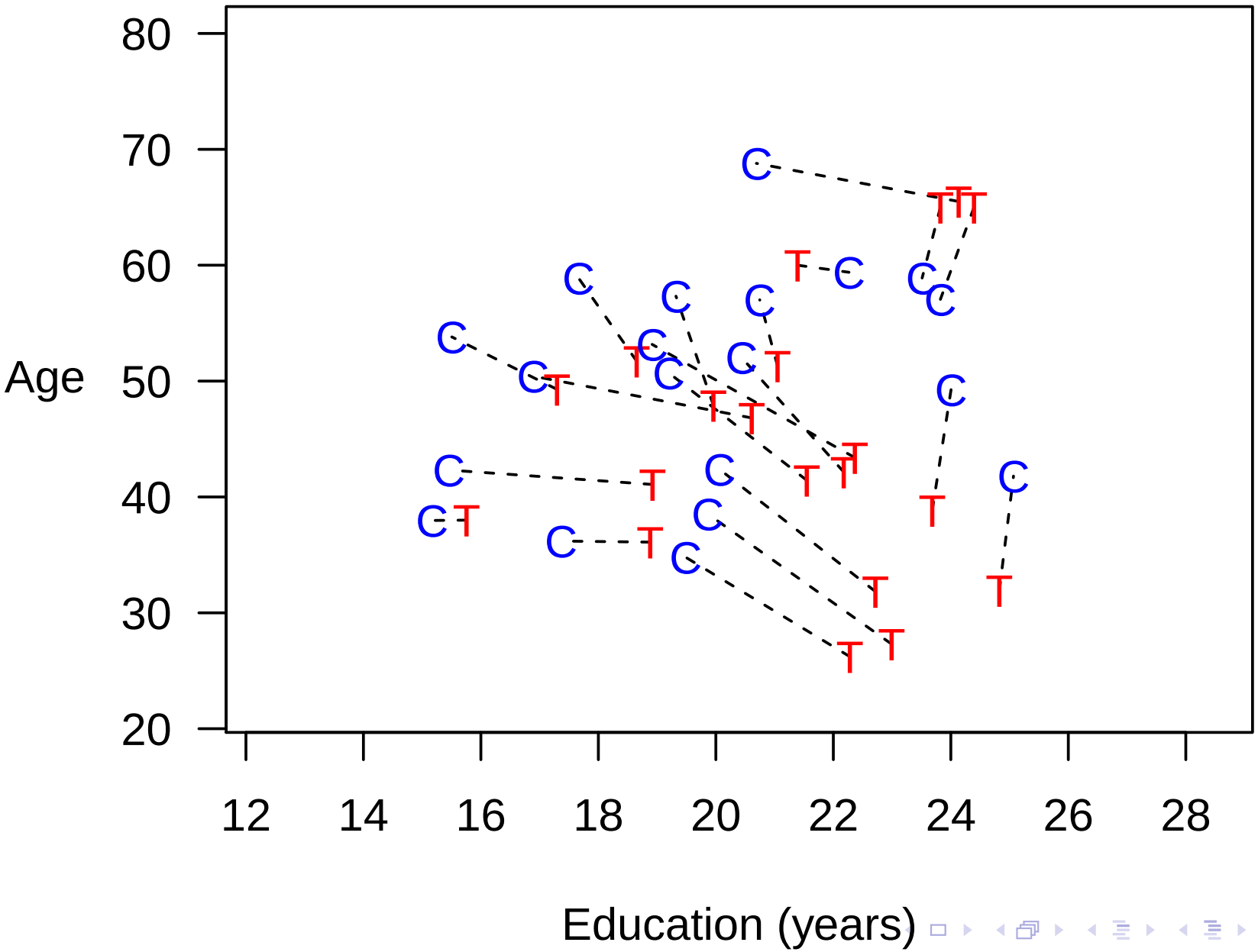
# Mahalanobis Distance Matching



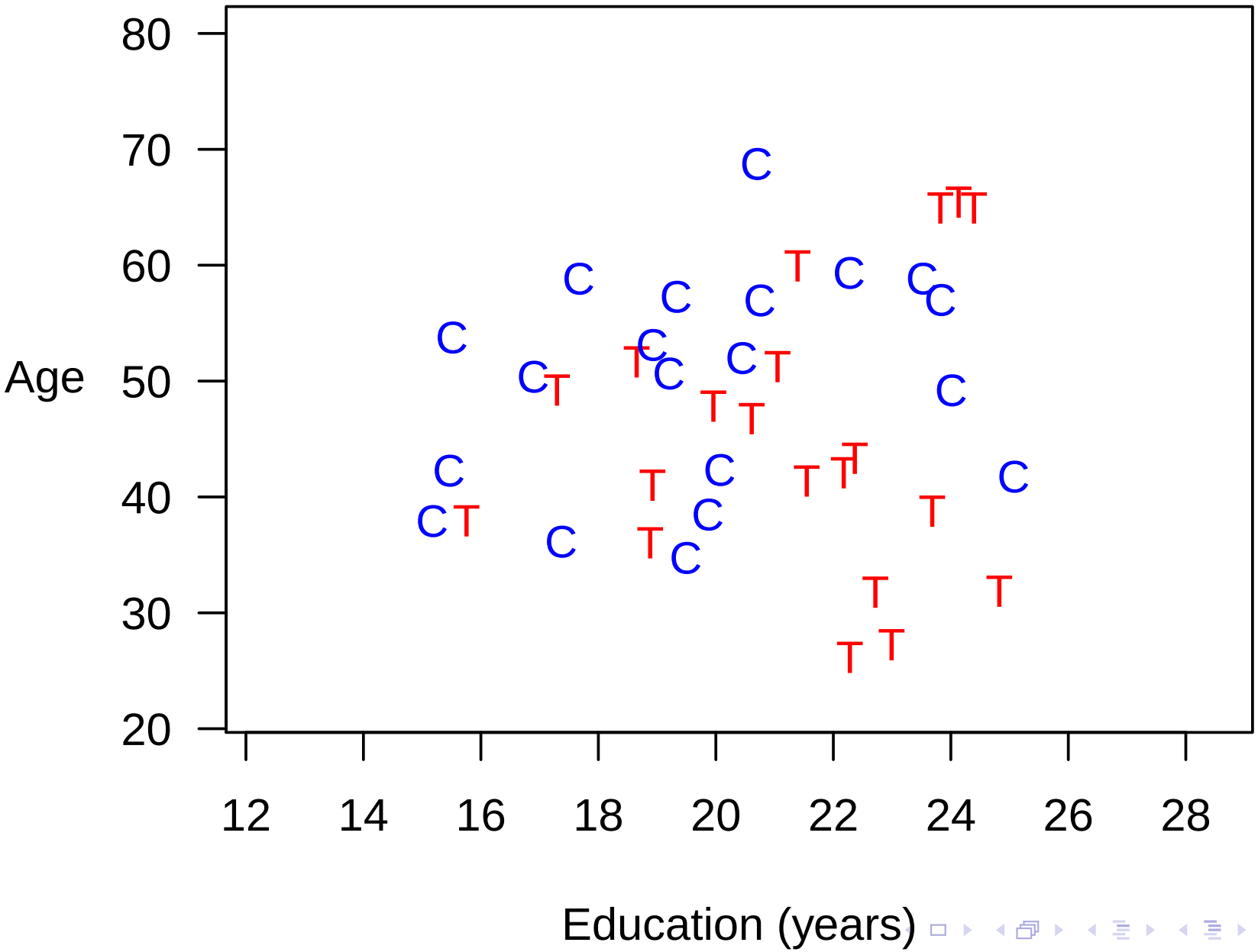
# Mahalanobis Distance Matching



# Mahalanobis Distance Matching



# Mahalanobis Distance Matching



# Method 2: Propensity Score Matching

# Method 2: Propensity Score Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model

# Method 2: Propensity Score Matching

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 2: Propensity Score Matching

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$(X_i, X_j) = |\pi_i - \pi_j|$

2. Estimation Difference in means or a model

# Method 2: Propensity Score Matching

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$(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit

2. Estimation Difference in means or a model

# Method 2: Propensity Score Matching

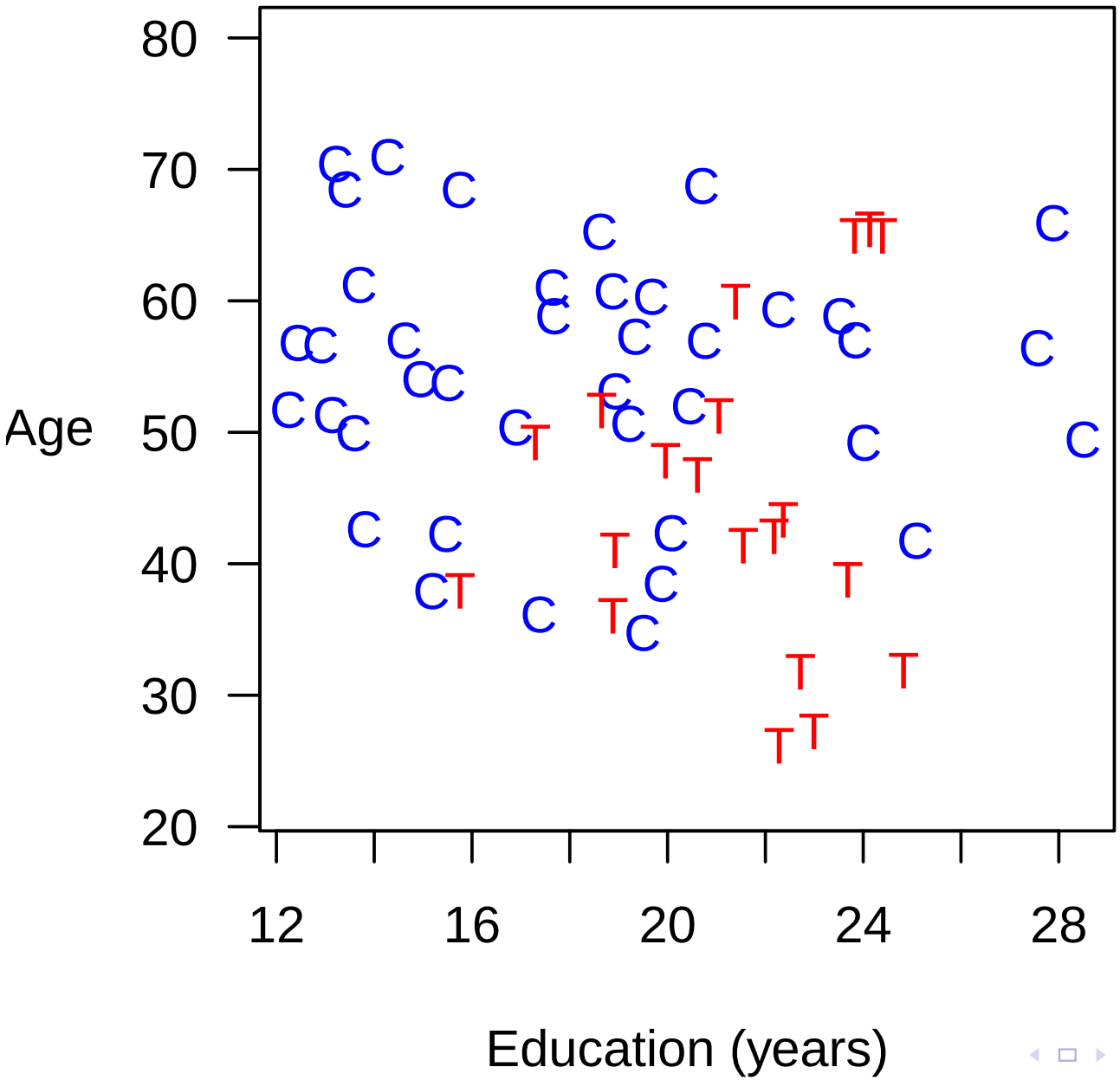
1. Preprocess (Matching)
   - Reduce $k$ elements of $X$ to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - Distance$(X_i, X_j) = |\pi_i - \pi_j|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. Estimation Difference in means or a model

