

Investigating Patterns of Fadeout using MERF, the Meta-Analysis of Educational RCTs with Follow-up

Emma R. Hart, PhD Candidate
Teachers College, Columbia University
May 17, 2024

Key Collaborators

Tyler Watts- Teachers College, Columbia

Drew Bailey- University of California, Irvine



Research Assistants

- | | | |
|--------------------|-----------------------|--------------------------|
| 1. Pritha Sengupta | 10. Gabby Lammano | 19. Camila Torres-Rivera |
| 2. Spruha Reddy | 11. Precious Elam | 20. Victoria Chang |
| 3. Sha Luo | 12. Fatmanur Ozay | 21. Yolanda Xu |
| 4. Simran Juneja | 13. John Schupbach | 22. Susan Kruglinski |
| 5. Jessica Sperber | 14. Opal Ofstedal | 23. Abby Lee |
| 6. Xinyu Pan | 15. Mindy Rosengarten | |
| 7. Josefa Wester | 16. Maddie Scricco | |
| 8. Siyu Liang | 17. Helen Ding | |
| 9. Leo Weaver | 18. Devon Turner | |

Theory & Background

Researchers and policymakers anticipate that educational interventions will:

1. Improve child outcomes at post-test
2. Set children on desirable trajectories
3. Have long-term impacts

Expectation of long-run effects

Theory:

- Rudimentary skills lay the foundation for advanced skills
- Developmental cascades
- Self- / cross-productivities
- Skills beget skills

Cunha & Heckman, 2007; Dodge et al., 1986;
Masten & Cicchetti, 2010; Sarama & Clements, 2009

Expectation of long-run effects

Correlational evidence:

1. Auto-regressive & cross-domain (Duncan et al., 2007; Moffitt et al., 2011; Davis-Kean et al., 2022)
2. Often interpreted to suggest that stronger earlier skills beget stronger later skills with long-run benefits on adult outcomes

Highly-influential RCT evidence:

1. Abecedarian (Campbell et al., 2002)
2. Perry (Elango et al., 2016)

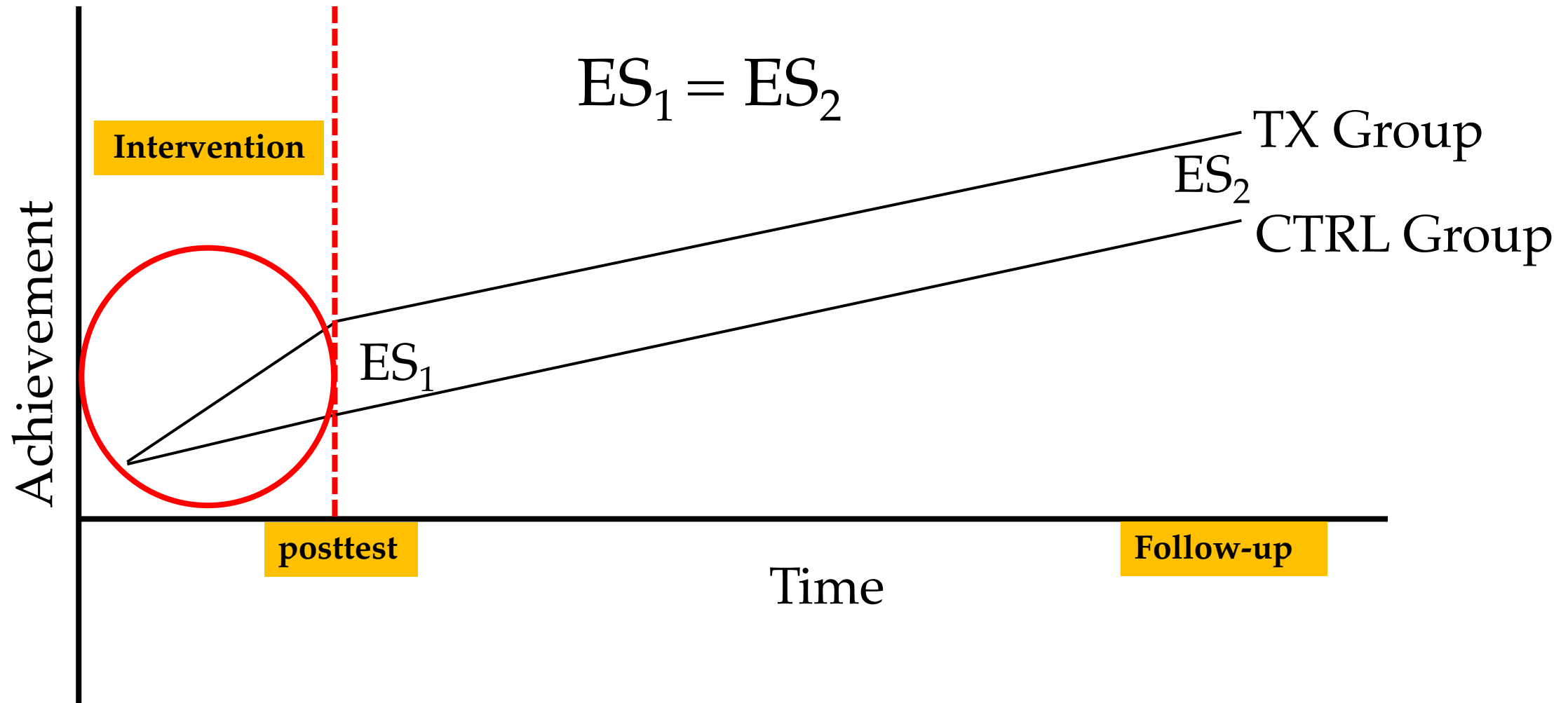
Skills beget skills theory
Correlational skill stability
+ Landmark RCTs

A plausible picture of development **whereby
stronger earlier skills lead to stronger later skills**

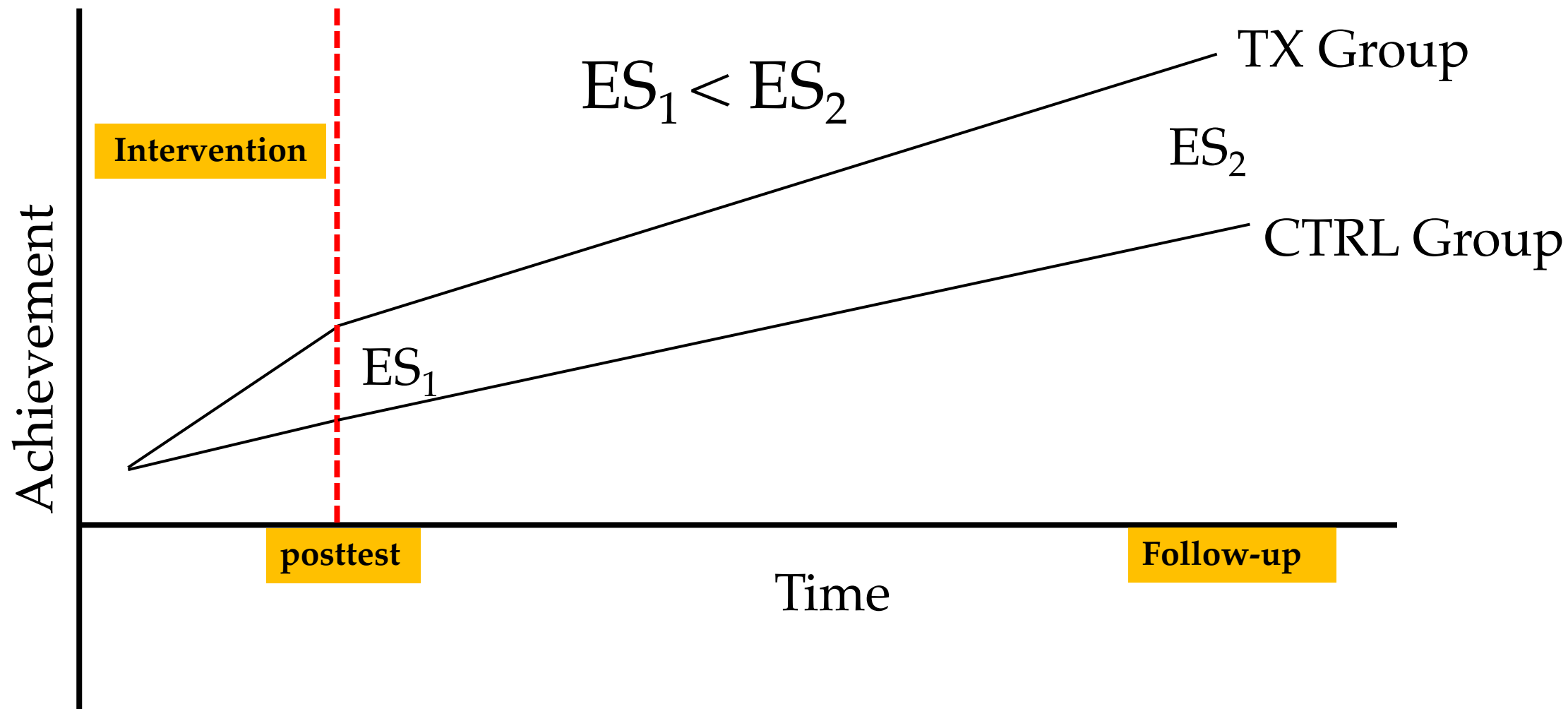


Intervention-driven boosts to child skills should
persist & give rise to long-run impacts on adult outcomes

Expectation of Persistence



Expectation of Persistence



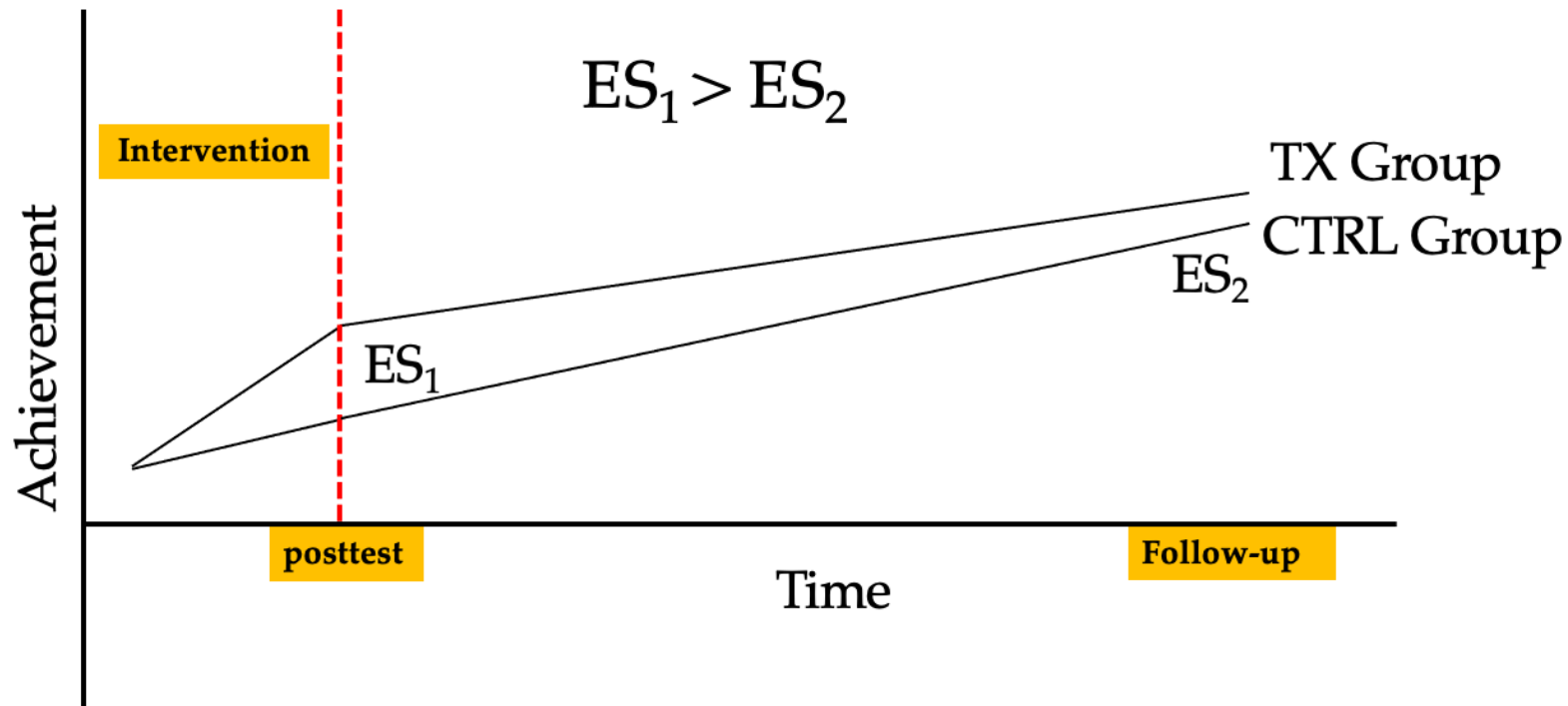
2 patterns in the literature that don't follow expectations:

1. Fadeout

2. Emergence

Fadeout

Intervention effects on cognitive skills tend to fade overtime (Bailey et al., 2017; 2020)



Emergence

Long-run emergent impacts on adult outcomes, *despite* fadeout
(Chetty et al., 2011; Gray-Lobe et al., 2023; Deming, 2009)

Fadeout & Emergence

Complicated to square with existing theory

- How broad of an issue is fadeout?
- What are the mechanisms through which long-run adult impacts emerge if not through sustained benefits on child skills?

Social-Emotional Skill Persistence

- Maybe persistent intervention impacts on social-emotional skills explain emergent long-run effects? (Chetty et al., 2011; Deming, 2009; Heckman & Kautz, 2012)
- Less is known about social-emotional skill fadeout (Abenavoli, 2019)
- Uniquely effective skills-beget-skills processes (Social Information Processing Theory; Dodge & Crick, 1990)

Do intervention impacts on social-emotional skills show more persistence than impacts on cognitive skills?

Better Interventions = More Persistence?

Maybe some interventions have “deeper” impacts, generating more persistence:

- ECE interventions
- Longer interventions
- Interventions targeting parents
- Interventions targeting many skills
- Interventions occurring under worse counterfactual conditions

Are there certain kinds of interventions that show greater persistence?

Meta-Science is Needed

- To date, most investigation occurs at the within-study level
 - Post hoc explanations
 - Possibility of under / overlearning from any one study
- Need to more systematically understand:
 - Breadth of issue
 - Theory → Do skill boosts change developmental trajectories?
 - Do larger post-test impacts on child skills predict larger follow-up effects on child skills?

Meta-analysis of Educational
RCTs with Follow-up
(**MERF**)

MERF Dataset

426 papers (from 8 published meta-analyses)

183 RCTs with cognitive or social-emotional outcomes

94 RCTs with follow-up on same students ≥ 6 mo after post-test

85 RCTs with usable data, double coded

726 post-test effects
1,247 follow-up effects

A variety of interventions such as:

1. Infant home visiting programs
2. Pre-k programs
3. Elementary-level curriculum
4. Adolescent substance use prevention

MERF Dataset

426 papers (from 8 published meta-analyses)

183 RCTs with cognitive or social-emotional outcomes

94 RCTs with follow-up on same students ≥ 6 mo after post-test

85 RCTs with usable data, double coded

726 post-test effects
1,247 follow-up effects

Cognitive outcomes examples:

- Math
- IQ
- Reading
- Working memory
- Academic achievement

Social-emotional outcomes examples:

- Externalizing problems
- Internalizing problems
- Prosocial behaviors
- School attendance
- Substance use

MERF Dataset

426 papers (from 8 published meta-analyses)



183 RCTs with cognitive or social-emotional outcomes



94 RCTs with follow-up on same students ≥ 6 mo after post-test



85 RCTs with usable data, double coded



726 post-test effects
1,247 follow-up effects

MERF Dataset

426 papers (from 8 published meta-analyses)



183 RCTs with cognitive or social-emotional outcomes



94 RCTs with follow-up on same students ≥ 6 mo after post-test



85 RCTs with usable data, double coded



726 post-test effects
1,247 follow-up effects

MERF Dataset

426 papers (from 8 published meta-analyses)



183 RCTs with cognitive or social-emotional outcomes



94 RCTs with follow-up on same students ≥ 6 mo after post-test



85 RCTs with usable data, double coded



726 post-test effects
1,247 follow-up effects

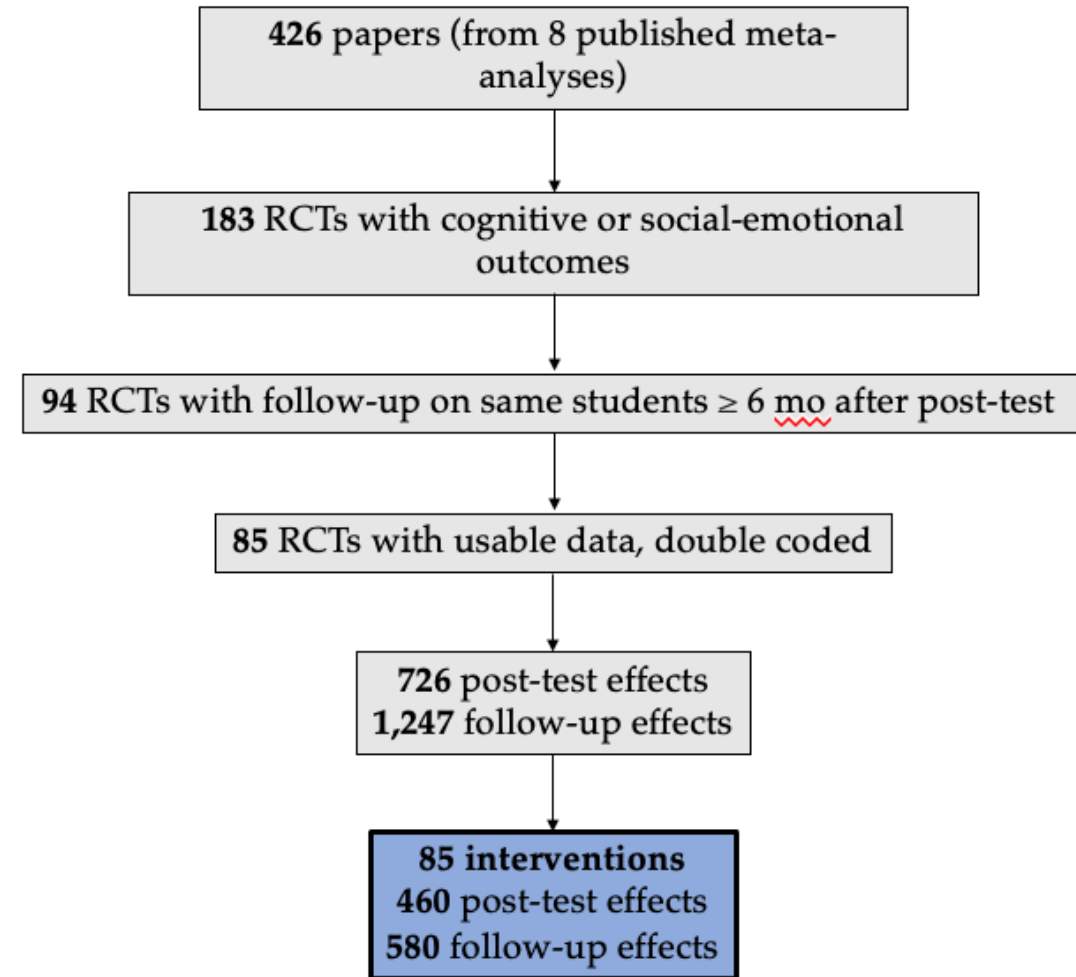
Analytic Groupings

Grouping the data:

- Same study
- Same construct
- Must be collected at post-test and at least 1 follow-up
- Same measure and subscale

This approach has major advantages:

- Clear definition of fadeout/persistence
- Avoids bias due to measure changes
- Measures researchers cared about enough to measure at multiple times



Analytic Grouping Example- Intervention X

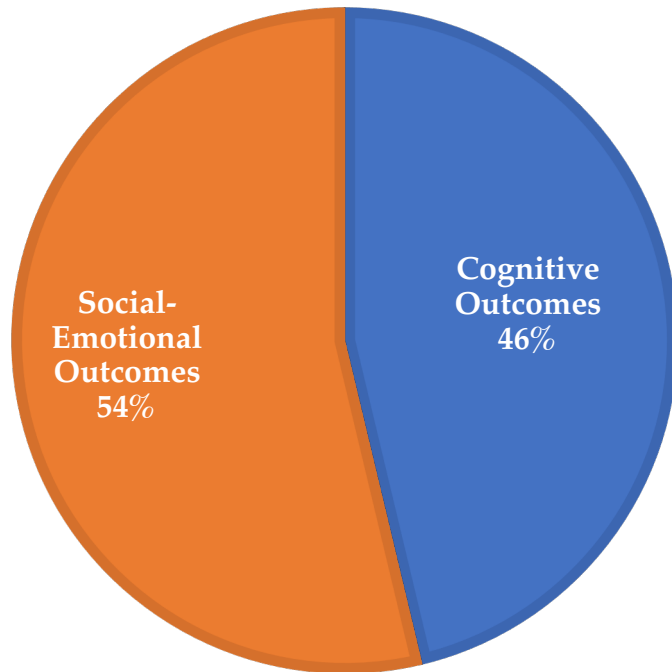
Time of Test	Construct	Measure
Post-test	Externalizing Problems	Strengths and Difficulties
6-mo Follow-up	Externalizing Problems	Strengths and Difficulties
2-yr Follow-up	Externalizing Problems	Strengths and Difficulties
Post-test	Depressive Symptoms	Children's Depression Inventory
Post-test	Vocabulary	Peabody Picture Vocabulary Test
2-yr Follow-up	Vocabulary	Stanford Achievement Test

