

Hoda S. Abdel Magid¹, Nathan Ryder², Brandon Gavett³, Katrina Kezios⁴, Adina Al Hazzouri⁵, Sarah Forrester⁶, Jingxuan Wang⁷, Xiaqing Jiang⁸, Alden L. Gross⁹, Emma Nichols¹

¹University of Southern California, Los Angeles, CA, USA, ²University of Illinois Chicago, Chicago, IL, USA, ³University of California Davis, Davis, CA, USA ⁴Boston University, Boston, MA, USA,

⁵Columbia University, New York City, NY, USA, ⁶UMass Chan Medical School, Worcester, MA, USA, ⁷Harvard University, Boston, MA, USA,

⁸University of California San Francisco, San Francisco, CA, USA, ⁹Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

Keck School of Medicine of USC

2026.12581852

Background

- ❖ Cognitive studies often use different cognitive tests, scales, and scoring methods, making exposure–cognition associations difficult to compare across cohorts.
- ❖ **Item-Response Theory (IRT)** co-calibration can place measures on a shared metric, but it requires item-level data, overlapping anchor items, and invariant linking items with no differential item functioning — conditions often difficult to meet in multi-cohort dementia research.
- ❖ **Cogxwalkr** is a newer effect-size crosswalking approach that translates associations from one outcome scale to another using a third dataset where both outcomes are observed. This may enable harmonization of published or restricted effect estimates without item-level data.
- ❖ However, its performance relative to IRT co-calibration remains unclear, motivating this simulation study.

Study Aim& Hypothesis

- ❖ We evaluated whether cogxwalkr effect-size crosswalking can recover a target education–cognition association with performance comparable to IRT co-calibration under realistic cognitive measurement conditions.
- ❖ **Hypothesis:** cogxwalkr shows comparable performance to co-calibration when both cognitive outcomes were observed in a third dataset, while performance may vary under weaker harmonization conditions.

Data and Simulation Framework

- ❖ We simulated cognitive data using memory-test item parameters estimated from the **Alzheimer's Disease Neuroimaging Initiative**.
- ❖ Each simulation included two main groups with different cognitive measures and a third group in which both measures were observed for developing the cogxwalkr conversion..
- ❖ **Education**, defined as college+ vs. lower education, was used as the example exposure, with a known true effect on latent cognition of **d = 0.2 or d = 0.4**.
- ❖ Cognitive item responses were generated using ADNI-derived memory-test parameters, with weak or strong differential item functioning (DIF) added in selected scenarios to test performance under common harmonization challenges.

Simulation Scenarios

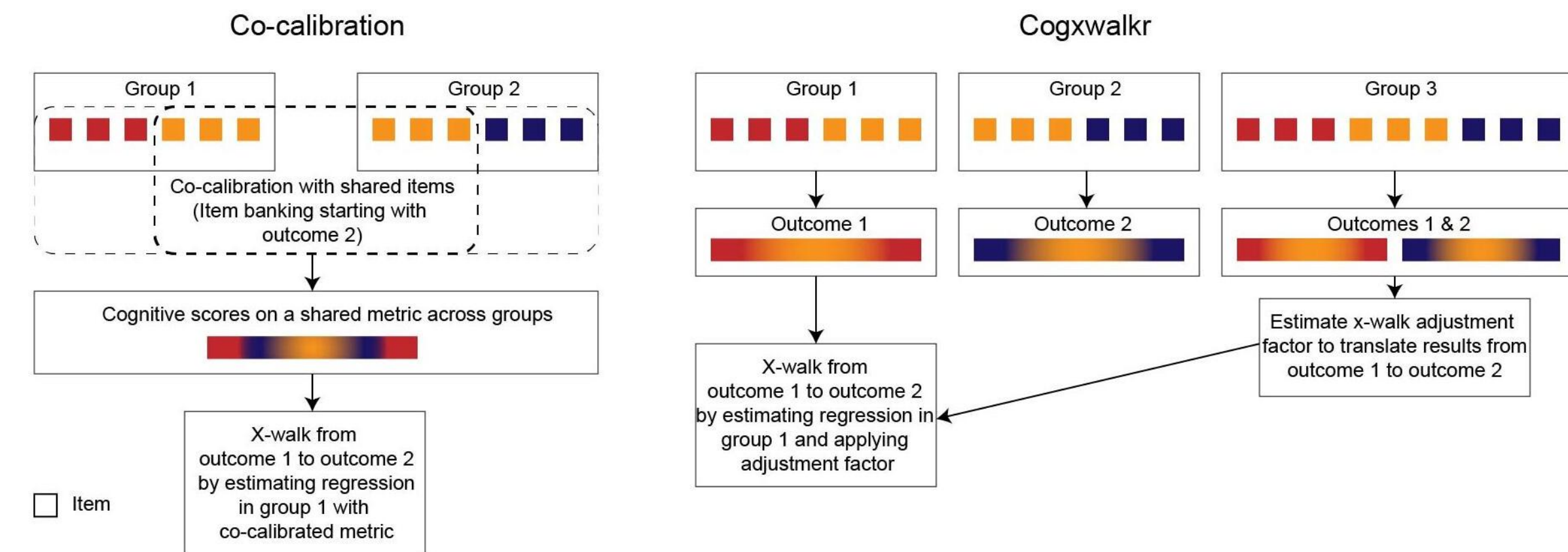
- ❖ We evaluated performance across four harmonization conditions that varied anchor-item quality and differential item functioning. These scenarios represent increasingly challenging conditions for linking cognitive measures across cohorts.
- ❖ Anchor items are shared cognitive test items used to link different measures onto a common scale.

