

# MODULE 8

## EFFECT SIZE

# OUTLINE

- 1 Effect size overview
- 2 Standardized mean difference
- 3 Within-subjects application
- 4 Between-subjects application
- 5 Study questions

# OUTLINE

- 1 Effect size overview
- 2 Standardized mean difference
- 3 Within-subjects application
- 4 Between-subjects application
- 5 Study questions

# EFFECT SIZE

---

- Effect size refers to a family of measures that quantify the magnitude of an effect, independent of sample size
- The goal is to express how big an effect is, not just whether it likely exists at the population level
- Common families: standardized mean difference, correlation, variance explained (R-square)

# WHY NOT JUST P-VALUES?

---

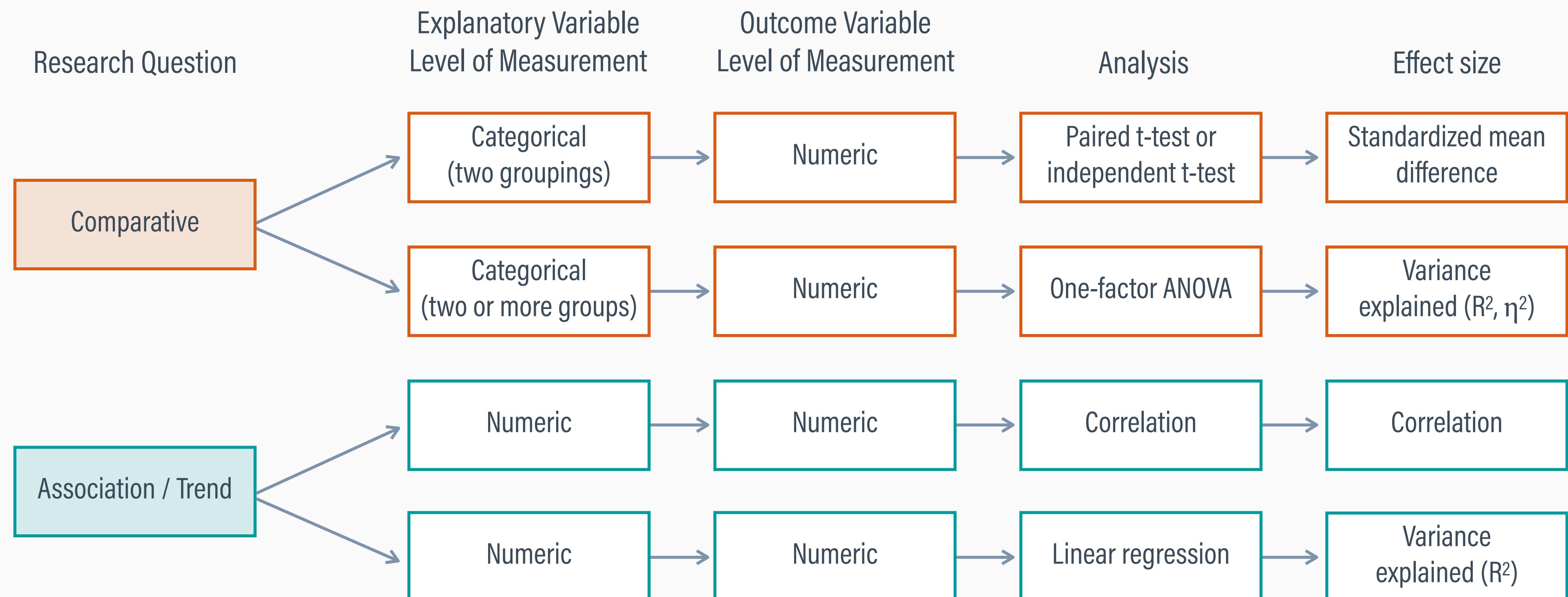
- p-values tell us how unusual the data are if the null is true
- With a large enough sample, even trivially small effects can be “statistically significant”; conversely, large effects can be “non-significant” with small samples
- Effect sizes convey the practical significance of an effect, independent of the N

## APA RECOMMENDATIONS

---

- APA began emphasizing effect size more than 20 years ago
- The current 7th edition manual specifies to always report effect sizes alongside significance tests, and include confidence intervals for effect sizes when possible
- This recommendation enhances transparency, interpretation of results, and supports replication

