MLDS CENTER

Maryland Longitudinal Data System

Better Data • Informed Choices • Improved Results

Program Evaluation in the Absence of Randomization: How Can Propensity Score Methods Help?

Angela K. Henneberger, Ph.D.
MLDS Center &
University of Maryland

Presented at Maryland Connections Summit
June 6, 2018

https://mldscenter.maryland.gov/

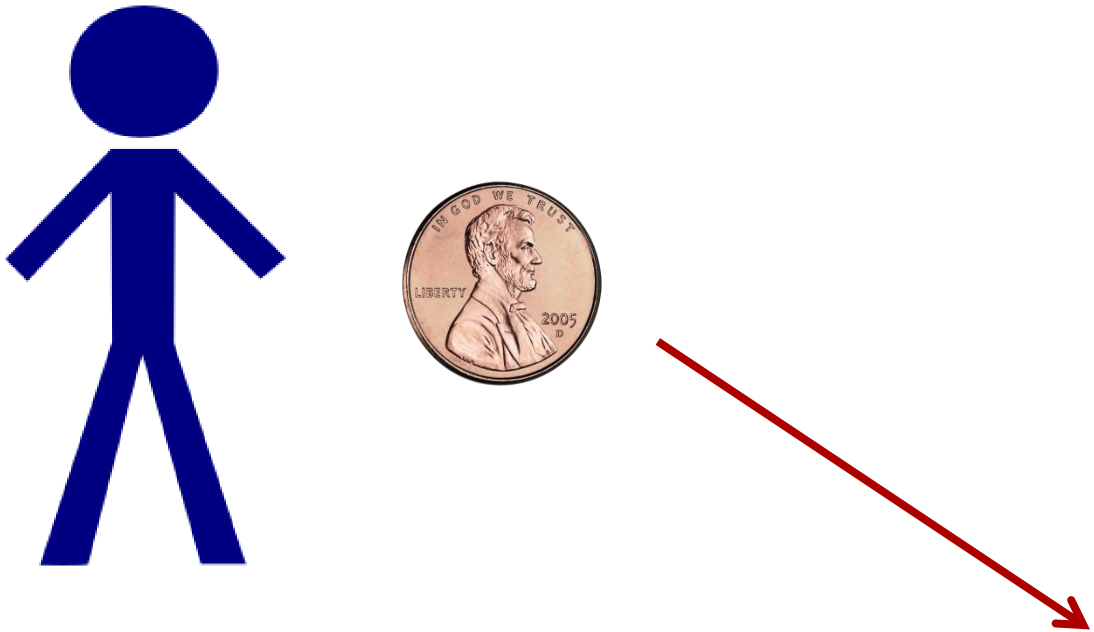
Program and Policy Evaluation

- Policy makers and practitioners are often interested in the effect of a program or policy on student outcomes
- Difficult to examine in the absence of randomization to treatment and control groups
- Quasi-experimental designs can be used to statistically mimic randomization (Cook, Campbell, & Shadish, 2002)
  - With specific assumptions
  - Internal validity
  - External validity

The Gold Standard Randomized Controlled Trial (RCT)

- Randomize students to participate in the treatment or receive no treatment (control)
- In this design, each student has a 50% chance to be in the treatment group

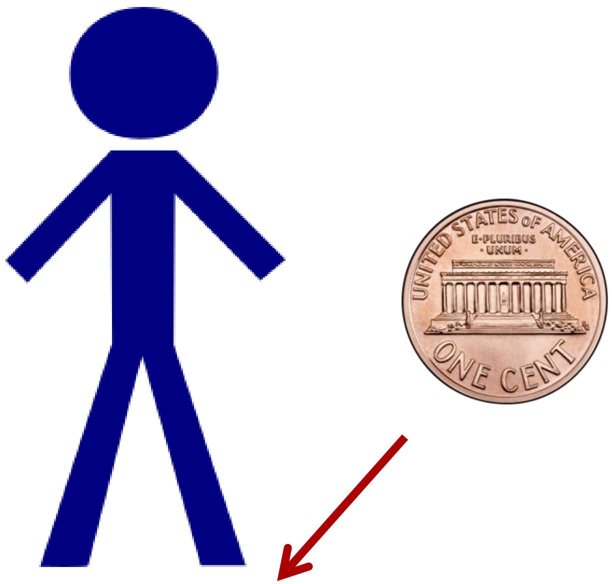
The Gold Standard RCT



Treatment

Control

The Gold Standard RCT

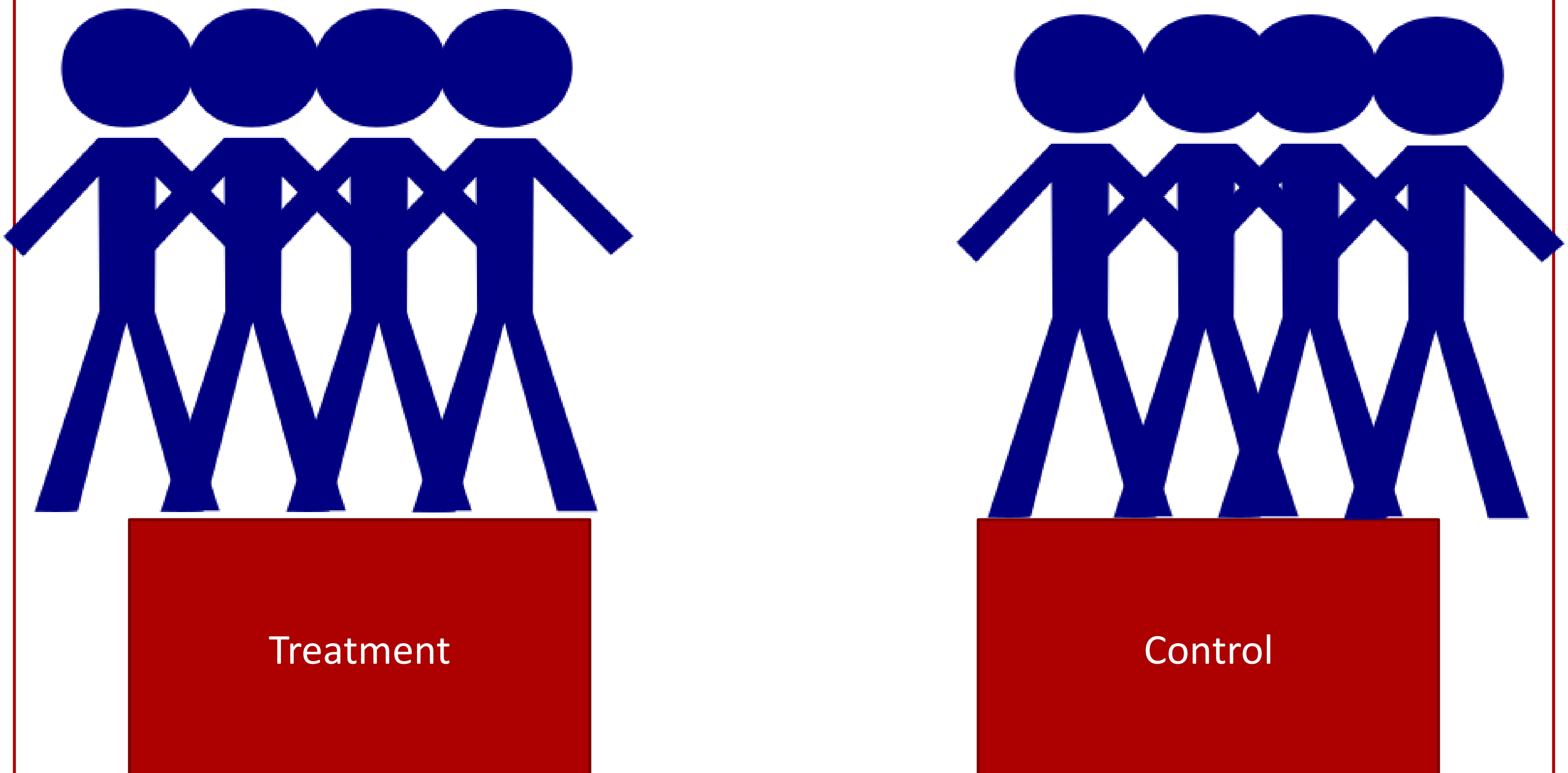


Treatment

Control

MLDS Center
Maryland Longitudinal Data System

The Gold Standard RCT



The Gold Standard RCT

- Randomize students to participate in the treatment or receive no treatment (control)
- In this design, each student has a 50% chance to be in the treatment group
- When sample sizes are large, confounders should be balanced across groups (testable)
- Power analysis can help to determine sample size
- RCT measures the causal effect of a treatment on an outcome (the gold standard)
- High internal validity, external validity varies