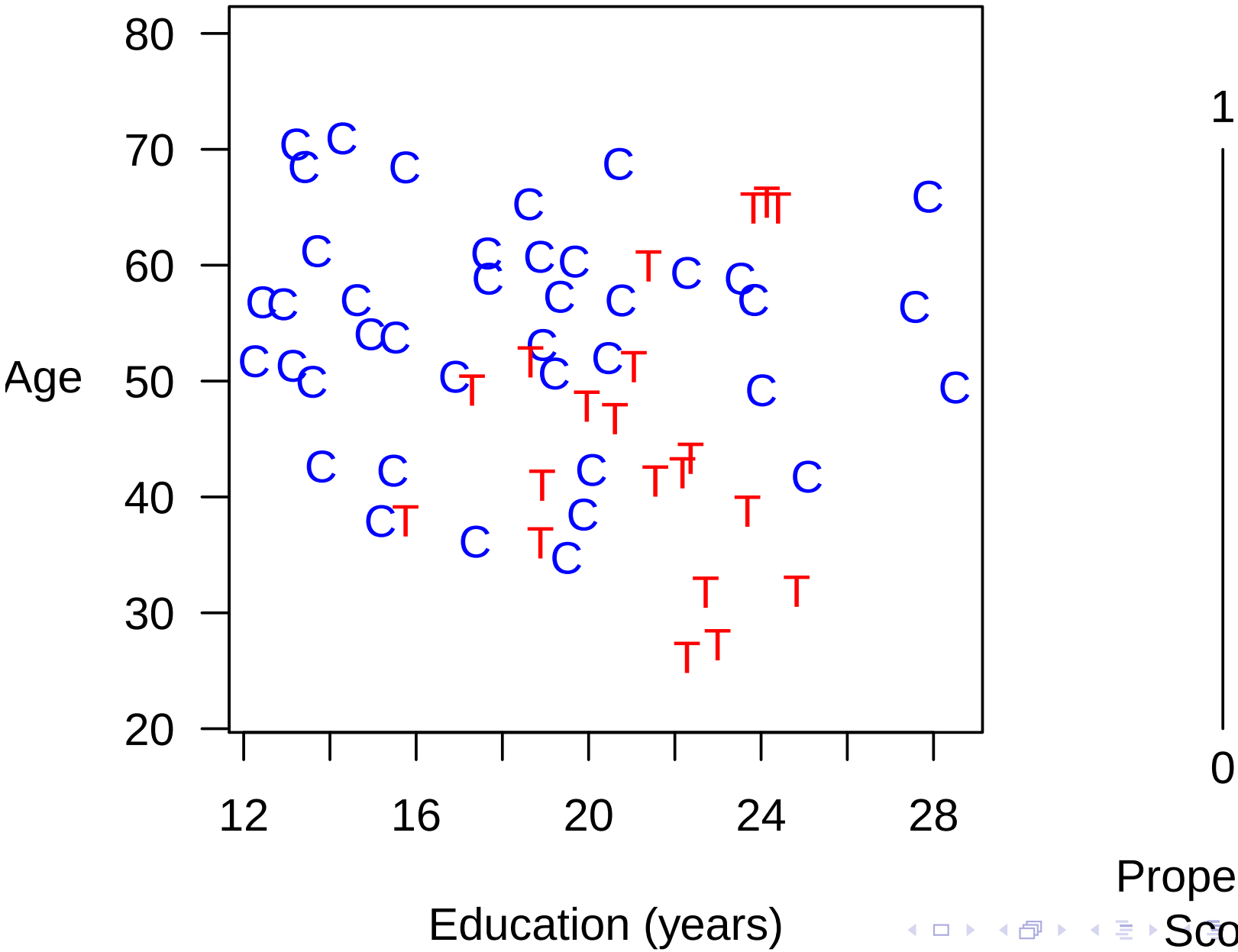
# Method 2: Propensity Score Matching

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$(X_i, X_j) = |\pi_i - \pi_j|$
   - 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