# STATISTICAL ORG CHART



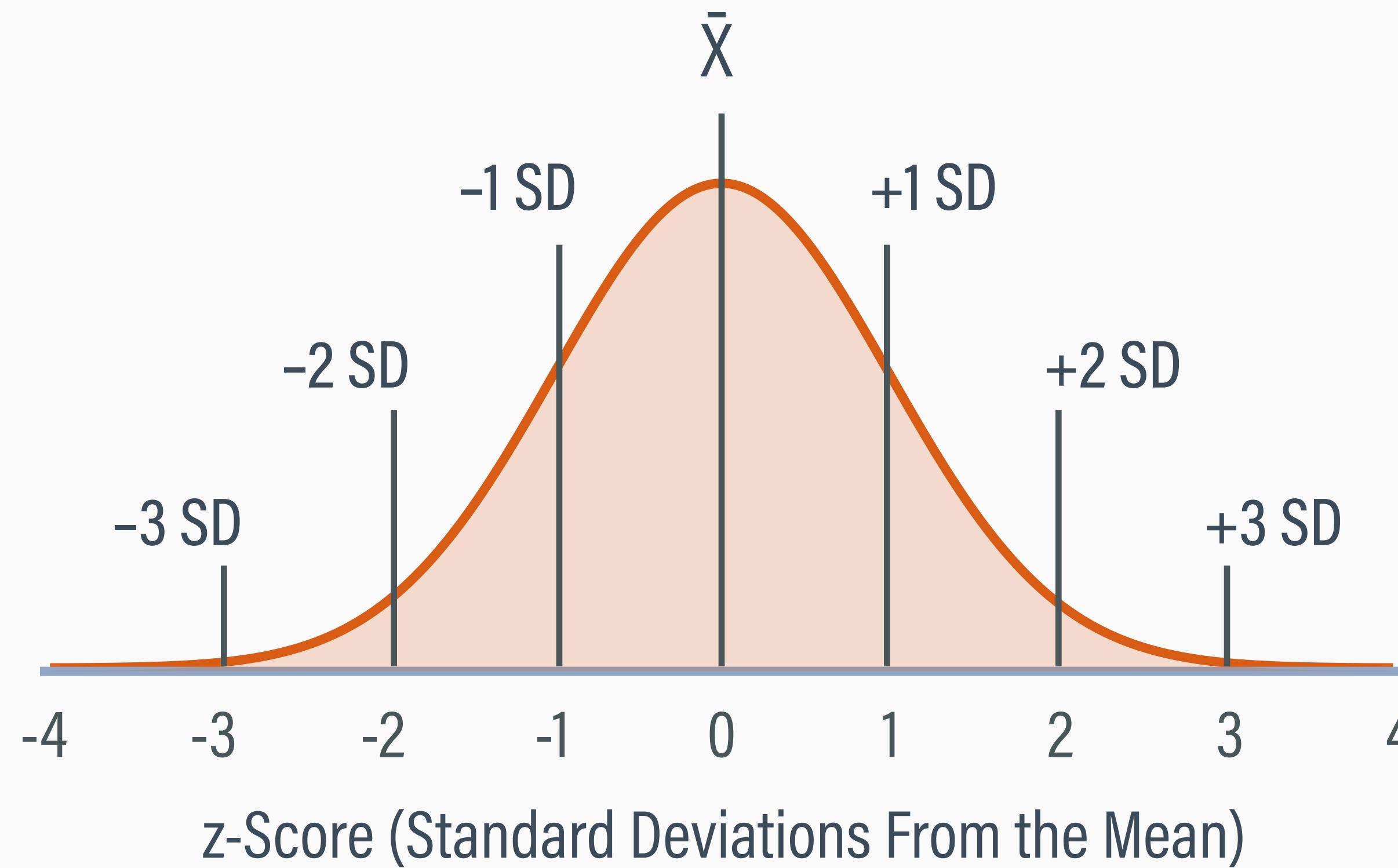
# OUTLINE

- 1 Effect size overview
- 2 Standardized mean difference
- 3 Within-subjects application
- 4 Between-subjects application
- 5 Study questions

# z-SCORES

---

- The **z-score** scale is a common standardized metric that expresses scores as standard deviation units from the mean



# z-SCORE FORMULA

---

- The numerator captures the raw score's distance (deviation) from the sample mean  $\bar{X}$

$$z = \frac{X - \bar{X}}{s} = \frac{\text{distance from mean}}{\text{standard deviation}}$$

- Dividing by the standard deviation converts the distance to standard deviation units

# COMPARATIVE RESEARCH QUESTIONS

---

- Comparative research questions ask whether two groups differ from one another
- Psychologists typically use the mean for comparative research questions involving a numeric dependent variable
- Two common scenarios: (a) compare two or more groups, or (b) assess change between pretest and posttest

# STANDARDIZED MEAN DIFFERENCE

---

- The standardized mean difference expresses the raw mean difference in standard deviation (z-score) units

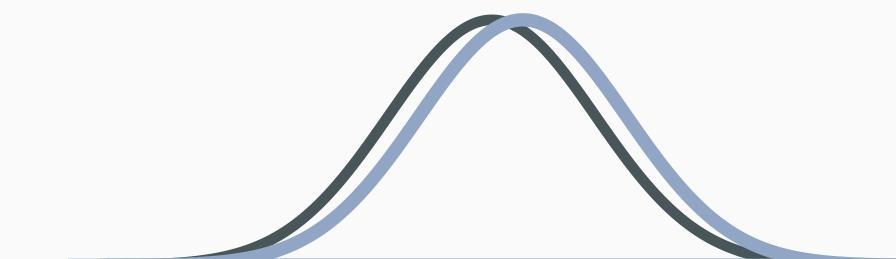
$$d = \frac{\bar{X}_1 - \bar{X}_2}{s} = \frac{\text{distance between means}}{\text{standard deviation}}$$

- Often referred to as **Cohen's d effect size**, although d is really a family of related standardized mean difference statistics