Scenario	Measurement challenge
1. No DIF + strong anchors	Ideal linking condition
2. No DIF + weak anchors	Poorer anchor quality
3. Weak DIF + strong anchors	Mild item non-invariance
4. Strong DIF + strong anchors	More severe item non-invariance

Harmonization Approaches Compared

IRT Co-calibration	Cogxwalkr Crosswalk
Used shared anchor items and item banking to estimate harmonized cognitive scores on a common metric across groups	Estimated an adjustment factor in the third group, where both cognitive measures were observed, and applied this factor to translate regression coefficients from one cognitive outcome scale to another
Approximates a psychometric “gold standard,” but requires item-level data and valid linking items.	Can be applied to effect estimates and does not require item-level data from all studies.

Figure 1. Comparison of the co-calibration and cogxwalkr approaches to crosswalking cognitive outcomes.



Statistical Analyses

For each scenario, we estimated the association between education and cognition using three outcomes:

- ❖ The target cognitive outcome on the desired metric
- ❖ The IRT co-calibrated cognitive score
- ❖ The cogxwalkr-crosswalked effect estimate

We compared each method's ability to recover the target education–cognition coefficient across simulation replications. Performance was summarized using bias, mean bias, and variability across replications.

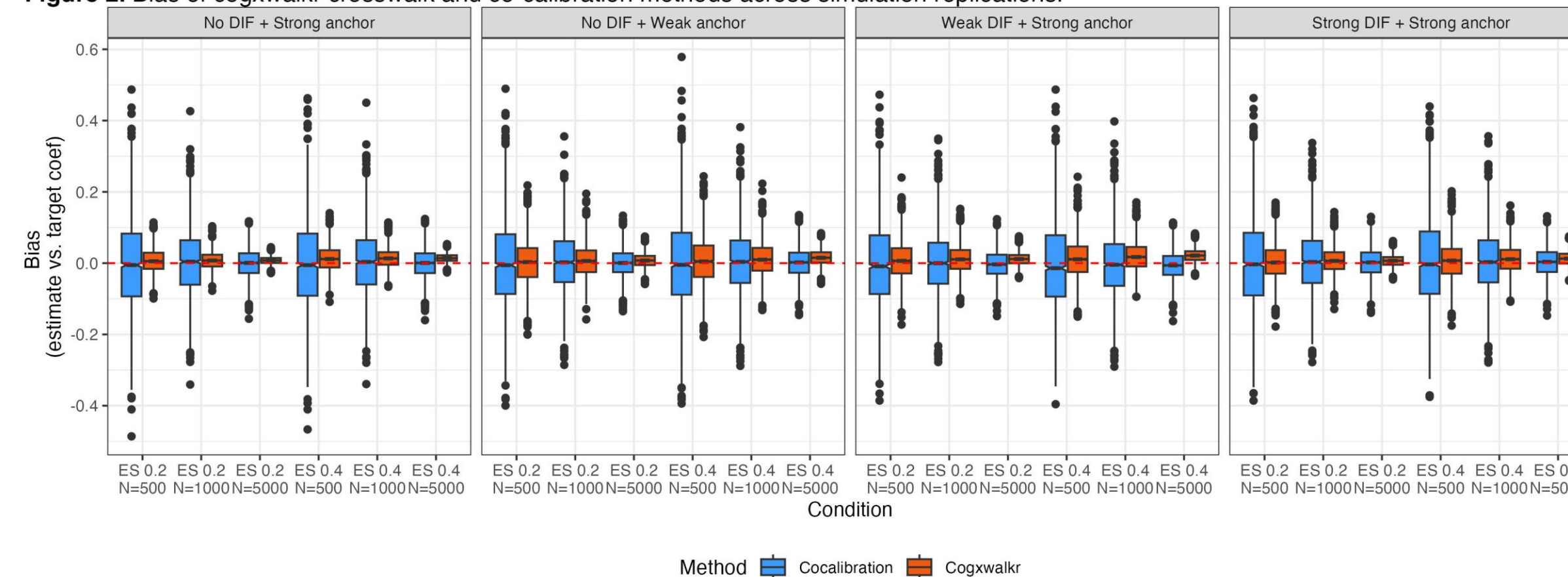
Results

Result 1: Co-calibration showed greater variability

Across simulation replications, co-calibration estimates showed greater variability than cogxwalkr estimates, with wider distributions of bias and more extreme values, particularly at smaller sample sizes. This pattern was generally consistent across DIF conditions, anchor strength, effect size, and sample size (Figure 2).

