AIR®
Advancing Evidence.
Improving Lives.

Evidence Quality and Reach Hub: Propensity Score Weighting/Matching

Session 2

August 20, 2024

Virtual Meeting/Conference Recording Notice

The American Institutes for Research® (AIR®) allows for the recording of audio, visuals, participants, and other information sent, verbalized, or utilized during business-related meetings. By joining a meeting, you automatically consent to such recordings. Any participant who prefers to participate via audio only should disable their video camera so only their audio will be captured. Video and/or audio recordings of any AIR session shall not be transmitted to an external third party without the permission of AIR.

In addition, AIR does not permit participation in AIR meetings through the use of AI bots such as Otter.ai or other AI platforms to record or transcribe conversations for AIR meetings, webinars, or virtual events in lieu of in-person attendance, unless requested as a reasonable accommodation. Any participants who attempt to use AI software for meeting participation will be denied admission, or their session will be terminated once it becomes apparent such software is in use.

AIR Inclusive Meeting Guidelines

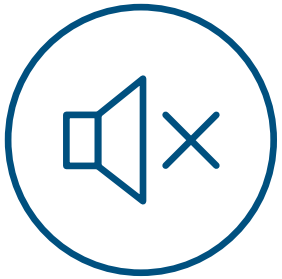
Hosting and Participating in Meetings

ENGAGE EVERYONE

BE HEARD AND SEEN

ACKNOWLEDGE SPEAKER

MAXIMIZE MICROPHONES



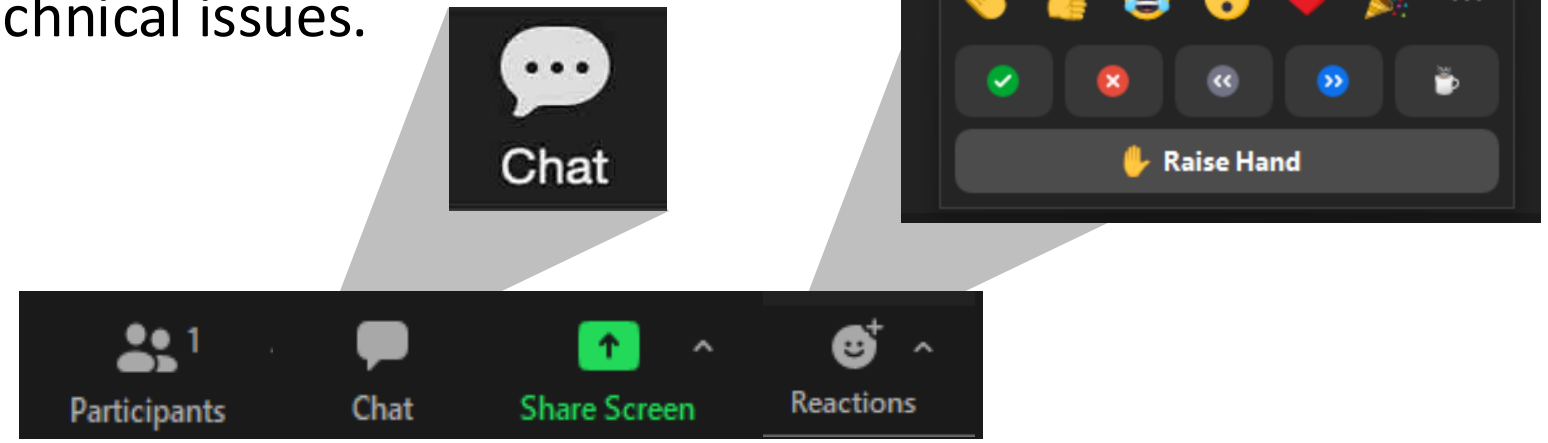
MINIMIZE NOISE

MAXIMIZE VISUAL DISPLAYS

How to Use Zoom

Click on the **Chat icon** to ask questions, share your thoughts and ideas when prompted, and let us know about any technical issues.

Click on **Reactions** to **Raise Hand** to ask a question during Q&A or to use one of the other reaction icons.



How to Use Zoom

Please participate in **Zoom polls** when prompted.

