Why Propensity Scores Should Not Be Used For Matching

Gary King\textsuperscript{1} \hspace{1cm} Richard Nielsen\textsuperscript{2}

Institute for Quantitative Social Science
Harvard University \hspace{1cm} MIT

(Talk at the International Methods Colloquium, 9/11/2015)

\textsuperscript{1}GaryKing.org
\textsuperscript{2}www.mit.edu/~rnielsen
The Scholarly Influence of Propensity Score Matching

• The most commonly used matching method
  • In 49,600 articles! (according to Google Scholar)
  • Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)

⇝ This paper is about: propensity score matching, as used in practice.
Not implicated by our results:
• Other uses of propensity scores: E.g., regression adjustment, inverse weighting, stratification, pscores used in other methods
• The mathematical theorems about propensity scores
The Scholarly Influence of Propensity Score Matching

- The most commonly used matching method
The Scholarly Influence of Propensity Score Matching

- The most commonly used matching method
- In 49,600 articles! (according to Google Scholar)
The Scholarly Influence of Propensity Score Matching

- The most commonly used matching method
- In 49,600 articles! (according to Google Scholar)
- Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)
The Scholarly Influence of Propensity Score Matching

- The most commonly used matching method
- In 49,600 articles! (according to Google Scholar)
- Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)
- This paper is about: propensity score matching,
The Scholarly Influence of Propensity Score Matching

• The most commonly used matching method
• In 49,600 articles! (according to Google Scholar)
• Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)
• This paper is about: propensity score matching, as used in practice.
The Scholarly Influence of Propensity Score Matching

- The most commonly used matching method
- In 49,600 articles! (according to Google Scholar)
- Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)
- ⇝ This paper is about: propensity score matching, as used in practice. Not implicated by our results:
The Scholarly Influence of Propensity Score Matching

- The most commonly used matching method
- In **49,600** articles! (according to Google Scholar)
- Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)

⇝ This paper is about: propensity score matching, as used in practice. Not implicated by our results:
  - Other uses of propensity scores: E.g.,
The most commonly used matching method

In 49,600 articles! (according to Google Scholar)

Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)

This paper is about: propensity score matching, as used in practice. Not implicated by our results:

Other uses of propensity scores: E.g., regression adjustment,
The Scholarly Influence of Propensity Score Matching

- The most commonly used matching method
- In 49,600 articles! (according to Google Scholar)
- Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)
- This paper is about: propensity score matching, as used in practice. Not implicated by our results:
  - Other uses of propensity scores: E.g., regression adjustment, inverse weighting,
The Scholarly Influence of Propensity Score Matching

- The most commonly used matching method
- In 49,600 articles! (according to Google Scholar)
- Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)
- This paper is about: propensity score matching, as used in practice. Not implicated by our results:
  - Other uses of propensity scores: E.g., regression adjustment, inverse weighting, stratification,
• The most commonly used matching method
• In 49,600 articles! (according to Google Scholar)
• Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)
• This paper is about: propensity score matching, as used in practice. Not implicated by our results:
  • Other uses of propensity scores: E.g., regression adjustment, inverse weighting, stratification, pscores used in other methods
The Scholarly Influence of Propensity Score Matching

- The most commonly used matching method
- In 49,600 articles! (according to Google Scholar)
- Maybe even “the most developed and popular strategy for causal analysis in observational studies” (Pearl, 2010)
- This paper is about: propensity score matching, as used in practice. Not implicated by our results:
  - Other uses of propensity scores: E.g., regression adjustment, inverse weighting, stratification, pscores used in other methods
  - The mathematical theorems about propensity scores
Matching to Reduce Model Dependence
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, Political Analysis)
Matching to Reduce Model Dependence

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

Education (years) vs. Outcome

Graph showing the relationship between education (years) and outcome, with marked points and a trend line.
The Problems Matching Solves

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

• Choosing based on “plausibility” is probably worse

• Conscientious effort doesn’t avoid biases (Banaji 2013)

• People do not have easy access to their own mental processes
  or feedback to avoid the problem (Wilson and Brekke 1994)

• Experts overestimate their ability to control personal biases
  more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)

• “Teaching psychology is mostly a waste of time” (Kahneman 2011)
The Problems Matching Solves