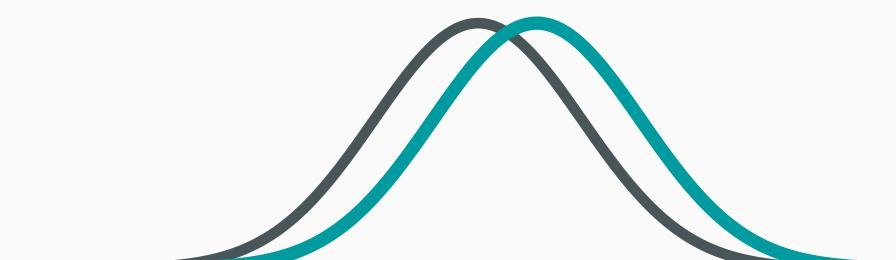
# COHEN'S GUIDELINES

---

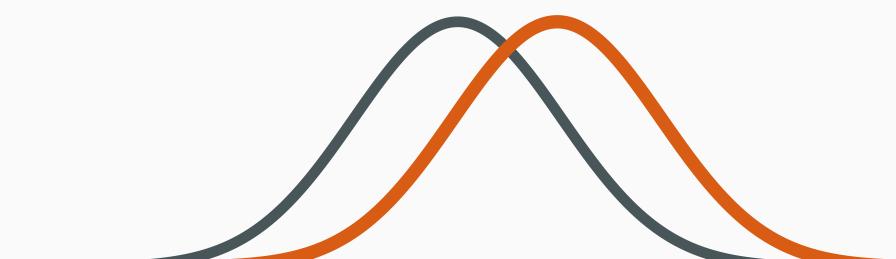
Negligible = less than  $| .20 |$



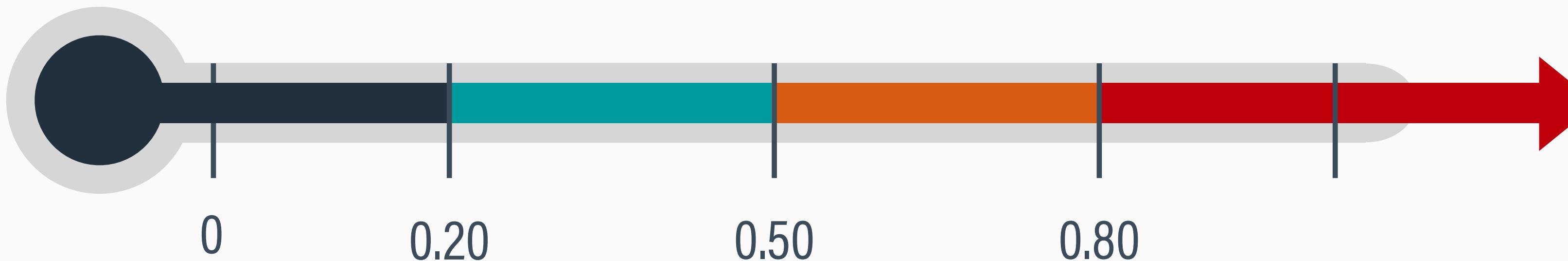
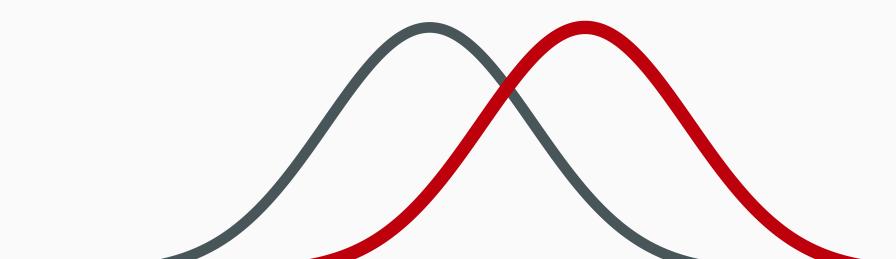
Small =  $| .20 \text{ to } .50 |$



Moderate =  $| .50 \text{ to } .80 |$



Large = greater than  $| .80 |$



# OUTLINE

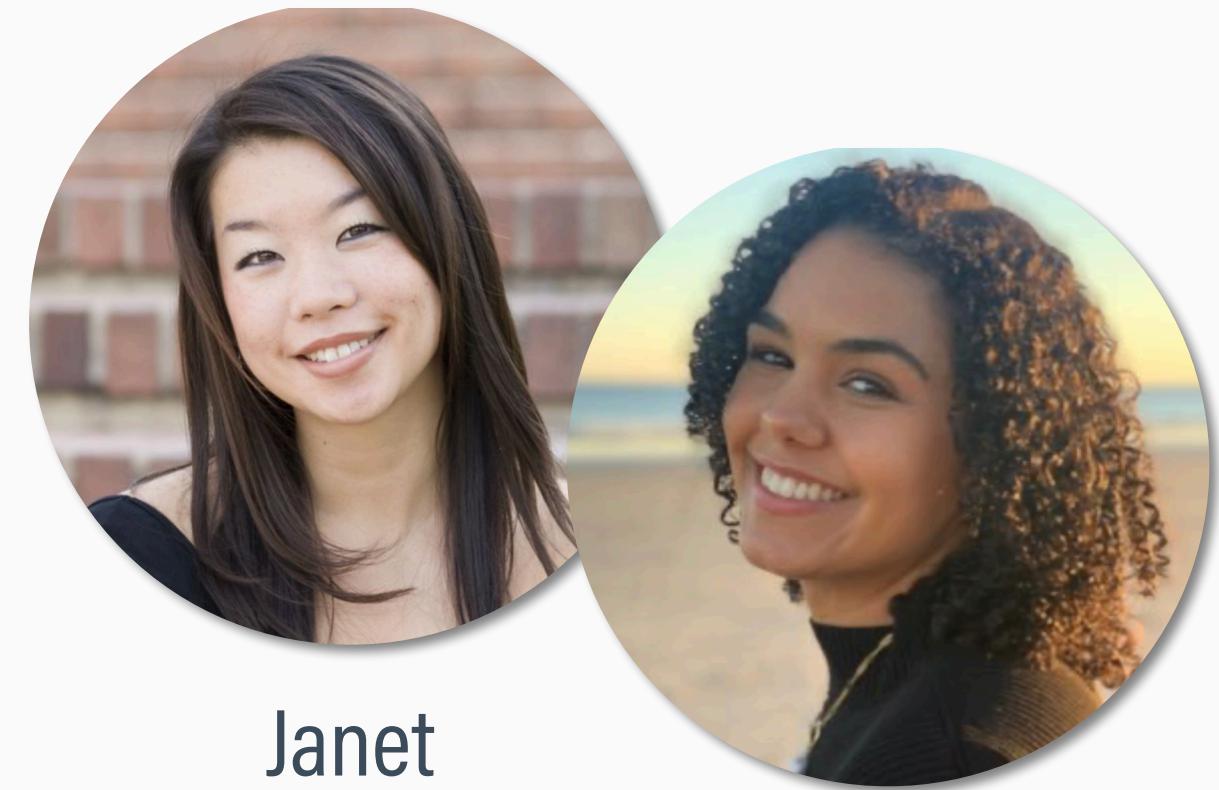
- 1 Effect size overview
- 2 Standardized mean difference
- 3 Within-subjects application
- 4 Between-subjects application
- 5 Study questions

