Research Design
Beyond Randomized Control Trials

Jody Worley, Ph.D.
College of Arts & Sciences
Human Relations
Introduction to the series

• Day 1: Nonrandomized Designs
• Day 2: Sampling Strategies
• Day 3: Matching Techniques for Balanced Designs
Matching Methods

• What are matching methods?
• Why use matching methods?
• How to use matching methods?
What are Matching Methods?

• A method to improve causal inferences in observational data.
  Ho et al., 2007; Morgan and Winship, 2014

• The goal of matching is to reduce imbalance.
  Stuart, 2010, p.13

• Lowering imbalance reduces the degree of model dependence, and, as a result, reduces inefficiency and bias.
  Ho et al., 2007; Imai, King and Stuart, 2008; Iacus, King and Porro, 2011

• Generally speaking, matching can be thought of as a technique for finding ideal experimental data hidden within an observational data set.
Why Use Matching Methods?

- The resulting process amounts to a search for a data set that might have resulted from a randomized experiment but is hidden in an observational data set.

- When matching can reveal this “hidden experiment,” many of the problems of observational data analysis vanish.

Okay, great, but ...
If the goal of matching is to reduce imbalance, what is imbalance?

and...

If lowering imbalance reduces the degree of model dependence, what is model dependence?

Also, how does reducing imbalance reduce inefficiency and bias?
Model Dependence
Example from Ho, Imai, King and Stuart (2007) Figure 1, Political Analysis
Model Dependence

Example from Ho, Imai, King and Stuart (2007) Figure 1, *Political Analysis*
Model Dependence
Example from Ho, Imai, King and Stuart (2007) Figure 1, Political Analysis
What are the problems that matching solves?

**Without Matching**

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

- Qualitative choice from unbiased estimates $=\text{biased estimator}$
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on “plausibility” is probably worse

- Conscientious effort doesn’t avoid bias (Banaji, 2013).
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson & 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 that matching solves

Without Matching

balance ≈ Model Dependence ≈ Researcher discretion ≈ Bias

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

- \( Y_i \) dep var, \( T_i \) (1=treated, 0=control), \( X_i \) = confounders
- Treatment Effect for treated observation \( i \):
  \[
  \text{TE}_i = Y_i - Y_i (0)
  \]
  = 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:
     \[
     \text{SATT} = \text{Mean} \left( \text{TE}_i \right)
     \]
  2. FSATT: Feasible SATT (prune badly matched treated too)
    - **Big Convenience**: Follow preprocessing with whatever statistical method you would have used without matching
    - **Pruning nonmatches makes the control variables matter less.** reduces imbalance, model dependence, researcher discretion, and bias.
Classes of Matching Methods

Each method defined here represents one of the two existing classes of matching methods:

1. Equal Percent Bias Reducing (EBPR) class
2. Monotonic Imbalance Bounding (MIB) class

• **Mahalanobis Distance Matching (MDM)** is one of the longest standing matching methods that can fall within the Equal Percent Bias Reducing (EPBR) class
  
  Rubin, 1976; Rubin and Stuart, 2006

• **Coarsened Exact Matching (CEM)** is the leading example within the Monotonic Imbalance Bounding (MIB) class.

  Iacus, King and Porro, 2011

Note: Propensity Score Matching (PSM) can also be EPBR, if used with appropriate data.