Without Matching:

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

Without Matching:

Imbalance
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence
The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion
The Problems Matching Solves

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Problems Matching Solves

Without Matching:

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

- Qualitative choice from unbiased estimates = biased estimator
The Problems Matching Solves

Without Matching:

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

- Qualitative choice from unbiased estimates $= \text{biased estimator}$
  - e.g., Choosing from *results* of 50 randomized experiments
The Problems Matching Solves

Without Matching:

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

- Qualitative choice from unbiased estimates $=$ biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse
The Problems Matching Solves

Without Matching:

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

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse\[eff\]
- conscientious effort doesn’t avoid biases (Banaji 2013)\[acc\]
The Problems Matching Solves

Without Matching:

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

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from results of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse
- conscientious effort doesn’t avoid biases (Banaji 2013)
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)
The Problems Matching Solves

Without Matching:

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

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

Without Matching:

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

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

Without Matching:

Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Problems Matching Solves

Without Matching:

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

- Qualitative choice from unbiased estimates = biased estimator
  
  - e.g., Choosing from results of 50 randomized experiments
  
  - Choosing based on "plausibility" is probably worse

- Conscientious effort doesn't avoid biases (Banaji 2013)

- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)

- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)

- "Teaching psychology is mostly a waste of time" (Kahneman 2011)
The Problems Matching Solves

Without Matching:

Imbalance ⇝ Model Dependence ⇝ Researcher discretion ⇝ Bias
The Problems Matching Solves

Without Matching:

- Imbalance $\leadsto$ Model Dependence $\leadsto$ Researcher discretion $\leadsto$ Bias
The Problems Matching Solves

Without Matching:

Imbalance \leadsto Model Dependence \leadsto Researcher discretion \leadsto Bias
The Problems Matching Solves

Without Matching:

\[ \text{Imbalance} \leadsto \text{Model Dependence} \leadsto \text{Researcher discretion} \leadsto \text{Bias} \]

A central project of statistics: Automating away human discretion
What’s Matching?

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

- Treatment Effect for treated observation \( i \):
  \[
  \text{TE}_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}
  \]

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

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
     \[
     \text{SATT} = \text{Mean}_{i \in \{T_i = 1\}}(\text{TE}_i)
     \]
  2. FSATT: Feasible SATT (prune badly matched treateds too)

- Big convenience: Follow preprocessing with whatever statistical method you'd have used without matching
- Pruning nonmatches makes control vars matter less: reduces imbalance, model dependence, researcher discretion, & bias
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
What’s Matching?

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

\[
TE_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}
\]

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

Quantities of Interest:

1. SATT: Sample Average Treatment effect on the Treated:

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

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

Big convenience: Follow preprocessing with whatever statistical method you’d have used without matching

Pruning nonmatches makes control vars matter less: reduces imbalance, model dependence, researcher discretion, & bias
What’s Matching?

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

$$\text{TE}_i = Y_i(1) - Y_i(0)$$
What’s Matching?

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

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

= observed $-$ unobserved
What’s Matching?

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

$$TE_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}$$
What’s Matching?

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

$$\text{TE}_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}$$

- Estimate $Y_i(0)$ with $Y_j$ with a matched ($X_i \approx X_j$) control
What’s Matching?

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

$$TE_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}$$

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

1. SATT: Sample Average Treatment effect on the Treated:
   $$\text{SATT} = \text{Mean}_{i \in \{ T_i = 1 \}} (TE_i)$$

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

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

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

= observed – unobserved

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

$$SATT = \text{Mean }_{i \in \{T_i=1\}} (TE_i)$$
What’s Matching?

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

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

$$= \text{observed} - \text{unobserved}$$

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

$$\text{SATT} = \frac{1}{i \in \{T_i=1\}} \sum_{i \in \{T_i=1\}} (TE_i)$$

  2. FSATT: Feasible SATT (prune badly matched treateds too)
What’s Matching?

- $Y_i$ dep var, $T_i$ (1=treated, 0=control), $X_i$ confounders
- Treatment Effect for treated observation $i$:
  \[
  \text{TE}_i = Y_i - Y_i(0) = \text{observed} - \text{unobserved}
  \]
- Estimate $Y_i(0)$ with $Y_j$ with a matched ($X_i \approx X_j$) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:
     \[
     \text{SATT} = \text{Mean}_{i \in \{T_i=1\}} (\text{TE}_i)
     \]
  2. FSATT: Feasible SATT (prune badly matched treateds too)
- Big convenience: Follow preprocessing with whatever statistical method you’d have used without matching
What’s Matching?