- ❖ Both methods were generally centered near zero bias
- ❖ Co-calibration (blue boxes) showed wider variability
- ❖ Cogxwalkr was more stable across replications
- ❖ Larger sample sizes improved performance

Figure 2. Bias of cogxwalkr crosswalk and co-calibration methods across simulation replications.



Note: ES = effect size.

Results

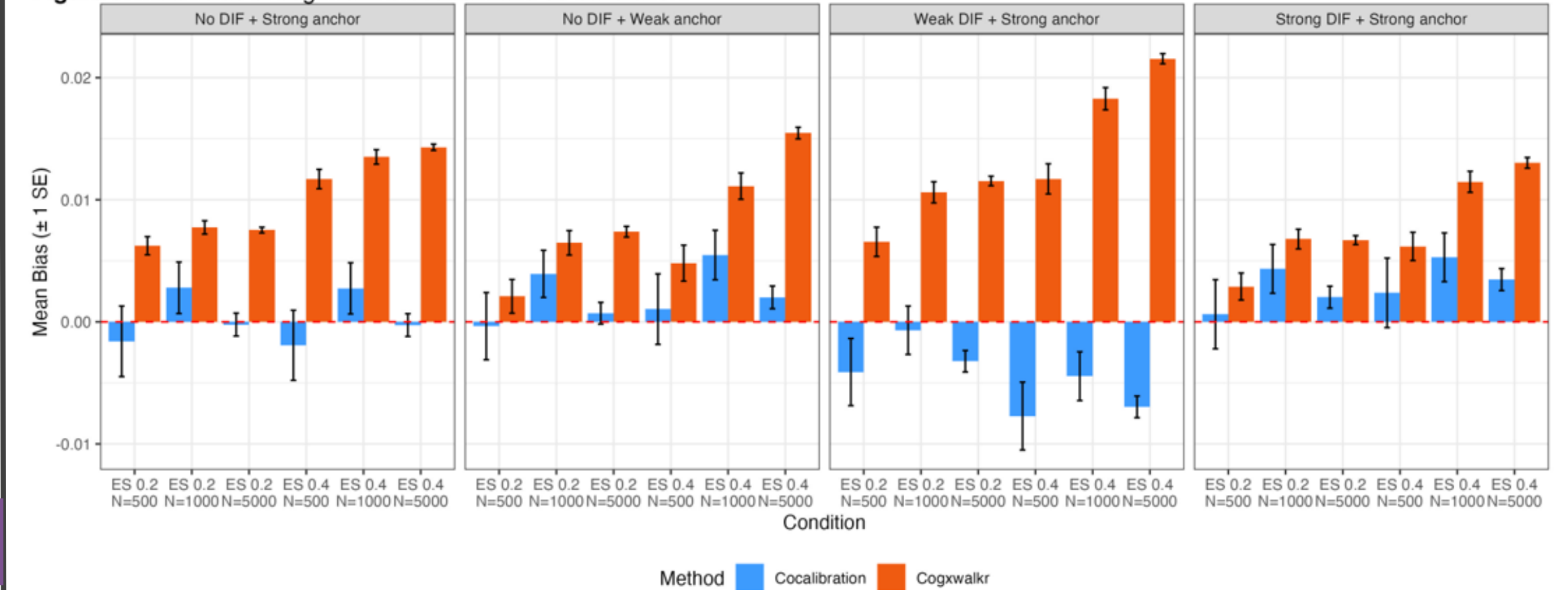
Result 2: Precision improved with larger sample size

- ❖ For both approaches, bias distributions narrowed as sample size increased, with the most stable performance observed at N = 5000. However, co-calibration remained more variable than cogxwalkr across most conditions.

Result 3: Mean bias was small but differed by method and scenario

- ❖ Across simulation conditions, both cogxwalkr and co-calibration generally recovered the target education–cognition association with small mean bias (Figure 3).
- ❖ Co-calibration estimates were often closer to zero on average, although negative bias appeared under the weak DIF scenario.
- ❖ Cogxwalkr showed modest positive bias across most conditions, with larger bias when the true education effect was stronger and under weak DIF.

Figure 3. Mean bias of cogxwalkr crosswalk and co-calibration.



Conclusions

- ❖ Cogxwalkr effect-size crosswalking performed similarly to IRT co-calibration in recovering target education–cognition associations across simulated cognitive measurement scenarios.
- ❖ This suggests that cogxwalkr may offer a practical tool for harmonizing cognitive effect estimates when item-level data or overlapping anchor items are unavailable.
- ❖ Future work should test performance under less idealized conditions and in real-world multi-cohort dementia studies.

Reference

- ❖ Ackley SF, et al. Methods to crosswalk between cognitive test scores using data from the Alzheimer's Disease Neuroimaging Cohort. *Alzheimer's Dement.* 2025;21:e14597.

Contact: Hoda Abdel Magid;
Email: hmagid.usc.edu