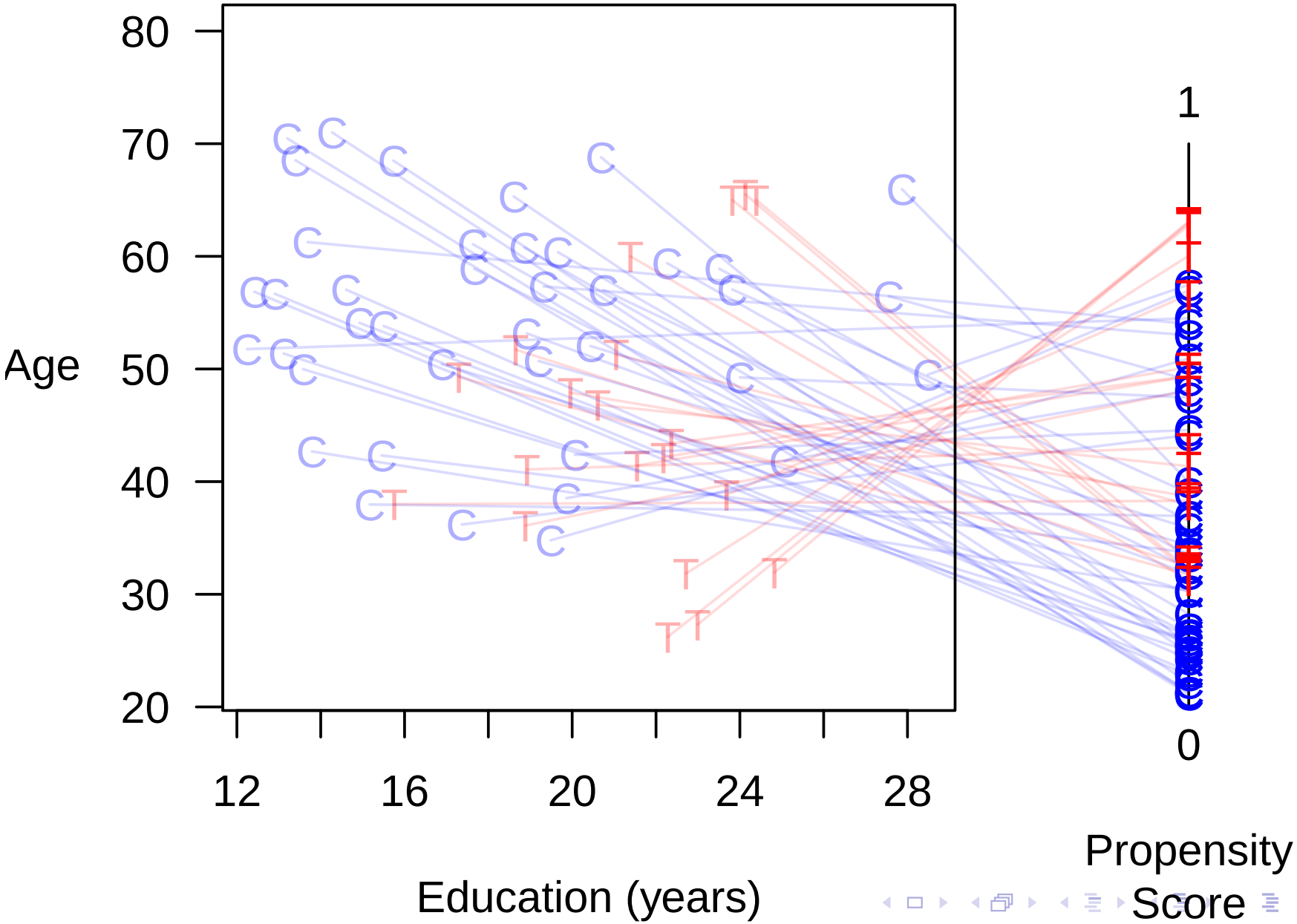
# Propensity Score Matching



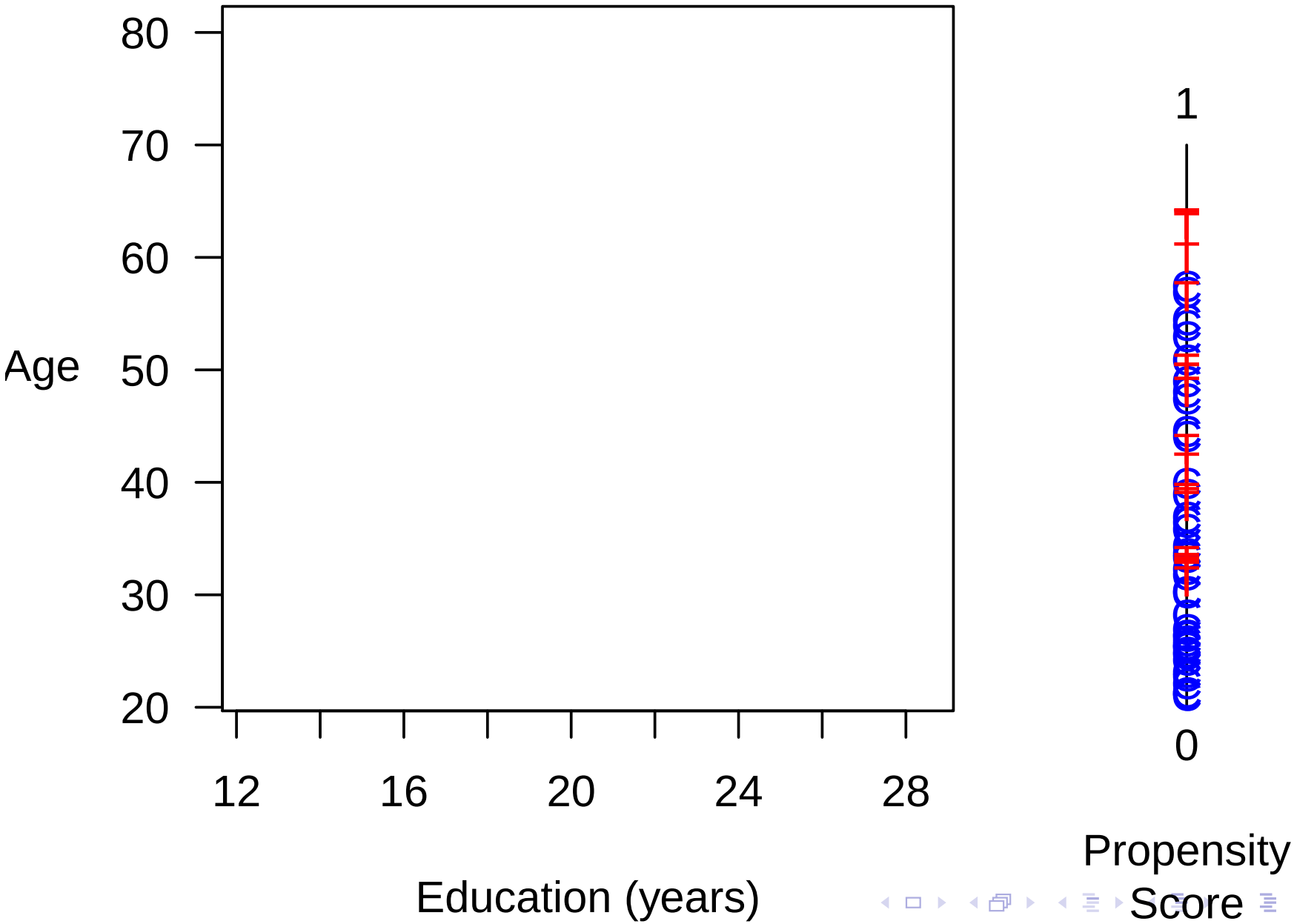
# Propensity Score Matching



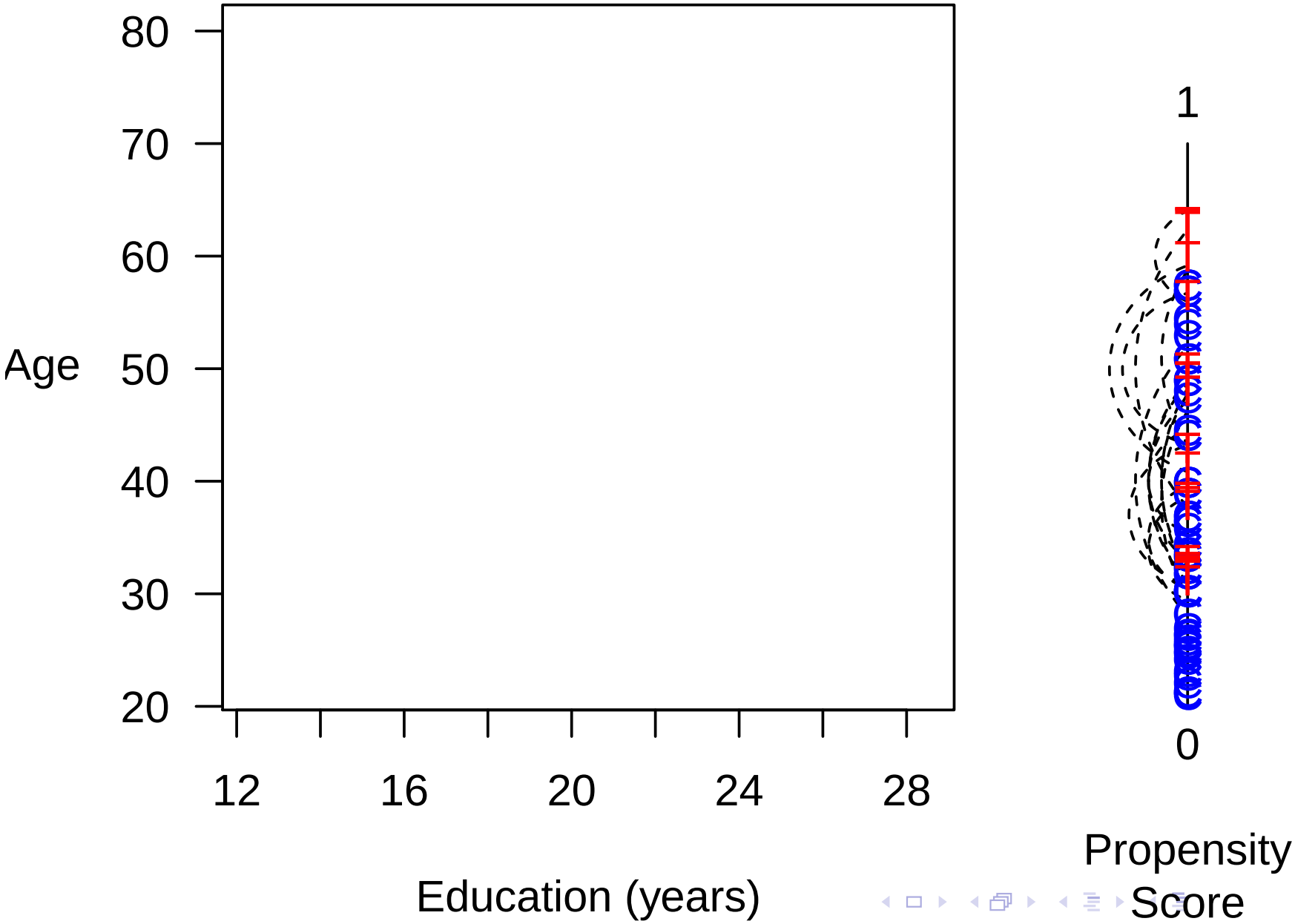
# Propensity Score Matching



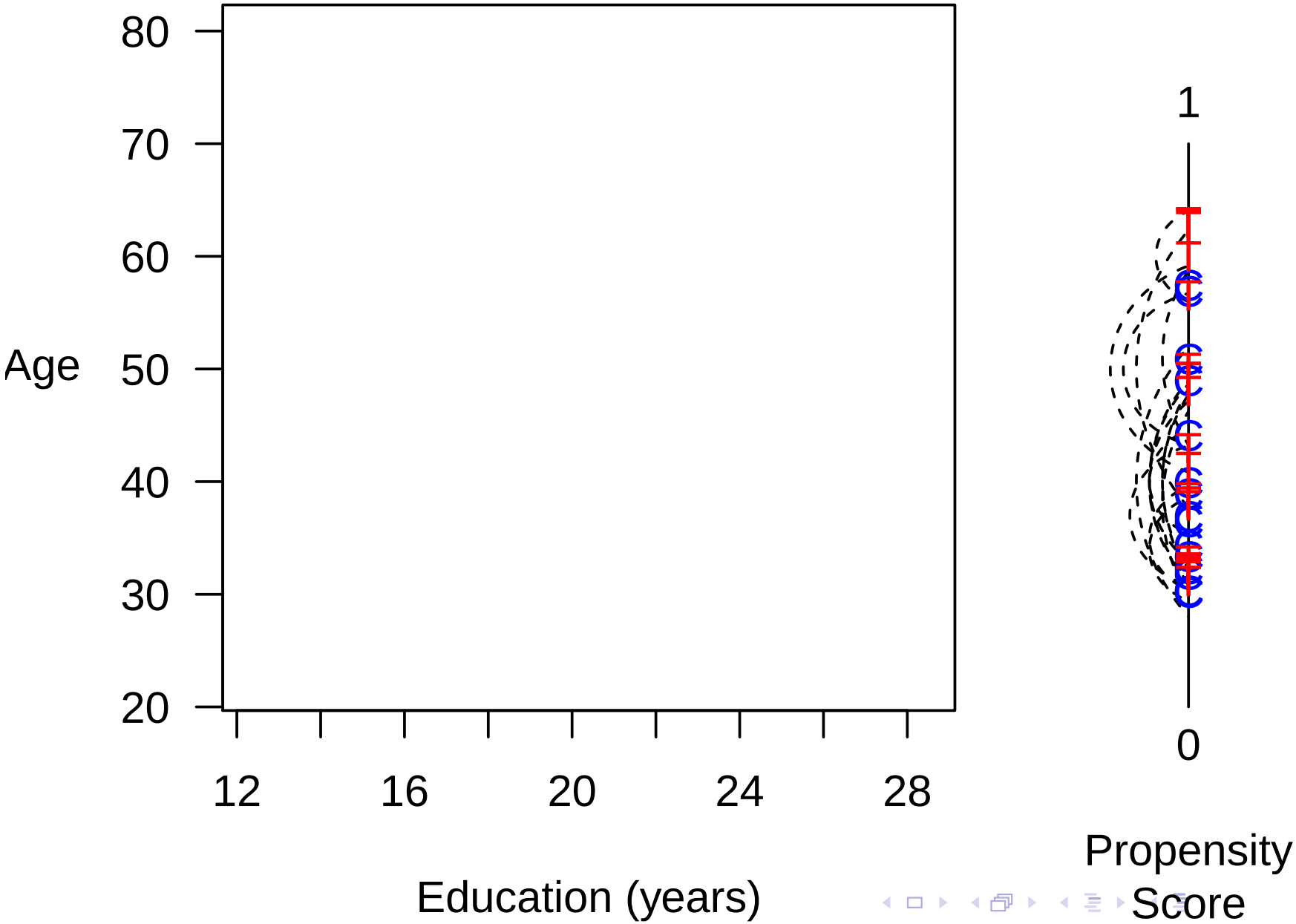
# Propensity Score Matching



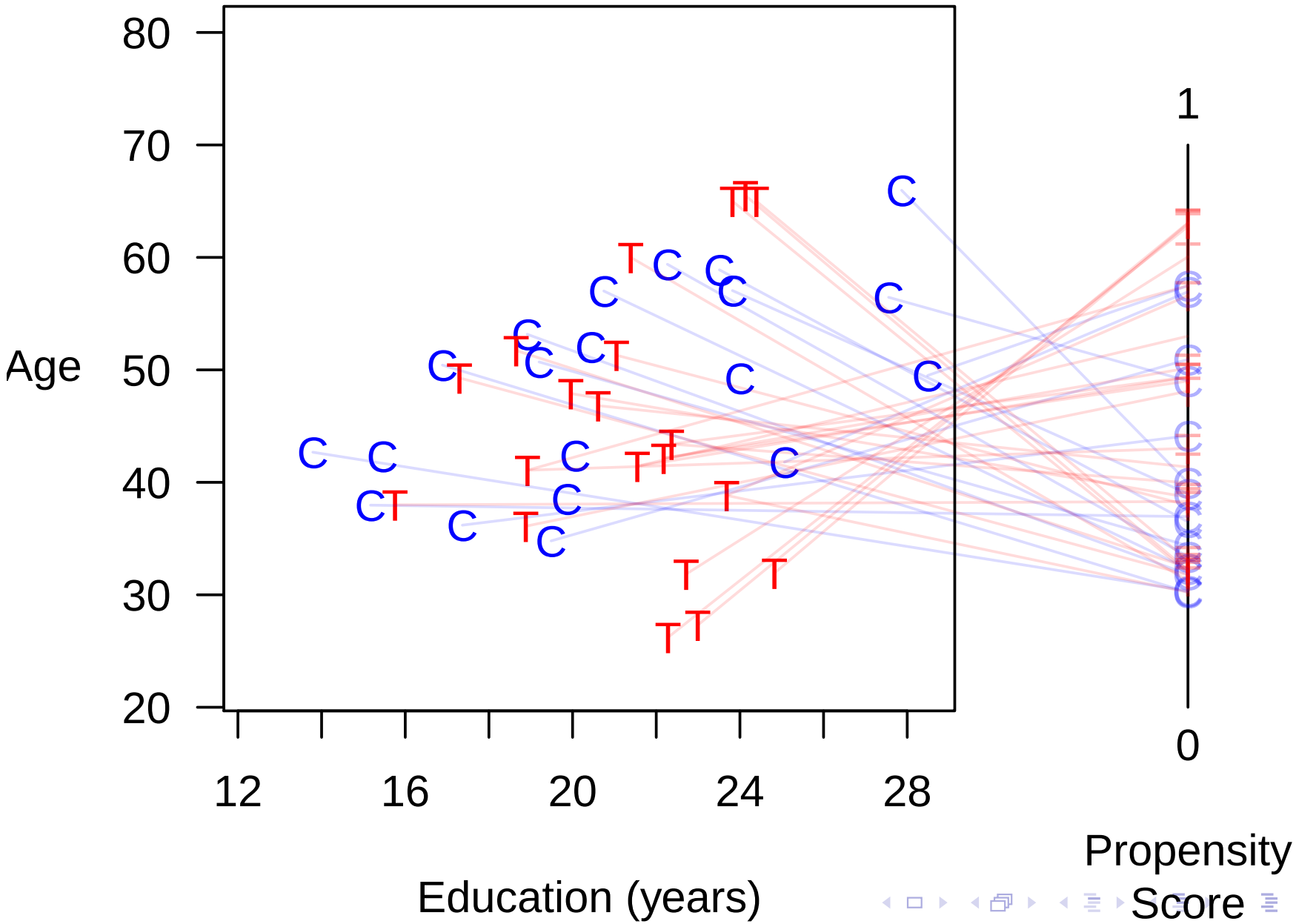
# Propensity Score Matching



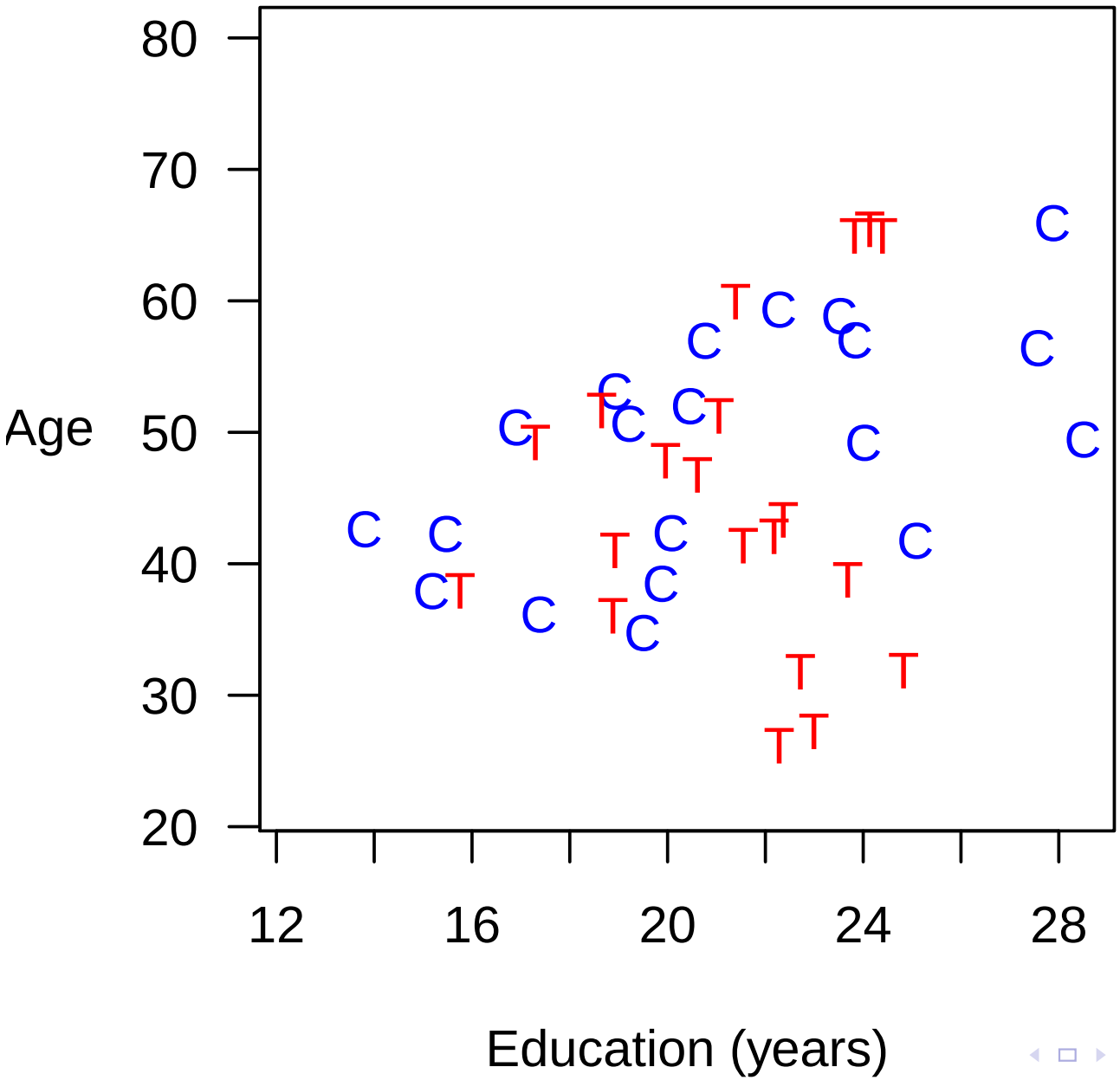
# Propensity Score Matching



# Propensity Score Matching



# Propensity Score Matching



# Method 3: Coarsened Exact Matching

# Method 3: Coarsened Exact Matching

1. Preprocess (Matching)

2. Estimation Difference in means or a model

# Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you're willing

2. Estimation Difference in means or a model

# Method 3: Coarsened Exact Matching

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 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram

2. Estimation Difference in means or a model

# Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened $X$, $C(X)$

2. Estimation Difference in means or a model

# Method 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - 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 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - 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 3: Coarsened Exact Matching

1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - 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 3: Coarsened Exact Matching

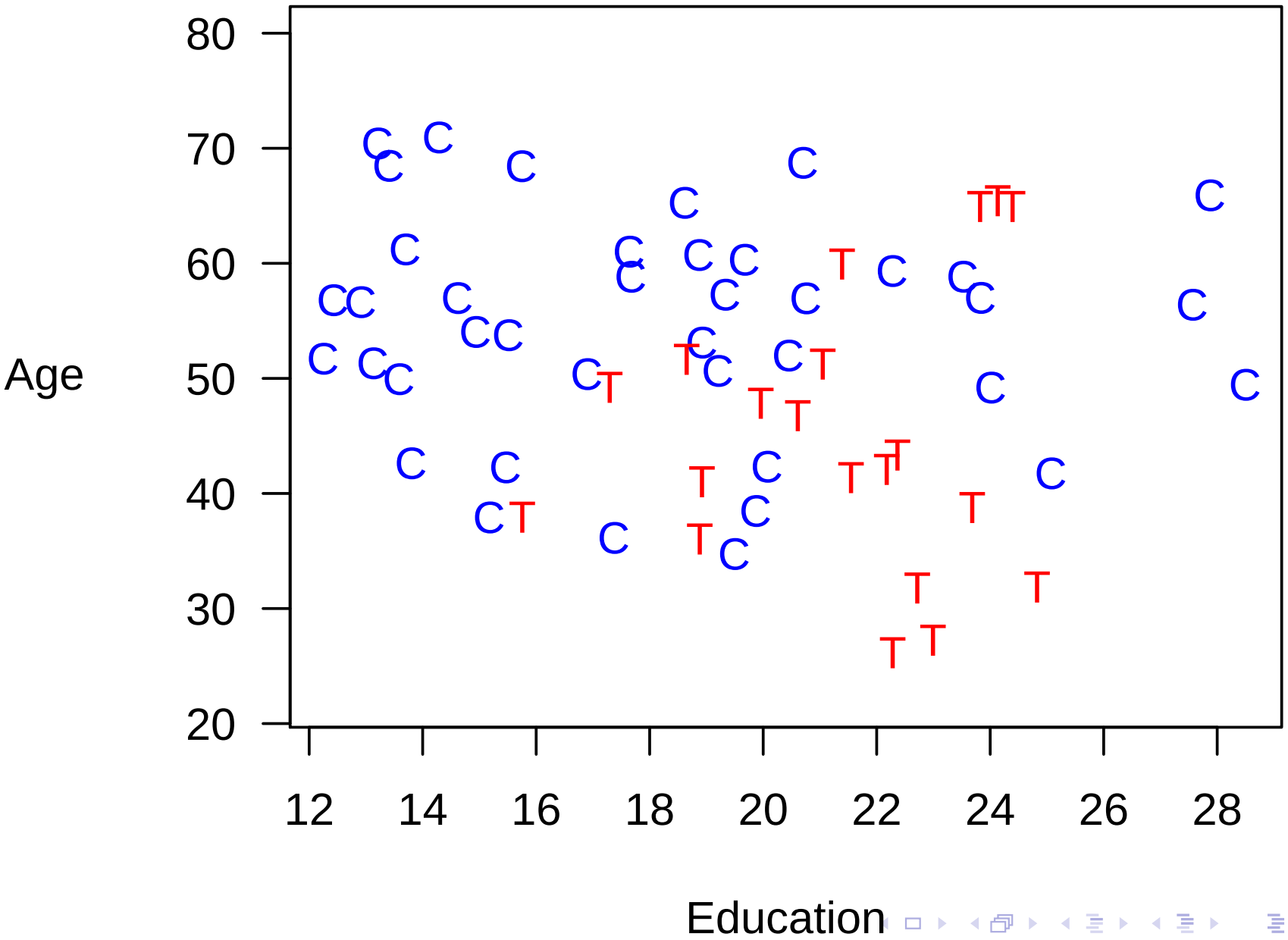
1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into 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
   - Need to weight controls in each stratum to equal treateds

