AIR®
Advancing Evidence.
Improving Lives.

Evidence Quality and Reach Hub: Propensity Score Analysis and Estimation

Session 1

August 13, 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

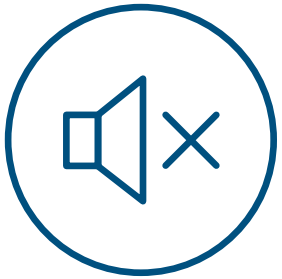
Hosting and Participating in Meetings

ENGAGE EVERYONE

BE HEARD AND SEEN

ACKNOWLEDGE SPEAKER

MAXIMIZE MICROPHONES



MINIMIZE NOISE

MAXIMIZE VISUAL DISPLAYS

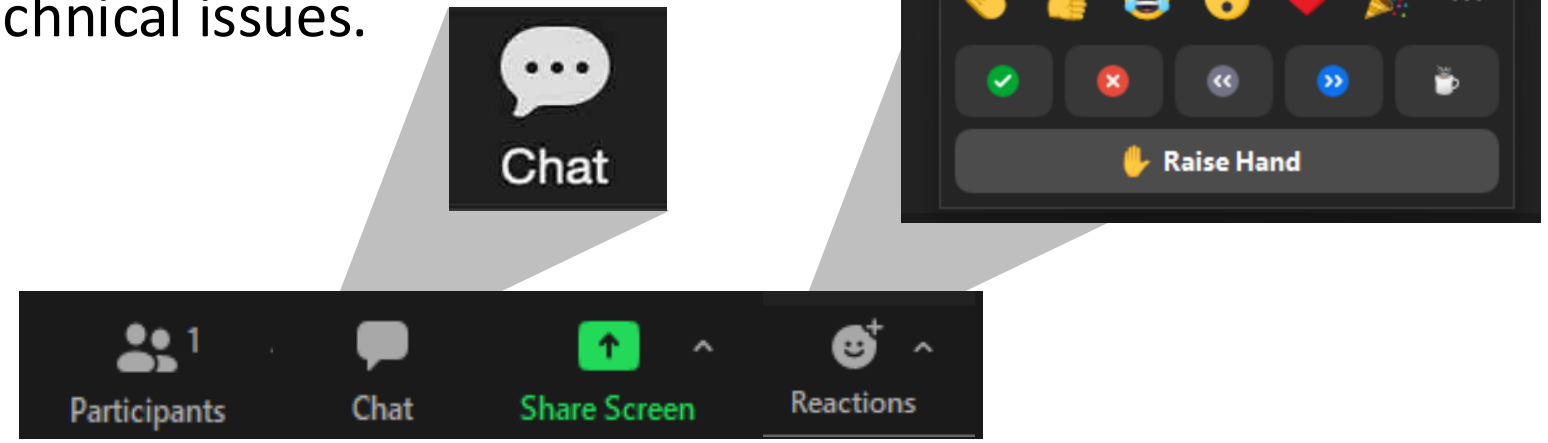
These guidelines are intended to improve the meeting experience for virtual participants, as well as people with hearing loss, visual impairment, and those for whom English is an additional language.

Developed by the Access AIR and AIR CREW Employee Resource Groups with support from the AIR Diversity and Inclusion Office.

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.