Analytic Grouping Example- Intervention X

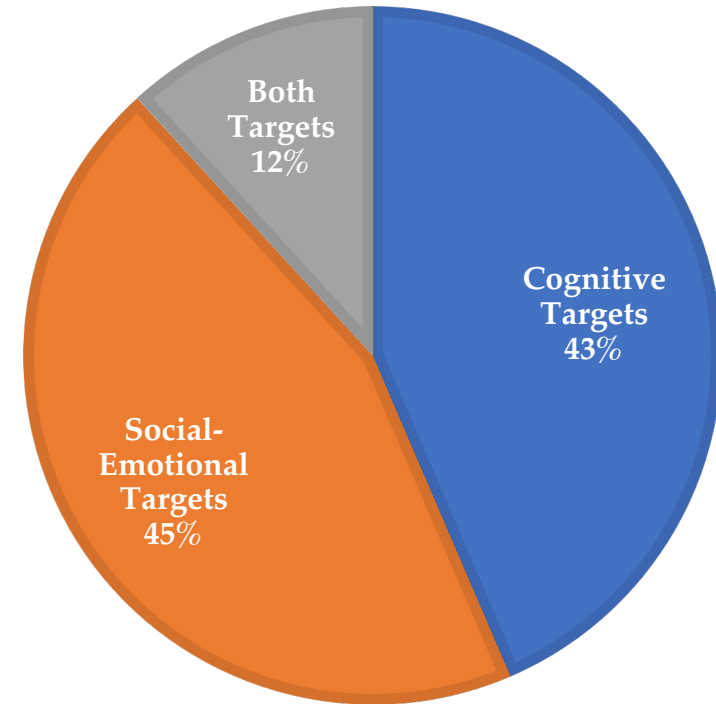
Time of Test	Construct	Measure
Post-test	Externalizing Problems	Strengths and Difficulties
6-mo Follow-up	Externalizing Problems	Strengths and Difficulties
2-yr Follow-up	Externalizing Problems	Strengths and Difficulties
Post-test	Depressive Symptoms	Children's Depression Inventory
Post-test	Vocabulary	Peabody Picture Vocabulary Test
2-yr Follow-up	Vocabulary	Stanford Achievement Test

Composition of Aligned Groups

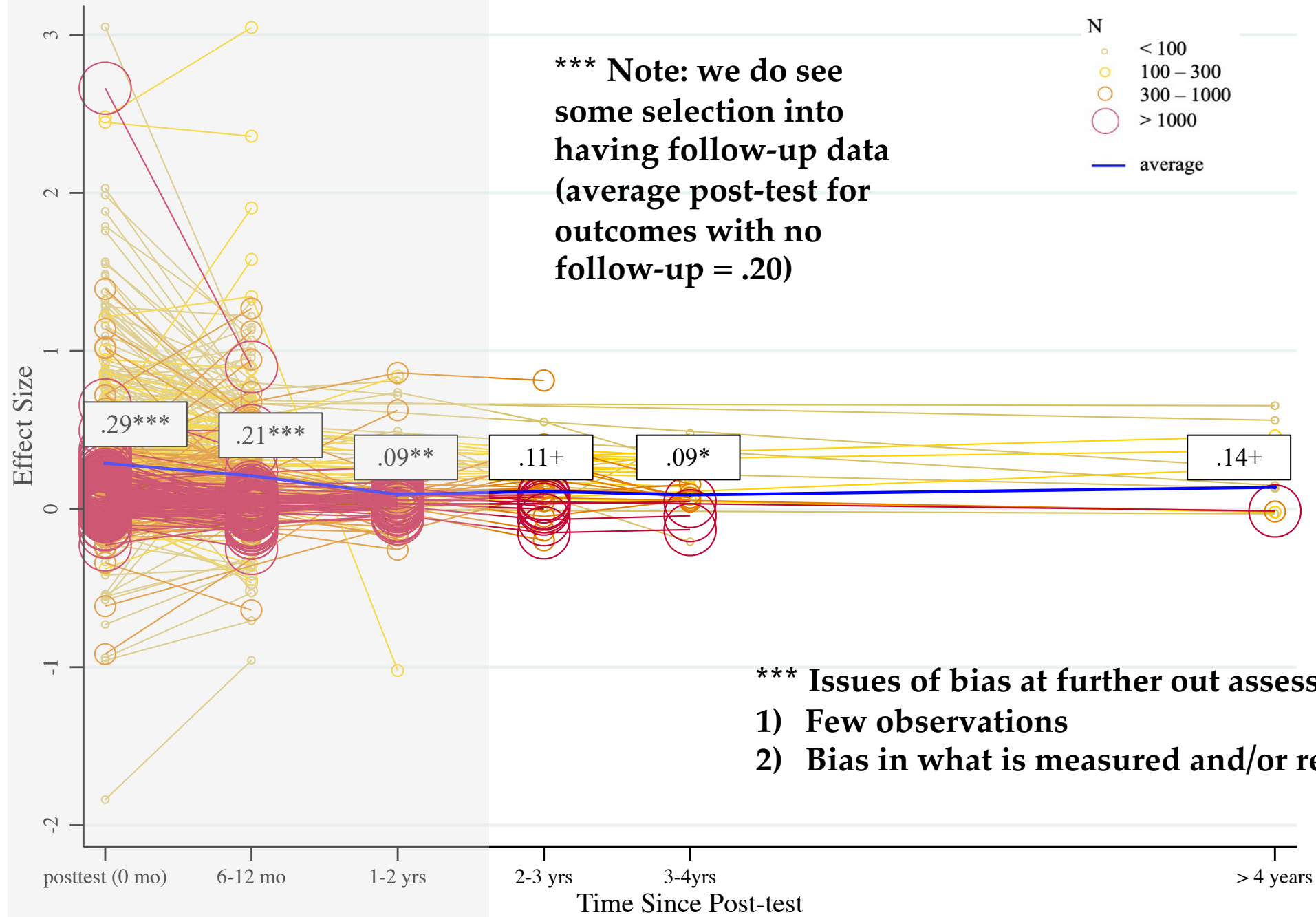
OUTCOMES



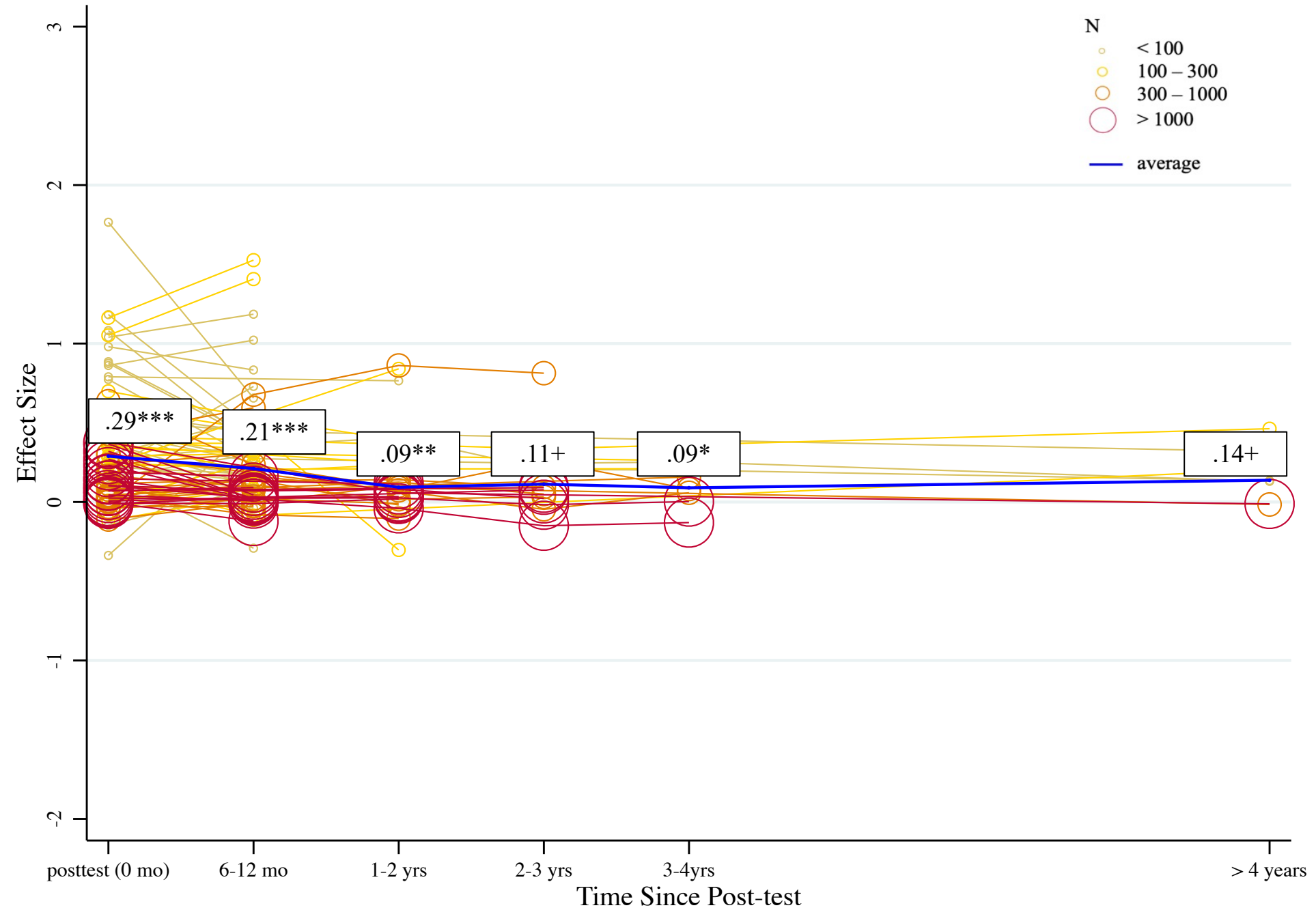
INTERVENTION TARGETS



Aligned Group Effect Size Trajectories



Average Aligned Group Trajectories by Study



To test theory about the extent to which post-test boosts persist, we need to look at trajectories within aligned groups