# SKIN COLOR SATISFACTION AND BINGE EATING

---

Although it has been demonstrated that (a) body dissatisfaction and internalization of societal appearance standards contribute to disordered eating and (b) that internalization of societal appearance standards leads to decreased skin color satisfaction among Black women, it has not been established whether skin color dissatisfaction contributes to disordered eating among Black women or girls. The objective of the present study is to determine the influence of skin color satisfaction as a potential predictor for binge eating, and its effect through body image in Black girls during the vulnerable developmental period of adolescence.

Parker, J.E., Enders, C.K., Mujahid, M.S., Laraia, B.A., Epel, E.S., Tomiyama, A.J. (2022). Prospective relationships between skin color satisfaction, body satisfaction, and binge eating in Black girls. *Body Image*, 41, 342-353.

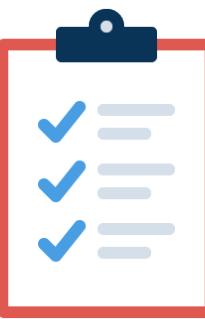


Janet  
Tomiyama

Jordan  
Parker

# KEY VARIABLES

---



## Body Satisfaction

Body satisfaction is the facet of self-concept associated with weight, and includes the attitudes, evaluations, and feelings an individual holds about his or her own body.



## Age

The grouping variable was age. Participants were followed longitudinally, with dependent variable measured at ages 10 and 18.

# RESEARCH QUESTION

---

- Question: Do Black girls experience a change in body satisfaction during adolescence from age 10 to 18?
- The explanatory (independent) variable, age, consists of two occasions: ages 10 and 18
- The outcome (dependent) variable, body satisfaction, is a numeric scale derived from several questionnaire items

# CHANGE (DIFFERENCE) SCORES

---

- The paired-samples t-test converts two measurements into a single column of change (difference) scores

$$\text{BodySatCha} = \text{BodySat18} - \text{BodySat10}$$

- A positive change score indicates that scores increased, a negative scores conveys a decrease

ID	BodySat10	BodySat18	BodySatCha
1	32	32	0
2	29	24	-5
3	23	26	3
...	...	...	...
881	28	26	-2
882	25	26	1

# STANDARDIZED MEAN DIFFERENCE

---

- One of several SMD formulas for within-subjects designs

$$d = \frac{\bar{X}_{18} - \bar{X}_{10}}{s_{\text{diff}}} = \frac{\text{sample mean difference}}{\text{standard deviation of difference scores}}$$

- This version's denominator is the standard deviation of the difference score variable, computed by subtracting each person's age 18 and age 10 scores

# EFFECT SIZE EXAMPLE

- The effect size expresses the mean difference ( $\bar{X}_{\text{diff}} = -3.05$ ) on a standardized metric
- The mean difference of 3.05 on the body satisfaction scale equates to 0.47 standard deviation units
- The two means differ by 0.47 z-score units

	N	$\bar{X}$	SD	SE
Age 10	882	28.49	5.14	0.17
Age 18	882	25.44	6.06	0.20
Difference	882	-3.05	6.44	0.22

$$d = \frac{\bar{X}_{18} - \bar{X}_{10}}{S_{\text{diff}}}$$

$$= \frac{25.44 - 28.49}{6.44} = |-0.47| = 0.47$$

# R OUTPUT

---

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Participant	1	882	441.50	254.76	441.50	441.50	326.91	1.00	882.00	881.00	0.00	-1.20	8.58
ParentEduc*	2	882	2.27	0.80	2.00	2.34	1.48	1.00	3.00	2.00	-0.52	-1.26	0.03
ParentIncome*	3	882	2.38	1.11	2.00	2.35	1.48	1.00	4.00	3.00	0.12	-1.33	0.04
BMI10	4	882	19.60	4.22	18.48	19.11	3.57	12.37	35.16	22.79	1.10	0.94	0.14
SkinColorSat10	5	882	3.59	0.64	4.00	3.70	0.00	1.00	4.00	3.00	-1.58	2.41	0.02
SkinColorSat18	6	882	3.50	0.65	4.00	3.58	0.00	1.00	4.00	3.00	-1.36	2.36	0.02
BodySat10	7	882	28.49	5.14	28.00	28.83	5.93	9.00	36.00	27.00	-0.61	0.36	0.17
BodySat18	8	882	25.44	6.06	26.00	25.64	5.93	9.00	36.00	27.00	-0.33	0.10	0.20
BingeEatDisorder10	9	882	2.59	1.82	2.00	2.49	1.48	0.00	8.00	8.00	0.45	-0.35	0.06
BingeEatDisorder18	10	882	1.68	1.45	1.00	1.52	1.48	0.00	8.00	8.00	0.88	0.84	0.05
BodySatCha	11	882	-3.05	6.45	-3.00	-3.09	5.93	-27.00	24.00	51.00	0.11	0.73	0.22