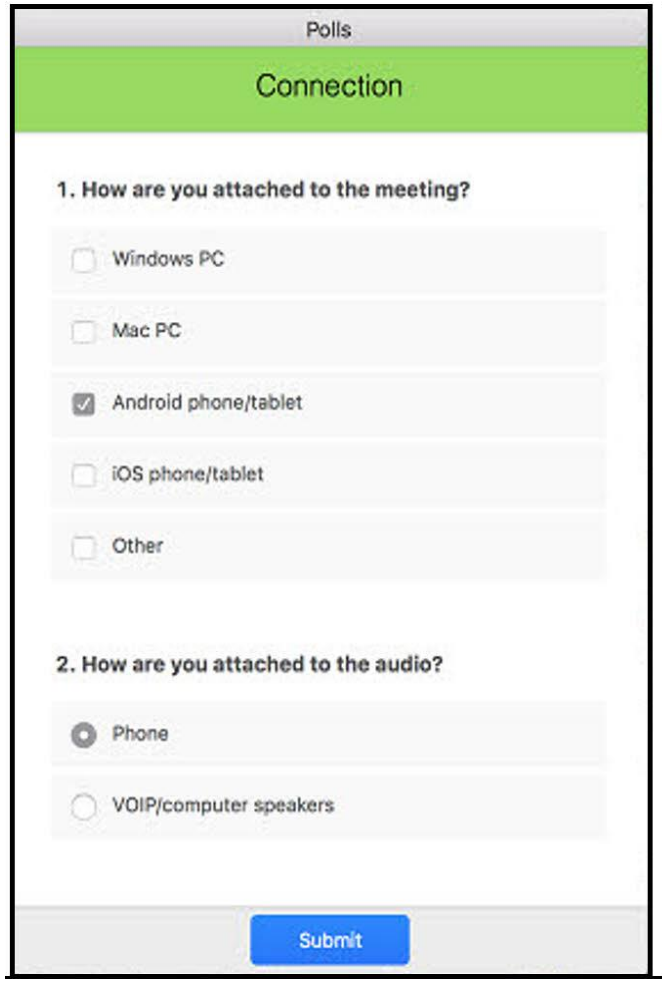
How to Use Zoom

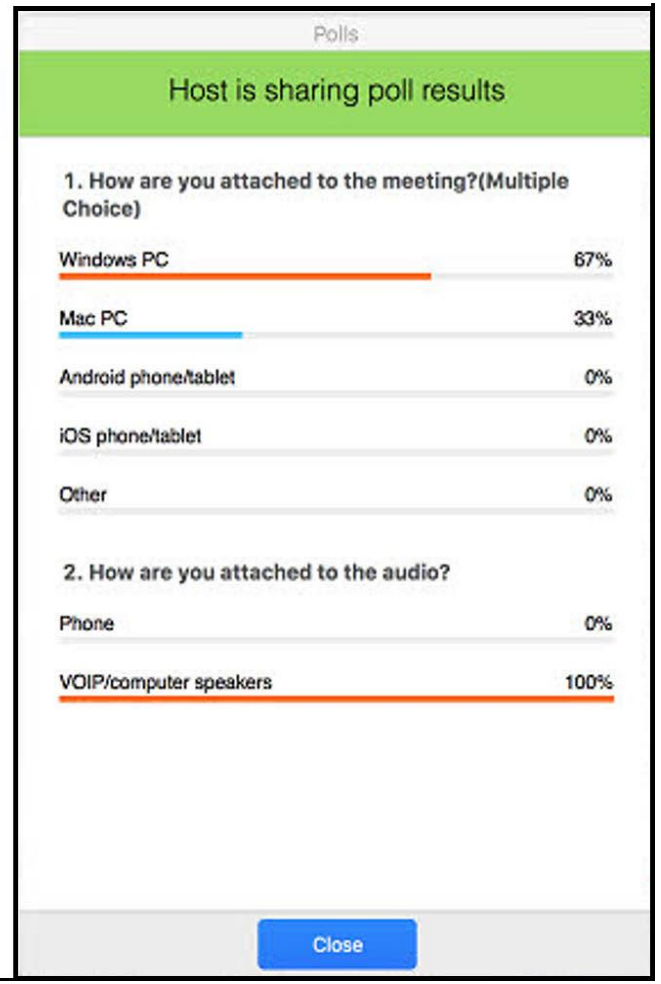
Please participate in **Zoom polls** when prompted.

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**.





Introductions

Time: 3 minutes

Instructions:

In the chat, share how you are feeling today!

Image from https://www.sofiafeldmann.com/social-emotional-check-ins

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**.

Meet the Presenter

Alberto Guzman-Alvarez, PhD

AIR

Research Data Scientist

- Located in Davis, California
- Specializes in quasi-experimental research design and quantitative methodology
- Focus on evaluating the effectiveness of educational policies and interventions
- PhD, Learning Sciences and Policy—Quantitative Methods, from the University of Pittsburgh

Community of Practice Sessions

Session 1: Propensity Score Analysis and Estimation

Objectives:

- Understand the concept and purpose of propensity score analysis.
- Understand how to estimate propensity scores.

Session 2: Propensity Score Weighting/Matching

Session 3: Propensity Scores and Outcome Modeling

Session 4: Office Hour

Poll

Time: 1 minutes

Have you heard of or used propensity score analysis before?