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

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

$= \text{observed} - \text{unobserved}$

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

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

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

Types of Experiments

- Balance
- Covariates:
  - Complete Randomization
  - Fully Blocked

- Observed On average
- Unobserved On average

⇝ Fully blocked dominates complete randomization for:
  - imbalance,
  - model dependence,
  - power,
  - efficiency,
  - bias,
  - research costs,
  - robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

- Balance
  - Covariates:
    - Complete Randomization
    - Fully Blocked
  - Observed
    - On average
  - Unobserved
    - On average

$\Rightarrow$ Fully blocked dominates complete randomization for:
- Imbalance,
- Model dependence,
- Power,
- Efficiency,
- Bias,
- Research costs,
- Robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

- Complete Randomization
- Fully Blocked
- Observed
- Unobserved

\[ \Rightarrow \]

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

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

- Complete Randomization
- Fully Blocked
- Observed On average
- Unobserved On average

\[\Rightarrow\] Fully blocked dominates complete randomization for:
- imbalance,
- model dependence,
- power,
- efficiency,
- bias,
- research costs,
- robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

- PSM: complete randomization
- Other methods: fully blocked
- Other matching methods dominate PSM
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
</table>

- Matching: Finding Hidden Randomized Experiments
- Types of Experiments

- Complete Randomization
- Fully Blocked
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance Covariates:</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On average, fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!
### Matching: Finding Hidden Randomized Experiments

#### Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td></td>
<td>On average</td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- To some extent, fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness.
- E.g., Imai, King, Nall 2009: SEs 600% smaller!
### Matching: Finding Hidden Randomized Experiments

#### Types of Experiments

<table>
<thead>
<tr>
<th>Balance Covariates:</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td></td>
</tr>
</tbody>
</table>

On average, fully blocked designs dominate complete randomization for:
- imbalance,
- model dependence,
- power,
- efficiency,
- bias,
- research costs,
- robustness.

E.g., Imai, King, Nall 2009: SEs 600% smaller!
## Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Matching: Finding Hidden Randomized Experiments

### Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete</th>
<th>Fully</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>Randomization</td>
<td>Blocked</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!
### Matching: Finding Hidden Randomized Experiments

#### Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td><em>Observed</em></td>
<td></td>
<td>On average</td>
</tr>
<tr>
<td><em>Unobserved</em></td>
<td>On average</td>
<td></td>
</tr>
</tbody>
</table>

\[ \leadsto \text{Fully blocked dominates complete randomization} \]
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

→ Fully blocked dominates complete randomization for:
### Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed</strong></td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td><strong>Unobserved</strong></td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

→ *Fully blocked* dominates *complete randomization* for: imbalance,
Types of Experiments

<table>
<thead>
<tr>
<th>Balance Covariates:</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

$\Rightarrow$ Fully blocked dominates complete randomization for: imbalance, model dependence,
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>On average</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

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

E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\[\Rightarrow\] Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency,
## Matching: Finding Hidden Randomized Experiments

### Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>Unobserved</td>
<td>On average</td>
</tr>
</tbody>
</table>

\[\Rightarrow \text{Fully blocked dominates complete randomization for:}\]
imbalance, model dependence, power, efficiency, bias,
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete</th>
<th>Fully</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>Randomization</td>
<td>Blocked</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

$\Rightarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs,
### Matching: Finding Hidden Randomized Experiments

#### Types of Experiments

<table>
<thead>
<tr>
<th>Balance Covariates</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\[\sim\text{ Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness.}\]
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\(\Rightarrow\) *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>Unobserved</td>
<td>On average</td>
</tr>
</tbody>
</table>

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

Goal of Each Matching Method (in Observational Data)
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

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

Goal of Each Matching Method (in Observational Data)

- PSM: complete randomization
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance Covariates:</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