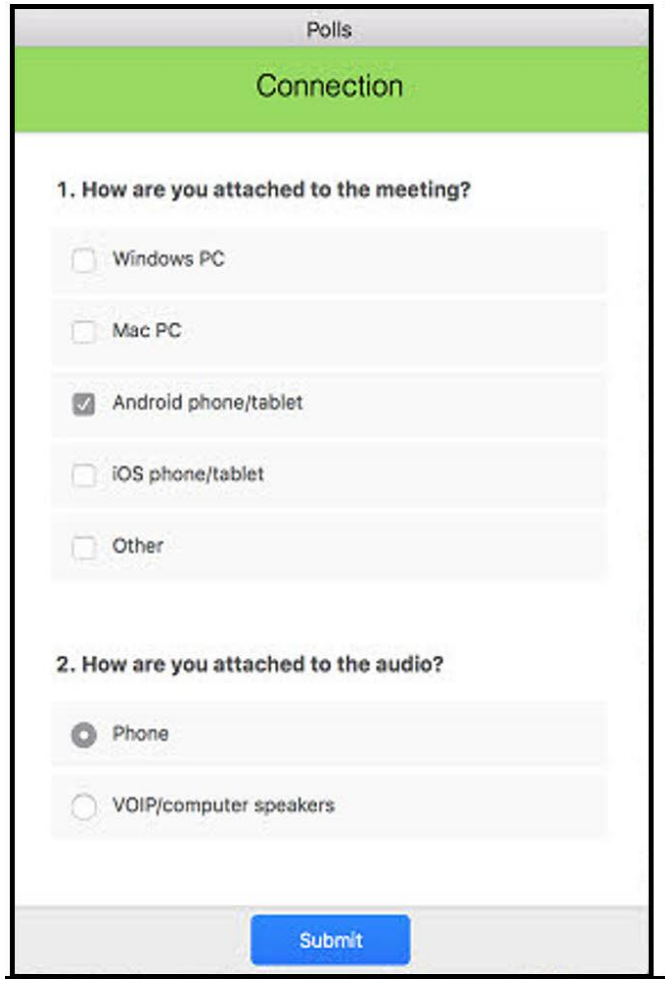
The poll will appear on your screen.

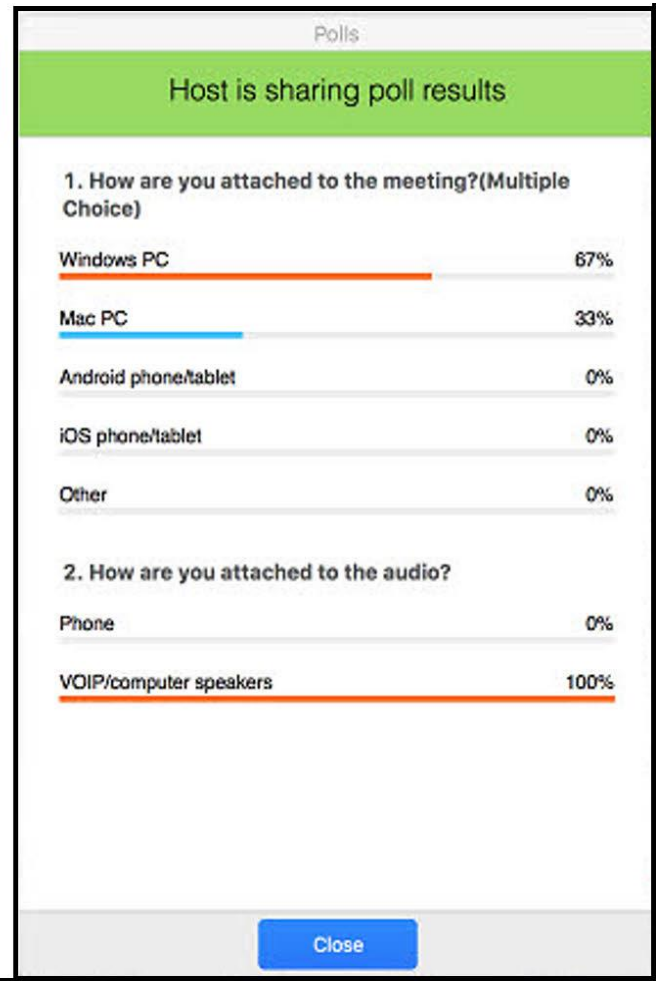
Respond to the question by selecting a response(s).

Then click **Submit**.

Results will be shared on screen.

Responses are **anonymous**.