- Yes
- No

Session 1 Agenda

1. Randomized Controlled Trials vs. Observational Studies
2. Propensity Score Analysis Overview
3. Let’s Code!
4. Q&A
5. Next Steps

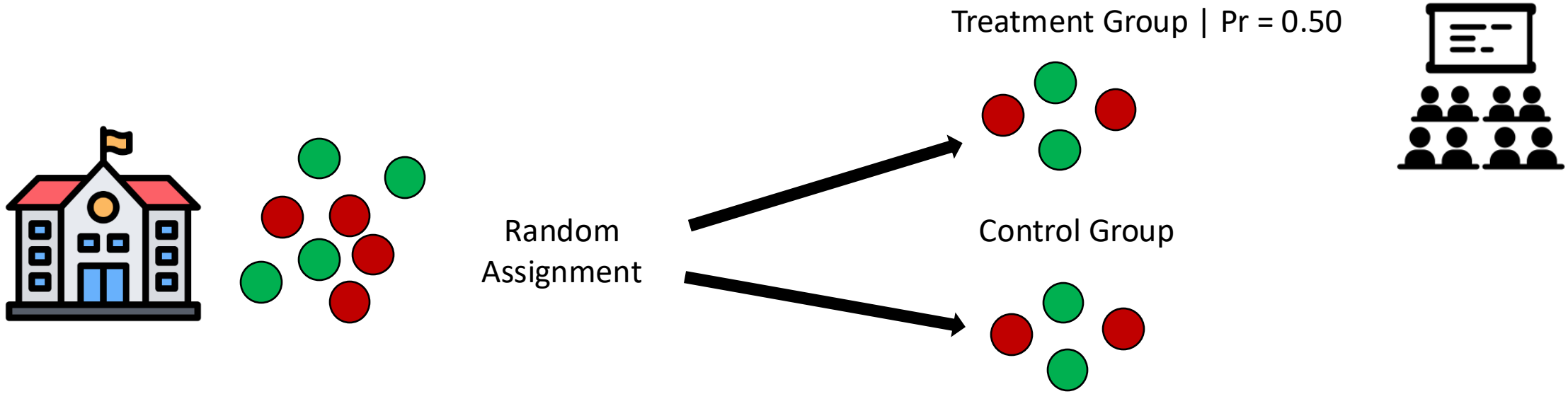
AIR®
Advancing Evidence.
Improving Lives.

Randomized Controlled Trials vs. Observational Studies

Randomized Controlled Trials

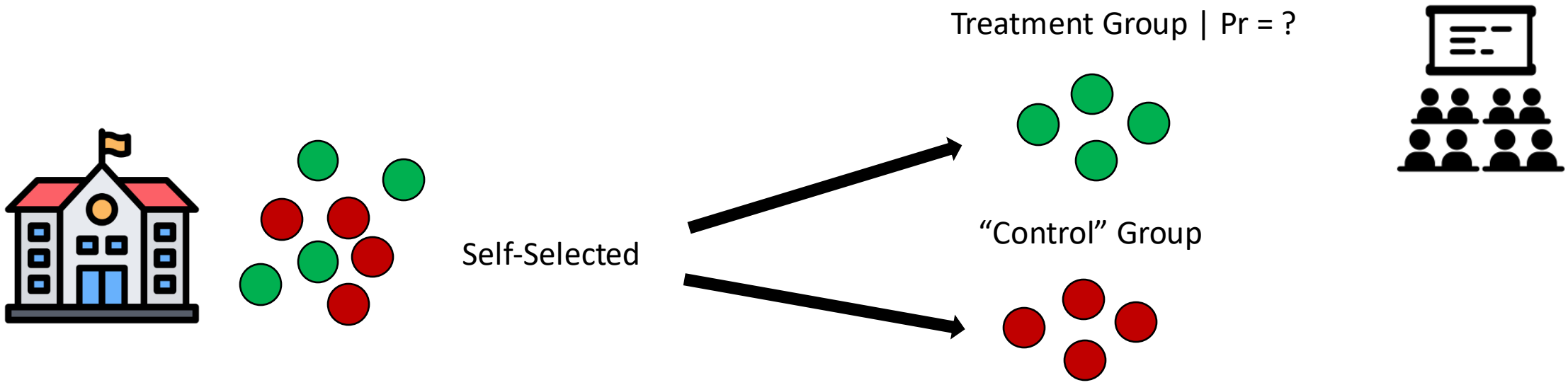
- Randomized controlled trials (RCTs) are considered the gold standard in causal research design (Rosenbaum & Rubin, 1983).
  - Random assignment *can* ensure that our estimated treatment effect is **unbiased**
    - » Randomization on average makes treatment and control groups equivalent on **observed** and **unobserved** characteristics.
- **In social science research, barriers—ethical, cost, or practical—can prevent a researcher from conducting an RCT** (Pan & Bai, 2018).