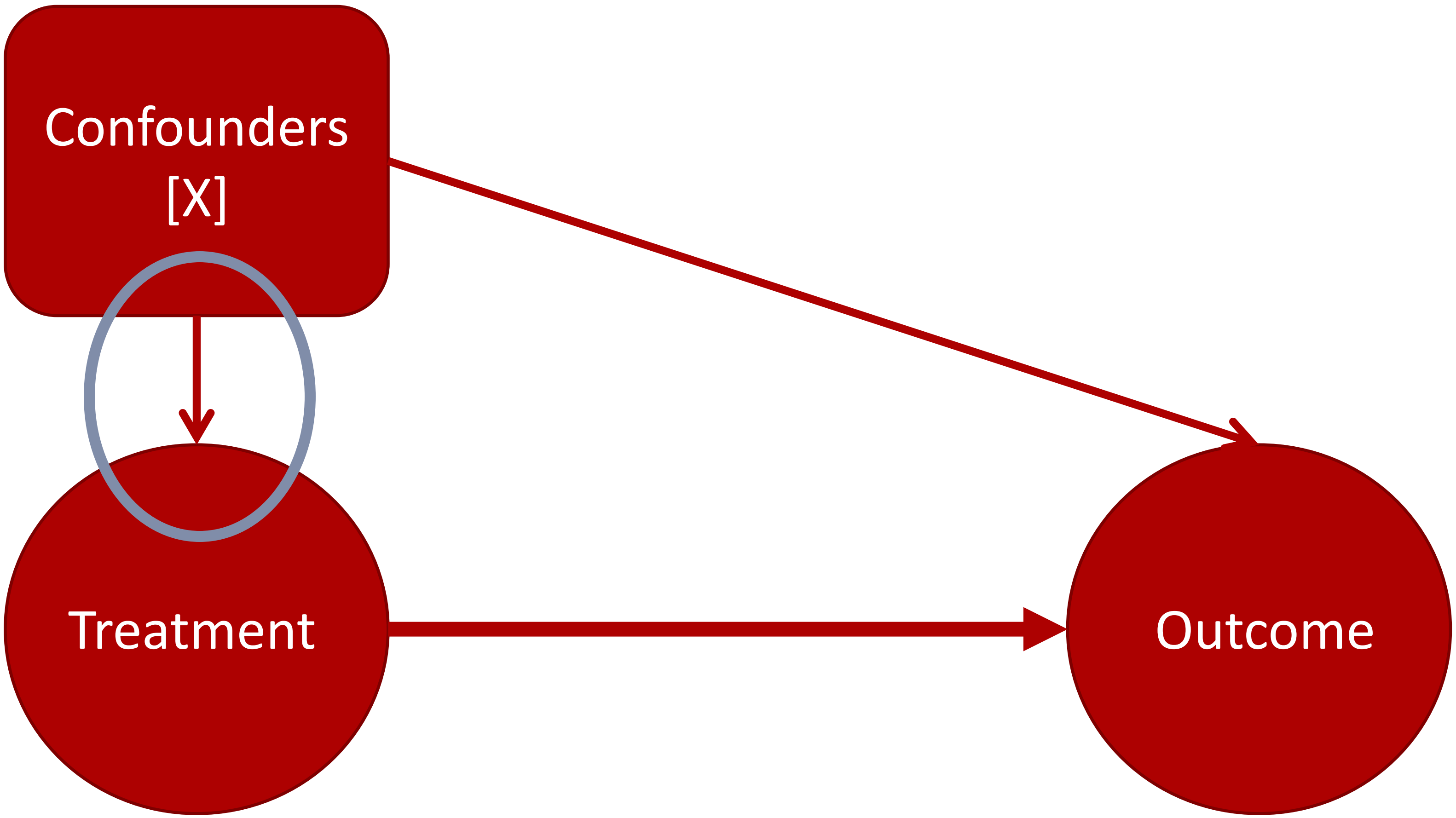
Limitations of the RCT

- Difficult to implement in the “real world”
- Costly
- Time-consuming
- Sometimes randomization is not feasible
- Sometimes randomization is unethical

The Real World

- Local, state, and federal agencies are often using observational (correlational) data
- Observations are collected on the same individuals over time
  - E.g., each school year, each fiscal year, each semester
- No randomization to treatment and control groups
- However, the interest in evaluating the causal effect of a program or policy remains…

The Problem: Confounders



Propensity Score Methods

- Modern causal inference techniques can be used to account for the absence of random assignment (Schafer & Kang, 2008).
- Propensity Score Methods
  - Propensity score is the conditional probability of experiencing the "treatment" given individual's values on confounders (Rosenbaum & Rubin, 1983).
  - The propensity score estimates the probability to participate in the "treatment".
  - Range 0-1; higher = greater likelihood to participate in the "treatment"
- Improves the ability to make causal inferences about program participation in the absence of randomization

MLDS Center
Maryland Longitudinal Data System

Propensity Score Weighting

- Treatment and control groups are not simple random samples from the population
  - Treatment group has an oversampling of people with high propensities
  - Control group has an oversampling of people with low propensities
- Inverse probability of treatment weighting (IPTW) can provide an unbiased effect estimate for the population (with assumptions):
  - Down-weight oversampled cases
  - Up-weight under-sampled cases

Calculating Weights

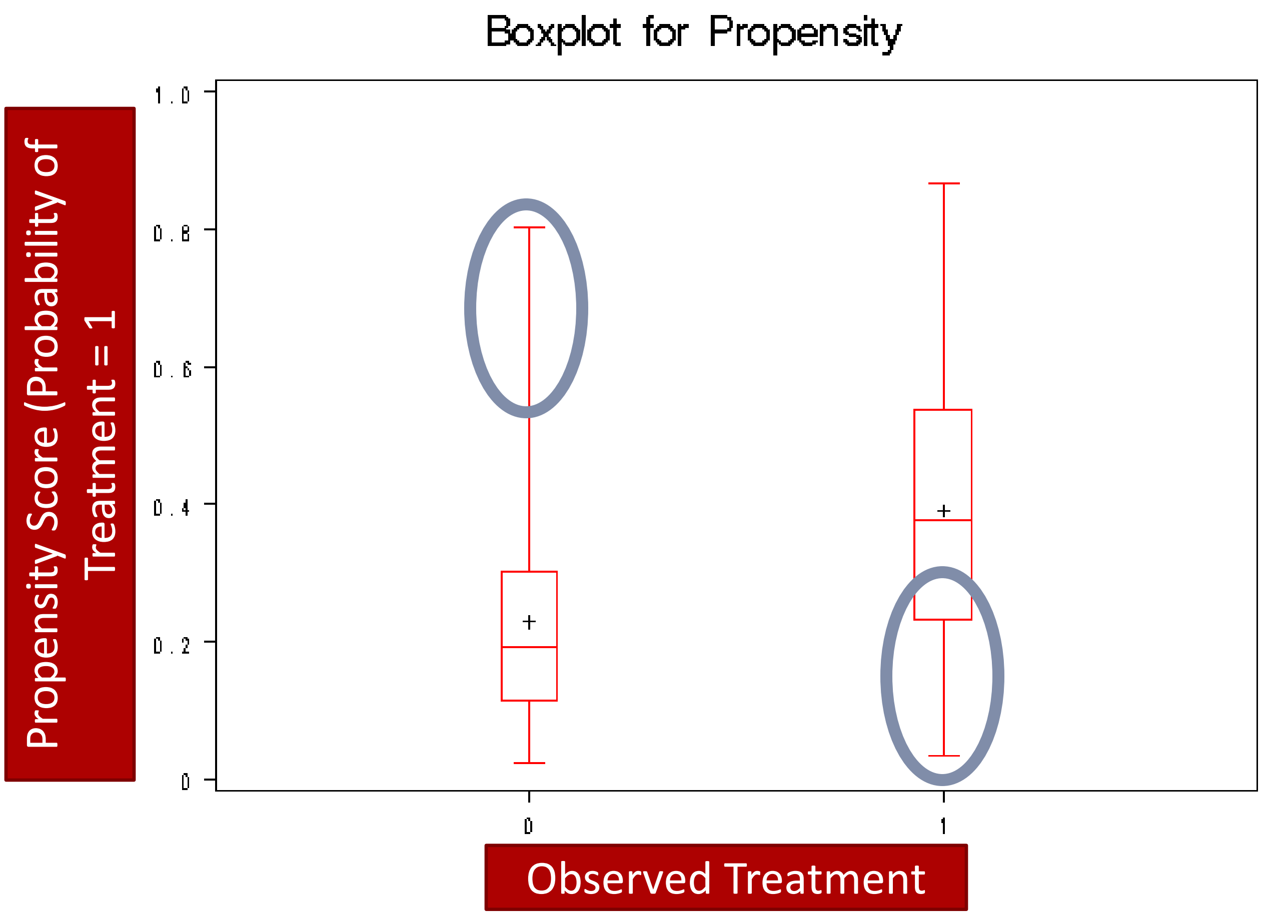
Observed Treatment = 1
(Treatment group)