EQR Hub

The Evidence Quality and Reach (EQR) Hub will provide current and aspiring STEM education researchers with **targeted learning opportunities regarding research methods; knowledge translation; and diversity, equity, and inclusion**. The hub will develop and implement virtual webinars and workshops for researchers in the Discovery Research PK–12 (DRK–12) community, convene communities of practice, and engage in individualized consultations with DRK–12 projects.

This work is made possible by the National Science Foundation (NSF) under Grant No. 2101162.

Community of Practice Process and Expectations

During the live sessions, learners will receive direct instruction from experts and participate in breakout activities and group discussions.

Between sessions, learners will engage in peer learning through asynchronous discussions.

Learners are expected to **attend all sessions**.

Community of Practice Sessions

~~**Session 1:** Propensity Score Analysis and Estimation~~

Session 2: Propensity Score Weighting/Matching

- Objective:
  - Understand how and when to use propensity score weighting or matching.

Session 3: Propensity Scores and Outcome Modeling

Session 4: Propensity Score Analysis in R

Activity: Poll

Time: 2 minutes

How are we feeling about the material in Session 1?

- **Excited and ready to dive into Session 2!**
- **Feeling curious but a bit overwhelmed**
- **Lost in a sea of statistics**

Any questions about Session 1?

Session 2 Agenda

1. Session 1 Recap
2. Overview of Propensity Score Matching and Weighting
3. Propensity Score Matching
4. Propensity Score Weighting
5. Let’s Code!
6. Q&A
7. Next Steps

Session 1 Recap

Propensity Score Analysis Overview

- Propensity score (PS) analysis **attempts** to replicate the reliability of the results of a randomized experiment.
- PS is the conditional **probability** that a student would have been in the treatment group (Z) based on **observed** covariates ($\boldsymbol{X}$; Rosenbaum & Rubin, 1983, 1984):

$$PS = \Pr(Z = 1|\boldsymbol{X})$$

- In randomized controlled trials (RCTs), PS = 0.5 in observational data that need to be estimated.
- Typically, PS is estimated using logistic regression or machine-learning approaches.
- Conditioning on the PS allows us to account for ***some*** of the selection bias in our estimated treatment effects.

Estimate PS Model

- In observational studies, we typically don't know the "true" propensity score.
- Logistic regression of treatment *Z* on observed predictors ***W*** (covariates ***X*** or transformations thereof—e.g., polynomials, log, interactions):
  - Include confounding variables
    - » Variables related to treatment assignment and outcome

```
m.out0 <- matchit(treat ~ age + educ + race + married +
                    nodegree + re74 + re75, data = lalonde,
                  method = NULL, distance = "glm")
```

- The estimated PS is the expected value from the logistic regression.

Propensity Score Analysis Overview

1. Examine initial imbalance: T vs. C
2. Estimate PS model: Choose covariates
3. **Condition on the PS: Matching/weighting**
4. Check balance (iterate): T vs. C
5. Estimate program impacts

Propensity Score Analysis Overview

ID	Age	Female	Took Alg2	Program	…	PS
1	16	1	1	1		0.9
2	17	1	0	1		0.8
3	16	0	1	0		0.9
4	16	1	0	0		0.2

Research question (RQ): What is the impact of participating in an afterschool math program on students' math achievement scores?

Activity: Small-Group Share-Out

Time: 5 minutes

Instructions: Raise your hand or type in chat:

What is an example in your own work of a study or evaluation where you could use propensity analysis?

- **What are you evaluating?**
- **Why can't you use randomization?**

AIR®
Advancing Evidence.
Improving Lives.

Overview of Propensity Score Matching and Weighting

Overview of Matching and Weighting

PS Matching

- Match treated student to comparison student based on their propensity scores.
- In the matched sample, the distribution of observed characteristics is similar between treated and matched students.
- **Helpful:** Large pool of comparison students