# Method 3: Coarsened Exact Matching

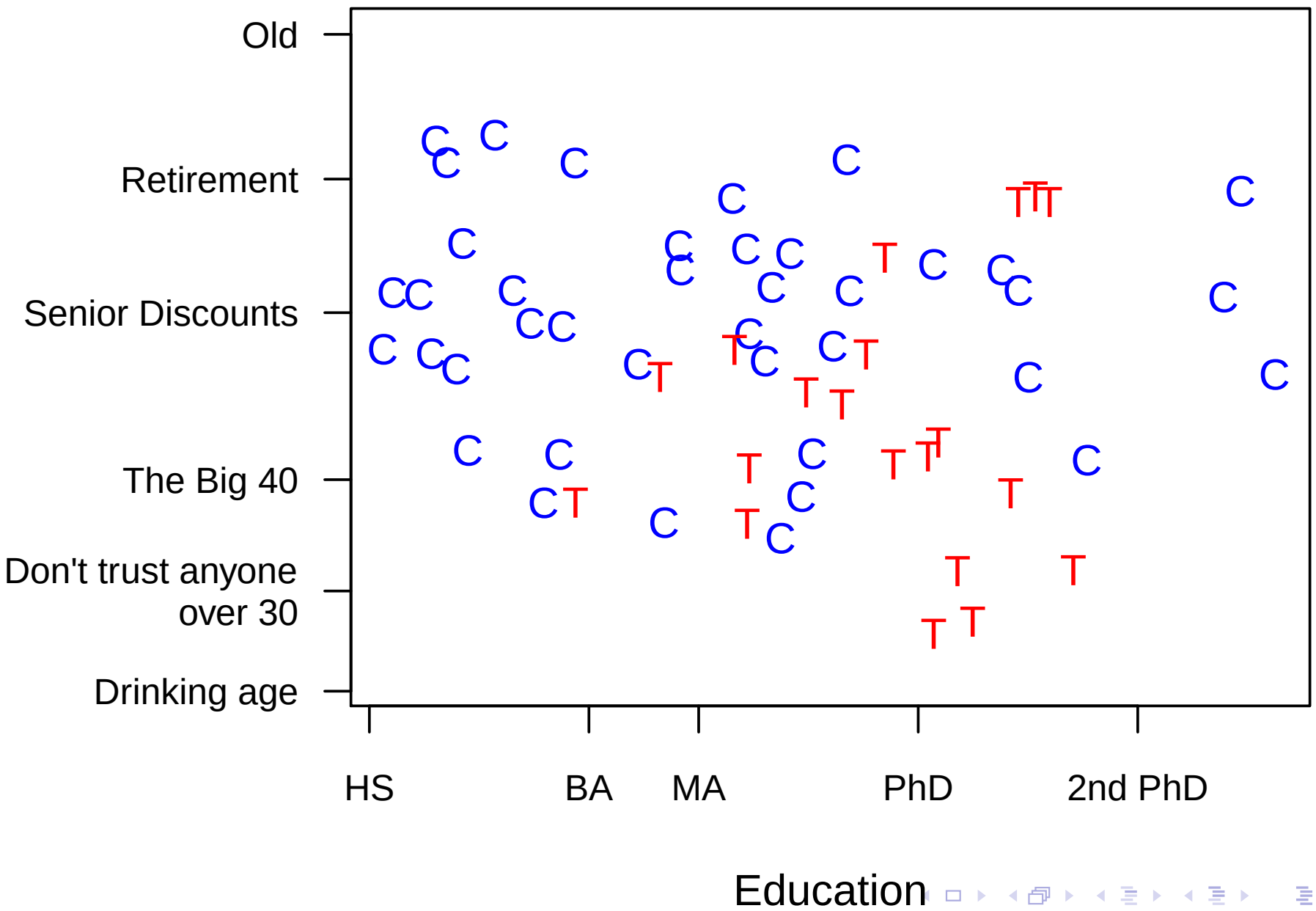
1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
     - Easy to understand, or can be automated as for a histogram
   - Apply exact matching to the coarsened $X$, $C(X)$
     - Sort observations into 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
   - Need to weight controls in each stratum to equal treateds
   - Can apply other matching methods within CEM strata (inherit CEM's properties)