- $w_i = 1/p(x_i)$
- If $p(x_i) = .75$
- $w_i = 1/.75 = 1.33$

Observed Treatment = 0
(Control group)

- $w_i = 1/1\text{-}p(x_i)$
- $p(x_i) = .75$
- $w_i = 1/1\text{-}.75 = 1/.25 = 4$

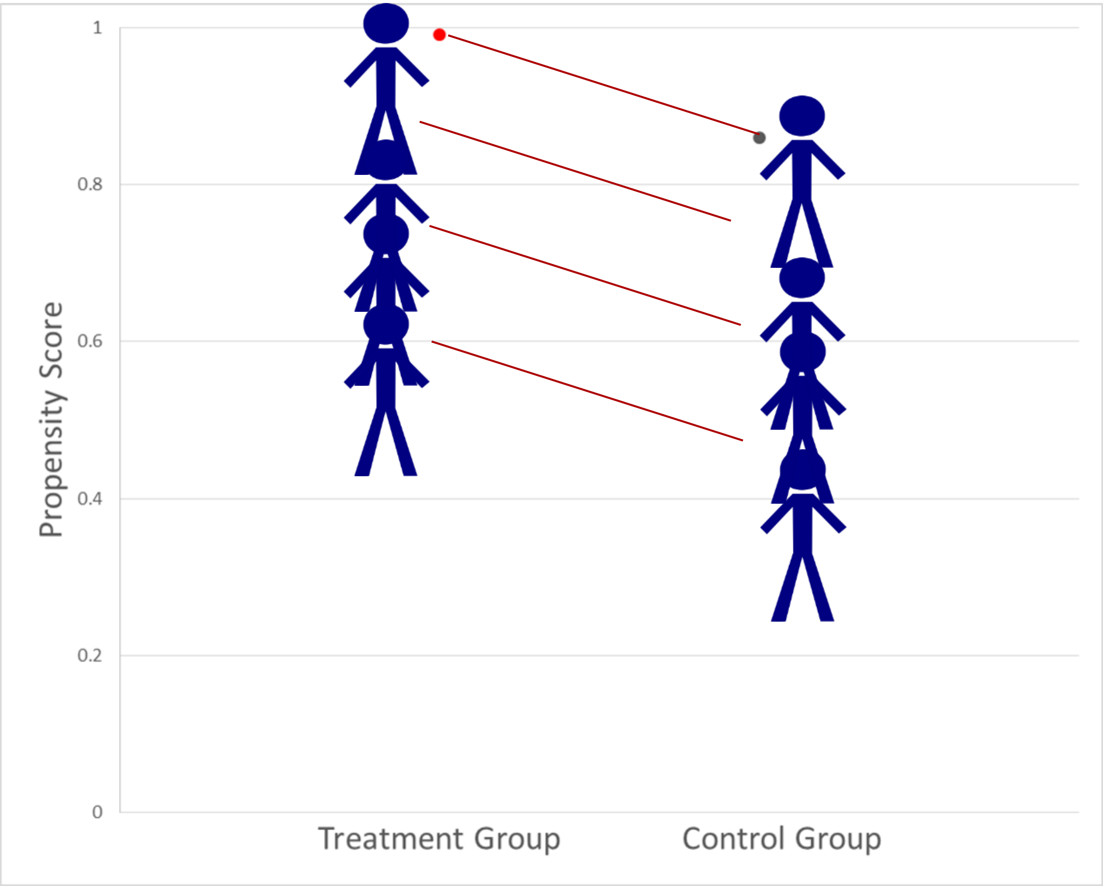
Inverse Probability of Treatment Weighting (IPTW)



Propensity Score Matching

- Treatment and control groups are not equivalent due to confounding variables
- Matching individuals in the treatment group to individuals in the control group based on propensity score can provide a causal estimate (with assumptions)
  - Estimate propensity score
  - Match students within a certain range of propensity score (e.g., caliper = 0.2)
  - Run outcome analyses with matched sample