Randomization



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

Randomized vs. Observational Studies



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

Chat Activity

Time: 2 minutes

In the chat, please share which type of study design do you most frequently propose or use in your own research?

Propensity Score Analysis Overview

Propensity Score Analysis Overview

- If we rely on observational data where students self-select or are placed into a "treatment":
  - Naïve comparison would lead to **biased** treatment effect.
- Propensity score (PS) analysis **attempts** to replicate a randomized experiment:
  - Conditioning on **observed** characteristics, *not* **unobserved** characteristics.

Propensity Score Analysis Overview

- 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):

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

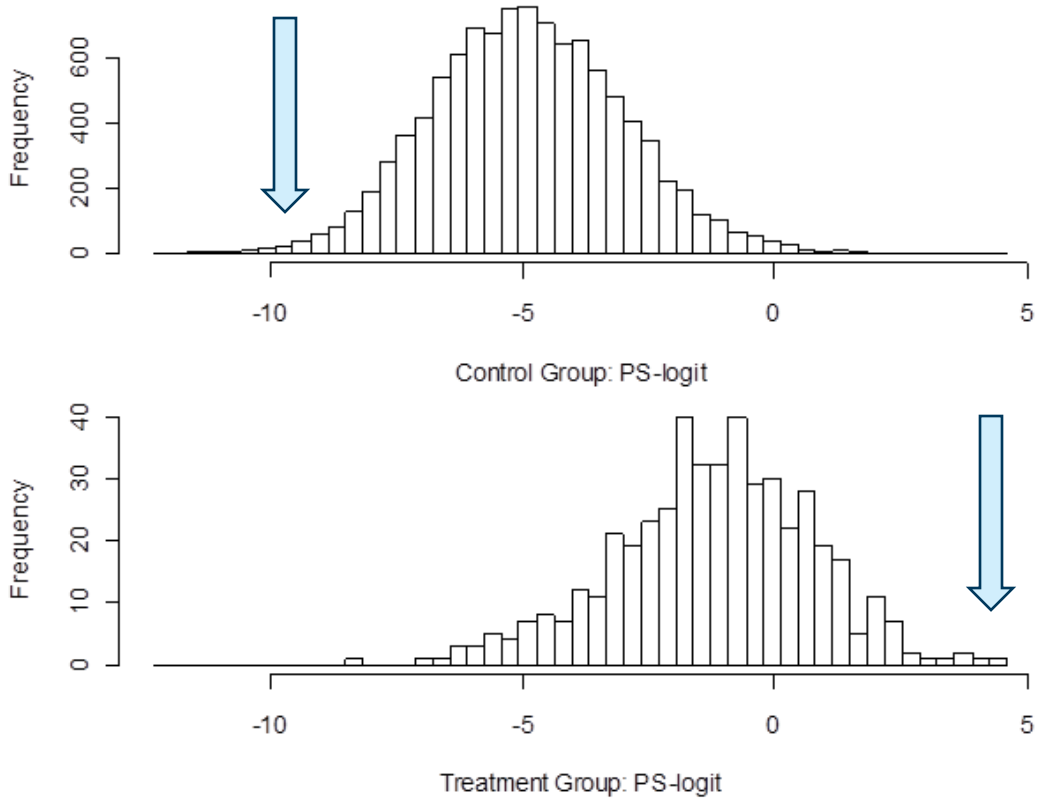
- **PS = Probability**
- **RCT PS = 0.5**
    - Students with similar PSs are similar on their underlying **observed** characteristics.