## R OUTPUT

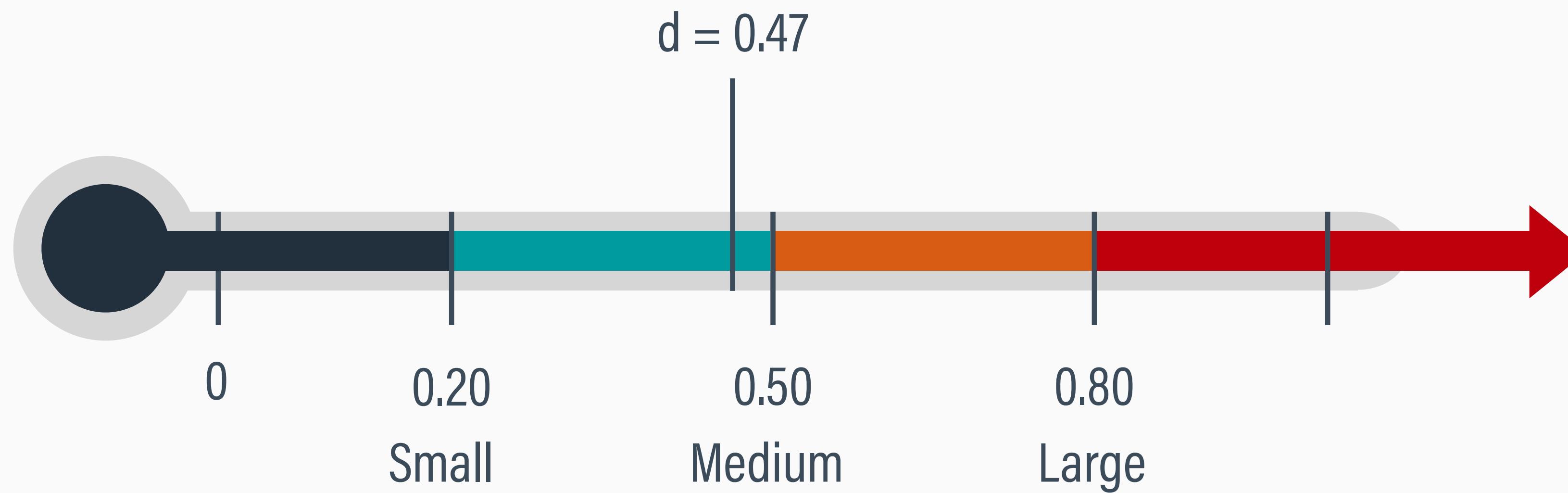
---

standardized mean difference effect size: -0.4726633

# COHEN'S BENCHMARKS

---

- A 0.47 standard deviation difference is in the small effect size range, very close to the medium cutoff