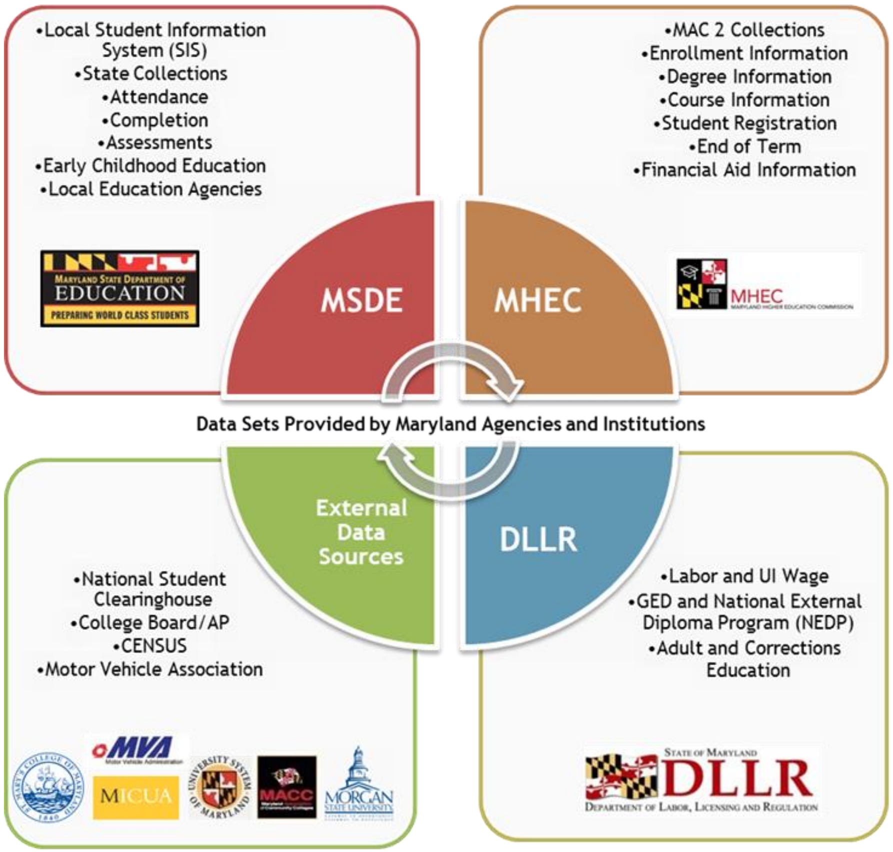
Propensity Score Matching



A Real World Example from the MLDS Center

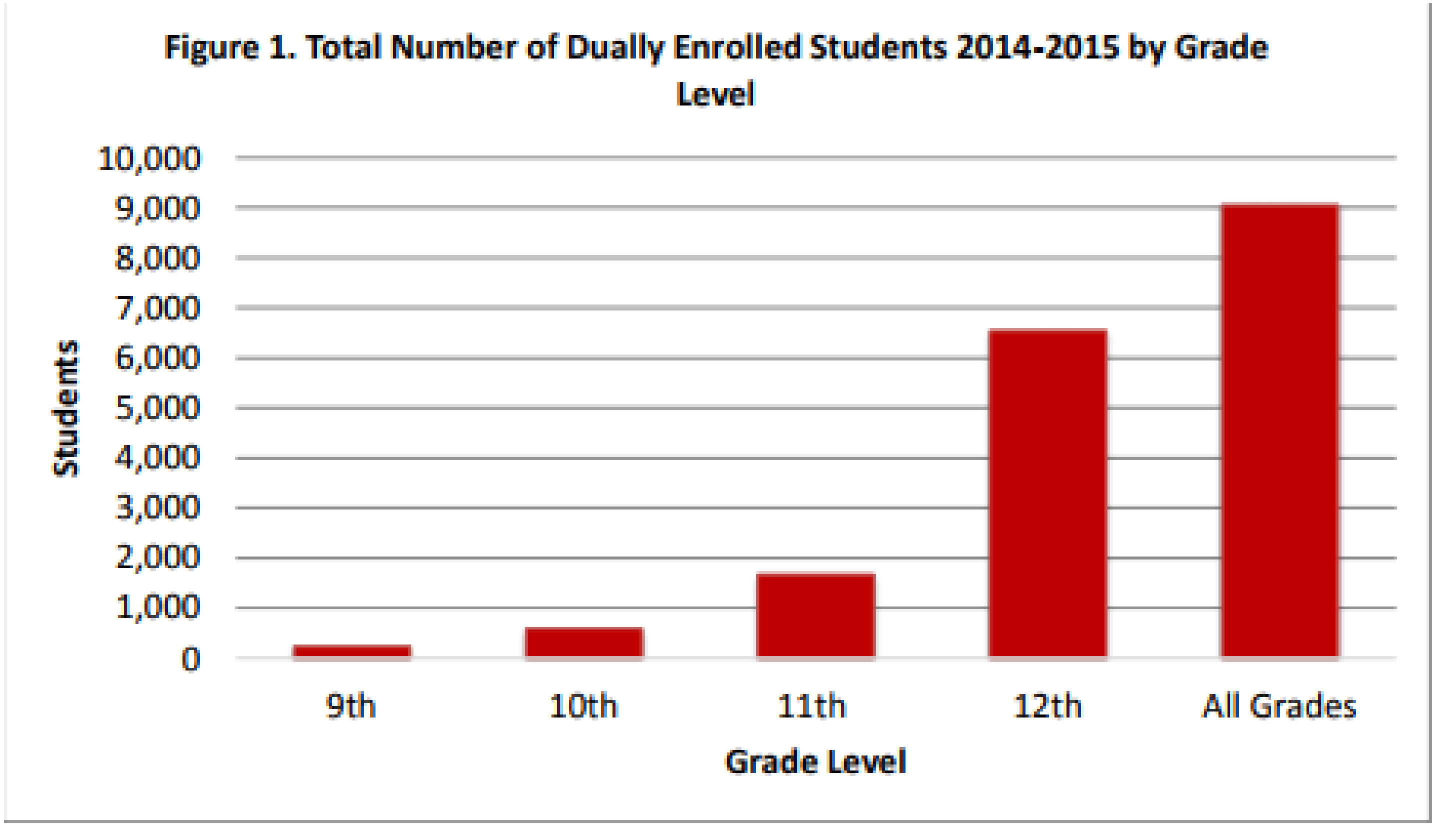
- What is the MLDS Center?
  - Independent unit of State government
  - ***Purpose***: generate timely and accurate information about student performance that can be used to improve the State's education system and guide decision makers at all levels
  - ***The MLDS Center partners with the University of Maryland*** to conduct advanced statistical analyses and policy evaluation to provide actionable information for policy and practice

The MLDS Data



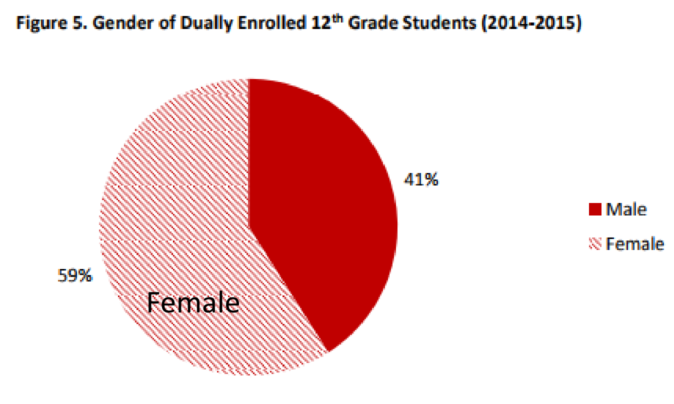
Maryland's Dual Enrollment Report

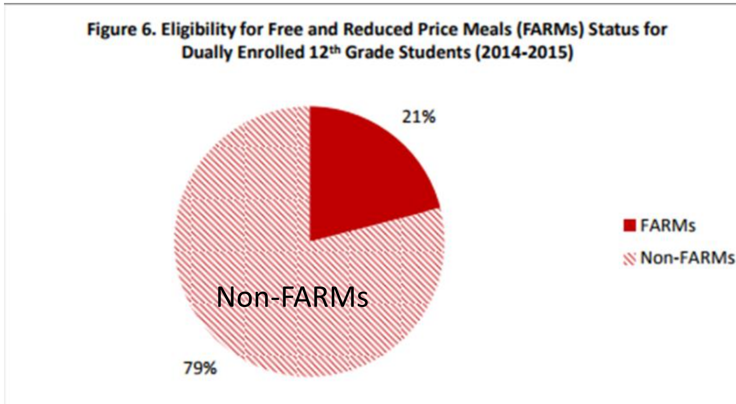
Dual Enrollment in Maryland

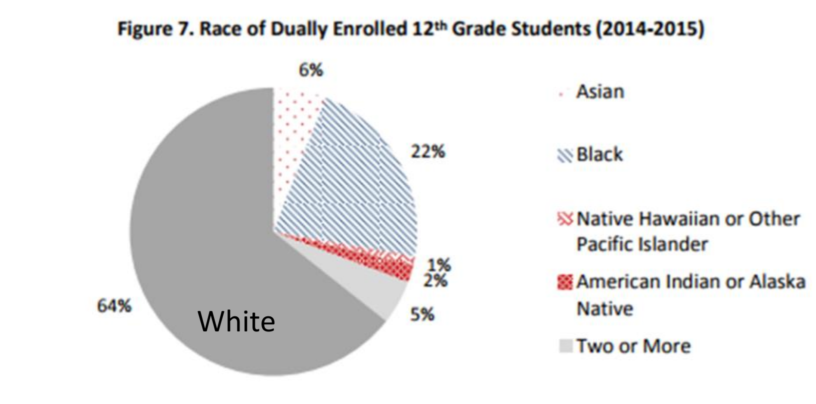


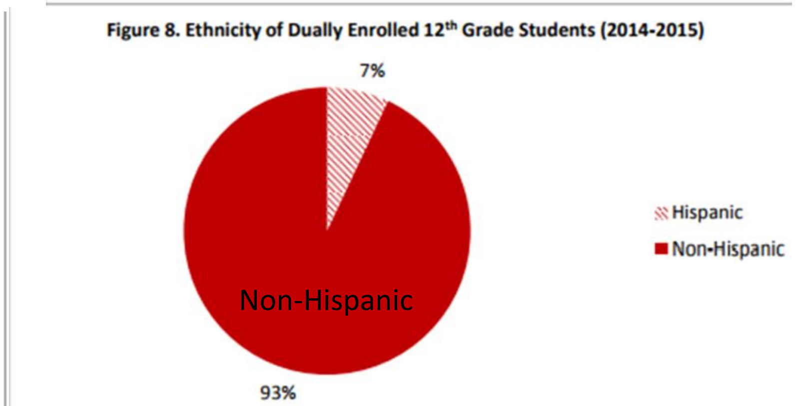
Source: Henneberger, Cohen, Shipe, & Shaw, 2016