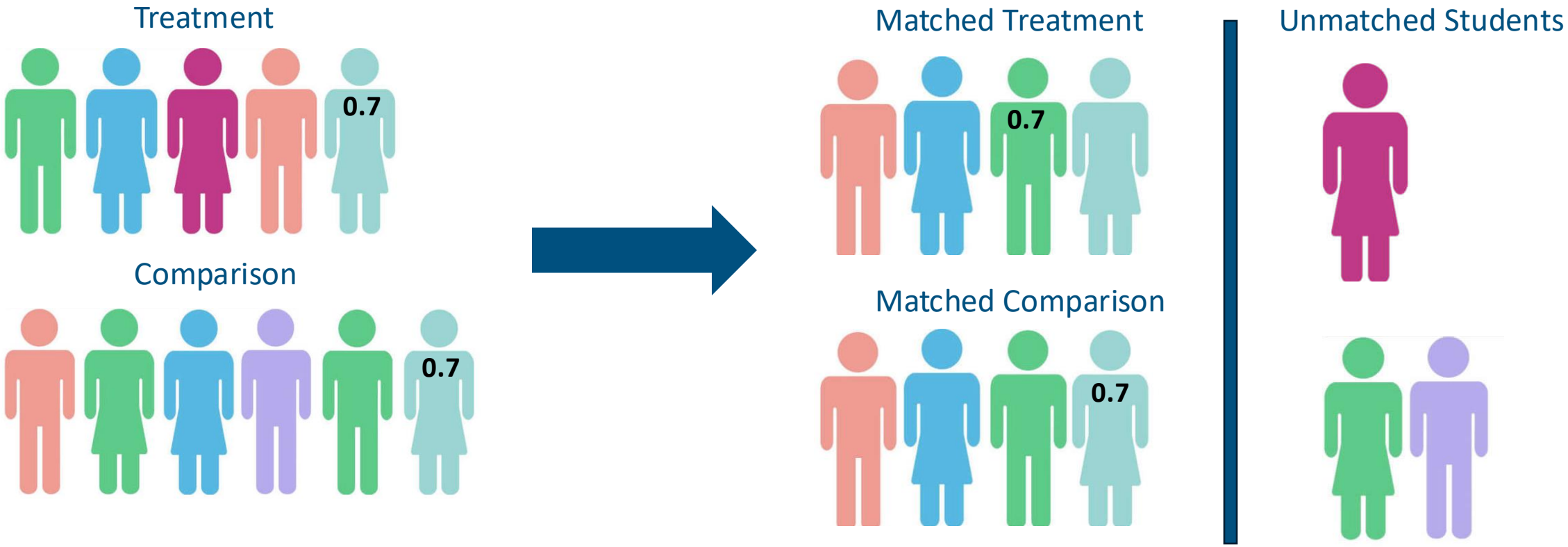
PS Weighting

- Reweight treated and and comparison students based on their propensity scores (i.e., survey weights).
- The weighted sample distribution of observed characteristics is similar between treated and comparison students.
- **Helpful:** Small pool of comparison students

AIR®
Advancing Evidence.
Improving Lives.

Propensity Score Matching

Propensity Score Matching



RQ: What is the impact of participating in an afterschool math program on students' math achievement scores?

Matching Methods

Nearest Neighbor Matching (i.e., greedy matching)

- Pairs each treated student with one or more comparison student based on proximity of PS
  - Caliper: Starting point to 0.25 *SD* of the PS (max allowable difference between matched PS)
  - Can be done with or without replacement

Optimal Matching (i.e., nongreedy matching)