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

Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*
- Other methods: *fully blocked*
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariates:</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Observed</td>
<td>On average</td>
<td></td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

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

Goal of Each Matching Method (in Observational Data)

- PSM: complete randomization
- Other methods: fully blocked
- Other matching methods dominate PSM
Matching: Finding Hidden Randomized Experiments

Types of Experiments

<table>
<thead>
<tr>
<th>Balance Covariates:</th>
<th>Complete Randomization</th>
<th>Fully Blocked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>On average</td>
<td>Exact</td>
</tr>
<tr>
<td>Unobserved</td>
<td>On average</td>
<td>On average</td>
</tr>
</tbody>
</table>

\[\rightsquigarrow\text{Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600\% smaller!}\]

Goal of Each Matching Method (in Observational Data)

- PSM: \textit{complete randomization}
- Other methods: \textit{fully blocked}
- Other matching methods dominate PSM (wait, it gets worse)
Method 1: Mahalanobis Distance Matching

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

2. Estimation Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

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

2. Estimation
   Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)

2. Estimation Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. **Preprocess** (Matching)
   - Distance($X_c, X_t$) = $\sqrt{(X_c - X_t)'S^{-1}(X_c - X_t)}$

2. **Estimation** Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance\( \left( X_c, X_t \right) = \sqrt{(X_c - X_t)'S^{-1}(X_c - X_t)} \)
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)

2. Estimation Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

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

2. Estimation Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

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

2. Estimation Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

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

2. **Estimation** Difference in means or a model
Method 1: Mahalanobis Distance Matching
(Approximates Fully Blocked Experiment)

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

2. Estimation Difference in means or a model
Mahalanobis Distance Matching

Age

Education (years)
Mahalanobis Distance Matching

Education (years) vs. Age

Age

12 14 16 18 20 22 24 26 28

Education (years)
Mahalanobis Distance Matching

Age

Education (years)
Best Case: Mahalanobis Distance Matching
Best Case: Mahalanobis Distance Matching

![Graph showing data points in a scatter plot with axes labeled Education (years) and Age.]
Best Case: Mahalanobis Distance Matching

![Graph showing correlation between education and age](image-url)
Method 2: Coarsened Exact Matching

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

2. Estimation
   - Difference in means or a model
   - Weight controls in each stratum to equal treateds
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

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

2. Estimation Difference in means or a model
   - Weight controls in each stratum to equal treateds
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)

2. Estimation Difference in means or a model
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

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

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Temporarily coarsen $X$ as much as you’re willing
     - e.g., Education (grade school, high school, college, graduate)

2. Estimation Difference in means or a model
Method 2: Coarsened Exact Matching  
(Approximates Fully Blocked Experiment)

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

2. **Estimation**  Difference in means or a model
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

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

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching  
(Approximates Fully Blocked Experiment)

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

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

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

2. **Estimation** Difference in means or a model
Method 2: Coarsened Exact Matching
(Approximates Fully Blocked Experiment)

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

2. Estimation Difference in means or a model
   - Weight controls in each stratum to equal treateds
Coarsened Exact Matching
Coarsened Exact Matching

Age

Education

12 14 16 18 20 22 24 26 28

20

30

40

50

60

70

80

CCC C
CC CC
C CC C CCC CCCC CCC CC CCC CCCCCC
C CCC CC C
T T
T T
TT TT T TT
TTT TT
T
T TT
Coarsened Exact Matching

Coarsened education:
- HS
- BA
- MA
- PhD
- 2nd PhD

Drinking age:
- Don’t trust anyone over 30
- The Big 40

Senior Discounts:
- Old
- Retirement
- The Big 40
- Don’t trust anyone over 30
- Drinking age

Education levels:
- HS
- BA
- MA
- PhD
- 2nd PhD
Coarsened Exact Matching

Education

Drinking age
Don't trust anyone over 30
The Big 40
Senior Discounts
Retirement
Old

C C C
CC
CC C CC
C
TT T T TT
TTT TT
T
T TT

HS BA MA PhD 2nd PhD
Best Case: Coarsened Exact Matching
Best Case: Coarsened Exact Matching

Education

Age

12 14 16 18 20 22 24 26 28

20

30

40

50

60

70

80

CC CCC CCC CC CC CCCC CCC CC C C CCCCC CCC CC CC ...