Characteristics of Dually Enrolled Students







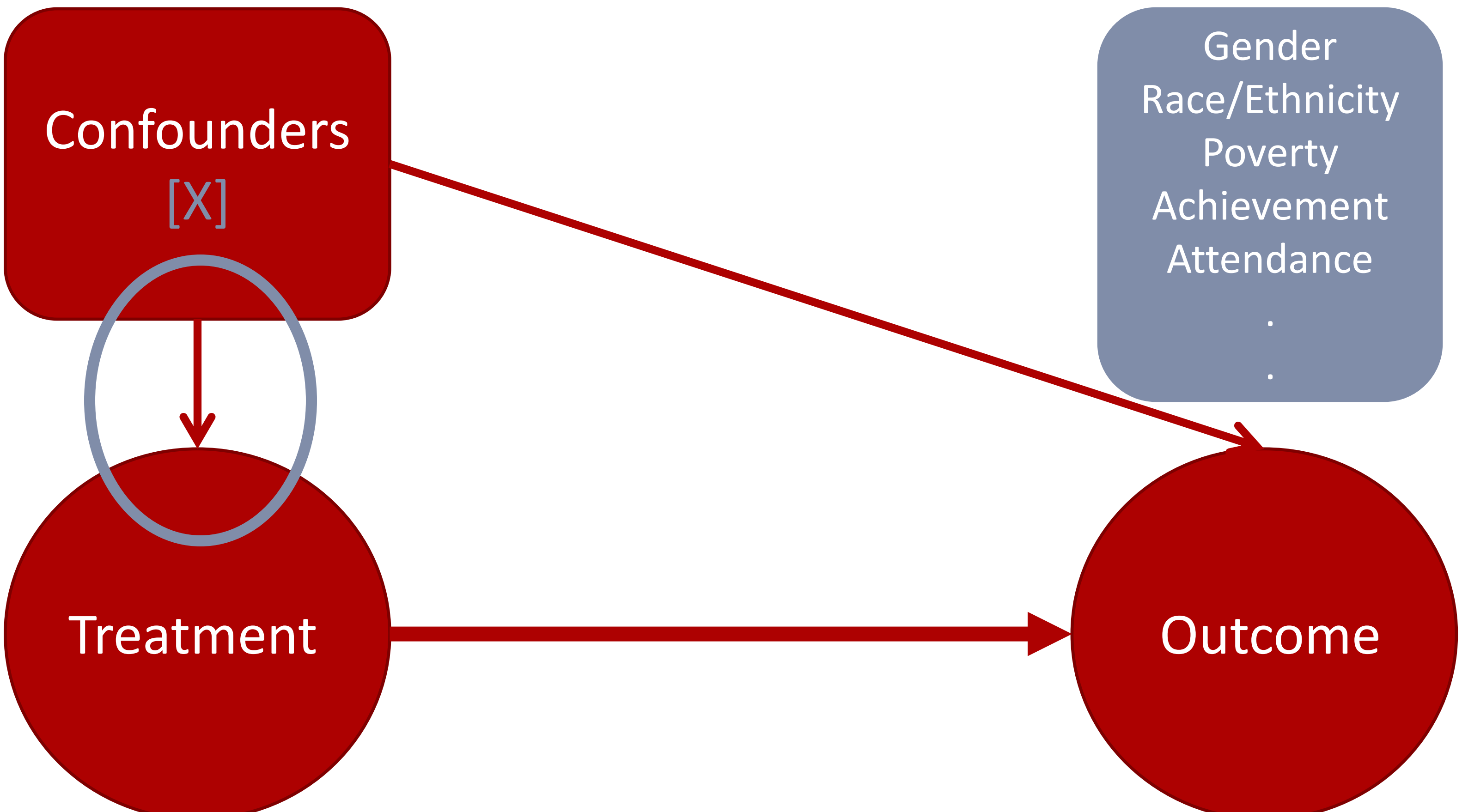


Source: Henneberger, Cohen, Shipe, & Shaw, 2016

Research Question and Motivation

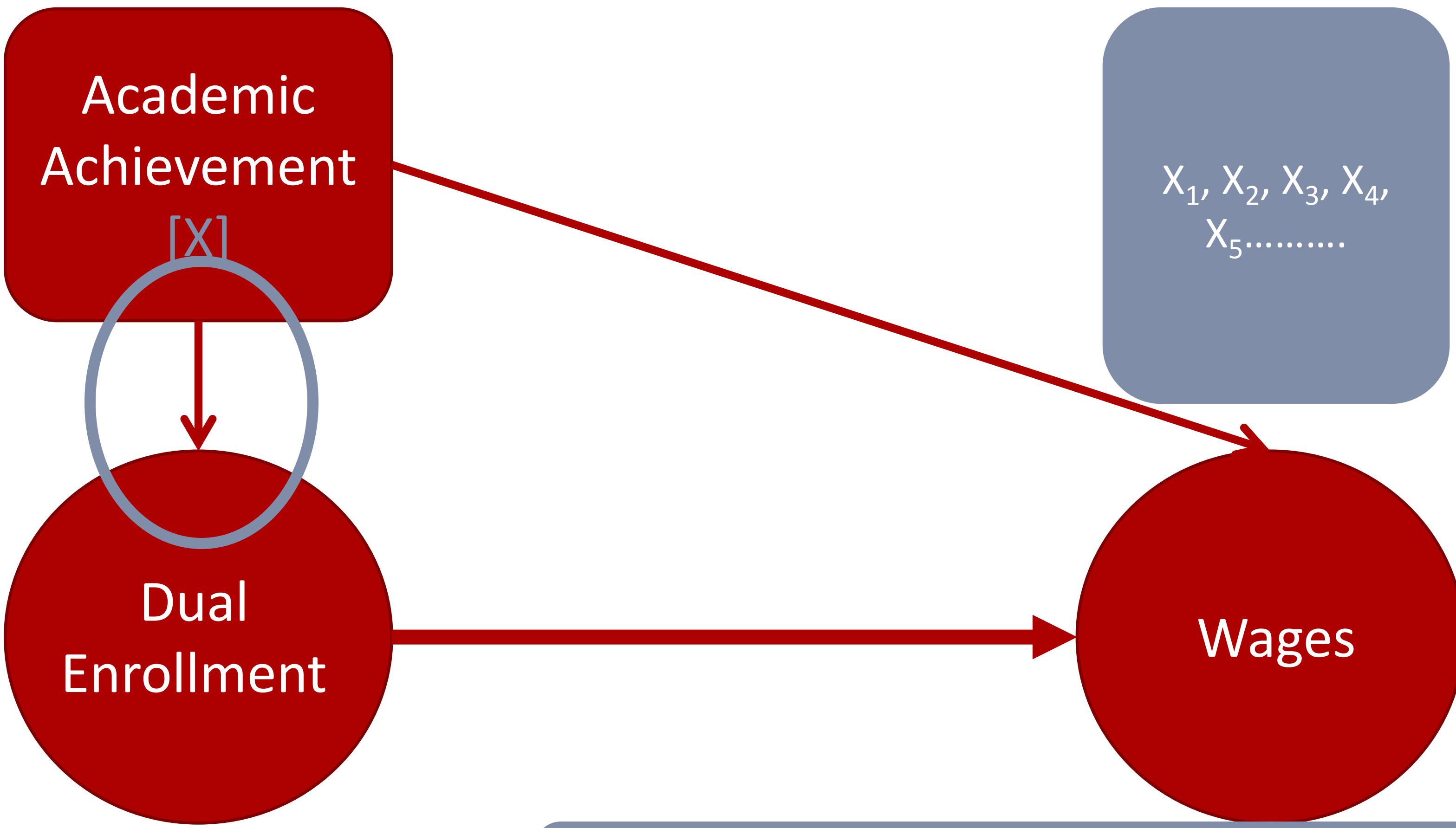
- Motivating Research Question:
  - What is the ***effect*** of dual enrollment program participation in high school on college enrollment outcome, degree attainment, and earnings?
  - ***Effect*** implies a causal design where dual enrollment ***causes*** a change in outcomes.
- Ideal design = randomization to dual enrollment program and control (Cook, Campbell, & Shadish, 2002)
  - But…. Our data are correlational.