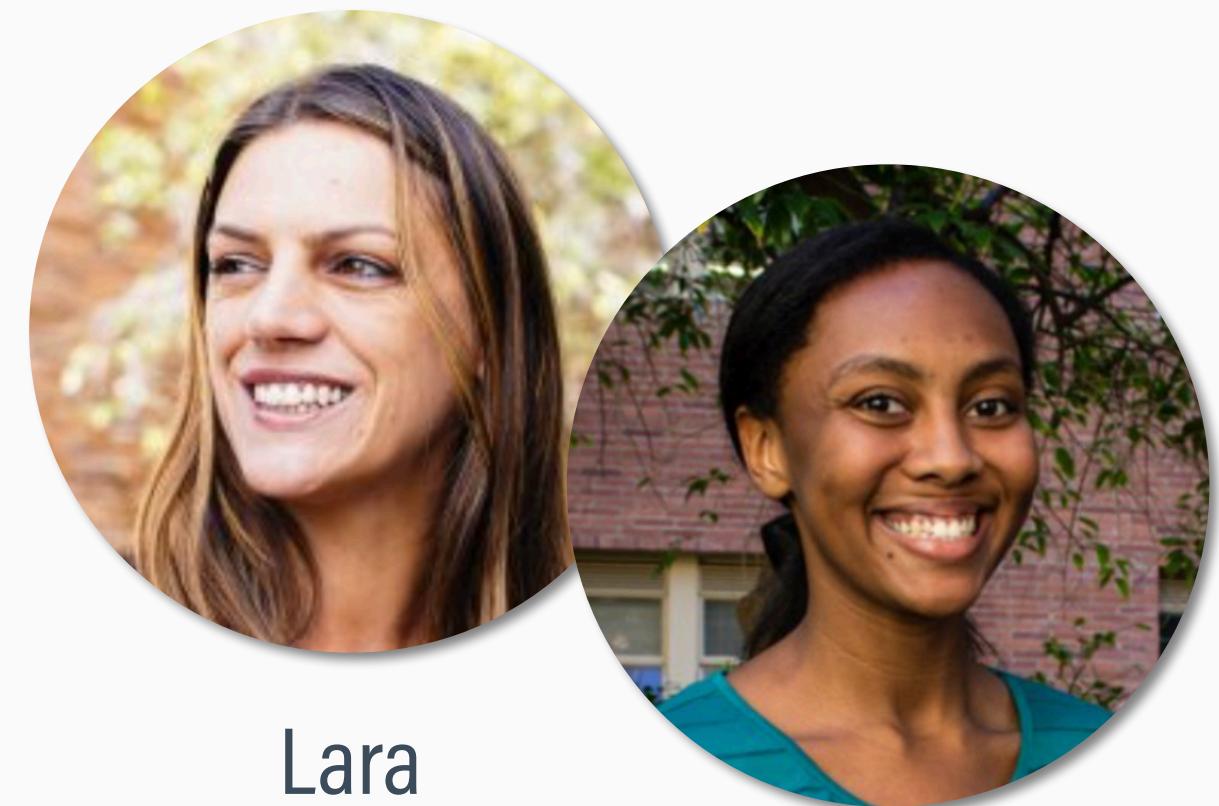
# OUTLINE

- 1 Effect size overview
- 2 Standardized mean difference
- 3 Within-subjects application
- 4 Between-subjects application
- 5 Study questions

# SMOKING AND DRINKING CESSATION TRIAL

---

Pharmacological treatments that can concomitantly address cigarette smoking and heavy drinking stand to improve health care delivery for these highly prevalent co-occurring conditions. This superiority trial compared the combination of varenicline and naltrexone against varenicline alone for smoking cessation and drinking reduction among heavy-drinking smokers.



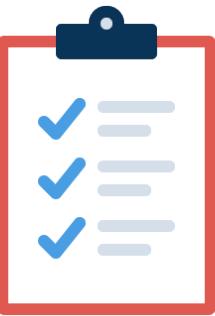
Lara  
Ray

ReJoyce  
Green

Ray, L.A., Green, R., Enders, C., et al. (2021). Efficacy of combining varenicline and naltrexone for smoking cessation and drinking reduction: A randomized clinical trial. *American Journal of Psychiatry*, 178, 818–828.

# KEY VARIABLES

---



## Breath (alveolar) carbon monoxide

A measure of carbon monoxide in the lungs. Breath carbon monoxide is a biomarker of smoking behavior common in clinical trials. Higher scores reflect more frequent smoking.



## Medication arm

Participants were randomly assigned to receive one of two meds: varenicline plus naltrexone or varenicline plus placebo pills