T TTTT T T TTT TT T T TT TTTT T T TTT TT T TT TTTT TTT TT TTT TT
TT TT TT
Best Case: Coarsened Exact Matching
Best Case: Coarsened Exact Matching

Age

Education

12 14 16 18 20 22 24 26 28

20 30 40 50 60 70 80

C C C C C C C C C C C C C C C C C C C C

T T T T T T T T T T T T T T T T T T T T
Method 3: 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($X_c, X_t$) = $|\pi_c - \pi_t|$
   • Match each treated unit to the nearest control unit
   • Control units: not reused; pruned if unused
   • Prune matches if Distance $> \text{caliper}$
   • (Many adjustments available to this basic method)

2. Estimation Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

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

2. Estimation Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)

2. Estimation Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. **Preprocess (Matching)**
   - Reduce $k$ elements of $X$ to scalar
     \[
     \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-x_i\beta}}
     \]

2. **Estimation** Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

1. **Preprocess (Matching)**
   - Reduce $k$ elements of $X$ to scalar
     \[ \pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}} \]
   - Distance($X_c$, $X_t$) = $|\pi_c - \pi_t|$

2. **Estimation** Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

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

2. **Estimation** Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

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

2. Estimation Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

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

2. **Estimation** Difference in means or a model
Method 3: Propensity Score Matching
(Approximates Completely Randomized Experiment)

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

2. **Estimation** Difference in means or a model
Propensity Score Matching

Education (years)

Age

12 16 20 24 28

20

30

40

50

60

70

80

C

C

CC

C

C

CC

C

C

CC

C

CC

C

CC

C

C

C

C

C

CC

CC

C

C

T TTT T

T

T

T

T

T

T

T

TT

T

T

T

T

T
<table>
<thead>
<tr>
<th>Education (years)</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>C</td>
</tr>
<tr>
<td>16</td>
<td>C</td>
</tr>
<tr>
<td>20</td>
<td>C</td>
</tr>
<tr>
<td>24</td>
<td>C</td>
</tr>
<tr>
<td>28</td>
<td>C</td>
</tr>
<tr>
<td>30</td>
<td>C</td>
</tr>
<tr>
<td>40</td>
<td>C</td>
</tr>
<tr>
<td>50</td>
<td>C</td>
</tr>
<tr>
<td>60</td>
<td>C</td>
</tr>
<tr>
<td>70</td>
<td>C</td>
</tr>
<tr>
<td>80</td>
<td>C</td>
</tr>
</tbody>
</table>

Propensity Score Matching

Propensity Score (0-1)
Propensity Score Matching

Propensity Score

Age

Education (years)
Propensity Score Matching

Age

Education (years)

Propensity Score

0
1

14/23
<table>
<thead>
<tr>
<th>Education (years)</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
</tr>
</tbody>
</table>

Propensity Score Matching
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching
Best Case: Propensity Score Matching is Suboptimal
Random Pruning Increases Imbalance

• "Random pruning": pruning process is independent of $X$

• Discrete example

• Sex-balanced dataset: treated $M_t$, $F_t$, controls $M_c$, $F_c$

• Randomly prune 1 treated & 1 control $\Rightarrow$ 4 possible datasets:
  - 2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
  - 2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$

• $\Rightarrow$ random pruning increases imbalance

• Continuous example

• Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units

• Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$

• $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)

• Random pruning $\Rightarrow n$ declines $\Rightarrow E(d^2)$ increases

• $\Rightarrow$ random pruning increases imbalance
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

• “Random pruning”: pruning process is independent of $X$

**Discrete example**

- Sex-balanced dataset: treated $M_t, F_t$, controls $M_c, F_c$
- Randomly prune 1 treated & 1 control $\Rightarrow$ 4 possible datasets:
  - 2 balanced: $\{M_t, M_c\}, \{F_t, F_c\}$
  - 2 imbalanced: $\{M_t, F_c\}, \{F_t, M_c\}$

• $\Rightarrow$ random pruning increases imbalance

**Continuous example**

- Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
- Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
- $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
- Random pruning $\Rightarrow n$ declines $\Rightarrow E(d^2)$ increases