The Problem: Confounders



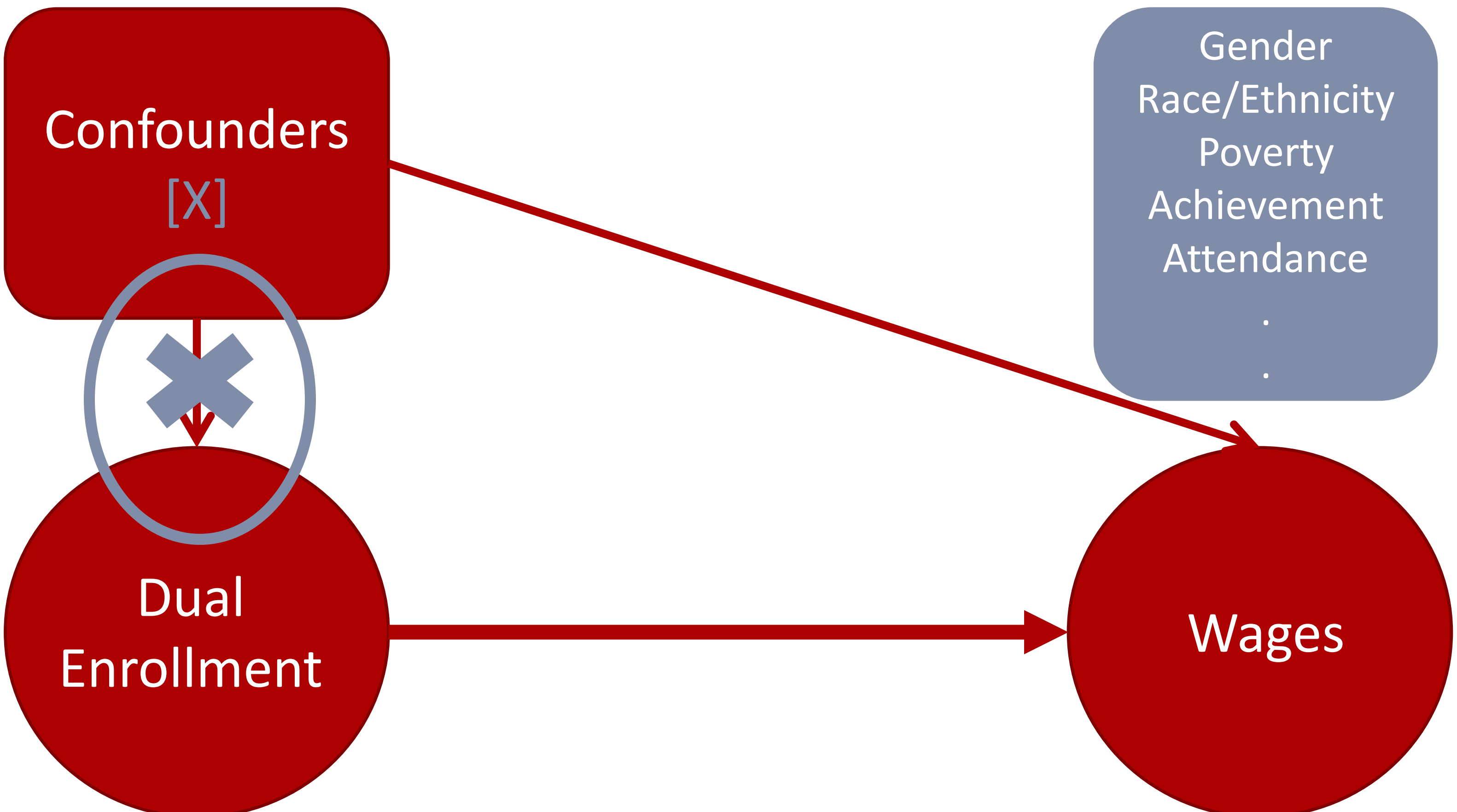
Confounders make it hard to know whether any relationship found is due to the treatment itself or due to confounding variables.

Example of the Problem: Academic Achievement



Is the relationship between dual enrollment and wages due to dual enrollment itself or due to academic achievement (selection bias)?

The Solution: Propensity Score Methods



What is the ***causal effect*** of dual enrollment program participation on workforce wages?

Method: Study Sample

- Student identified as dually enrolled if:
  - Overlapping enrollment dates in MD public high school and MD college
- Population for 2009-2010 cohort:
  - 63,000 12th grade students (2009-2010)
  - 4,200 were dually enrolled
  - Outcomes: college enrollment, degree completion, wages 6 years after high school graduation

Method: Confounders

Confounders Predicting Dual Enrollment Program Participation (0/1)
Demographic Variables: Gender, Race, Ethnicity
Program Participation: Eligibility for Free and Reduced Price Meals (FARMS), Special Education, Homelessness
Academic Indicators: High School Assessment (HSA) Algebra, English, Biology (Presence of score * score), Number of Advanced Placement (AP) tests taken (by subject), 3.0 GPA indicator, Weeks Absent
Distance of high school to nearest 2-year college
Local School System: to account for differences between school systems that may make students in some local school systems more likely to dually enroll (e.g., course offerings, incentives, district agreements with community colleges)
Matching implemented in R; nearest neighbor match; 1:1; Caliper = 0.2

MLDS Center
Maryland Longitudinal Data System

Method: Analyses

$$ATT = E[Y_1 - Y_0 \mid D = 1, X]$$

ATT = Average treatment on the treated

D = Treatment status

X = Vector of covariates

(Rosenbaum & Rubin, 1983)

Method: Assumptions

- *Unconfoundedness*: Conditional on propensity score (and thus covariates), assignment to treatment is independent of outcomes.

$$(Y_0, Y_1) \perp D \mid P(X)$$

- *Overlap:* The probability of being treated is bounded away from 0 or 1.