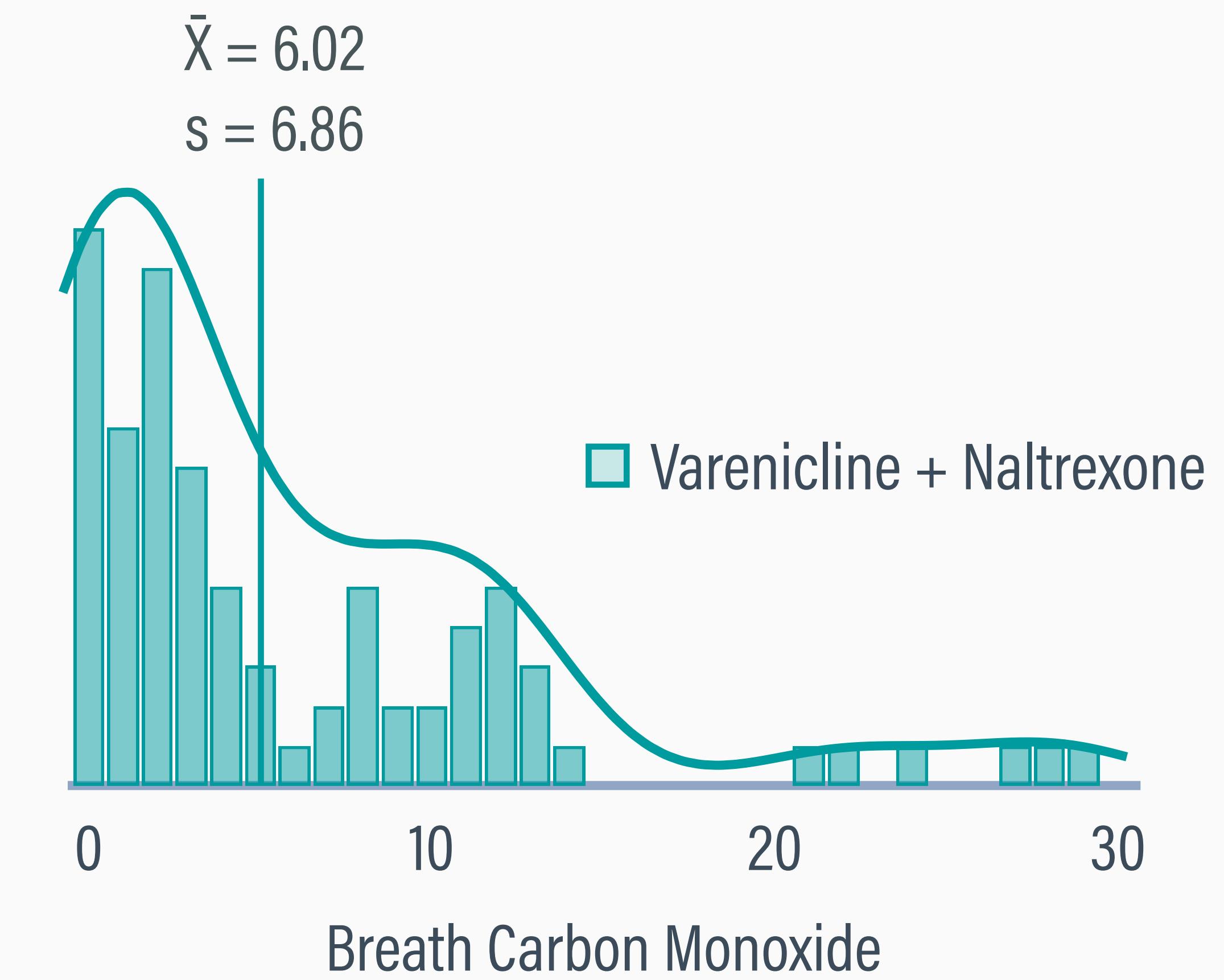
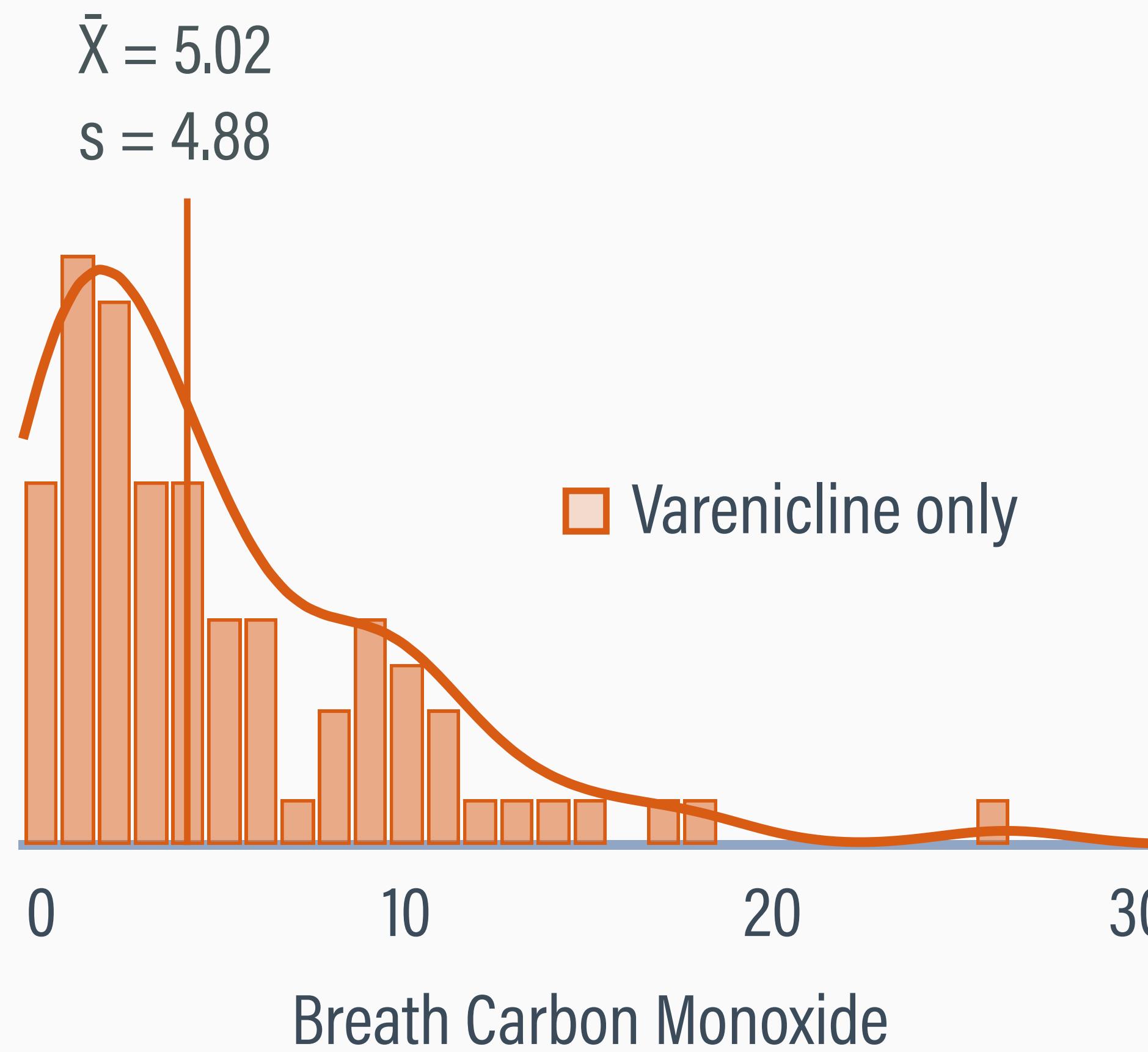
# RESEARCH QUESTION

---

- Question: Does smoking intensity differ between participants receiving a combination of two medications and those receiving a single medication alone?
- The explanatory (independent) variable, medication arm, consists of two groups: varenicline only and varenicline plus naltrexone
- The outcome (dependent) variable, breath carbon monoxide, is a numeric biomarker of smoking intensity

# GROUP STATISTICS

---



# STANDARDIZED MEAN DIFFERENCE

---

- One of several SMD formulas for between-subjects designs

$$d = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} = \frac{|\bar{X}_1 - \bar{X}_2|}{s_{\text{pooled}}} = \frac{\text{sample mean difference}}{\text{average standard deviation}}$$

- This version's denominator is the average (pooled) standard deviation of the two groups