$$ES_{f si} = \beta_0 + \beta_1 ES_{psi} + \varepsilon_{si}$$

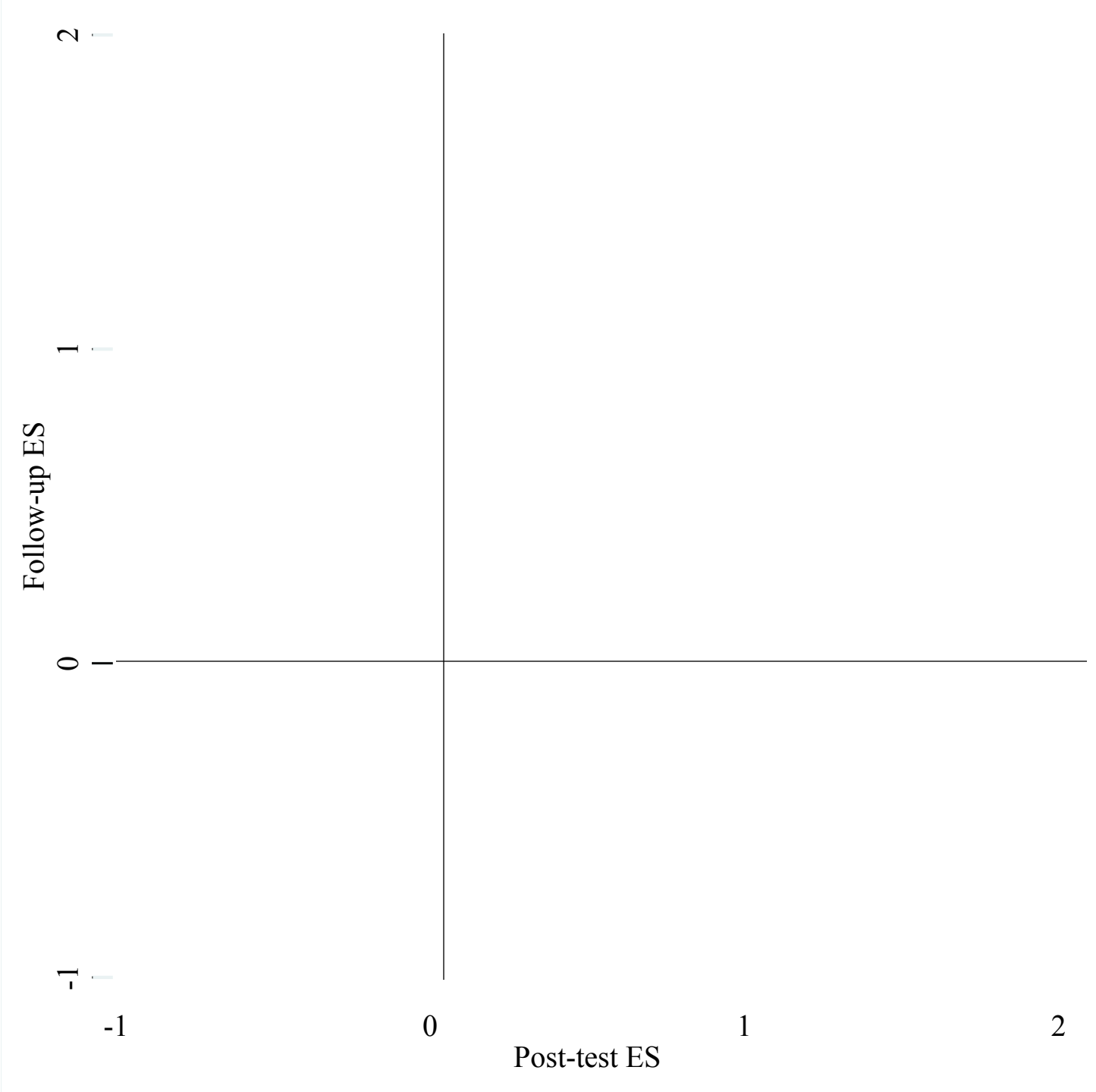
i = analytic grouping

s = study

f = follow-up

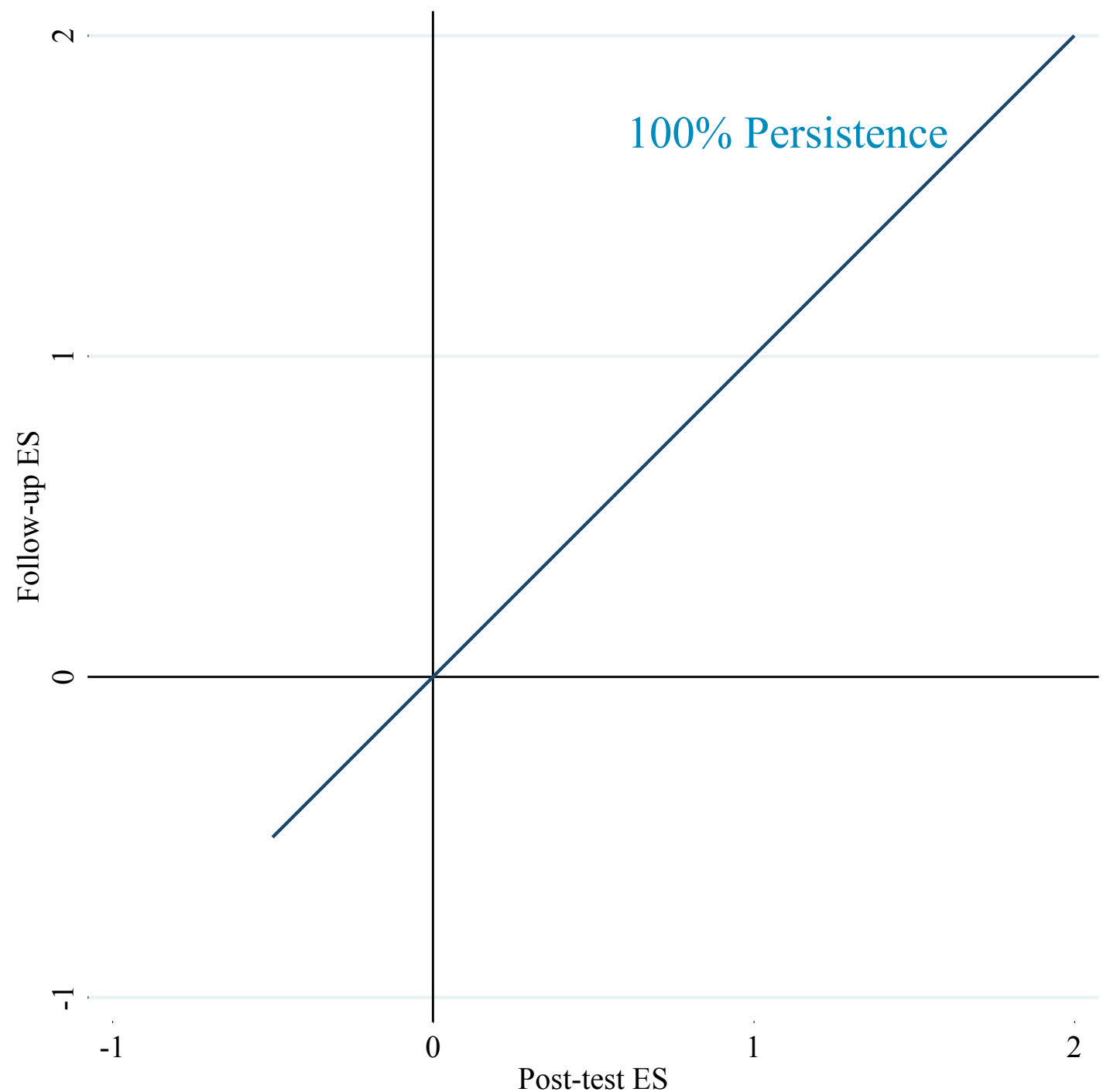
p = post-test

$$ES_{fsi} = \beta_0 + \beta_1 ES_{psi} + \epsilon_{si}$$



$$ES_{fsi} = \beta_0 + \beta_1 ES_{psi} + \epsilon_{si}$$

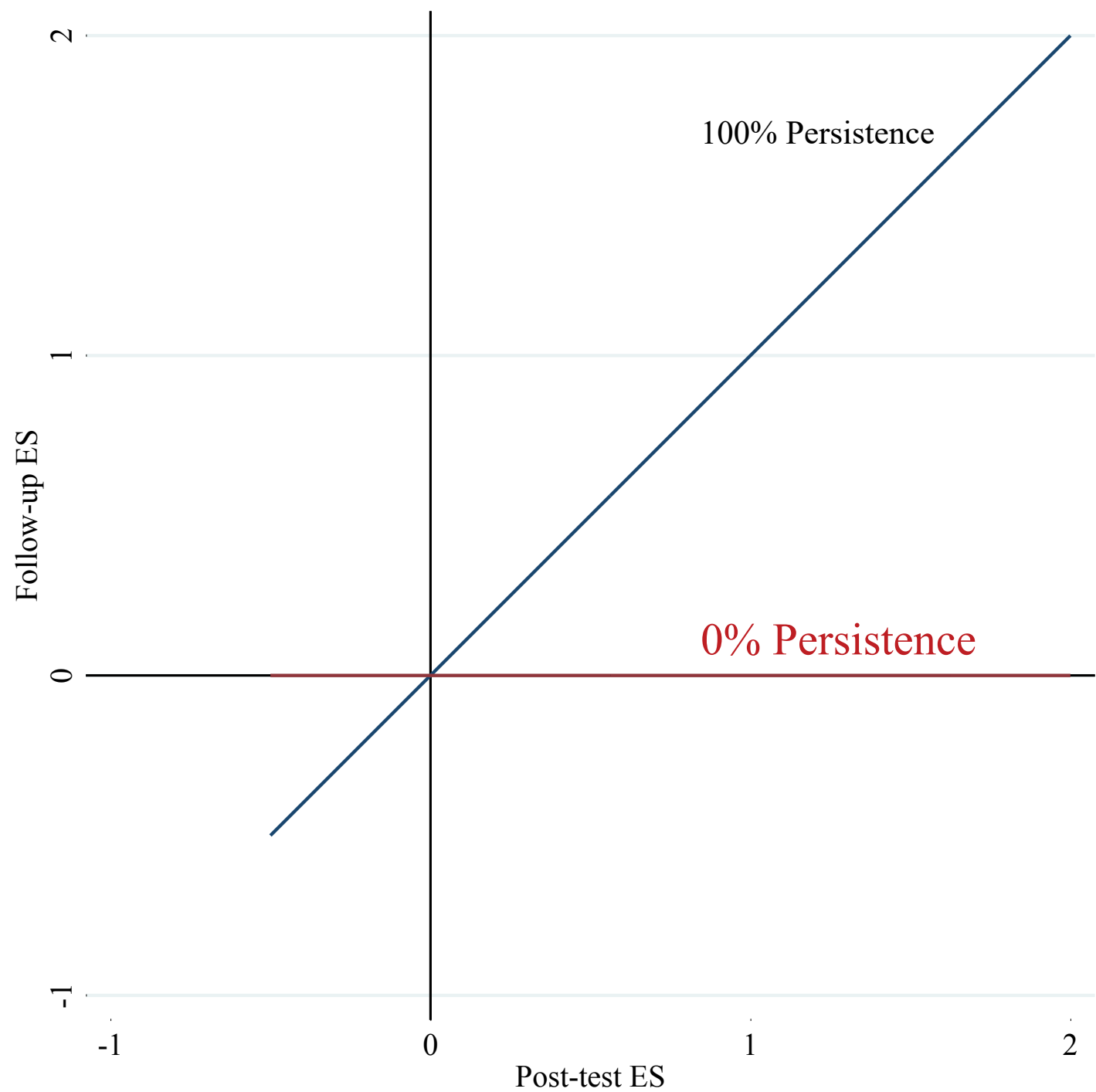
$$ES_{fsi} = (0) + (1)ES_{psi} + \epsilon_{si}$$



$$ES_{fsi} = \beta_0 + \beta_1 ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (1)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (0)ES_{psi} + \varepsilon_{si}$$

