

Causally Invalid Methods Make Misleading Arguments for Surrogate Outcomes

Michael D. Flanders, Michelle Caunca, Renaud La Joie, Lon S. Schneider, Sarah F. Ackley



IPD from RCTs is our best bet to identify useful surrogate outcomes – but requires rigorous causal analysis to avoid misleading regulators, patients and providers

BACKGROUND

Individual-level data (IPD) from RCTs present a rare opportunity to rigorously assess candidate surrogate outcomes

A recently reported analytic approach¹, applied to IPD from TRAILBLAZER-ALZ 2, found a strong relationship between post-treatment amyloid and cognitive change, and argued in support of amyloid as a surrogate endpoint

OBJECTIVE

Characterize the newly reported merge-and-aggregate (quantile aggregation) method

Test: does it provide reliable evidence in support of amyloid or other candidate surrogate outcomes?

APPROACH

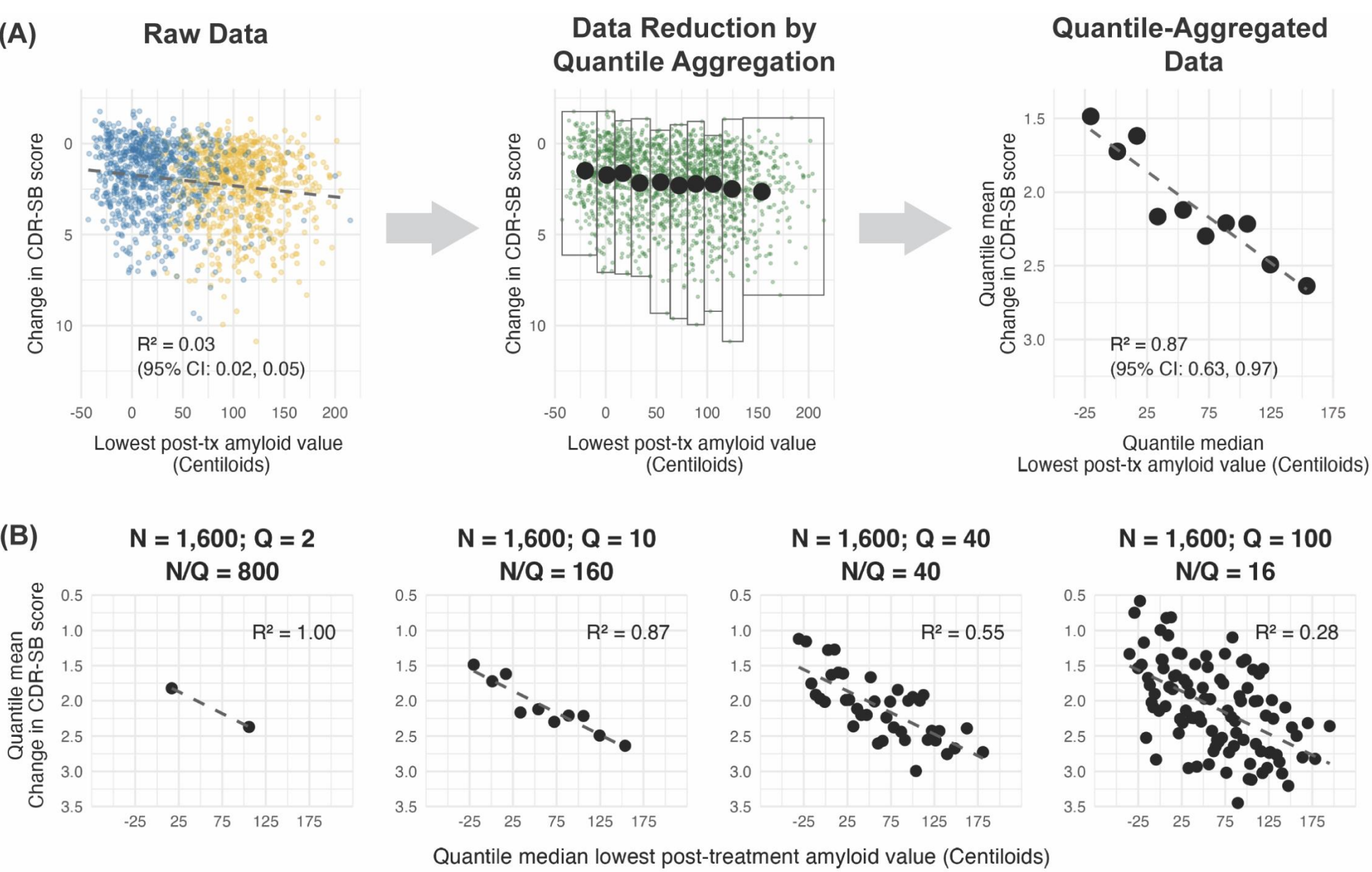
Reconstruct the recent merge-and-aggregate procedure¹:

- **Pool** trial arms (treatment + control)
- **Rank** all participants by post-treatment amyloid
- **Partition** into quantiles
- **Summarize** each by median amyloid and mean cognitive change
- **Regress** the quantile summaries to estimate the amyloid–cognition association (R^2)