# Coarsened Exact Matching

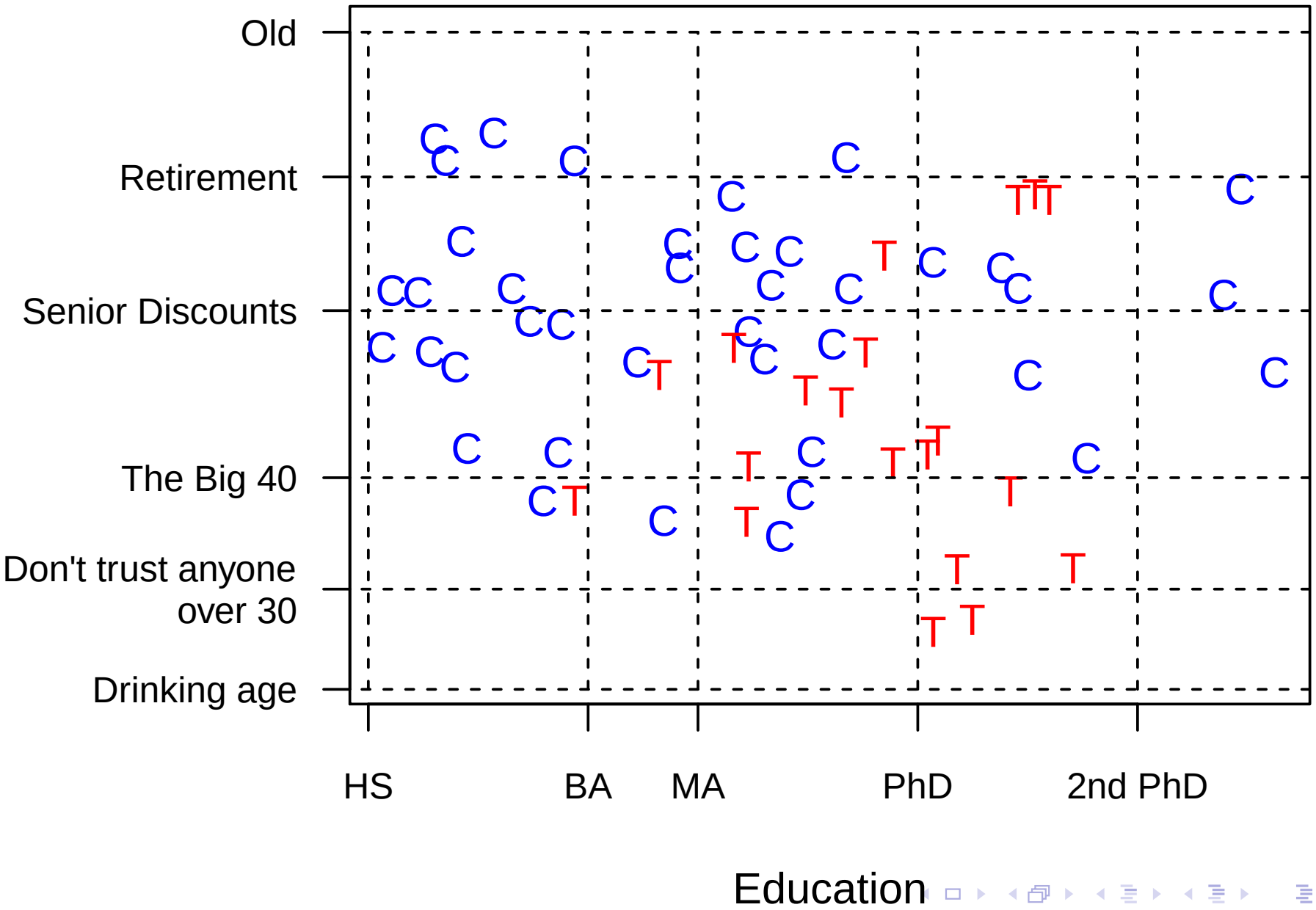
# Coarsened Exact Matching



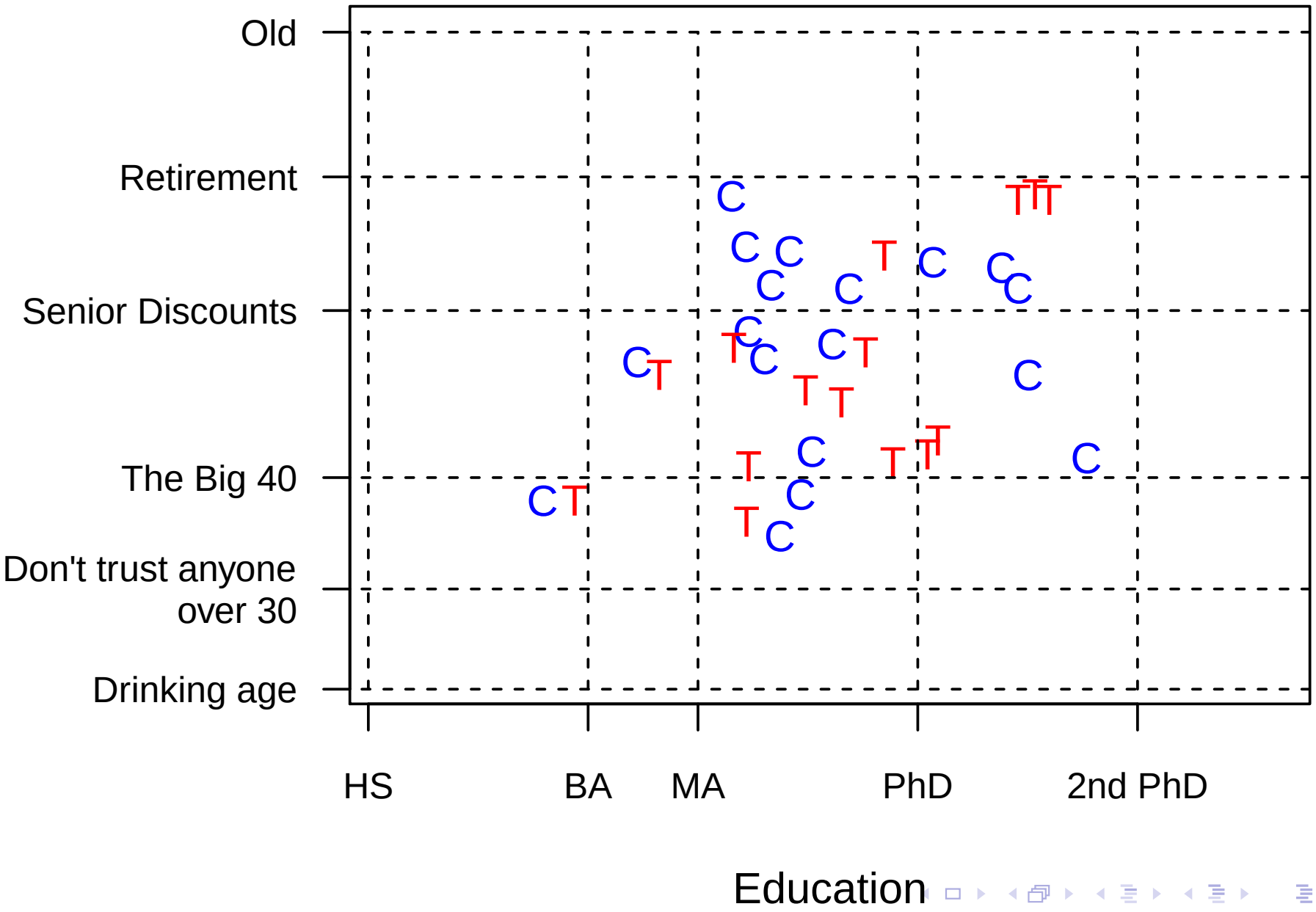
# Coarsened Exact Matching



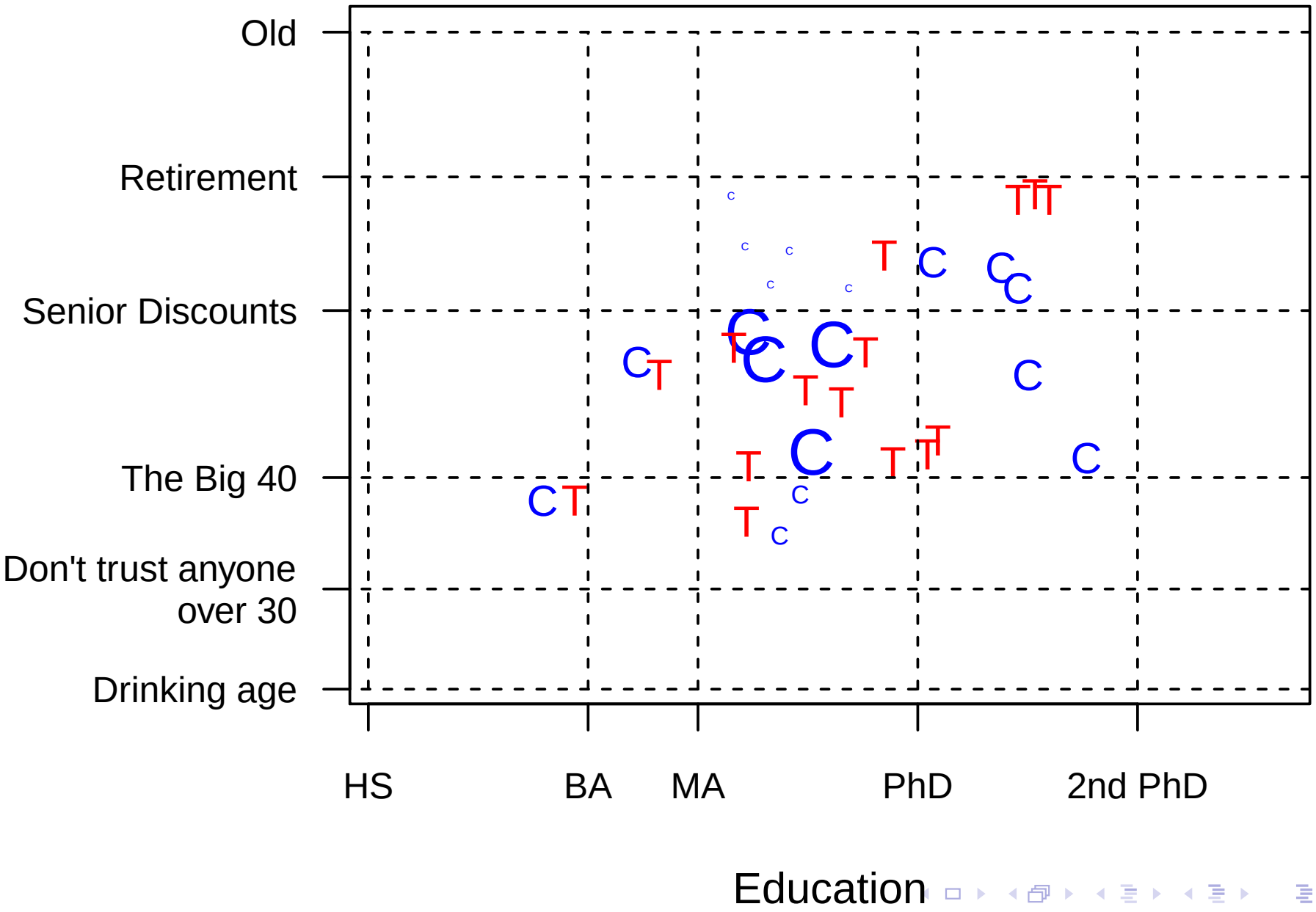
# Coarsened Exact Matching



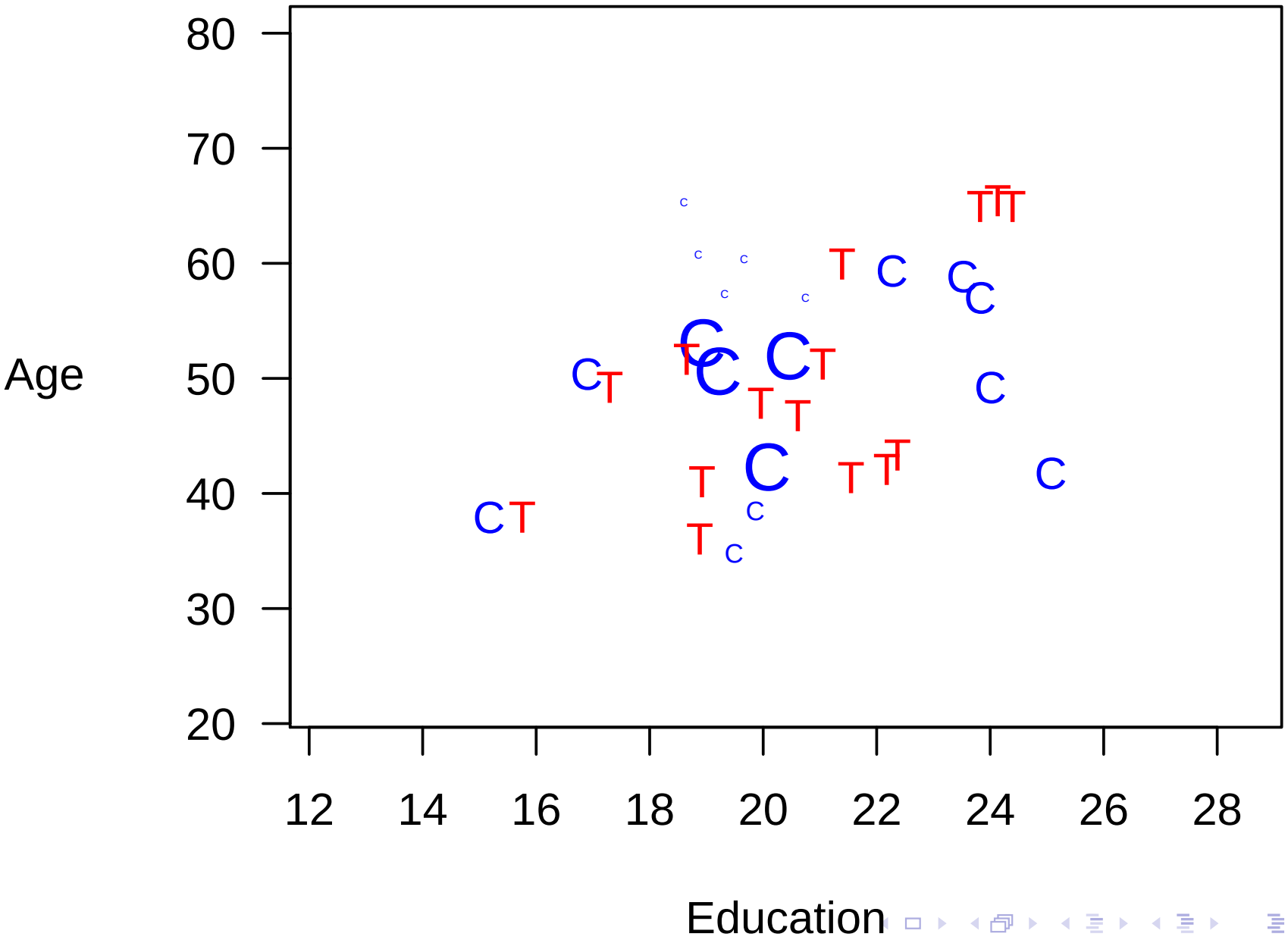
# Coarsened Exact Matching



# Coarsened Exact Matching



# Coarsened Exact Matching



# The Bias-Variance Trade Off in Matching

# The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f$(imbalance, importance, estimator)
  $\rightsquigarrow$ we measure imbalance instead

# The Bias-Variance Trade Off in Matching

- Bias (& model dependence) = $f$(imbalance, importance, estimator)
  $\rightsquigarrow$ we measure imbalance instead
- Variance = $f$(matched sample size, estimator)
  $\rightsquigarrow$ we measure matched sample size instead

# The Bias-Variance Trade Off in Matching

- Bias (& model dependence) = $f$(imbalance, importance, estimator)
  $\rightsquigarrow$ we measure imbalance instead
- Variance = $f$(matched sample size, estimator)
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off

# The Bias-Variance Trade Off in Matching

- Bias (& model dependence) = $f$(imbalance, importance, estimator)
  $\rightsquigarrow$ we measure imbalance instead
- Variance = $f$(matched sample size, estimator)
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off
- Measuring Imbalance

# The Bias-Variance Trade Off in Matching

- Bias (& model dependence) = $f$(imbalance, importance, estimator)
  $\rightsquigarrow$ we measure imbalance instead
- Variance = $f$(matched sample size, estimator)
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off
- Measuring Imbalance
  - Classic measure: Difference of means (for each variable)

# The Bias-Variance Trade Off in Matching

- Bias (& model dependence) $= f$(imbalance, importance, estimator)
  $\rightsquigarrow$ we measure imbalance instead
- Variance $= f$(matched sample size, estimator)
  $\rightsquigarrow$ we measure matched sample size instead
- Bias-Variance trade off $\rightsquigarrow$ Imbalance-$n$ Trade Off
- Measuring Imbalance
  - Classic measure: Difference of means (for each variable)
  - Better measure: Difference of multivariate histograms,

$$\mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \cdots \ell_k \in H(\mathbf{X})} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

# Comparing Matching Methods

# Comparing Matching Methods

- Standard approach

# Comparing Matching Methods

- Standard approach
  - MDM & PSM: Choose matched $n$, match, check imbalance

# Comparing Matching Methods

- Standard approach
  - MDM & PSM: Choose matched $n$, match, check imbalance
  - CEM: Choose imbalance, match, check matched $n$

# Comparing Matching Methods

- Standard approach
  - MDM & PSM: Choose matched $n$, match, check imbalance
  - CEM: Choose imbalance, match, check matched $n$
  - Best practice: iterate

# Comparing Matching Methods

- Standard approach
  - MDM & PSM: Choose matched $n$, match, check imbalance
  - CEM: Choose imbalance, match, check matched $n$
  - Best practice: iterate (Ugh!)

# Comparing Matching Methods

- Standard approach
  - MDM & PSM: Choose matched *n*, match, check imbalance
  - CEM: Choose imbalance, match, check matched *n*
  - Best practice: iterate (Ugh!)
  - Choose matched solution & matching method becomes irrelevant

# Comparing Matching Methods

- Standard approach
  - MDM & PSM: Choose matched $n$, match, check imbalance
  - CEM: Choose imbalance, match, check matched $n$
  - Best practice: iterate (Ugh!)
  - Choose matched solution & matching method becomes irrelevant
- An alternative approach

# Comparing Matching Methods

- Standard approach
  - MDM & PSM: Choose matched $n$, match, check imbalance
  - CEM: Choose imbalance, match, check matched $n$
  - Best practice: iterate (Ugh!)
  - Choose matched solution & matching method becomes irrelevant
- An alternative approach
  - Compute lots of matching solutions,

# Comparing Matching Methods

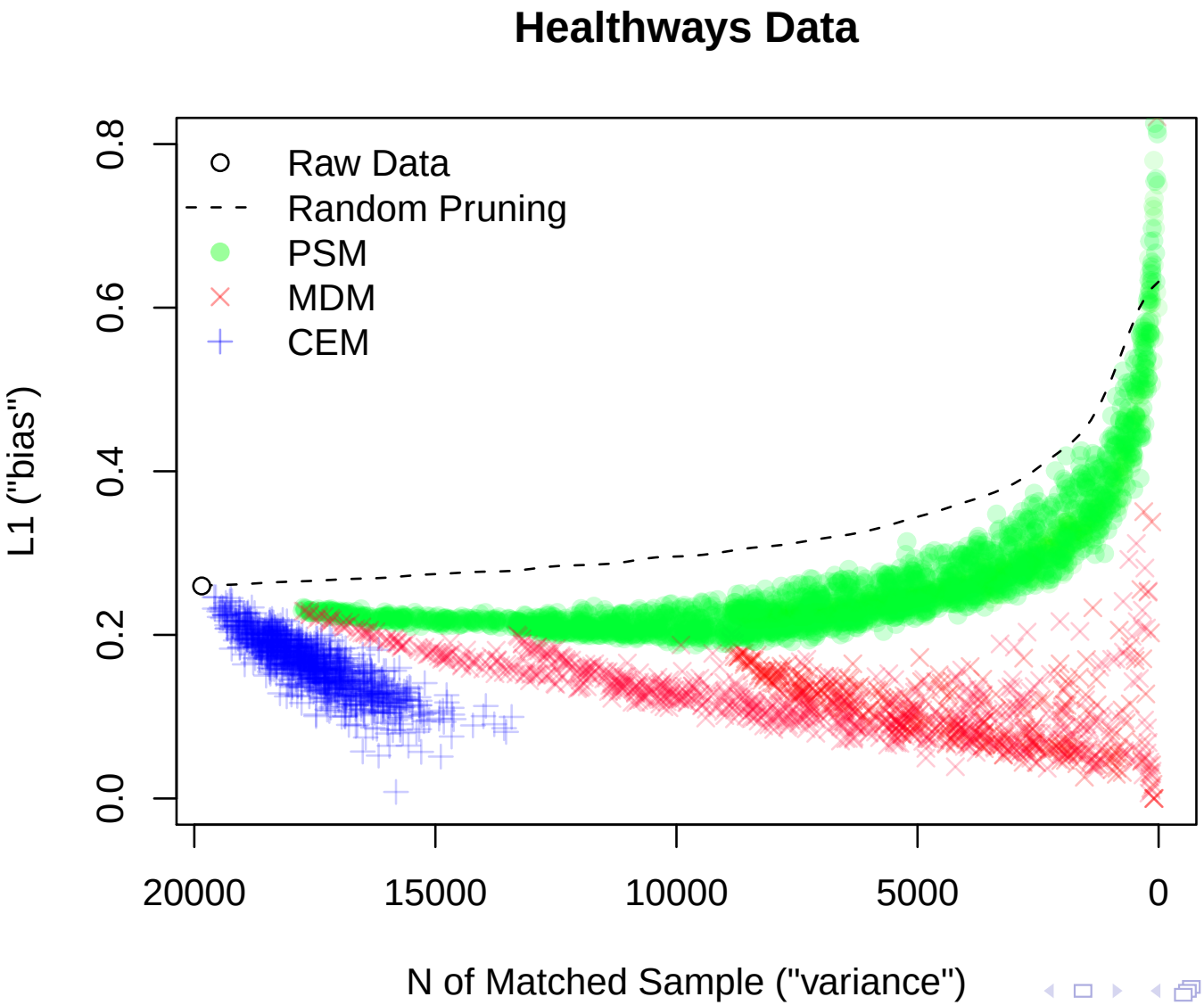
- Standard approach
  - MDM & PSM: Choose matched $n$, match, check imbalance
  - CEM: Choose imbalance, match, check matched $n$
  - Best practice: iterate (Ugh!)
  - Choose matched solution & matching method becomes irrelevant
- An alternative approach
  - Compute lots of matching solutions,
  - Identify the frontier of lowest imbalance for each given $n$, and