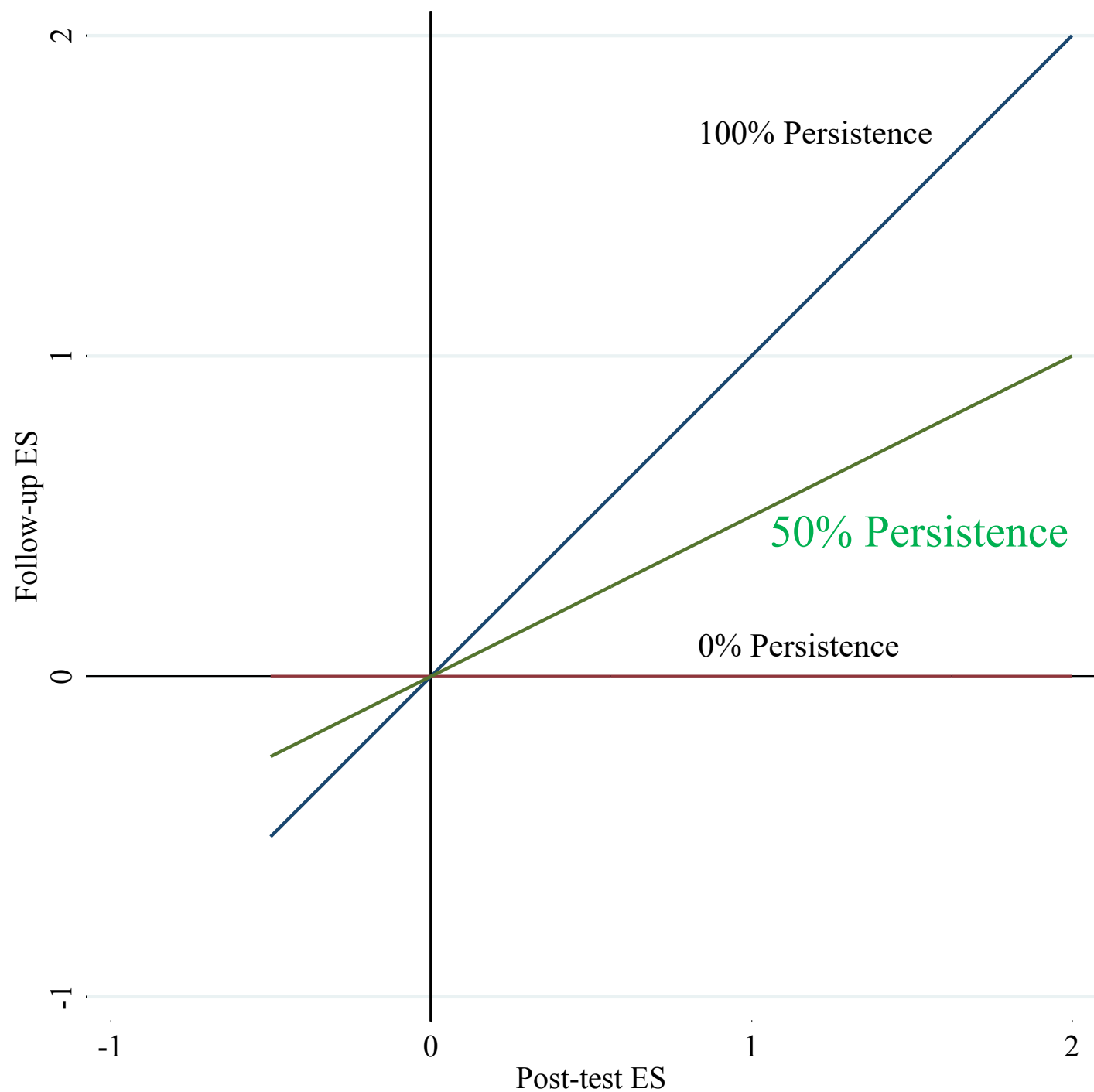


$$ES_{fsi} = \beta_0 + \beta_1 ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (1)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (0)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (.5)ES_{psi} + \varepsilon_{si}$$



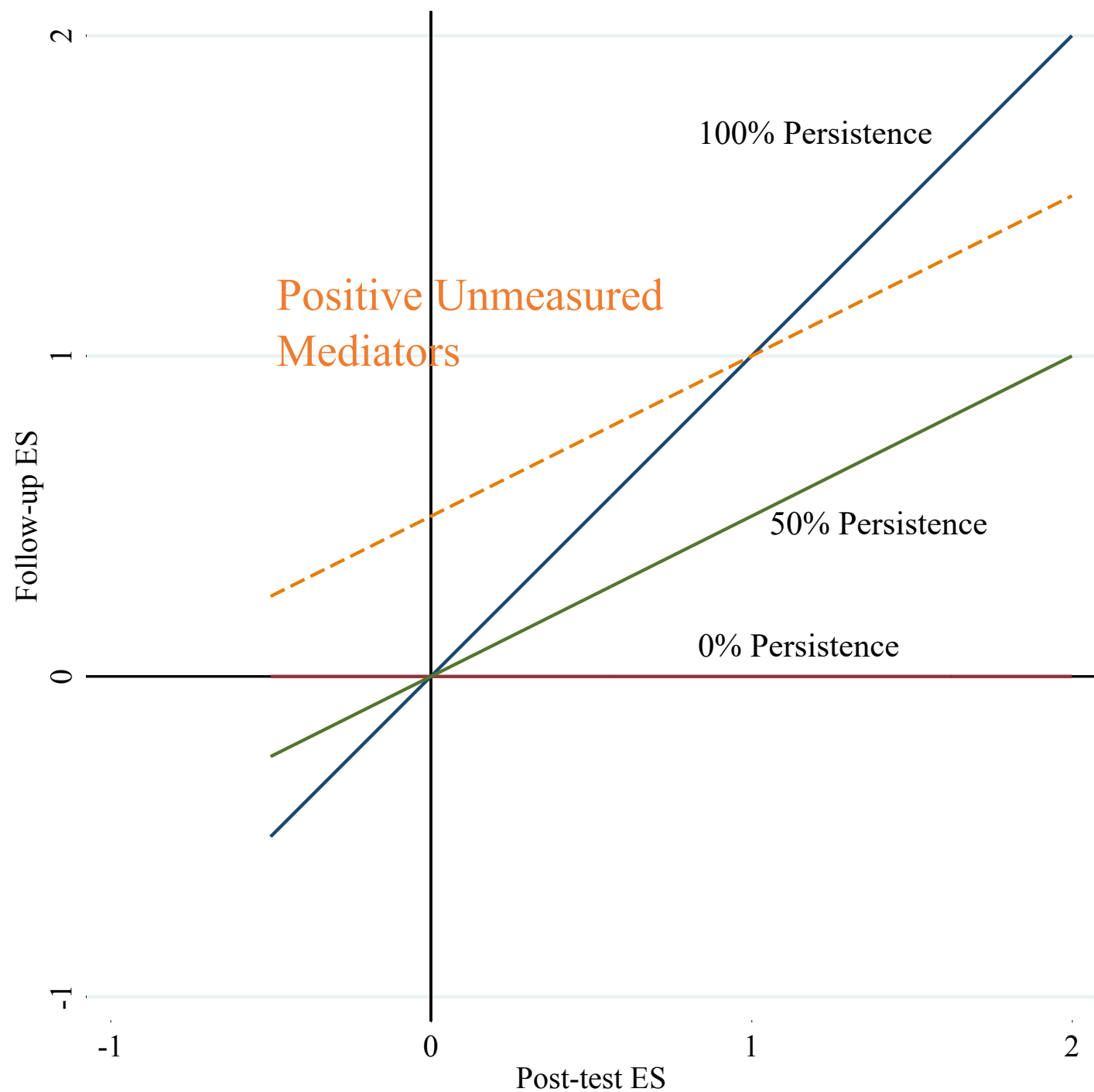
$$ES_{fsi} = \beta_0 + \beta_1 ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (1)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (0)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (.5)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (.5) + (.5)ES_{psi} + \varepsilon_{si}$$



$$ES_{fsi} = \beta_0 + \beta_1 ES_{psi} + \varepsilon_{si}$$

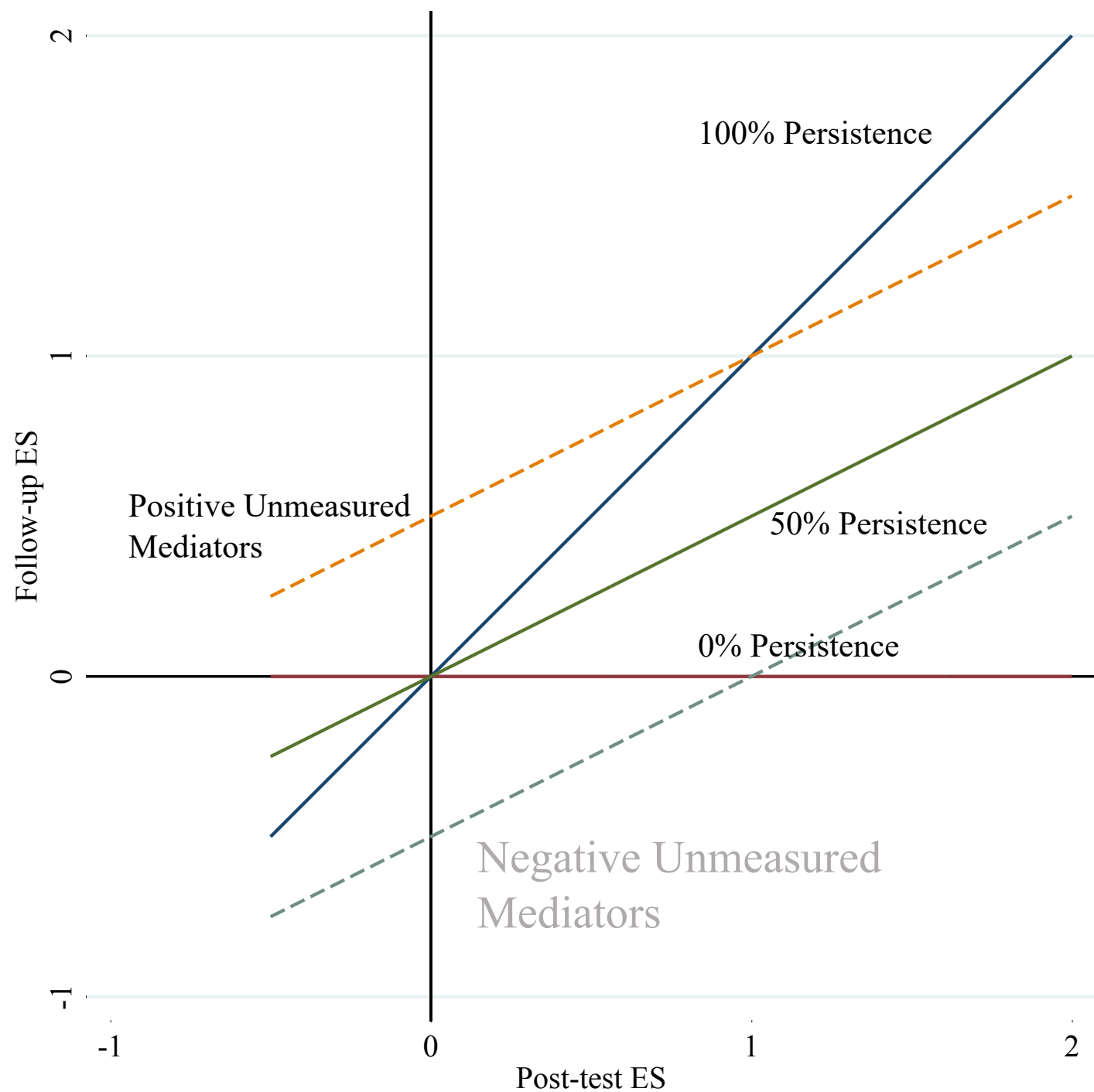
$$ES_{fsi} = (0) + (1)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (0)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (0) + (.5)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (.5) + (.5)ES_{psi} + \varepsilon_{si}$$

$$ES_{fsi} = (-.5) + (.5)ES_{psi} + \varepsilon_{si}$$



Do social-emotional impacts persist **more** than cognitive impacts?