• $\Rightarrow$ random pruning increases imbalance
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- Discrete example

- Sex-balanced dataset: treated $M_t$, $F_t$, controls $M_c$, $F_c$
  - Randomly prune 1 treated & 1 control $\Rightarrow$ 4 possible datasets:
    - 2 balanced: $\{M_t, M_c\}$, $\{F_t, F_c\}$
    - 2 imbalanced: $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\Rightarrow$ random pruning increases imbalance

- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)

- Random pruning $\Rightarrow$ $n$ declines $\Rightarrow E(d^2)$ increases
  - $\Rightarrow$ random pruning increases imbalance
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- Discrete example
  - Sex-balanced dataset: treated $M_t, F_t$, controls $M_c, F_c$

\[ \begin{align*}
\text{Randomly prune 1 treated & 1 control} & \rightarrow 4 \text{ possible datasets:} \\
& \{M_t, M_c\}, \{F_t, F_c\} \\
& \{M_t, F_c\}, \{F_t, M_c\} \\
\end{align*} \]

$= \Rightarrow \text{random pruning increases imbalance}$

- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)

$\Rightarrow \text{random pruning increases imbalance}$
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of \( X \)
- Discrete example
  - Sex-balanced dataset: treateds \( M_t, F_t \), controls \( M_c, F_c \)
  - Randomly prune 1 treated & 1 control \( \rightarrow \) 4 possible datasets:
    2 balanced \( \{M_t, M_c\}, \{F_t, F_c\} \)
    2 imbalanced \( \{M_t, F_c\}, \{F_t, M_c\} \)
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- **Discrete example**
  - Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\mapsto$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
- $\implies$ random pruning increases imbalance
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- Discrete example
  - Sex-balanced dataset: treated $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\leadsto$ 4 possible datasets:
    - 2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    - 2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- Continuous example
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- Discrete example
  - Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\leadsto$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
Random Pruning Increases Imbalance

Deleting data only helps if you’re careful!

• “Random pruning”: pruning process is independent of $X$

• Discrete example
  
  • Sex-balanced dataset: treated $M_t, F_t$, controls $M_c, F_c$
  • Randomly prune 1 treated & 1 control $\Rightarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  
  $\Rightarrow$ random pruning increases imbalance

• Continuous example
  
  • Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  • Measure of imbalance: squared difference in means $d^2$, where
    $d = \bar{X}_t - \bar{X}_c$
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

• “Random pruning”: pruning process is independent of $X$

• Discrete example
  • Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  • Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  • $\implies$ random pruning increases imbalance

• Continuous example
  • Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  • Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
  • $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- “Random pruning”: pruning process is independent of $X$
- **Discrete example**
  - Sex-balanced dataset: treated $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\sim$ 4 possible datasets:
    - 2 balanced $\{M_t, M_c\}, \{F_t, F_c\}$
    - 2 imbalanced $\{M_t, F_c\}, \{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- **Continuous example**
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
  - Random pruning $\sim n$ declines $\sim E(d^2)$ increases
Random Pruning Increases Imbalance
Deleting data only helps if you’re careful!

- “Random pruning”: pruning process is independent of $X$
- **Discrete example**
  - Sex-balanced dataset: treateds $M_t, F_t$, controls $M_c, F_c$
  - Randomly prune 1 treated & 1 control $\leadsto$ 4 possible datasets:
    - 2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    - 2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- **Continuous example**
  - Dataset: $T \in \{0, 1\}$ randomly assigned; $X$ any fixed variable; with $n$ units
  - Measure of imbalance: squared difference in means $d^2$, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
  - Random pruning $\leadsto n$ declines $\leadsto E(d^2)$ increases
  - $\implies$ random pruning increases imbalance
PSM’s Statistical Properties

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

2. The PSM Paradox:
   - When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning)
   - All $\hat{\pi} \approx 0.5$ (or constant within strata)
   - Pruning at random
   - Imbalance
   - Inefficency
   - Model dependence
   - Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem?
   - No.
   - The PSM Paradox gets worse with more covariates
   - What if I match on a few important covariates and then use PSM?
   - The low standards will be raised some, but the PSM Paradox will kick in earlier
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   
   ...
PSM’s Statistical Properties