# Comparing Matching Methods

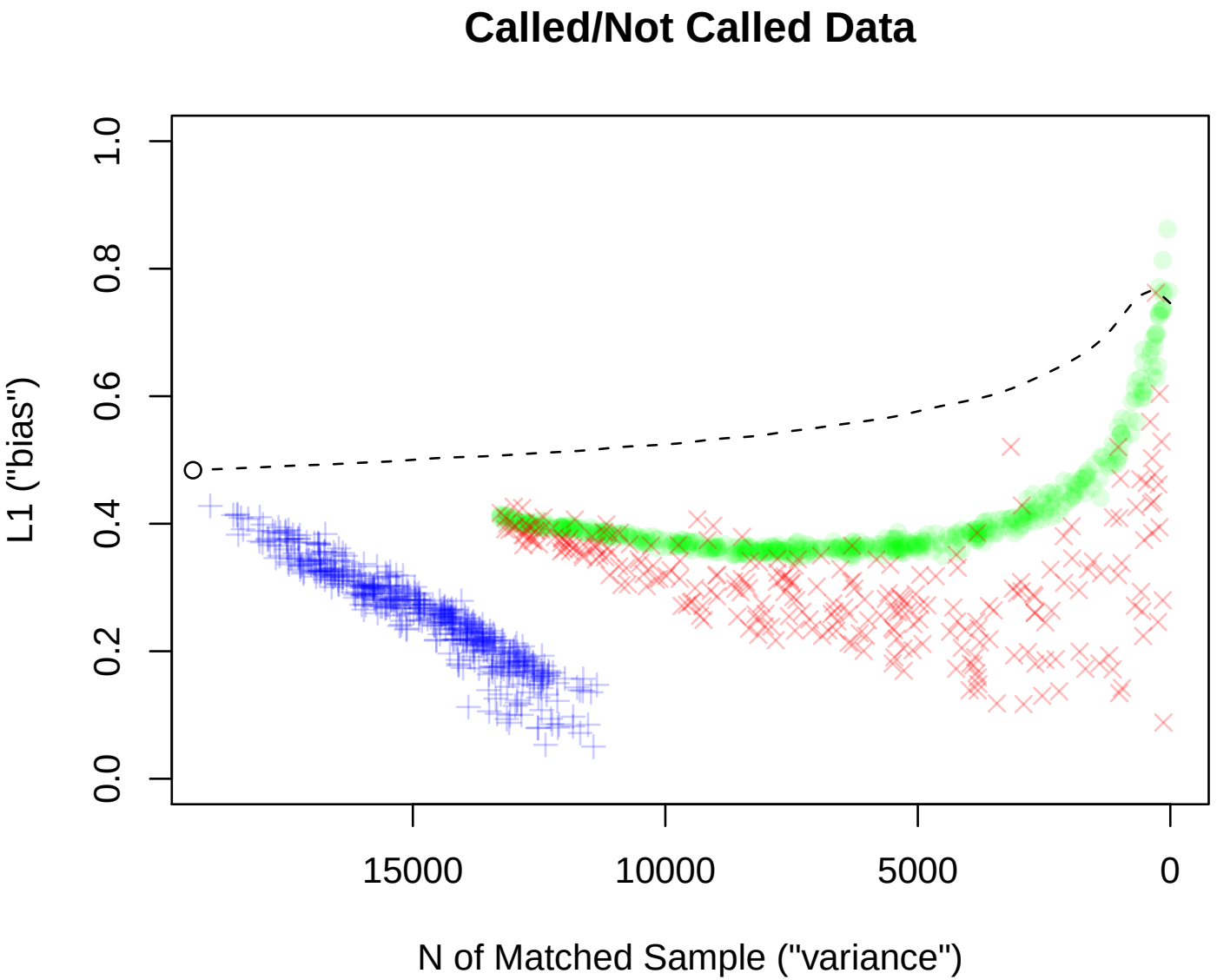
- Standard approach
  - MDM & PSM: Choose matched $n$, match, check imbalance
  - CEM: Choose imbalance, match, check matched $n$
  - Best practice: iterate (Ugh!)
  - Choose matched solution & matching method becomes irrelevant
- An alternative approach
  - Compute lots of matching solutions,
  - Identify the frontier of lowest imbalance for each given $n$, and
  - Choose a matching solution among those on the frontier

# A Space Graph: Real Data

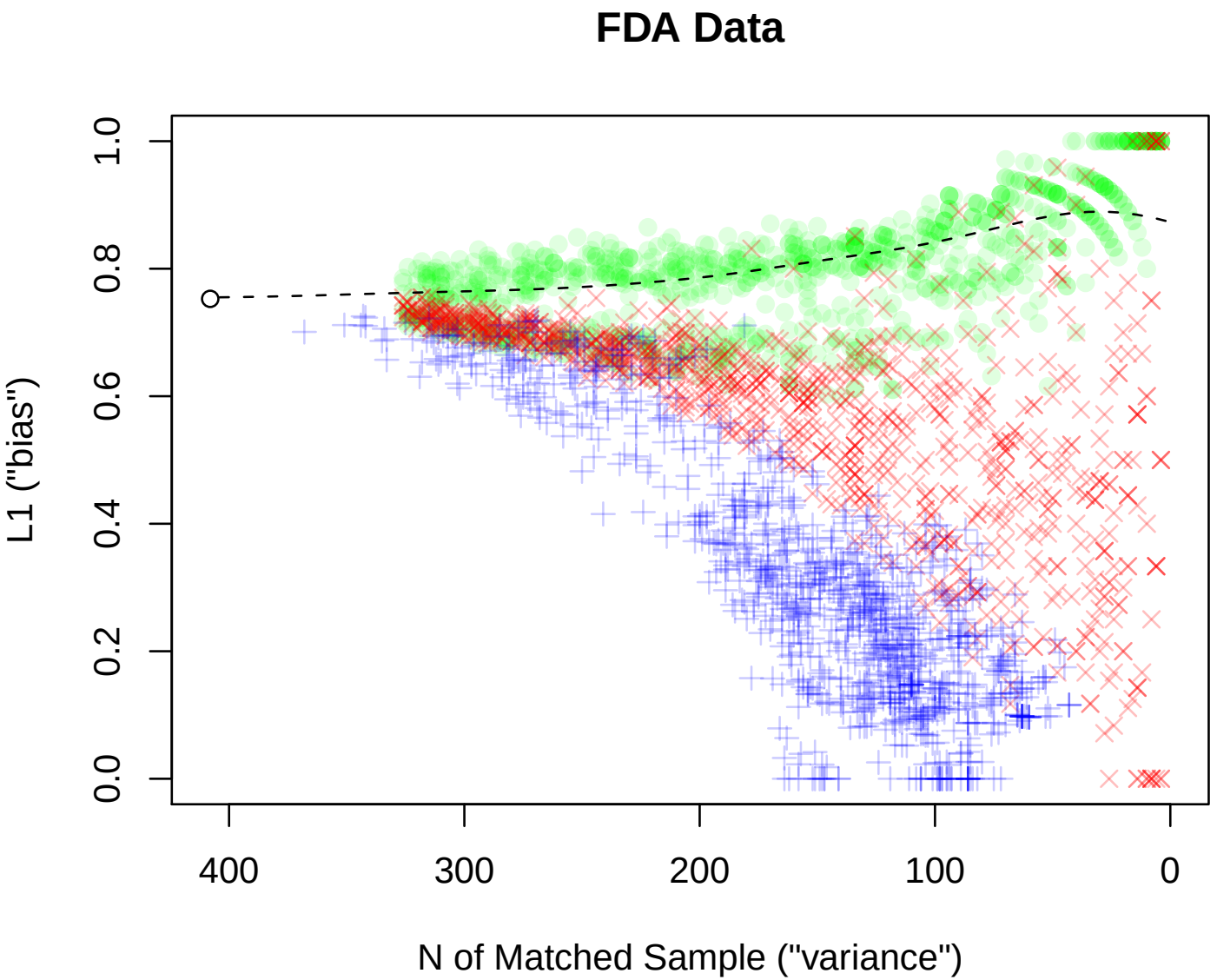
King, Nielsen, Coberley, Pope, and Wells (2011)