$$ES_{fsi} = \beta_{0s} + \beta_1 ES_{psi} + \beta_2 SOC_{si} + \beta_3 ES * SOC_{si} + \epsilon_{si}$$

$$\beta_{0s} = \gamma_{00} + \tau_{0s}$$

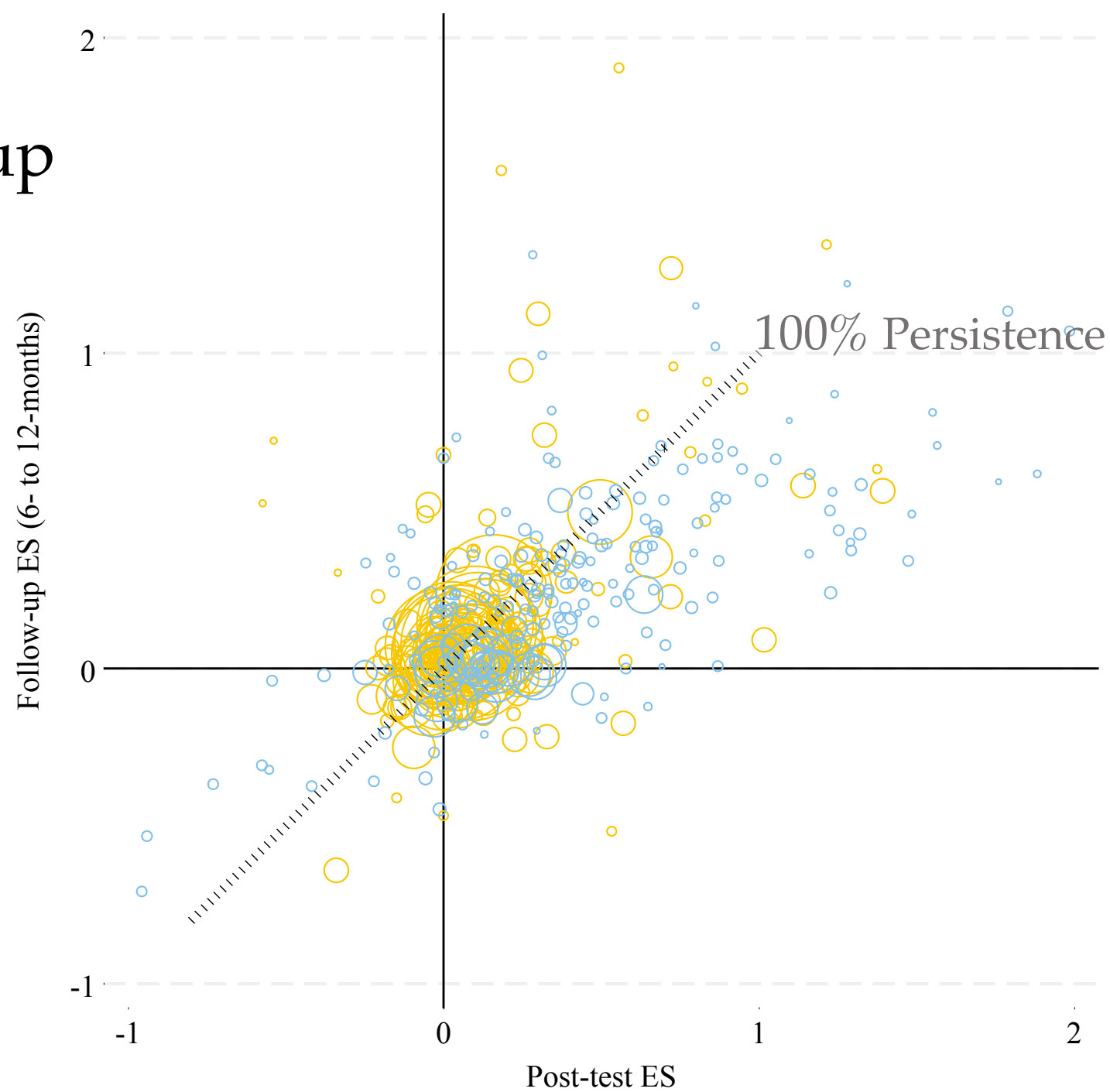
$$\beta_{1s} = \gamma_{10} + \tau_{1s}$$

Weighted by $\frac{1}{se^2}$

Cluster-robust standard errors
metafor

6- to 12-months Follow-up

- Social-emotional outcomes (SOC)
- Cognitive outcomes (COG)
- Coordinates weighted by $\frac{1}{se^2}$

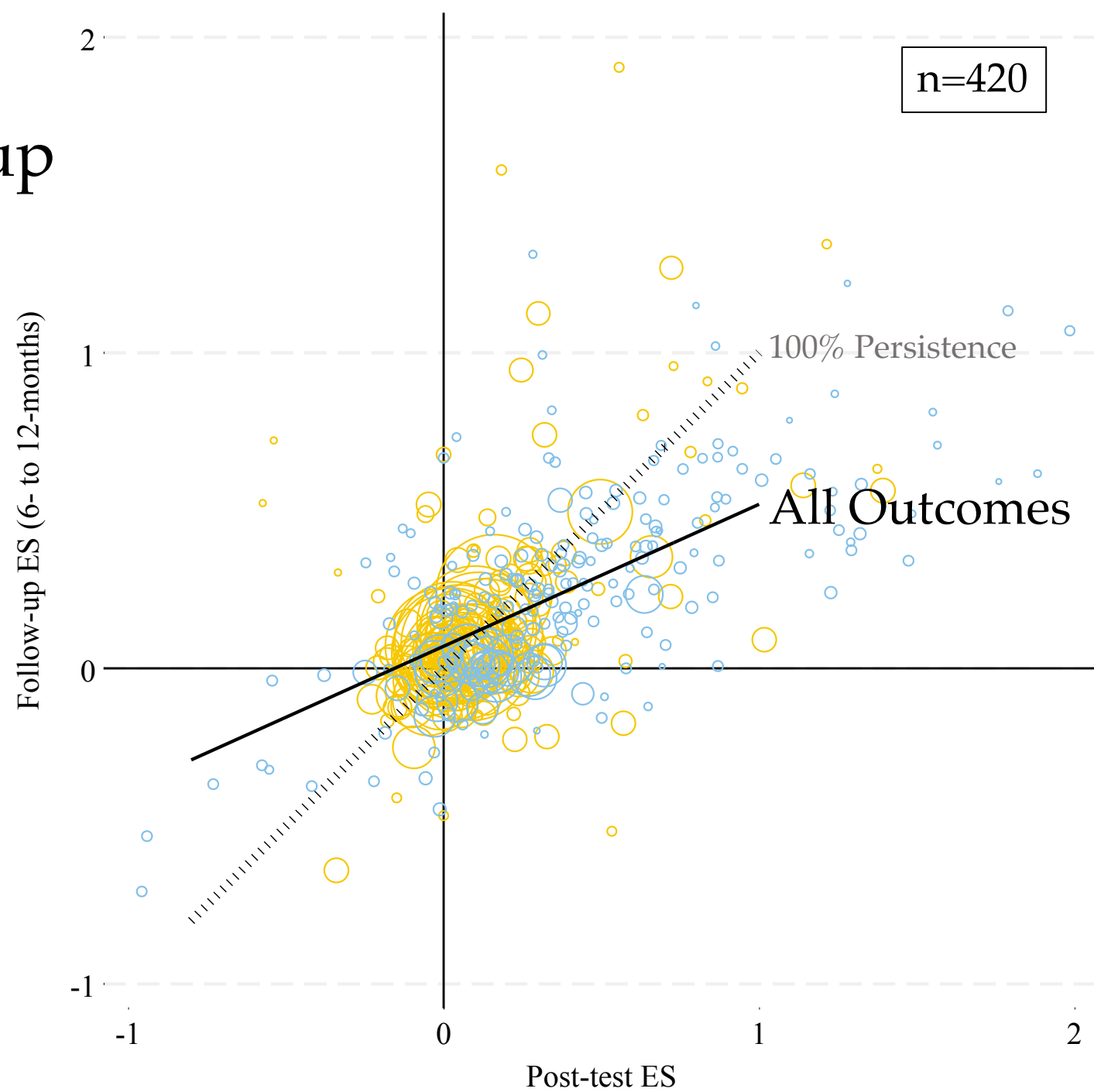


6- to 12-months Follow-up

All Outcomes:

$$\beta_0 = 0.07 *$$

$$\beta_1 = 0.45 ***$$

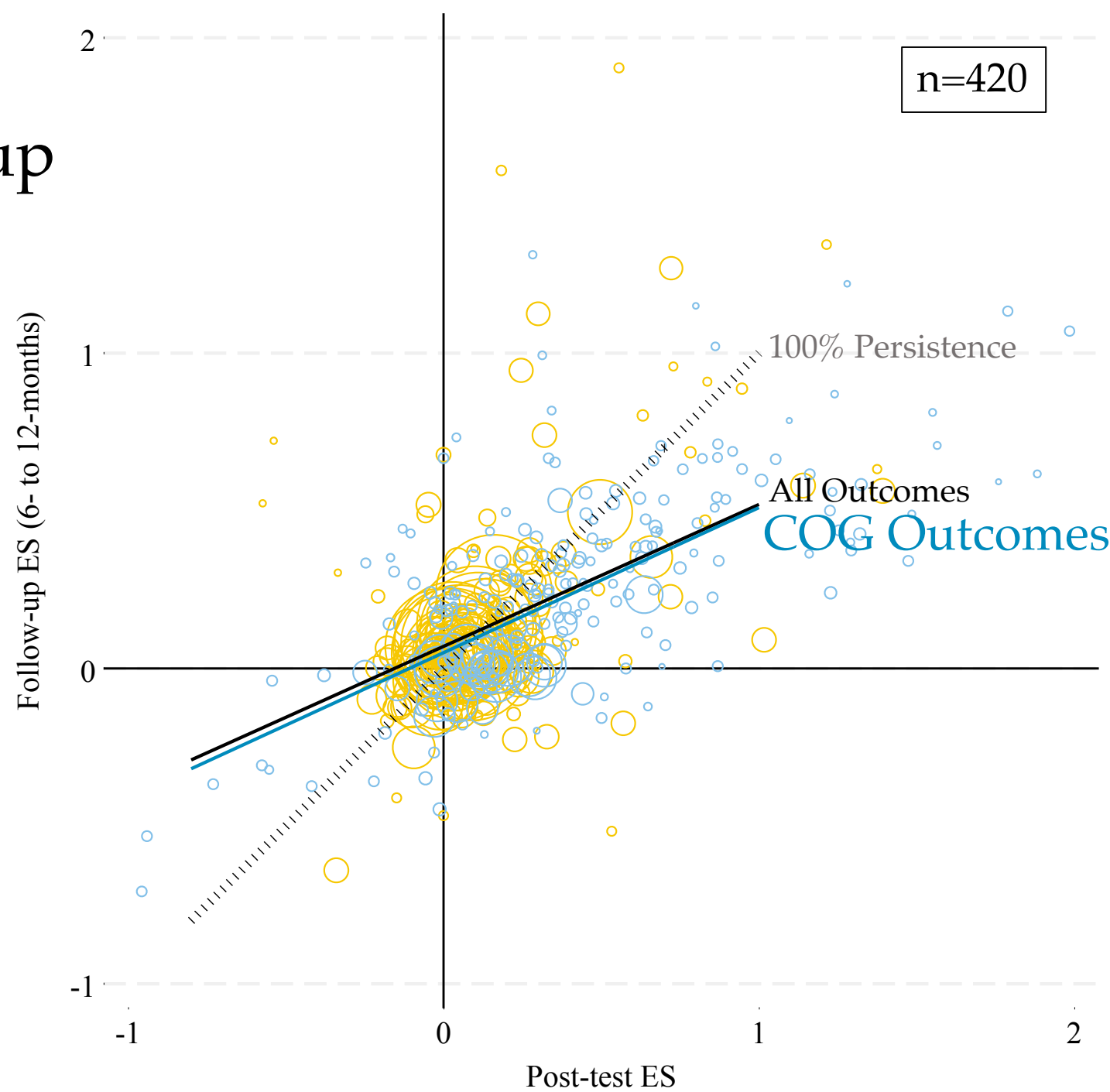


6- to 12-months Follow-up

Cognitive Outcomes:

$$\beta_0 = 0.05$$

$$\beta_1 = 0.46 ***$$



6- to 12-months Follow-up

Cognitive Outcomes:

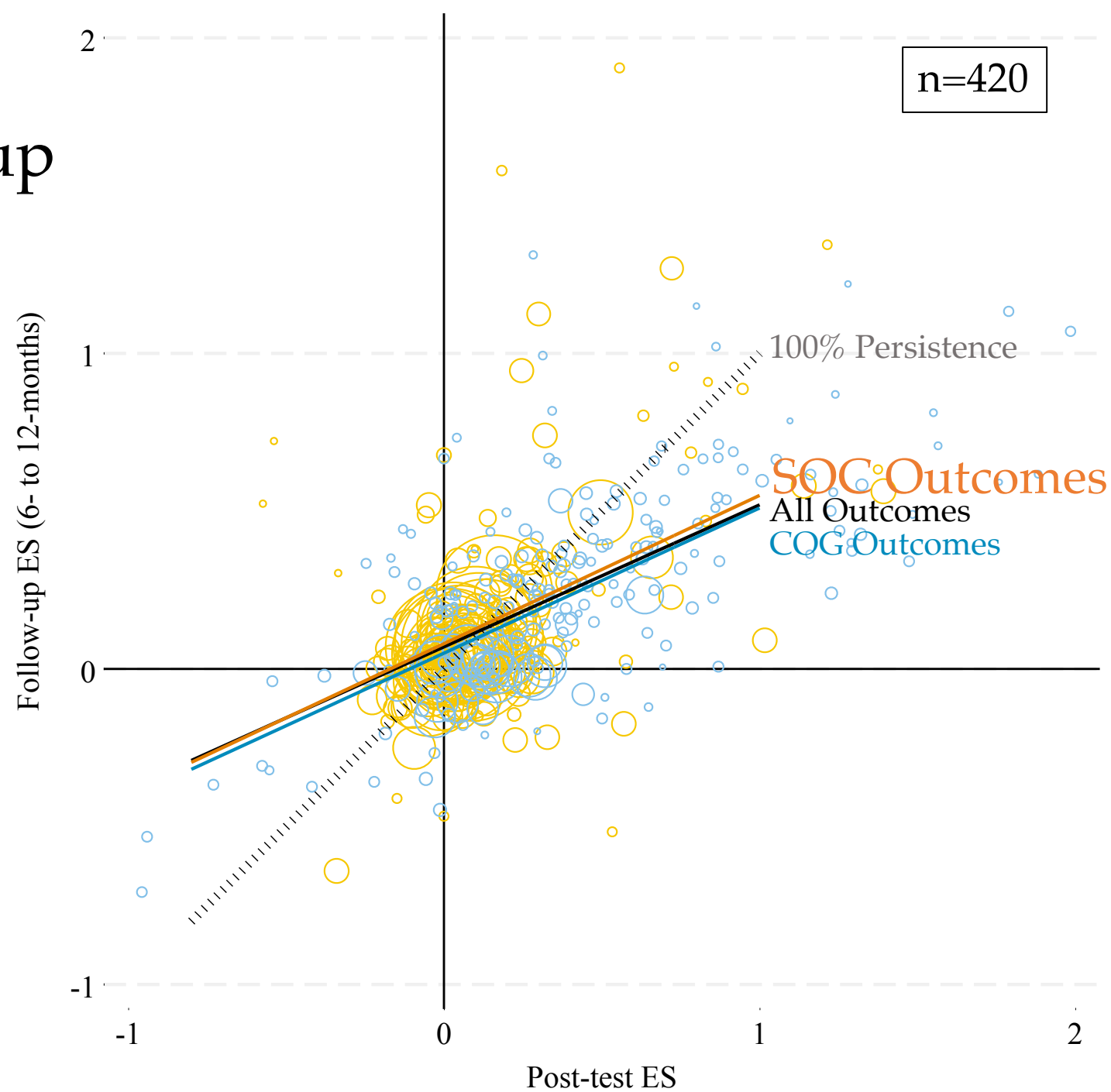
$$\beta_0 = 0.05$$

$$\beta_1 = 0.46 ***$$

Social-Emotional Outcomes:

$$\beta_0 = 0.08$$

$$\beta_1 = 0.47 ***$$



6- to 12-months Follow-up

Cognitive Outcomes:

$$\beta_0 = 0.05$$

$$\beta_1 = 0.46 ***$$

Social-Emotional Outcomes:

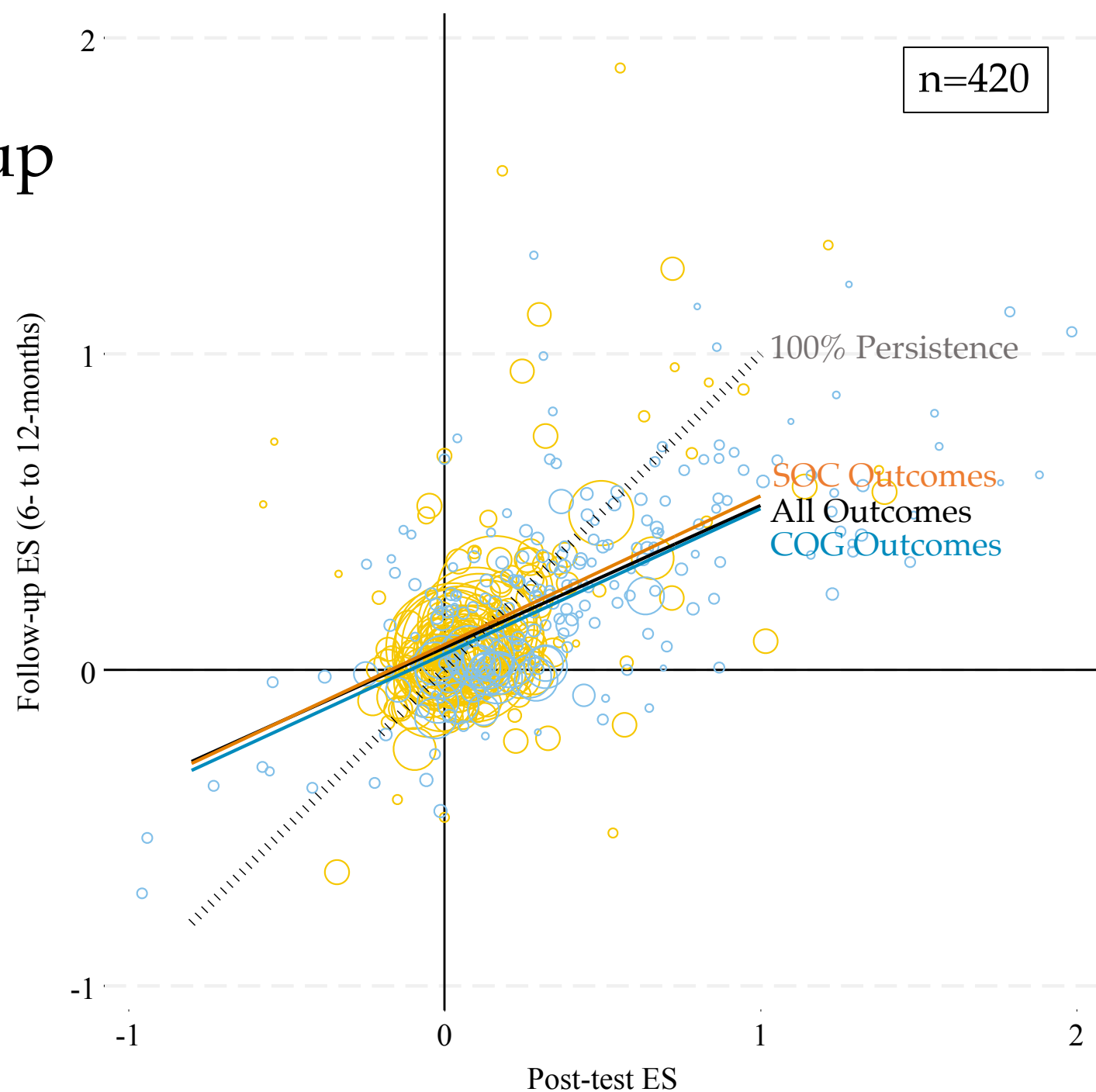
$$\beta_0 = 0.08$$

$$\beta_1 = 0.47 ***$$

Interaction is not statistically significant.

$$\tau_{intercept} = 0.11$$

$$\tau_{slope} = 0.24$$

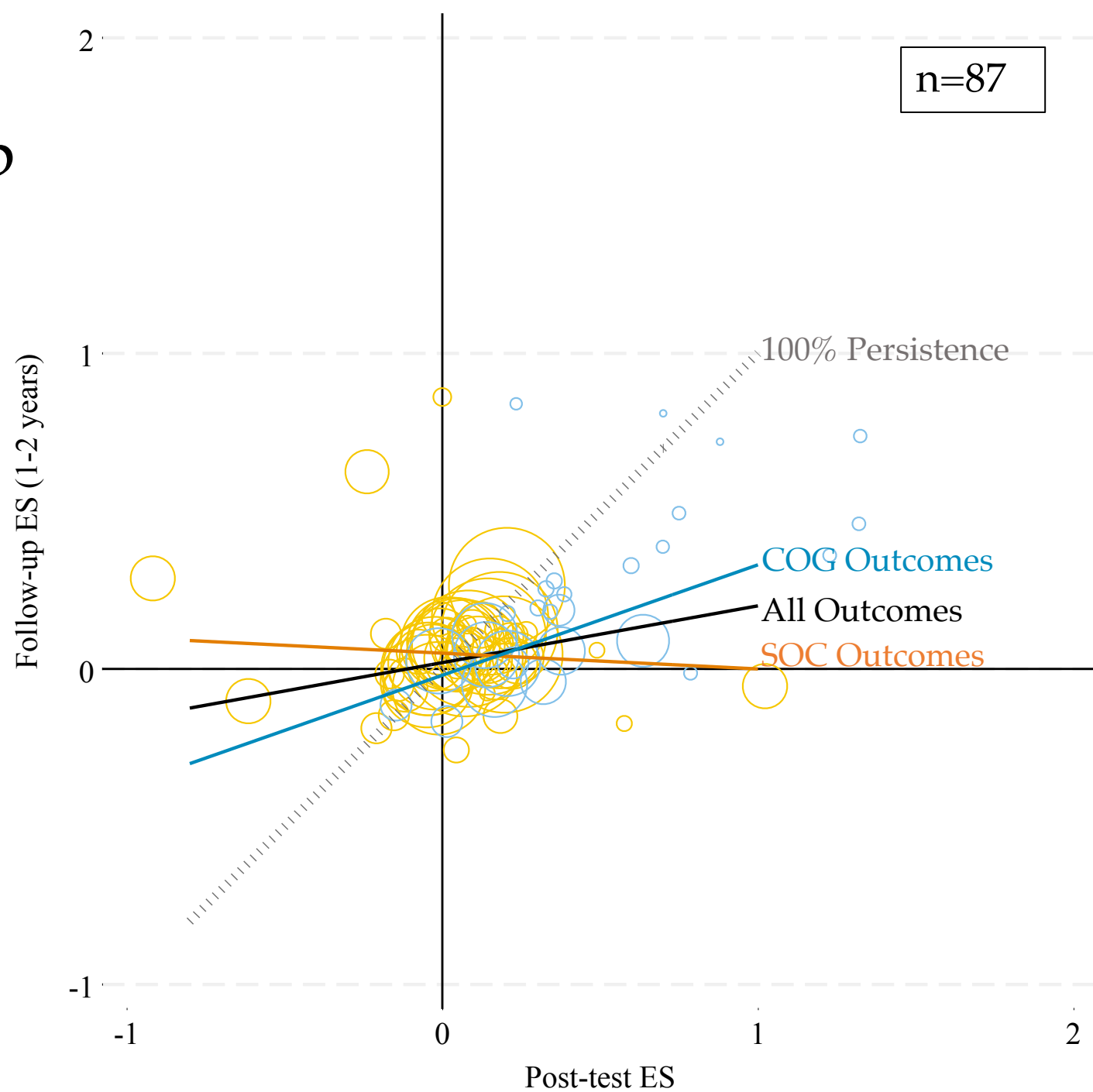


>1, up to 2 yrs Follow-up

All Outcomes:

$$\beta_0 = 0.02$$

$$\beta_1 = 0.18$$



>1, up to 2 yrs Follow-up

All Outcomes:

$$\beta_0 = 0.02$$

$$\beta_1 = 0.18$$

Cognitive Outcomes:

$$\beta_0 = -0.02$$

$$\beta_1 = 0.35 *$$

Social-Emotional Outcomes:

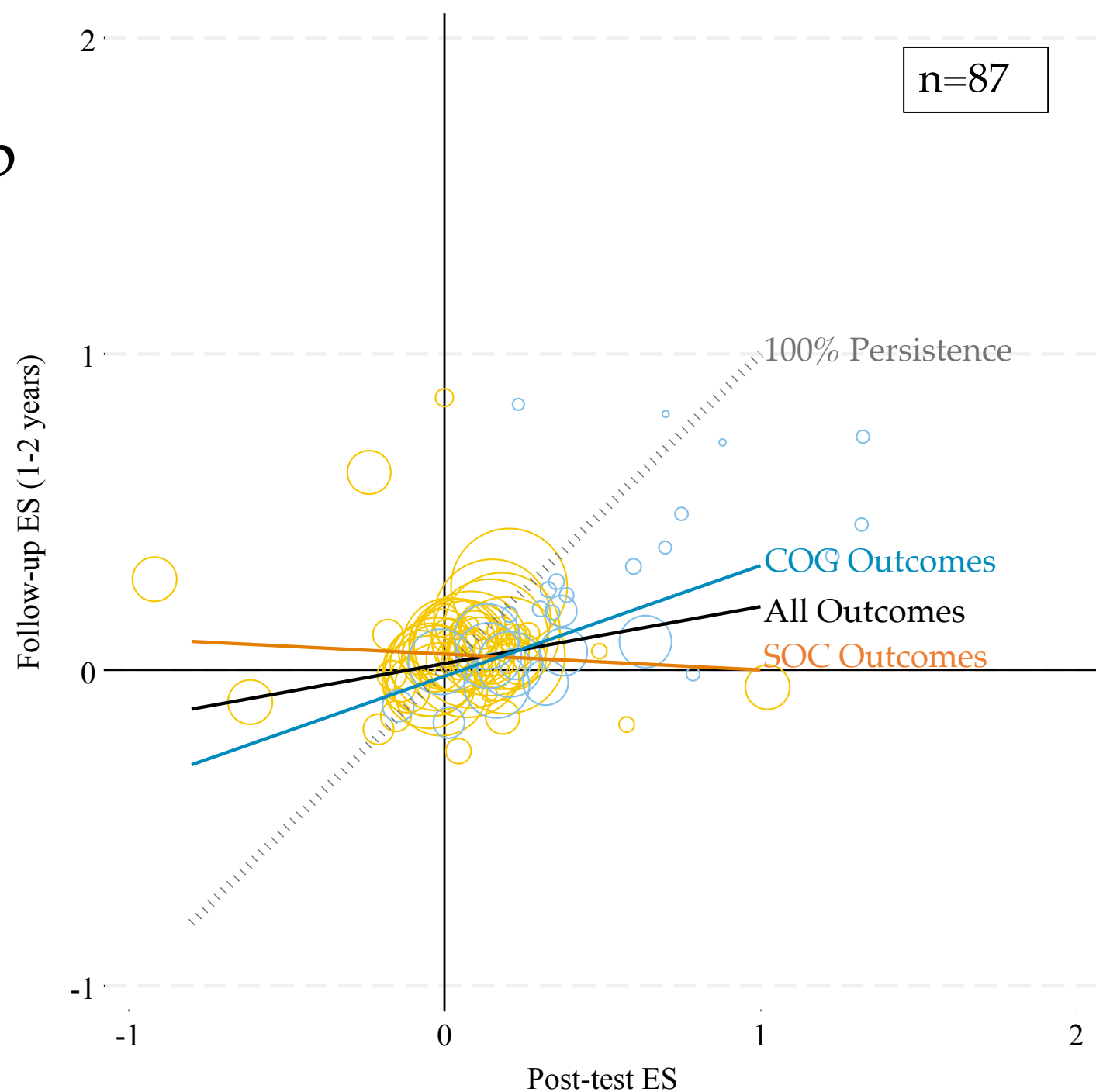
$$\beta_0 = 0.05$$

$$\beta_1 = -0.05$$

Interaction is not statistically significant.

$$\tau_{\text{intercept}} = 0.04$$

$$\tau_{\text{slope}} = 0.27$$



Robustness Checks

- Study fixed effects
- Correlated-and-Hierarchical Model
- Controls for intervention characteristics

Social-emotional and cognitive skill impacts fade

Social-emotional skills don't seem to be the single “missing link” that explains long-run emergent effects

Slope:

- Post-test impacts on child skills are meaningfully predictive of follow-up effects
- ~45% conditional persistence for both skills

Intercept:

- Small portion of follow-up effect unexplained by post-test effect
- Can be meaningful if post-test effects are small

Do theoretically-salient intervention features predict greater persistence?

ECE intervention

Intervention length

Parent targeted

Broad set of skills targeted

Intervention occurred under “worse” counterfactual conditions

Do interventions with certain features persist
more?

$$ES_{fsi} = \beta_{0s} + \beta_1 ES_{psi} + \beta_2 FEATURE_{si} + \beta_3 ES * FEATURE_{si} + \varepsilon_{si}$$

$$\beta_{0s} = \gamma_{00} + \tau_{0s}$$

$$\beta_{1s} = \gamma_{10} + \tau_{1s}$$

6- to 12-month follow-up

	Split Sample Models		Full Interaction Models		
	Moderator=1 (1)	Moderator=0 (2)	p-value (3)	Total Obs (4)	
ECE	<i>ECE</i>	<i>Non-ECE</i>			
Constant	0.11 (0.03)**	0.03 (0.02)	0.058	+	413
Post-test Effect Size	0.33 (0.07)***	0.56 (0.07)***	0.031	*	
Observations	178	235			
TX Duration	<i>Long</i>	<i>Short</i>			
Constant	0.03 (0.00)**	0.07 (0.05)	0.969		341
Post-test Effect Size	0.47 (0.08)***	0.53 (0.07)***	0.320		
Observations	178	163			
Parent Involvement	<i>Parents</i>	<i>No Parents</i>			
Constant	0.03 (0.00)*	0.06 (0.03)*	0.911		419
Post-test Effect Size	0.37 (0.11)*	0.48 (0.06)***	0.316		
Observations	78	341			
TX year > 2000	<i>> 2000</i>	<i><= 2000</i>			
Constant	0.08 (0.03)*	0.04 (0.03)	0.358		419
Post-test Effect Size	0.47 (0.09)***	0.44 (0.05)***	0.725		
Observations	217	202			
Broad TX	<i>Broad</i>	<i>Narrow</i>			
Constant	0.05 (0.03)	0.07 (0.02)**	0.786		419
Post-test Effect Size	0.15 (0.09)	0.49 (0.06)***	0.049	*	
Observations	43	376			

Greater conditional persistence for **non-ECE studies**

Greater conditional persistence for **narrow interventions**

1- to 2- year follow-up

	Split Sample Models		Full Interaction Models		
	Moderator=1 (1)	Moderator=0 (2)	p-value (3)	Total Obs (4)	
ECE	<i>ECE</i>	<i>Non-ECE</i>			
Constant	-0.02 (0.02)	0.08 (0.04)	0.073	84	+
Post-test Effect Size	0.32 (0.13)*	-0.06 (0.28)	0.189		
Observations	40	44			
TX Duration	<i>Long</i>	<i>Short</i>			
Constant	0.04 (0.01)+	0.03 (0.06)	0.788	75	
Post-test Effect Size	0.06 (0.1)	0.25 (0.28)	0.476		
Observations	41	34			
Parent Involvement	<i>Parents</i>	<i>No Parents</i>			
Constant	0.05 (0.01)*	0 (0.03)	0.508	90	
Post-test Effect Size	0.25 (0.13)	0.13 (0.17)	0.408		
Observations	50	40			
TX year > 2000	<i>> 2000</i>	<i><= 2000</i>			
Constant	0.01 (0.02)	0.17 (0.08)+	0.188	90	
Post-test Effect Size	0.15 (0.13)	-0.11 (0.3)	0.597		
Observations	60	30			
Broad TX	<i>Broad</i>	<i>Narrow</i>			
Constant	0.04 (0)*	0.03 (0.03)	0.641	90	
Post-test Effect Size	0.15 (0.14)	0.14 (0.16)	0.847		
Observations	31	59			

Flips

No difference

Intervention Characteristics

- We tested other moderators too– no differences in persistence
- It does not appear that there is one broad class of interventions for which fadeout is not an issue

Evidence of fadeout across a broad range of RCTs.

No evidence that fadeout looks substantially different for soc/cog skills*

* at least within the first few years after post-test

Limitations

- Limited power to precisely estimate trajectories of fadeout for follow-ups greater than 2 years after post-test
 - Issues in funding long-term data collection (Watts et al., 2019)
- Issues related to selection bias that we can't overcome: selection into follow-up

Implications

- It does not appear that we can confidently predict *a priori* which kinds of skills or interventions will show greater persistence
- Two things may be true: fadeout on insular skills and long-run emergent effects on outcomes in adulthood
- But, if so, how?
 - Dynamics are likely much more complicated than what is expected based on the '*plausible picture of development*' where boosted skills change developmental trajectories
 - Rippling effects through more complex skill networks

Future Directions

- If sustained skill impacts are not a necessary condition for long-run effects, is there any information that can be used to forecast which interventions are likely to have long-run effects?
- Do post-test impacts forecast long-run intervention effects?

Thank you! Questions?

erh2169@tc.columbia.edu