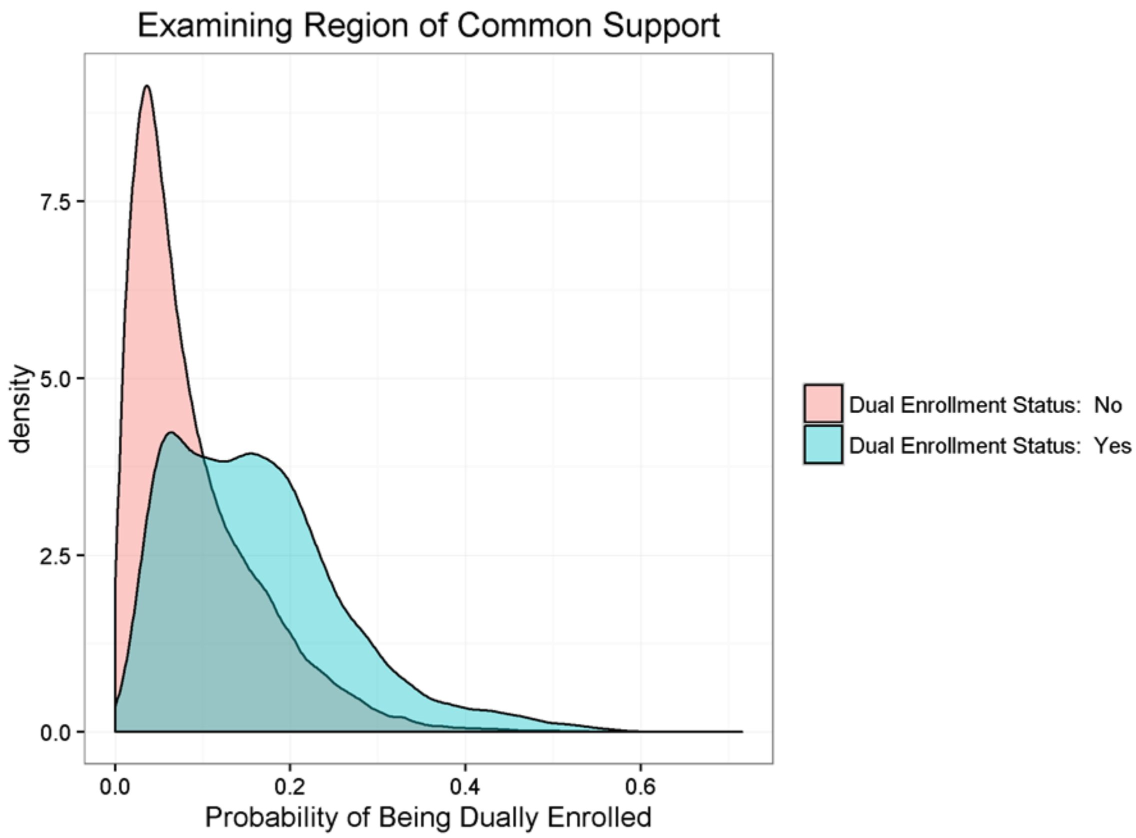
$$0 < P(X) < 1$$

- *No unmeasured confounders*

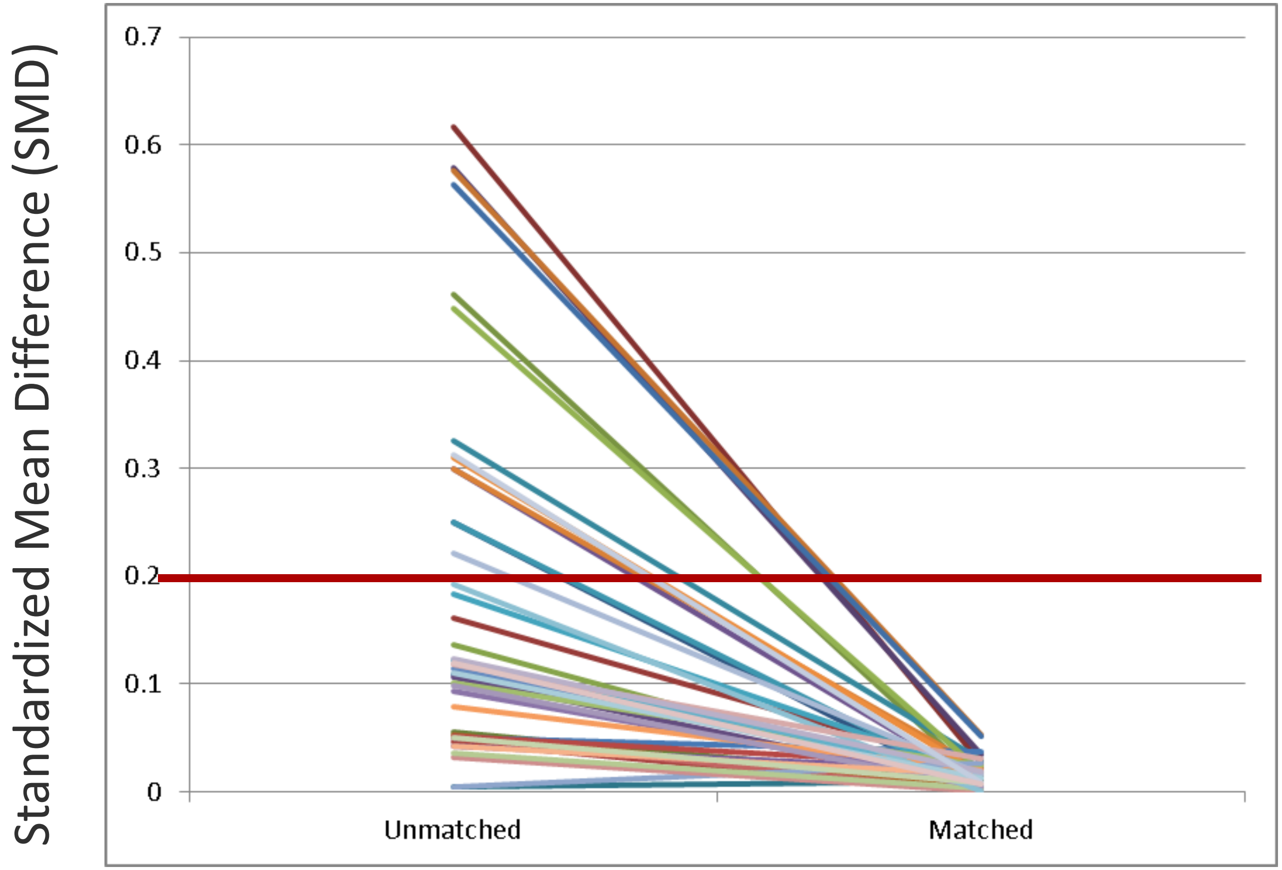
$$ATT = E[Y_1 - Y_0 \mid D = 1, X]$$

(Rosenbaum & Rubin, 1983)