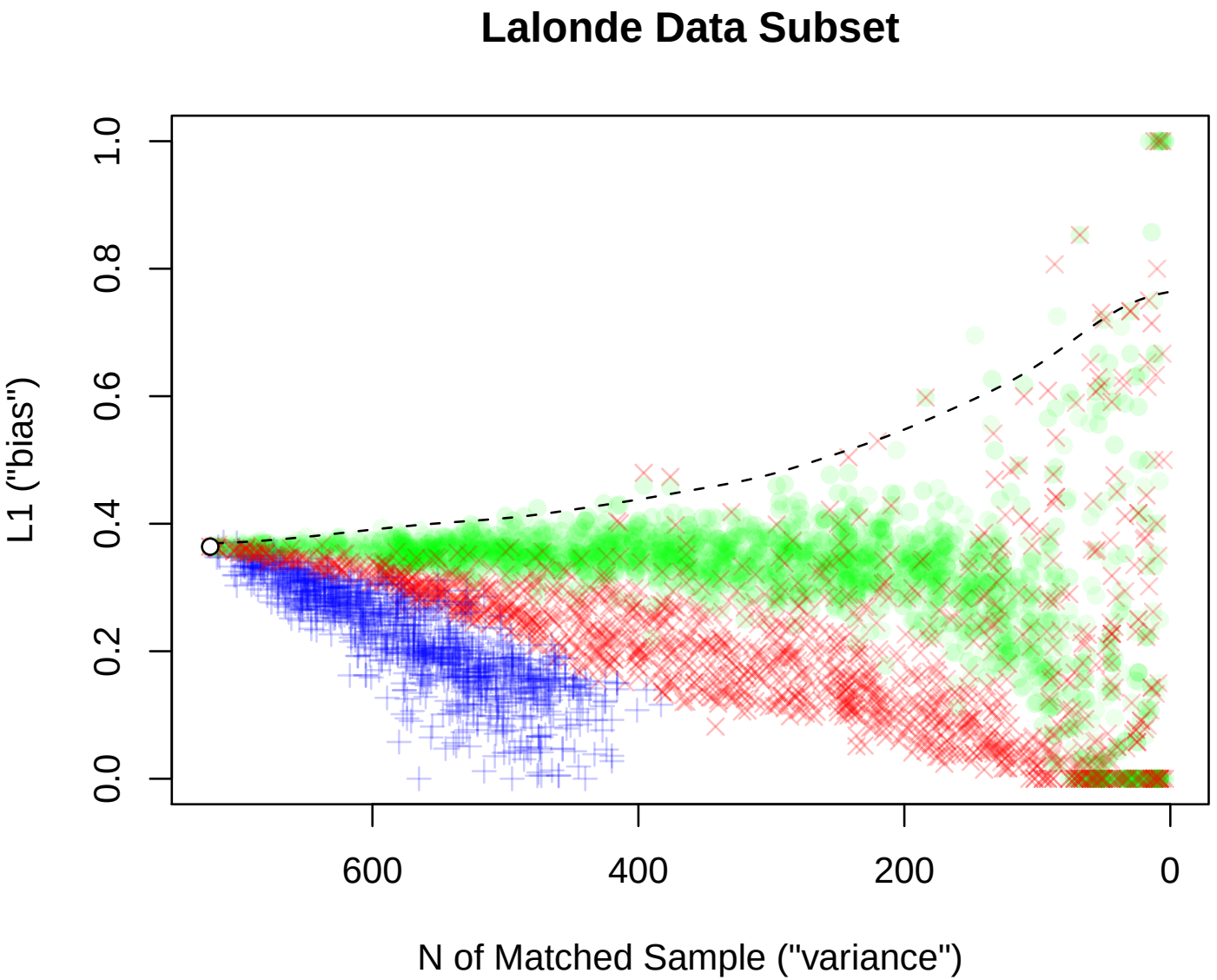
# A Space Graph: Real Data



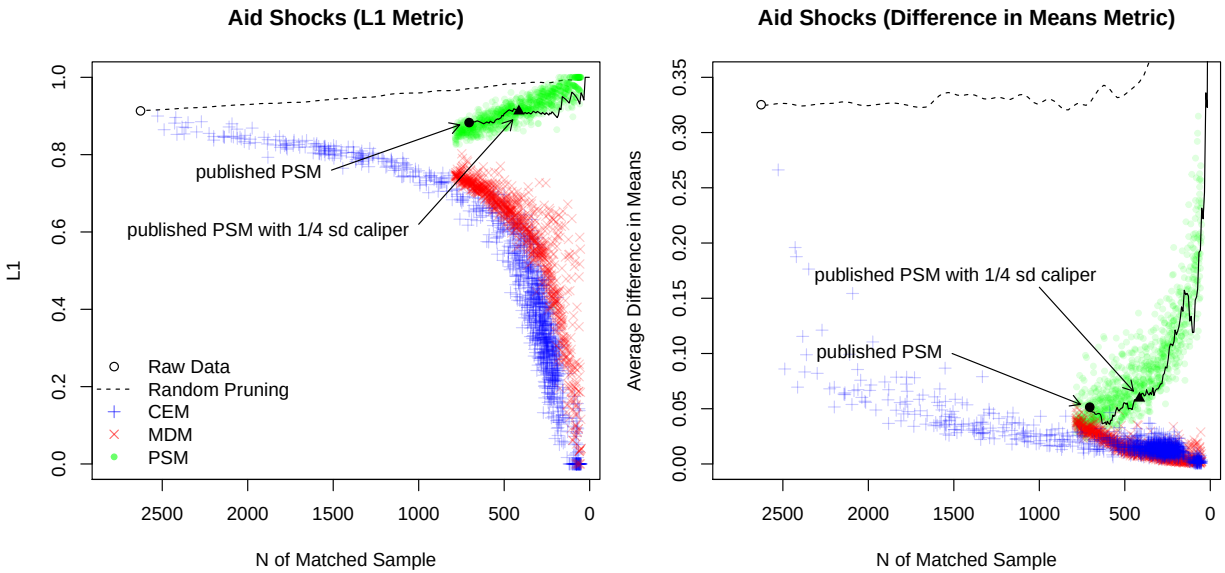
# A Space Graph: Real Data

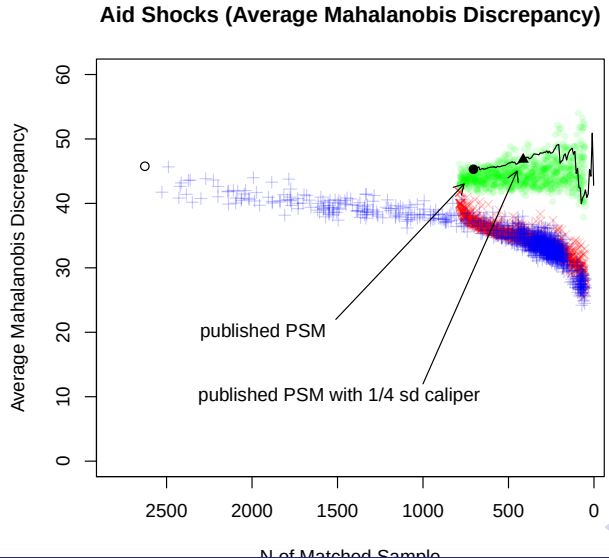


# A Space Graph: Real Data

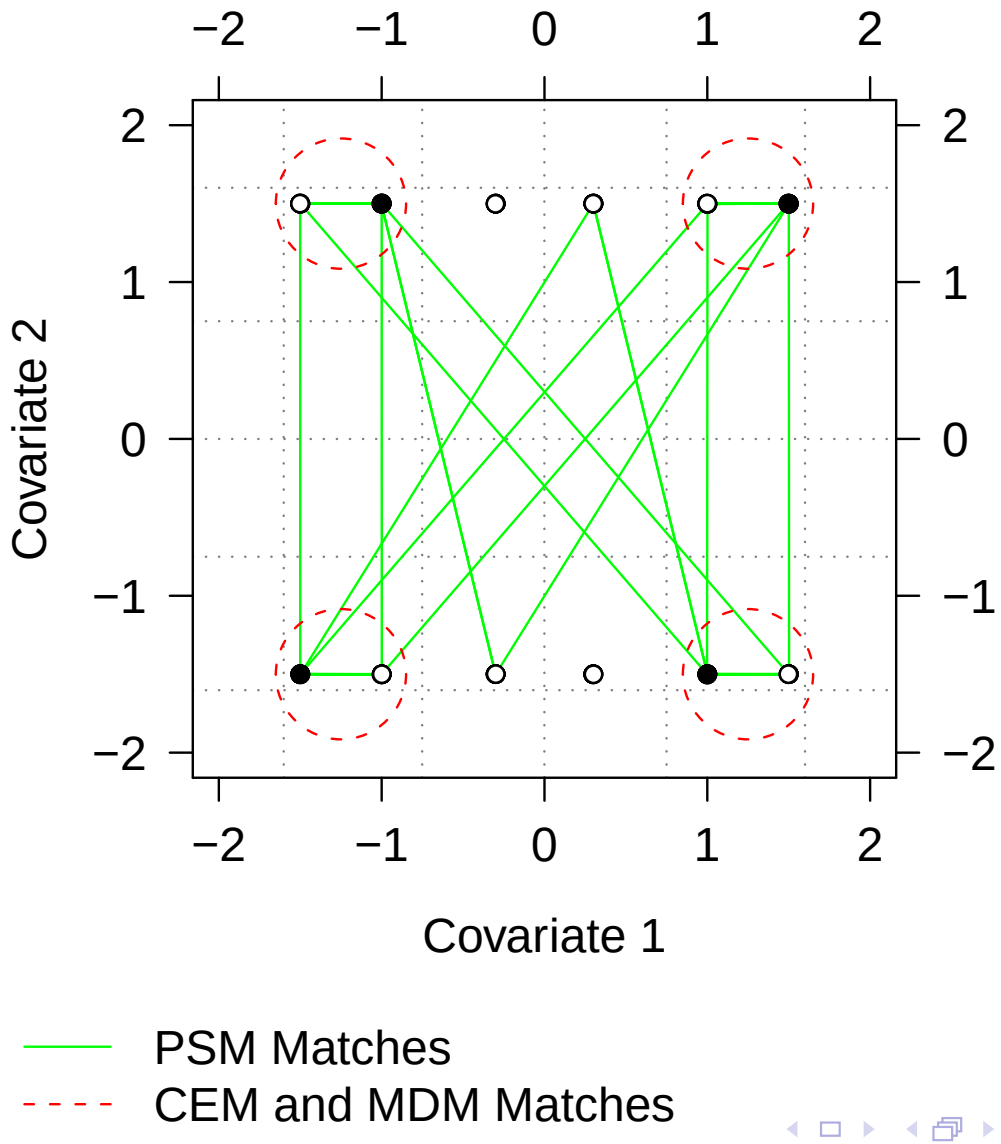


# Space Graphs: Different Imbalance Metrics

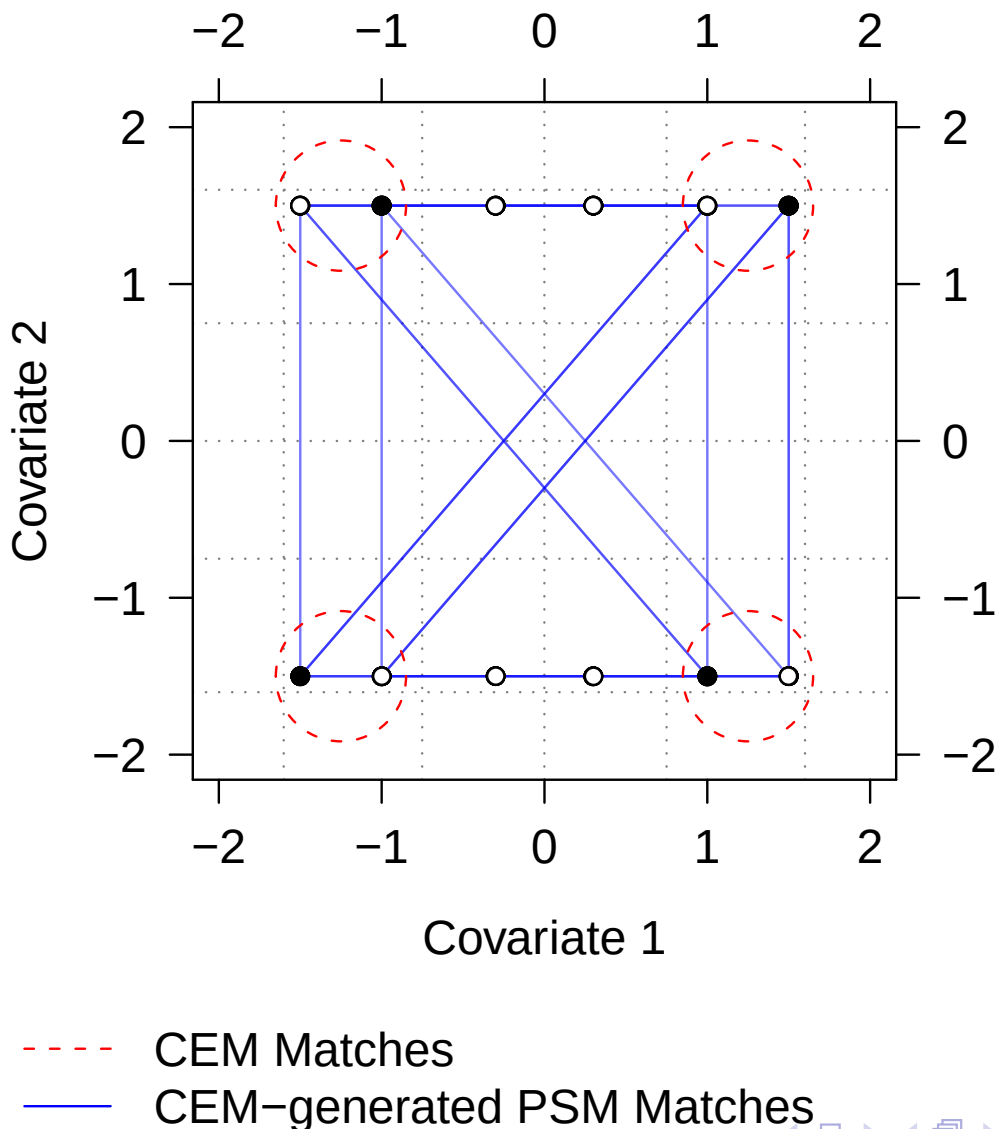


# PSM Approximates Random Matching in Balanced Data



# Destroying CEM with PSM's Two Step Approach



# Pause for Conclusions

# Pause for Conclusions

- Propensity score matching:

# Pause for Conclusions

- Propensity score matching:
  - The problem:

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM* is a mistake

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM* is a mistake
    - Adjusting experimental data *with PSM* is a mistake

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM* is a mistake
    - Adjusting experimental data *with PSM* is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a mistake

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM* is a mistake
    - Adjusting experimental data *with PSM* is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a mistake
- CEM and Mahalanobis do not have PSM's problems

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM* is a mistake
    - Adjusting experimental data *with PSM* is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a mistake
- CEM and Mahalanobis do not have PSM's problems
- CEM $>$ Mahalanobis $>$ Propensity Score (in many data sets and sims)

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM* is a mistake
    - Adjusting experimental data *with PSM* is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a mistake
- CEM and Mahalanobis do not have PSM's problems
- CEM $>$ Mahalanobis $>$ Propensity Score (in many data sets and sims)
- (Your performance may vary)

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM* is a mistake
    - Adjusting experimental data *with PSM* is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a mistake
- CEM and Mahalanobis do not have PSM's problems
- CEM $>$ Mahalanobis $>$ Propensity Score (in many data sets and sims)
- (Your performance may vary)
- You can easily check with the Space Graph

# Pause for Conclusions

- Propensity score matching:
  - The problem:
    - Imbalance can be worse than original data
    - Can increase imbalance when removing the worst matches
    - Approximates random matching in well-balanced data (Random matching increases imbalance)
  - The Cause: unnecessary 1st stage dimension reduction
  - Implications:
    - Balance checking required
    - Adjusting for potentially irrelevant covariates *with PSM* is a mistake
    - Adjusting experimental data *with PSM* is a mistake
    - Reestimating the propensity score after eliminating noncommon support may be a mistake
- CEM and Mahalanobis do not have PSM's problems
- CEM $>$ Mahalanobis $>$ Propensity Score (in many data sets and sims)
- (Your performance may vary)
- You can easily check with the Space Graph
- CEM is the easiest and most powerful; let's look more deeply. . .

# Problems With Matching Methods (other than CEM)

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance):

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES);

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$;

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$;

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance;

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice:

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check,

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check,

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check,

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check,

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice:

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose $n$,

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose $n$, match,

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose $n$, match, publish,

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose $n$, match, publish, STOP.

# Problems With Matching Methods (other than CEM)

- Don't eliminate extrapolation region
- Don't work with multiply imputed data
- Most violate the congruence principle
- Largest class of matching methods (EPBR, e.g., propensity scores, Mahalanobis distance): requires normal data (or DMPES); all $X$'s must have same effect on $Y$; $Y$ must be a linear function of $X$; aims only for expected (not in-sample) imbalance; $\rightsquigarrow$ in practice, we're lucky if mean imbalance is reduced
- Not well designed for observational data:
  - Least important (variance): matched $n$ chosen ex ante
  - Most important (bias): imbalance reduction checked ex post
- Hard to use: Improving balance on 1 variable can reduce it on others
  - Best practice: choose $n$-match-check, tweak-match-check, tweak-match-check, tweak-match-check, $\cdots$
  - Actual practice: choose $n$, match, publish, STOP. (Is balance even improved?)

# CEM as an MIB Method

# CEM as an MIB Method

- Coarsening determines the level of imbalance

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for:

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means,

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances,

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness,

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances,

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments,

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness,

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis,

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles,

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- What if coarsening is set ...

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- What if coarsening is set ...
  - too coarse?

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- What if coarsening is set ...
  - too coarse? $\rightsquigarrow$ You're left modeling remaining imbalances

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- What if coarsening is set ...
  - too coarse? $\rightsquigarrow$ You're left modeling remaining imbalances
  - not coarse enough?

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- What if coarsening is set ...
  - too coarse? $\rightsquigarrow$ You're left modeling remaining imbalances
  - not coarse enough? $\rightsquigarrow$ $n$ may be too small

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- What if coarsening is set ...
  - too coarse? $\rightsquigarrow$ You're left modeling remaining imbalances
  - not coarse enough? $\rightsquigarrow$ $n$ may be too small
  - as large as you're comfortable with, but $n$ is still too small?