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

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

   \[ \pi_c = \pi_t \Rightarrow \pi_c = \pi_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning)
     \[ \Rightarrow \hat{\pi} \approx 0 \text{(or constant within strata)} \]
     \[ \Rightarrow \text{pruning at random} \Rightarrow \text{Imbalance} \Rightarrow \text{Inefficency} \Rightarrow \text{Model dependence} \Rightarrow \text{Bias} \]
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem?
     - Nope.
   - The PSM Paradox gets worse with more covariates

   What if I match on a few important covariates and then use PSM?
   The low standards will be raised some, but the PSM Paradox will kick in earlier
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \]
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \not\implies X_c = X_t \]
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning)
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata)
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \sim \) pruning at random
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \rightsquigarrow \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \rightsquigarrow \) pruning at random \( \rightsquigarrow \) Imbalance
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \leadsto \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \leadsto \) pruning at random \( \leadsto \) Imbalance \( \leadsto \) Inefficiency
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \sim \) pruning at random \( \sim \) Imbalance \( \sim \) Inefficency \( \sim \) Model dependence
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \nLeftarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \leadsto \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \leadsto \) pruning at random \( \leadsto \) Imbalance \( \leadsto \) Inefficency \( \leadsto \) Model dependence \( \leadsto \) Bias
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ \pi_c = \pi_t \implies X_c = X_t \]
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \implies \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \implies \) pruning at random \( \implies \) Imbalance \( \implies \) Inefficency \( \implies \) Model dependence \( \implies \) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
PSM’s Statistical Properties

1. **Low Standards**: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox**: When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \leadsto \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \leadsto \) pruning at random \( \leadsto \) Imbalance \( \leadsto \) Inefficiency \( \leadsto \) Model dependence \( \leadsto \) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem?
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but } \pi_c = \pi_t \not\implies X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \rightsquigarrow \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \rightsquigarrow \) pruning at random \( \rightsquigarrow \) Imbalance \( \rightsquigarrow \) Inefficency \( \rightsquigarrow \) Model dependence \( \rightsquigarrow \) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem? Nope.
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \rightsquigarrow \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \rightsquigarrow \) pruning at random \( \rightsquigarrow \) Imbalance \( \rightsquigarrow \) Inefficency \( \rightsquigarrow \) Model dependence \( \rightsquigarrow \) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem? Nope. The PSM Paradox gets worse with more covariates.
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \leadsto \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \leadsto \) pruning at random \( \leadsto \) Imbalance \( \leadsto \) Inefficiency \( \leadsto \) Model dependence \( \leadsto \) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem? Nope. The PSM Paradox gets worse with more covariates
   - What if I match on a few important covariates and then use PSM?
PSM’s Statistical Properties

1. **Low Standards:** Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     \[ X_c = X_t \implies \pi_c = \pi_t \text{ but} \]
     \[ \pi_c = \pi_t \nRightarrow X_c = X_t \]

2. **The PSM Paradox:** When you do “better,” you do worse
   - When PSM approximates complete randomization (to begin with or, after some pruning) \( \sim \) all \( \hat{\pi} \approx 0.5 \) (or constant within strata) \( \sim \) pruning at random \( \sim \) Imbalance \( \sim \) Inefficency \( \sim \) Model dependence \( \sim \) Bias
   - If the data have no good matches, the paradox won’t be a problem but you’re cooked anyway.
   - Doesn’t PSM solve the curse of dimensionality problem? Nope. The PSM Paradox gets worse with more covariates
   - What if I match on a few important covariates and then use PSM? The low standards will be raised some, but the PSM Paradox will kick in earlier
PSM is Blind Where Other Methods Can See
PSM is Blind Where Other Methods Can See

\[ \begin{array}{c}
X_1 \\
X_2
\end{array} \]
PSM is Blind Where Other Methods Can See
What Does PSM Match?