Constructing Propensity Scores

ID	Age	Female	Took Alg1	Program	…	PS
1	16	1	1	1		0.9
2	17	1	0	1		0.8
3	17	1	1	0		0.9
4	16	1	0	0		0.2

Propensity Score Assumptions

- No unobserved confounders
  - Outcomes are independent of treatment assignment
    - » **Cannot be tested!**
- Common support assumption



Propensity Score Analysis Steps

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

1. Examine initial imbalance.

```
## Summary of Balance for All Data:
##            Means Treated Means Control Std. Mean Diff.
## distance          0.5774        0.1822          1.7941
## age              25.8162       28.0303         -0.3094
## educ             10.3459       10.2354          0.0550
## raceblack         0.8432        0.2028          1.7615
## racehispan        0.0595        0.1422         -0.3498
## racewhite         0.0973        0.6550         -1.8819
## married           0.1892        0.5128         -0.8263
## nodegree          0.7081        0.5967          0.2450
## re74           2095.5737     5619.2365         -0.7211
## re75           1532.0553     2466.4844         -0.2903
```

WWC standard states that the SMDs need to be < |.25| to consider a covariate balanced.

Note. SMD = standard mean difference; WWC = What Works Clearinghouse

2. Estimate PS model.

- In observational studies, we typically don't know the "true" propensity score:
  - It needs to be estimated by **observed** data.
- Logistic regression of treatment *Z* on observed predictors (covariates ***X*** or transformations thereof—e.g., polynomials, log, interactions):
  - Includes confounding variables
    - » Variables related to treatment assignment and outcome
- The estimated PS is the expected value from the logistic regression:

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

3. Condition on the PS (Session 2). | 4. Check balance.

```
## Summary of Balance for Matched Data:
##            Means Treated Means Control Std. Mean Diff. 
## distance          0.5773        0.5766          0.0036
## age              25.8162       25.6335          0.0255
## educ             10.3459       10.4590         -0.0562
## raceblack         0.8432        0.8389          0.0119
## racehispan        0.0595        0.0469          0.0532
## racewhite         0.0973        0.1142         -0.0571
## married           0.1892        0.1555          0.0860
## nodegree          0.7081        0.6711          0.0814
## re74           2095.5737     2108.4175         -0.0026
## re75           1532.0553     1557.1654         -0.0078
```

WWC standard states that the SMDs need to be < .25 to consider a covariate balanced.

5. Estimate program impacts (Session 3).

```
library("marginaleffects")

fit <- lm(re78 ~ treat * (age + educ + race + married + nodegree +
             re74 + re75), data = m.data, weights = weights)

avg_comparisons(fit,
                variables = "treat",
                vcov = ~subclass,
                newdata = subset(m.data, treat == 1),
                wts = "weights")
```

```
##
##   Term Contrast Estimate Std. Error   z Pr(>|z|) 2.5 % 97.5 %
##  treat    1 - 0     1912        765 2.5   0.0124   413   3411
##
## Columns: term, contrast, estimate, std.error, statistic, p.value, conf.low, conf.high
```

Poll

Time: 2 minutes

How are we feeling about the material?

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

Let’s Code!

R and STATA Packages

- R: https://bookdown.org/introrbook/intro2r/installing-r-and-rstudio.html
- MatchIt: https://kosukeimai.github.io/MatchIt/
- WeightIt: https://ngreifer.github.io/WeightIt/
- STATA: psmatch2: http://repec.org/bocode/p/psmatch2.html
- Python: Causalinference: https://causalinferenceinpython.org/

Let's Code!

- **Data set:** Evaluation 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

Live Coding Demo in R

Q&A

Group Discussion

Time: 5 minutes

Please come off mute and share what challenges do you foresee when implementing propensity score analysis in your own work?