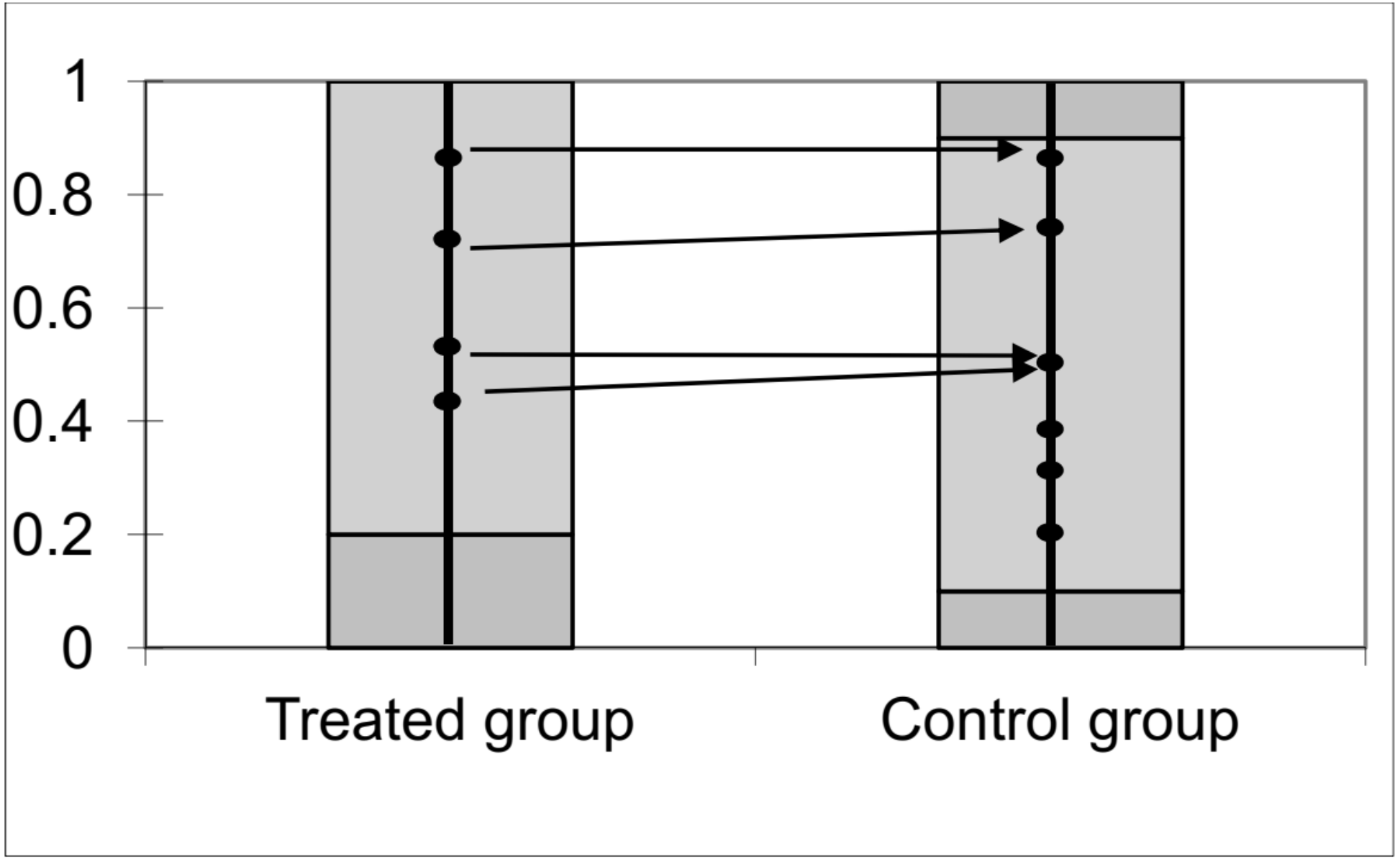
- Minimizes the overall difference in PS across all treated-comparison pairs

Exact Matching

- Matches based on exact values of predefined covariates—examples: state, school, age, gender

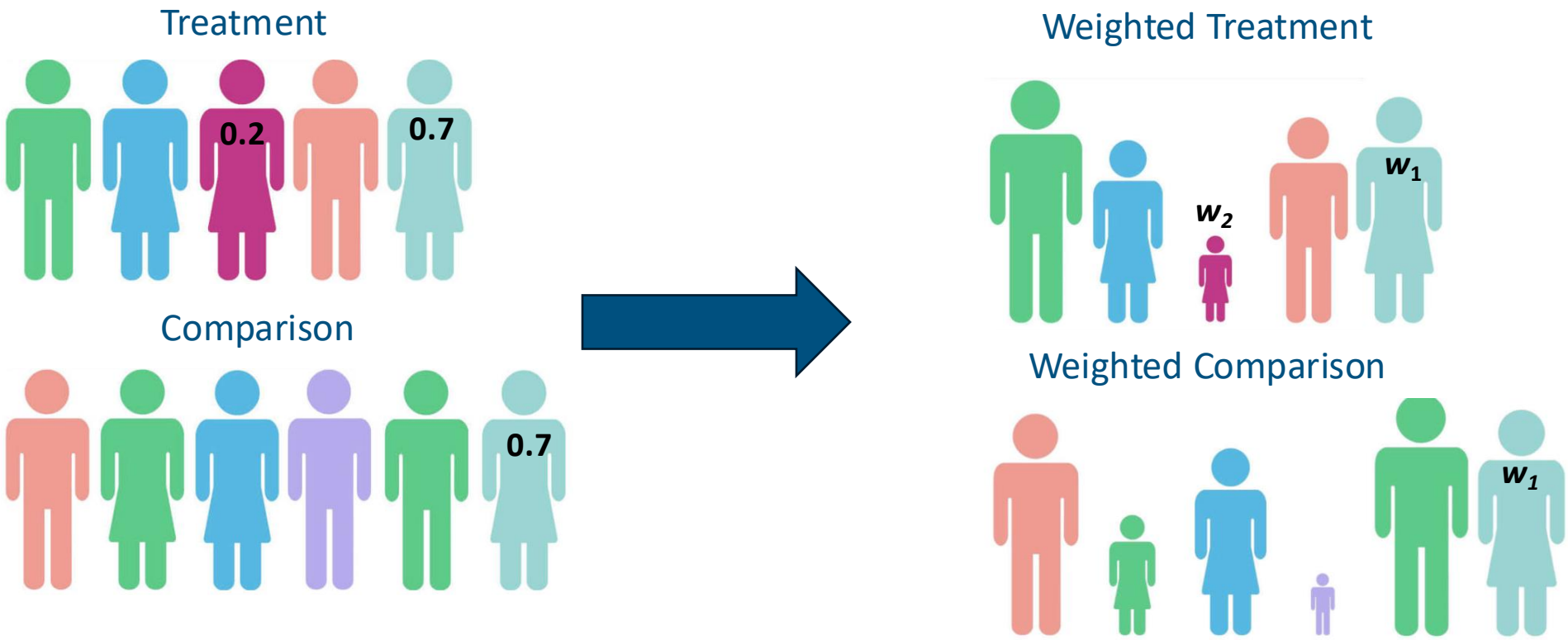
Advantages: Can result in tightly controlled matched pairs