# EFFECT SIZE EXAMPLE

- The effect size expresses the mean difference ( $\bar{X}_{\text{diff}} = -1$ ) on a standardized metric
- The mean difference of 1.0 on the breath CO scale equates to 0.17 standard deviation units
- The two means differ by 0.17 z-score units

Medication	$\bar{X}$	s	n
Varenicline	5.02	4.88	82
Varenicline + Naltrexone	6.02	6.86	83

$$d = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} =$$

$$\frac{|5.02 - 6.02|}{\sqrt{\frac{(82 - 1)4.88^2 + (83 - 1)6.86^2}{82 + 83 - 2}}} = 0.17$$

## R OUTPUT

---

Cohen d statistic of difference between two means

lower effect upper

COWeek8 -0.14    **0.17**    0.47

Multivariate (Mahalanobis) distance between groups

[1] 0.17

r equivalent of difference between two means

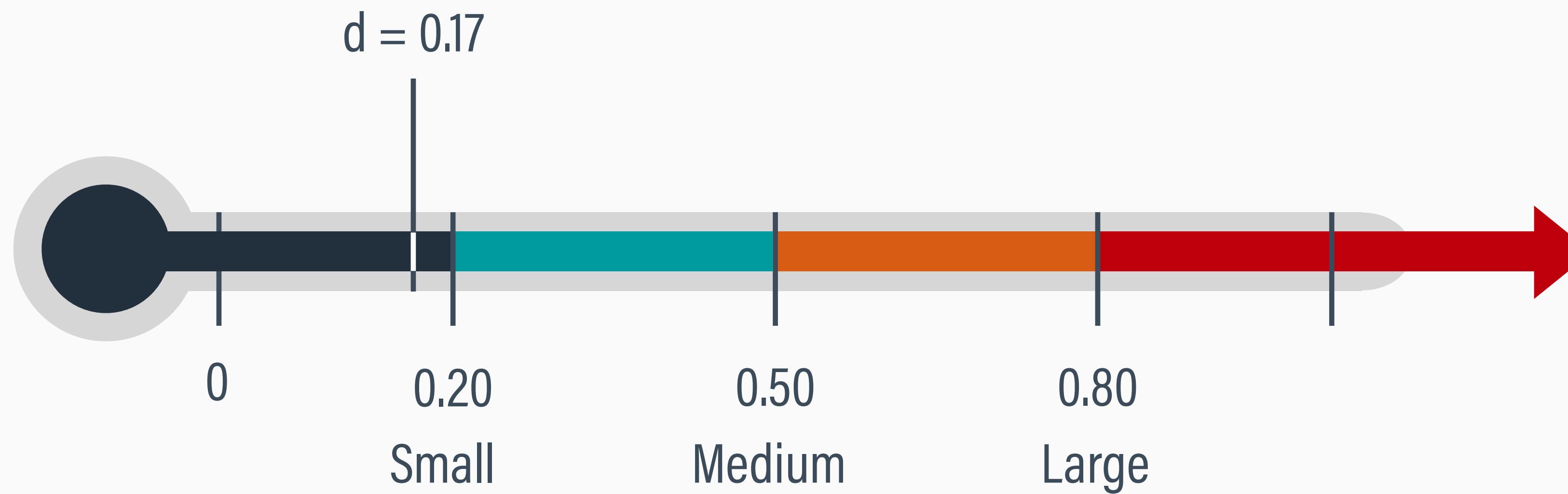
COWeek8

0.08

# COHEN'S BENCHMARKS

---

- A 0.17 standard deviation difference is very close to the small effect size cutoff



# OUTLINE

- 1 Effect size overview
- 2 Standardized mean difference
- 3 Within-subjects application
- 4 Between-subjects application
- 5 Study questions

# STUDY QUESTIONS (1)

---

1. In psychological research, statistical significance has traditionally been the main criterion for evaluating results. Explain why significance testing alone is insufficient. In your answer, describe what effect size contributes to the interpretation of findings, how it differs from p-values, and why APA requires researchers to report it.
  
2. A researcher analyzing a large national database ( $N = 10,000$ ) found a significant difference between two groups in their attitudes ( $p < .01$ ). The standardized mean difference effect size was only  $d = 0.05$ . How should these results be interpreted together? Why might statistical significance and effect size appear to send different messages, and what does each contribute to our understanding of the finding?

## STUDY QUESTIONS (2)

---

Researchers studying married couples want to determine whether husbands and wives differ in their relationship satisfaction. The sample mean difference from a sample of  $N = 100$  couples is  $\bar{X}_{\text{diff}} = \bar{X}_{\text{wives}} - \bar{X}_{\text{husbands}} = -5$ . The effect size is  $d = 0.25$ .

3. Provide an interpretation of the standardized mean difference effect size. Explain what the numeric value means and how it compares to Cohen's "off-the-shelf" benchmarks. How would you characterize the magnitude of improvement in pain scores?

## STUDY QUESTIONS (3)

---

A clinical psychologist wants to determine whether cognitive behavioral therapy decreases anxiety. She finds the treatment group mean of  $\bar{X} = 5$  ( $SD = 1$ ,  $n = 30$ ) and a control group mean of  $\bar{X} = 5.6$  ( $SD = 1$ ,  $n = 30$ ). The standardized mean difference effect size is  $d = 0.60$ .

4. Provide an interpretation of the standardized mean difference effect size. Explain what the numeric value means and how it compares to Cohen's "off-the-shelf" benchmarks.