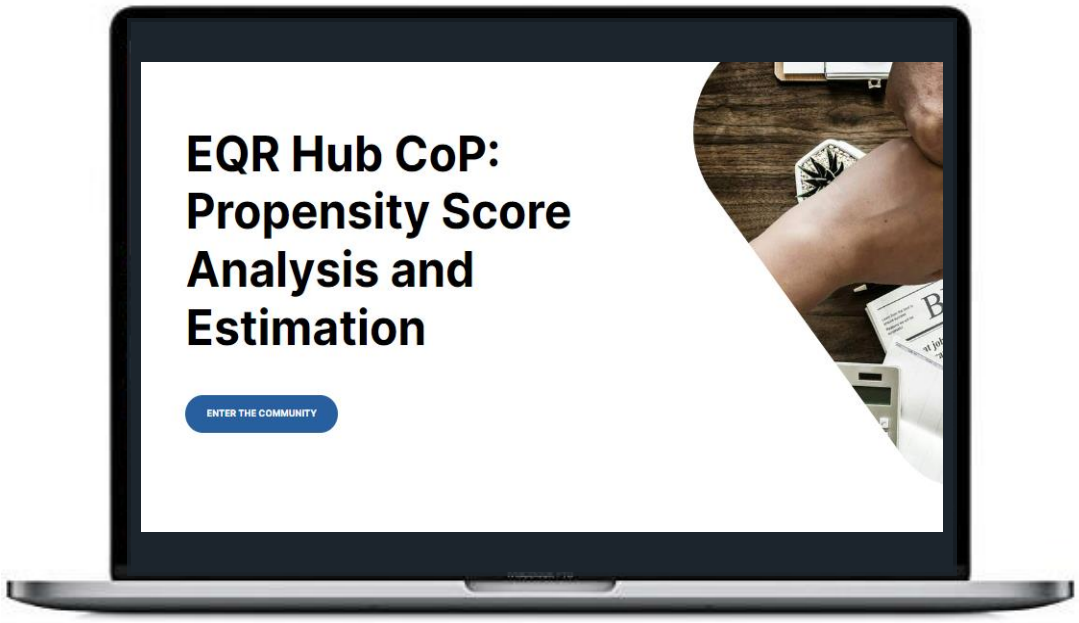
Next Steps

Homework Assignment

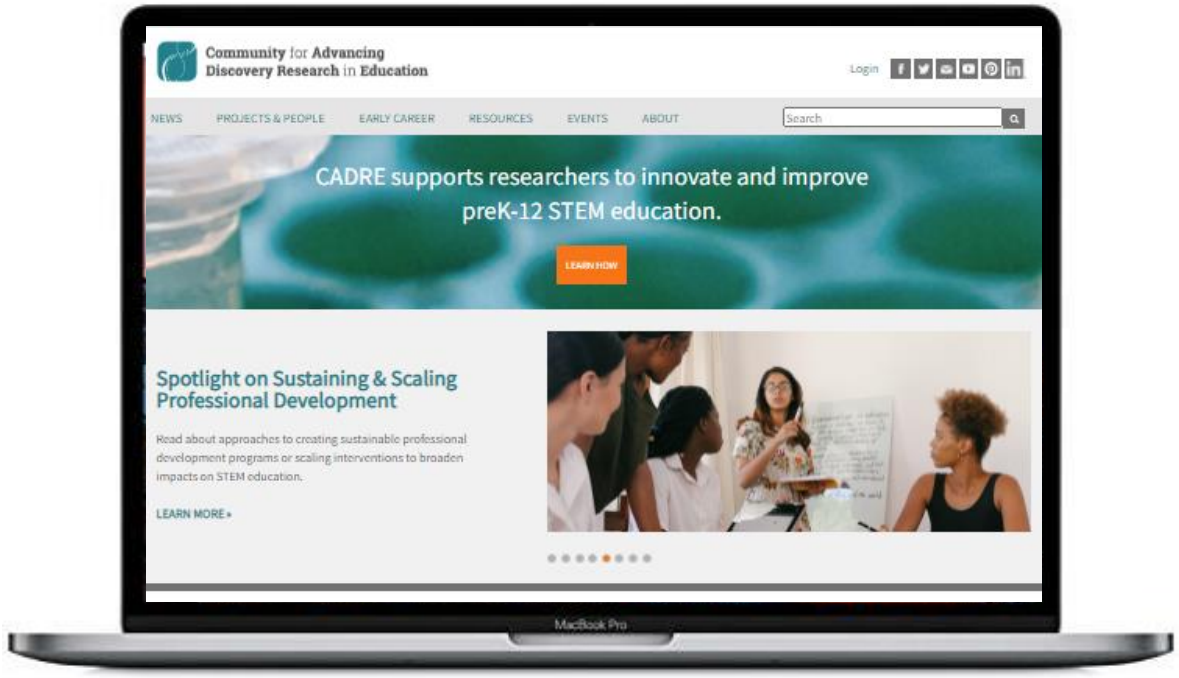
Time: Complete prior to Session 2

Instructions:

1. Review Session 1 slides.
2. Think through: “What are some evaluations/studies in my own work where I could use this method?” and share your thoughts on the RISE site
   - We will share out next week!



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

References

Pan & Bai, 2018

Rosenbaum & Rubin, 1983

Shadish et al., 2008