MDM Matches

PSM Matches

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

Treated: \(X_1, X_2 \sim \text{Uniform}(1,6)\)
**PSM Increases Model Dependence & Bias**

**Model Dependence**

- **Variance vs. Number of Units Pruned**

**Bias**

- **Maximum Coefficient across 512 Specifications**

\[ Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i \]

\[ \epsilon_i \sim N(0, 1) \]
The Propensity Score Paradox in Real Data

Finkel et al. (JOP, 2012)

Nielsen et al. (AJPS, 2011)
Conclusions

• Why propensity scores should not be used for matching
• Low Standards: sometimes helps, never optimizes
• The PSM Paradox: When you do “better,” you do worse
• Some mistakes with PSM: Controlling for irrelevant covariates; adjusting experimental data; reestimating propensity score after eliminating noncommon support; $\frac{1}{4}$ caliper on propensity score; not switching to other methods.

• A warning for any matching method:
  - Pruning discards information; you must overcome this.
  - Other methods can generate a “paradox” if you prune after approximating full blocking (rare, but possible)
• If you’re not doing positive good, you may be hurting yourself

Matching methods still highly recommended; choose one with higher standards.
Conclusions

- Why propensity scores should not be used for matching
Conclusions

• Why propensity scores should not be used for matching
  • Low Standards: sometimes helps, never optimizes
Conclusions

• Why propensity scores should not be used for matching
  • Low Standards: sometimes helps, never optimizes
  • The PSM Paradox: When you do “better,” you do worse
Conclusions

- Why propensity scores should not be used for matching
  - Low Standards: sometimes helps, never optimizes
  - The PSM Paradox: When you do “better,” you do worse
  - Some mistakes with PSM:
Conclusions

• Why propensity scores should not be used for matching
  • **Low Standards:** sometimes helps, never optimizes
  • **The PSM Paradox:** When you do “better,” you do worse
  • Some mistakes with PSM: Controlling for irrelevant covariates;
Conclusions

• Why propensity scores should not be used for matching
  • Low Standards: sometimes helps, never optimizes
  • The PSM Paradox: When you do “better,” you do worse
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data;
Conclusions

• Why propensity scores should not be used for matching
  • Low Standards: sometimes helps, never optimizes
  • The PSM Paradox: When you do “better,” you do worse
  • Some mistakes with PSM: Controlling for irrelevant covariates;
    Adjusting experimental data; Reestimating propensity score
    after eliminating noncommon support;
Conclusions

• Why propensity scores should not be used for matching
  • Low Standards: sometimes helps, never optimizes
  • The PSM Paradox: When you do “better,” you do worse
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score;
Conclusions

• Why propensity scores should not be used for matching
  • Low Standards: sometimes helps, never optimizes
  • The PSM Paradox: When you do “better,” you do worse
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
Conclusions

- Why propensity scores should not be used for matching
  - Low Standards: sometimes helps, never optimizes
  - The PSM Paradox: When you do “better,” you do worse
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

- A warning for any matching method:
Conclusions

• Why propensity scores should not be used for matching
  • Low Standards: sometimes helps, never optimizes
  • The PSM Paradox: When you do “better,” you do worse
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

• A warning for any matching method:
  • Pruning discards information; you must overcome this.
Conclusions

• Why propensity scores should not be used for matching
  • **Low Standards**: sometimes helps, never optimizes
  • **The PSM Paradox**: When you do “better,” you do worse
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

• A warning for any matching method:
  • Pruning discards information; you must overcome this.
  • Other methods can generate a “paradox” if you prune after approximating full blocking (rare, but possible)
Conclusions

• Why propensity scores should not be used for matching
  • Low Standards: sometimes helps, never optimizes
  • The PSM Paradox: When you do “better,” you do worse
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

• A warning for any matching method:
  • Pruning discards information; you must overcome this.
  • Other methods can generate a “paradox” if you prune after approximating full blocking (rare, but possible)
  • If you’re not doing positive good, you may be hurting yourself
Conclusions

• Why propensity scores should not be used for matching
  • Low Standards: sometimes helps, never optimizes
  • The PSM Paradox: When you do “better,” you do worse
  • Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

• A warning for any matching method:
  • Pruning discards information; you must overcome this.
  • Other methods can generate a “paradox” if you prune after approximating full blocking (rare, but possible)
  • If you’re not doing positive good, you may be hurting yourself

• Matching methods still highly recommended; choose one with higher standards
For more information, papers, & software

GaryKing.org
www.mit.edu/~rnielsen