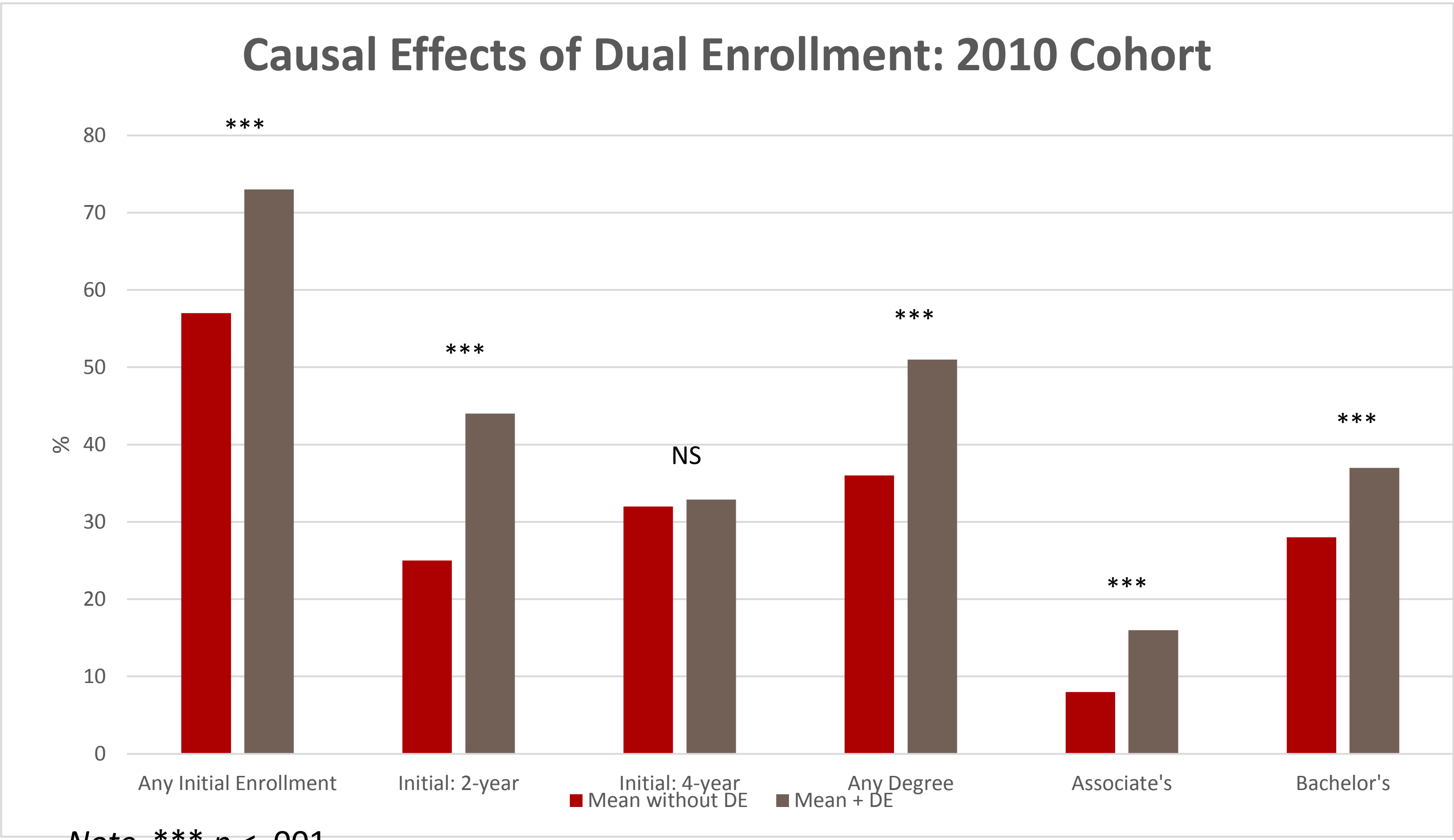
Method: Overlap



Method: Balance on Confounders

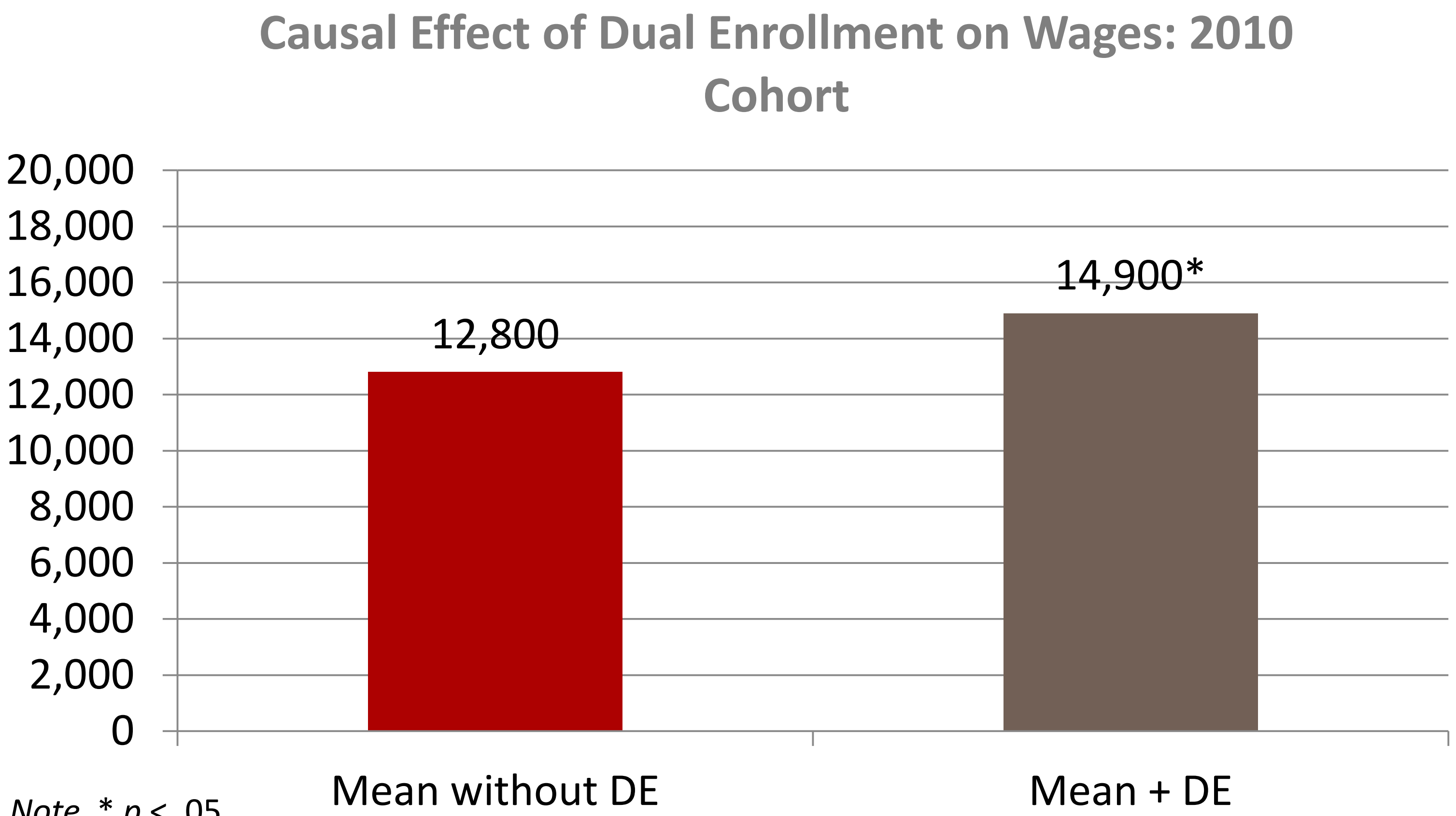


Results: College Enrollment and Degree



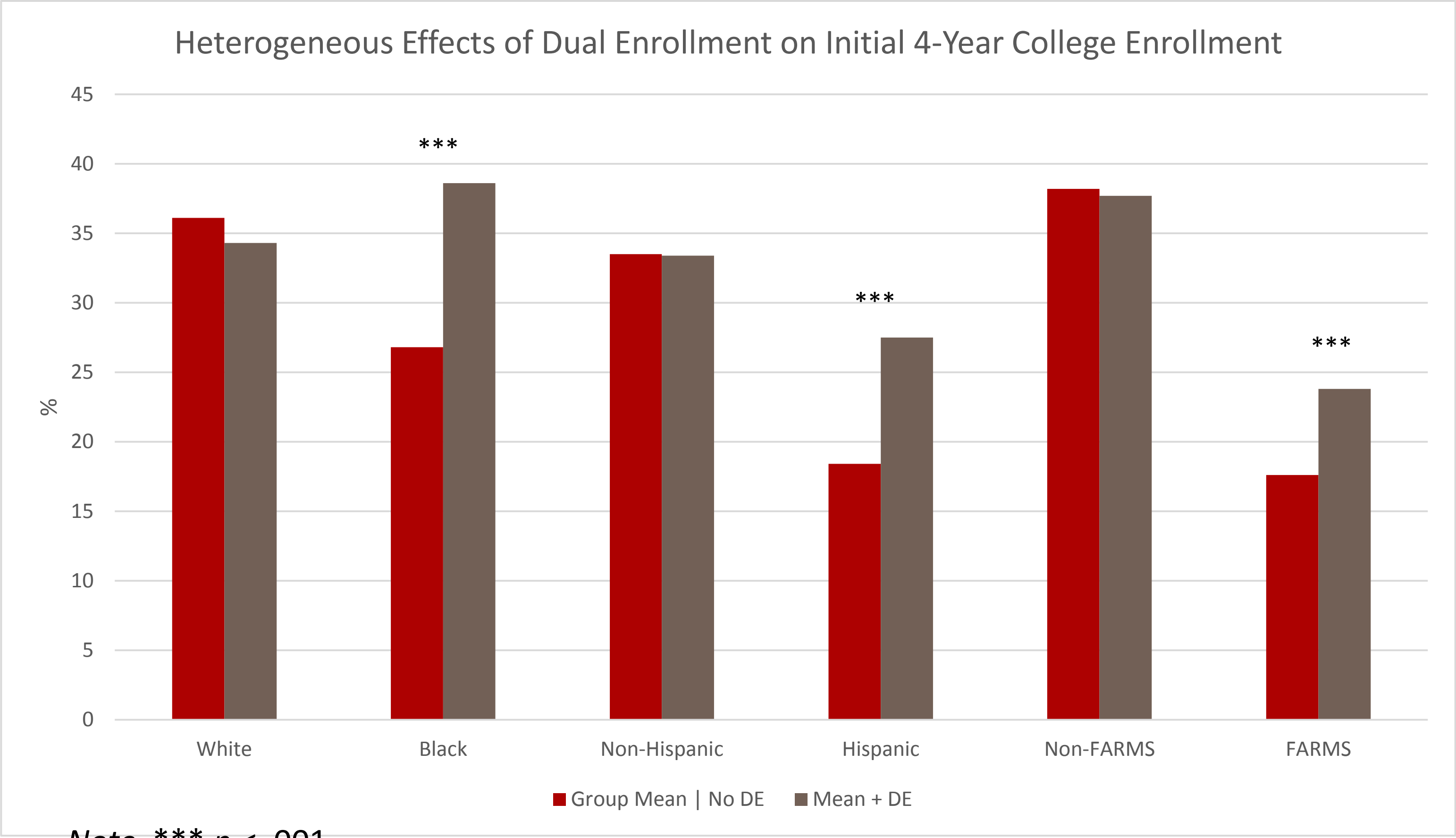
Note. *** $p < .001$

Results: Wages 6 Years after High School



Note. * $p < .05$

Results: Heterogeneity of Effects



Note. *** *p* < .001

Limitations

- Propensity score methods assume no unmeasured confounders—
  - Academic motivation
  - Behavioral problems
  - Etc.
- The MLDS data do not offer the granularity needed to provide more nuanced comparisons of types of dual enrollment program participation and outcomes (e.g., characteristics of district partnership; Early Middle College program).

MLDS Center
Maryland Longitudinal Data System

Strengths

- The ability to draw causal conclusions about the effect of dual enrollment participation is improved through using propensity score matching.
  - This approach gave us the ability to efficiently control for >25 confounding variables.
  - No assumption that confounders are linearly related to predictor and multicollinearity between confounders is not a factor.
  - Ability to examine diagnostics (balance and overlap/common support) to ensure the method worked.
- Propensity score matching is a powerful statistical tool that helps to answer research questions about the ***effect of a policy, practice, or program*** on outcomes in the absence of randomization.

MLDS Center
Maryland Longitudinal Data System

For More Information

https://mldscenter.maryland.gov/

Resources on Causal Inference

- Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55.
- Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods, 13*, 279–313.

Thank you!

Contact :
Angela K. Henneberger, Ph.D.
Principal Investigator and Research Director
MLDS Center
Angela.henneberger@maryland.gov

Acknowledgements: Thanks to Heath Witzen and Alison Preston, co-authors on this project.