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- What if coarsening is set ...
  - too coarse? $\rightsquigarrow$ You're left modeling remaining imbalances
  - not coarse enough? $\rightsquigarrow$ $n$ may be too small
  - as large as you're comfortable with, but $n$ is still too small?

    $\rightsquigarrow$ No magic method of matching can save you;

# CEM as an MIB Method

- Coarsening determines the level of imbalance
- Convenient monotonicity property: Reducing maximum imbalance on one $X$: no effect on others
- We Prove: setting $\epsilon$ bounds the treated-control group difference, within strata and globally, for: means, variances, skewness, covariances, comoments, coskewness, co-kurtosis, quantiles, and full multivariate histogram.

  $\Longrightarrow$ Setting $\epsilon$ controls all multivariate treatment-control differences, interactions, and nonlinearities, up to the chosen level (matched $n$ is determined ex post)
- What if coarsening is set ...
  - too coarse? $\rightsquigarrow$ You're left modeling remaining imbalances
  - not coarse enough? $\rightsquigarrow$ $n$ may be too small
  - as large as you're comfortable with, but $n$ is still too small?

    $\rightsquigarrow$ No magic method of matching can save you;

    $\rightsquigarrow$ You're stuck modeling or collecting better data

End of planned slides for today; others follow

# Other CEM properties we prove

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable's units

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable's units
  - E.g., calipers (strata centered on each unit):

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student;

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett
- Approximate invariance to measurement error:

| | CEM | pscore | Mahalanobis | Genetic |
|---|---|---|---|---|
| % Common Units | 96.5 | 70.2 | 80.9 | 80.0 |

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett
- Approximate invariance to measurement error:

| | CEM | pscore | Mahalanobis | Genetic |
|---|---|---|---|---|
| % Common Units | 96.5 | 70.2 | 80.9 | 80.0 |

- Fast and memory-efficient even for large $n$; can be fully automated

# Other CEM properties we prove

- Automatically eliminates extrapolation region (no separate step)
- Bounds model dependence
- Bounds causal effect estimation error
- Meets the congruence principle
  - The principle: data space = analysis space
  - Estimators that violate it are nonrobust and counterintuitive
  - CEM: $\epsilon_j$ is set using each variable's units
  - E.g., calipers (strata centered on each unit): would bin college drop out with 1st year grad student; and not bin Bill Gates & Warren Buffett
- Approximate invariance to measurement error:

| | CEM | pscore | Mahalanobis | Genetic |
|---|---|---|---|---|
| % Common Units | 96.5 | 70.2 | 80.9 | 80.0 |

- Fast and memory-efficient even for large $n$; can be fully automated
- Simple to teach: coarsen, then exact match

# Imbalance Measures

# Imbalance Measures

## Variable-by-Variable Difference in Global Means

$$I_1^{(j)} = \left| \bar{X}_{m_T}^{(j)} - \bar{X}_{m_C}^{(j)} \right|, \quad j = 1, \ldots, k$$

# Imbalance Measures

## Variable-by-Variable Difference in Global Means

$$I_1^{(j)} = \left| \bar{X}_{m_T}^{(j)} - \bar{X}_{m_C}^{(j)} \right|, \quad j = 1, \ldots, k$$

## Multivariate Imbalance: difference in histograms (bins fixed ex ante)

$$\mathcal{L}_1(f, g) = \sum_{\ell_1 \cdots \ell_k} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

# Imbalance Measures

## Variable-by-Variable Difference in Global Means

$$I_1^{(j)} = \left| \bar{X}_{m_T}^{(j)} - \bar{X}_{m_C}^{(j)} \right|, \quad j = 1, \ldots, k$$

## Multivariate Imbalance: difference in histograms (bins fixed ex ante)

$$\mathcal{L}_1(f, g) = \sum_{\ell_1 \cdots \ell_k} |f_{\ell_1 \cdots \ell_k} - g_{\ell_1 \cdots \ell_k}|$$

## Local Imbalance by Variable (given strata fixed by matching method)

$$I_2^{(j)} = \frac{1}{S} \sum_{s=1}^{S} \left| \bar{X}_{m_T^s}^{(j)} - \bar{X}_{m_C^s}^{(j)} \right|, \quad j = 1, \ldots, k$$

# CEM in Practice: EPBR-Compliant Data

# CEM in Practice: EPBR-Compliant Data

Monte Carlo:

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$.

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($l_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($l_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($L_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($I_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($L_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |
| PSC | .11 | .06 | .03 | .06 | .03 | .16 |

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($L_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |
| PSC | .11 | .06 | .03 | .06 | .03 | .16 |
| CEM | .04 | .02 | .06 | .06 | .04 | .08 |

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($I_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |
| PSC | .11 | .06 | .03 | .06 | .03 | .16 |
| CEM | .04 | .02 | .06 | .06 | .04 | .08 |

Local ($I_2$) and multivariate $\mathcal{L}_1$ imbalance:

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $\mathcal{L}_1$ |
|---|---|---|---|---|---|---|

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($I_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |
| PSC | .11 | .06 | .03 | .06 | .03 | .16 |
| CEM | .04 | .02 | .06 | .06 | .04 | .08 |

Local ($I_2$) and multivariate $\mathcal{L}_1$ imbalance:

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $\mathcal{L}_1$ |
|---|---|---|---|---|---|---|

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($I_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |
| PSC | .11 | .06 | .03 | .06 | .03 | .16 |
| CEM | .04 | .02 | .06 | .06 | .04 | .08 |

Local ($I_2$) and multivariate $\mathcal{L}_1$ imbalance:

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $\mathcal{L}_1$ |
|---|---|---|---|---|---|---|
| initial | | | | | | 1.24 |

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($I_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |
| PSC | .11 | .06 | .03 | .06 | .03 | .16 |
| CEM | .04 | .02 | .06 | .06 | .04 | .08 |

Local ($I_2$) and multivariate $\mathcal{L}_1$ imbalance:

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $\mathcal{L}_1$ |
|---|---|---|---|---|---|---|
| initial | | | | | | 1.24 |
| PSC | 2.38 | 1.25 | .74 | 1.25 | .74 | 1.18 |

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($I_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |
| PSC | .11 | .06 | .03 | .06 | .03 | .16 |
| CEM | .04 | .02 | .06 | .06 | .04 | .08 |

Local ($I_2$) and multivariate $\mathcal{L}_1$ imbalance:

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $\mathcal{L}_1$ |
|---|---|---|---|---|---|---|
| initial | | | | | | 1.24 |
| PSC | 2.38 | 1.25 | .74 | 1.25 | .74 | 1.18 |
| MAH | .56 | .36 | .29 | .36 | .29 | 1.13 |

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($I_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |
| PSC | .11 | .06 | .03 | .06 | .03 | .16 |
| CEM | .04 | .02 | .06 | .06 | .04 | .08 |

Local ($I_2$) and multivariate $\mathcal{L}_1$ imbalance:

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $\mathcal{L}_1$ |
|---|---|---|---|---|---|---|
| initial | | | | | | 1.24 |
| PSC | 2.38 | 1.25 | .74 | 1.25 | .74 | 1.18 |
| MAH | .56 | .36 | .29 | .36 | .29 | 1.13 |
| CEM | .42 | .26 | .17 | .22 | .19 | .78 |

# CEM in Practice: EPBR-Compliant Data

Monte Carlo: $\mathbf{X}_T \sim N_5(\mathbf{0}, \Sigma)$ and $\mathbf{X}_C \sim N_5(\mathbf{1}, \Sigma)$. $n = 2,000$, reps=5,000
Allow MAH & PSC to match with replacement; use automated CEM

Difference in means ($I_1$):

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | Seconds |
|---|---|---|---|---|---|---|
| initial | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| MAH | .20 | .20 | .20 | .20 | .20 | .28 |
| PSC | .11 | .06 | .03 | .06 | .03 | .16 |
| CEM | .04 | .02 | .06 | .06 | .04 | .08 |

Local ($I_2$) and multivariate $\mathcal{L}_1$ imbalance:

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $\mathcal{L}_1$ |
|---|---|---|---|---|---|---|
| initial | | | | | | 1.24 |
| PSC | 2.38 | 1.25 | .74 | 1.25 | .74 | 1.18 |
| MAH | .56 | .36 | .29 | .36 | .29 | 1.13 |
| CEM | .42 | .26 | .17 | .22 | .19 | .78 |

$\rightsquigarrow$ CEM dominates EPBR-methods in EPBR Data

# CEM in Practice: Non-EPBR Data

# CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

# CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

| | BIAS | SD | RMSE | Seconds | $\mathcal{L}_1$ |
|---|---|---|---|---|---|

# CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

| | BIAS | SD | RMSE | Seconds | $\mathcal{L}_1$ |
|---|---|---|---|---|---|

# CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

| | BIAS | SD | RMSE | Seconds | $\mathcal{L}_1$ |
|---|---|---|---|---|---|
| initial | −423.7 | 1566.5 | 1622.6 | .00 | 1.28 |

# CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

| | BIAS | SD | RMSE | Seconds | $\mathcal{L}_1$ |
|---|---|---|---|---|---|
| initial | −423.7 | 1566.5 | 1622.6 | .00 | 1.28 |
| MAH | 784.8 | 737.9 | 1077.2 | .03 | 1.08 |

# CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

| | BIAS | SD | RMSE | Seconds | $\mathcal{L}_1$ |
|---|---|---|---|---|---|
| initial | −423.7 | 1566.5 | 1622.6 | .00 | 1.28 |
| MAH | 784.8 | 737.9 | 1077.2 | .03 | 1.08 |
| PSC | 260.5 | 1025.8 | 1058.4 | .02 | 1.23 |

# CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

| | BIAS | SD | RMSE | Seconds | $\mathcal{L}_1$ |
|---|---|---|---|---|---|
| initial | −423.7 | 1566.5 | 1622.6 | .00 | 1.28 |
| MAH | 784.8 | 737.9 | 1077.2 | .03 | 1.08 |
| PSC | 260.5 | 1025.8 | 1058.4 | .02 | 1.23 |
| GEN | 78.3 | 499.5 | 505.6 | 27.38 | 1.12 |

# CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

| | BIAS | SD | RMSE | Seconds | $\mathcal{L}_1$ |
|---|---|---|---|---|---|
| initial | −423.7 | 1566.5 | 1622.6 | .00 | 1.28 |
| MAH | 784.8 | 737.9 | 1077.2 | .03 | 1.08 |
| PSC | 260.5 | 1025.8 | 1058.4 | .02 | 1.23 |
| GEN | 78.3 | 499.5 | 505.6 | 27.38 | 1.12 |
| CEM | .8 | 111.4 | 111.4 | .03 | .76 |

# CEM in Practice: Non-EPBR Data

Monte Carlo: Exact replication of Diamond and Sekhon (2005), using data from Dehejia and Wahba (1999). CEM coarsening automated.

| | BIAS | SD | RMSE | Seconds | $\mathcal{L}_1$ |
|---|---|---|---|---|---|
| initial | −423.7 | 1566.5 | 1622.6 | .00 | 1.28 |
| MAH | 784.8 | 737.9 | 1077.2 | .03 | 1.08 |
| PSC | 260.5 | 1025.8 | 1058.4 | .02 | 1.23 |
| GEN | 78.3 | 499.5 | 505.6 | 27.38 | 1.12 |
| CEM | .8 | 111.4 | 111.4 | .03 | .76 |

$\rightsquigarrow$ CEM works well in non-EPBR data too

# CEM Extensions I

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory Treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$ ($L_1$ is max difference across treatment groups)

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory Treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$ ($L_1$ is max difference across treatment groups)
- Continuous Treatments: Coarsen treatment and apply CEM as usual

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory Treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$ ($L_1$ is max difference across treatment groups)
- Continuous Treatments: Coarsen treatment and apply CEM as usual
- Blocking in Randomized Experiments: no modification needed; randomly assign $T$ within CEM strata

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory Treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$ ($L_1$ is max difference across treatment groups)
- Continuous Treatments: Coarsen treatment and apply CEM as usual
- Blocking in Randomized Experiments: no modification needed; randomly assign $T$ within CEM strata
- Automating user choices

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory Treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$ ($L_1$ is max difference across treatment groups)
- Continuous Treatments: Coarsen treatment and apply CEM as usual
- Blocking in Randomized Experiments: no modification needed; randomly assign $T$ within CEM strata
- Automating user choices Histogram bin size calculations,

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory Treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$ ($L_1$ is max difference across treatment groups)
- Continuous Treatments: Coarsen treatment and apply CEM as usual
- Blocking in Randomized Experiments: no modification needed; randomly assign $T$ within CEM strata
- Automating user choices Histogram bin size calculations, *Estimated* SATT error bound,

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory Treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$ ($L_1$ is max difference across treatment groups)
- Continuous Treatments: Coarsen treatment and apply CEM as usual
- Blocking in Randomized Experiments: no modification needed; randomly assign $T$ within CEM strata
- Automating user choices Histogram bin size calculations, *Estimated* SATT error bound, Progressive Coarsening

# CEM Extensions I

- CEM and Multiple Imputation for Missing Data
  1. put missing observation in stratum where plurality of imputations fall
  2. pass on uncoarsened imputations to analysis stage
  3. Use the usual MI combining rules to analyze
- Multicategory Treatments: No modification necessary; keep all strata with $\geq 1$ unit having each value of $T$ ($L_1$ is max difference across treatment groups)
- Continuous Treatments: Coarsen treatment and apply CEM as usual
- Blocking in Randomized Experiments: no modification needed; randomly assign $T$ within CEM strata
- Automating user choices Histogram bin size calculations, *Estimated* SATT error bound, Progressive Coarsening
- Detecting Extreme Counterfactuals

# For papers, software (for R and Stata), tutorials, etc.

http://GKing.Harvard.edu/cem