Limitations: Can discard unmatched students, may lead to statistical power issues


1
0.8
0.6
0.4
0.2
0
Treated group
Control group

Propensity Score Weighting

Propensity Score Weighting



RQ: What is the impact of participating in an afterschool math program on students' math achievement scores?

Weighting Methods

Inverse Probability of Treatment Weights (IPTW)

Weight Calculation

- For average treatment effect (ATE):
  - Treated: 1/PS
  - Comparison: 1/(1-PS)
- For average treatment effect on the treated (ATT):
  - Treated: 1
  - Comparison: PS/(1-PS)

Advantages: It retains all observations in the data set.

Limitations: Extreme weights can introduce high variance, potentially making estimates less stable.

Activity: Q&A Chat

Time: 3 minutes

Instructions: Raise your hand or type in chat any questions you might have.

Let’s Code!

Let's Code!

- **Data set:** Experiment from a Midwestern university (Shadish et al., 2008)
- **Sample:** College students
- **Intervention:** Math or vocabulary training (self-selected)
- **Baseline Covariates:** Measured before training choice
  - **Total:** 23 covariates
    - » Pretest of outcome
    - » Prior academic achievement
    - » Topic preference
    - » Demographics
- **Outcome:** Posttest

Q&A

Next Steps

Homework Assignment

Time: Complete prior to Session 3

Instructions:

1. Review Session 2 slides.
2. Think through: “Any concepts where you are struggling?”

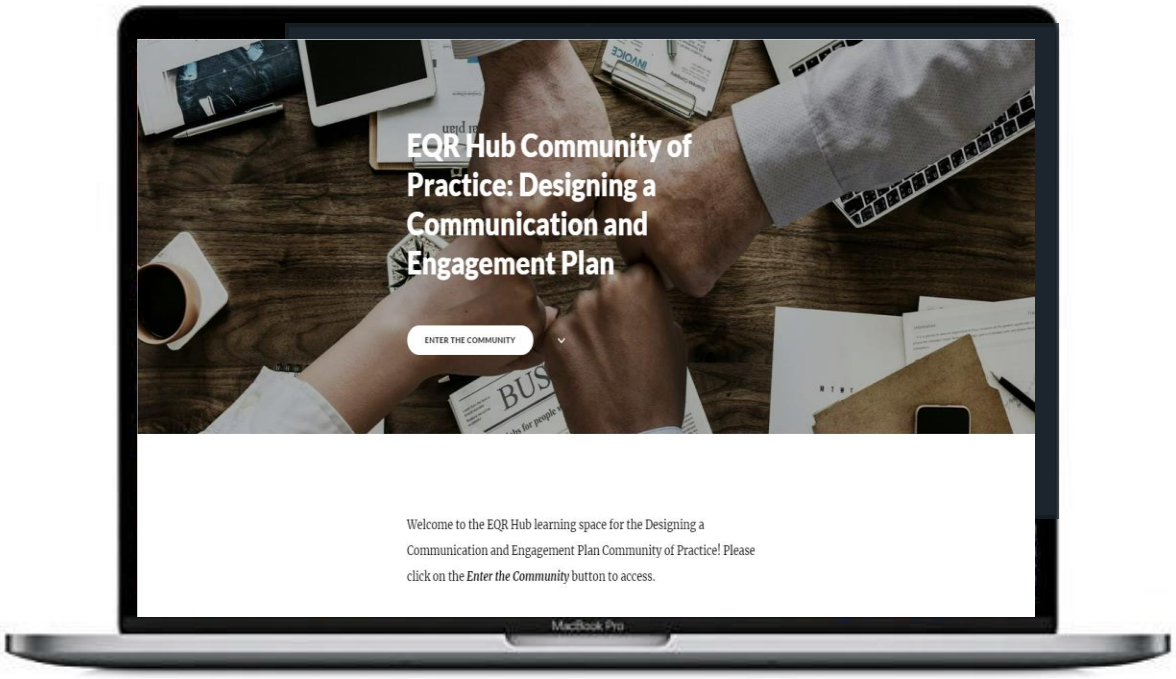
References

Rosenbaum & Rubin, 1983

Rosenbaum & Rubin, 1984

Shadish et al., 2008

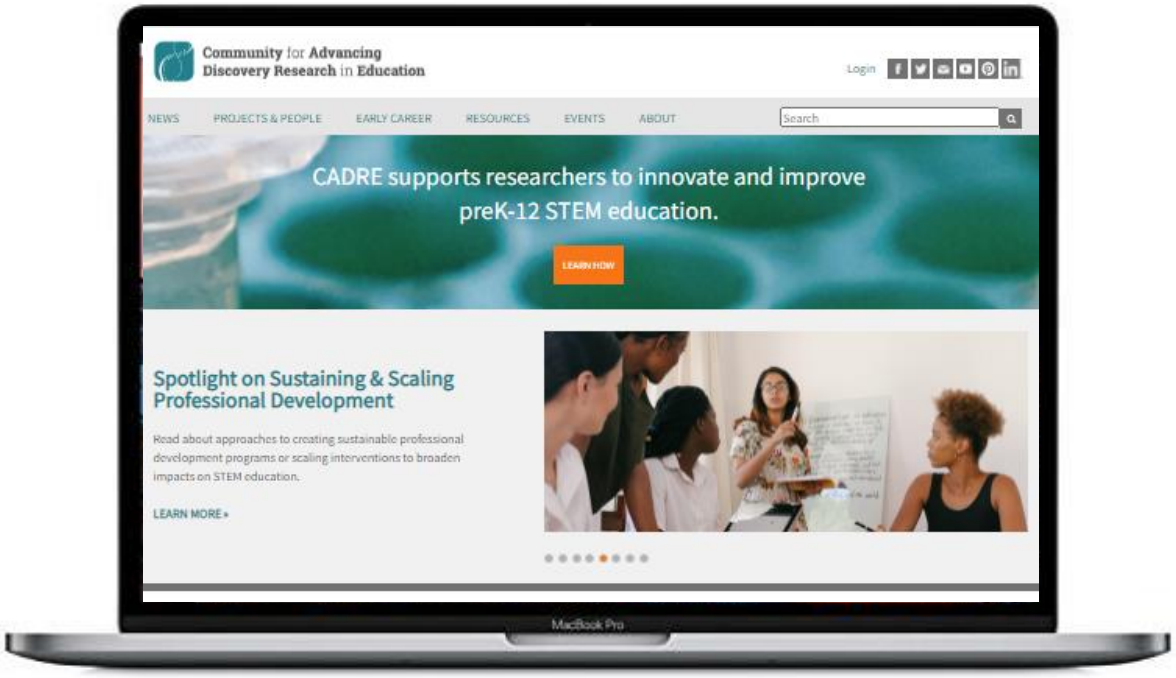
Visit the community of practice learning space.



Develop a conceptual framework centered on diversity, equity, and inclusion.

CoP Learning Space

Stay connected.



Visit the CADRE resources

https://cadrek12.org/**resources**

and EQR Hub page

http://cadrek12.org/**eqr-hub**

AIR®
Advancing Evidence.
Improving Lives.

Thanks for attending!

25615_08/24