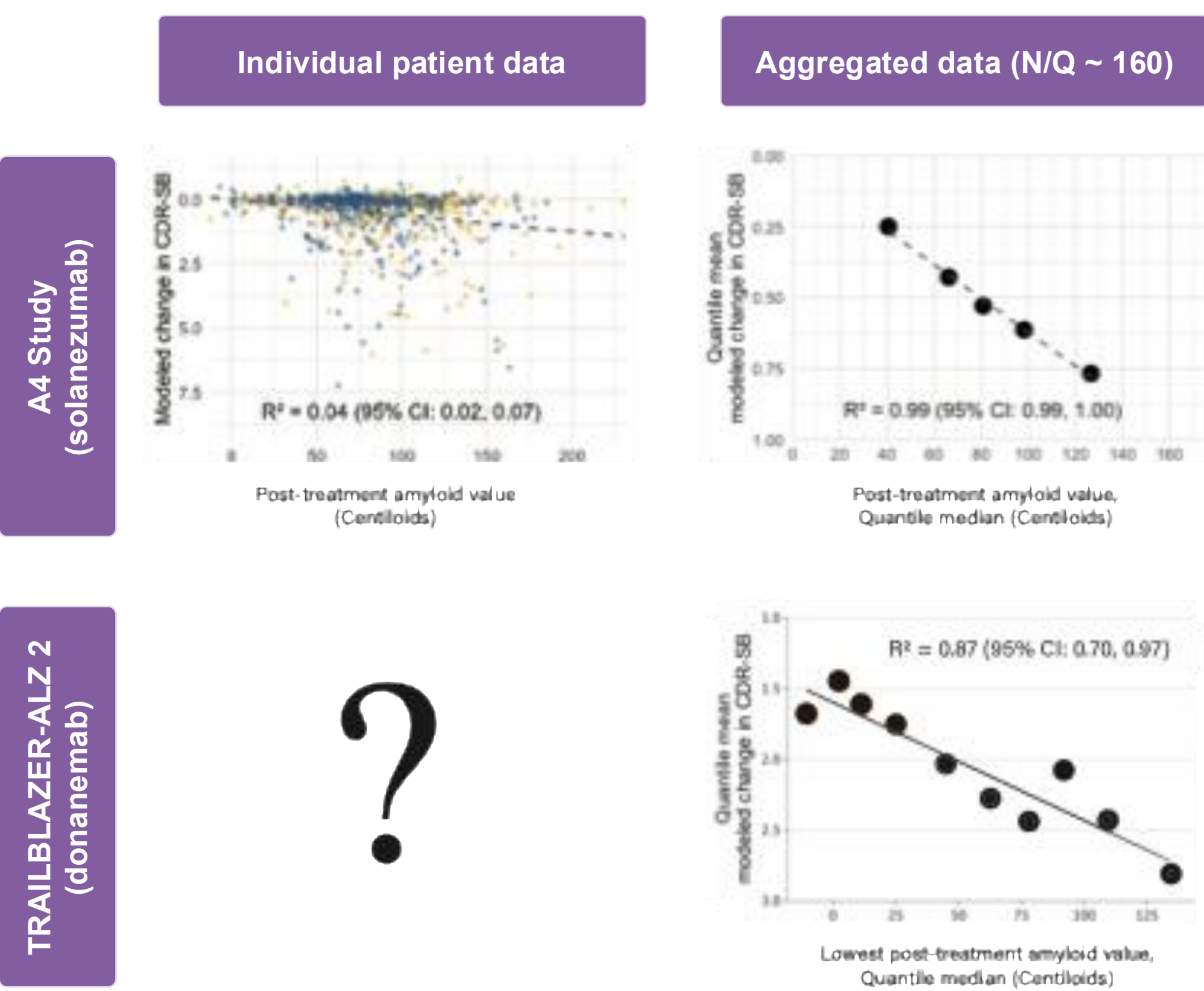
Test the approach to see if it yields interpretable evidence in:

- **Simulated trial data:** 1,600 participants, calibrated to recent trials, with only a weak amyloid–cognition relationship built in by design — so we know the ground truth
- **RCT data:** individual-level data from the A4 trial of solanezumab, which showed no cognitive benefit

(1) DEMONSTRATE AND TEST IN SIMULATED TRIAL DATA



(2) HOW DOES IT PERFORM IN DATA FROM REAL ANTI-AMYLOID TRIALS?



Discarding treatment allocation data limits any causal interpretation. With merge-and-aggregate, a causally clear RCT becomes an observational study of unclear meaning [1A]

Grouping patients by values of a post-treatment variable (amyloid) makes this approach a variety of ecological study, subject to confounding [1A, 2]

Reported R^2 values do not reflect the strength of an amyloid-cognition relationship in the underlying IPD; the variance in cognition apparently explained by amyloid is a statistical artifact of the aggregation procedure itself [B]

This approach produces a similar correlation in the A4 Study– despite clear lack of treatment benefit with solanezumab

The method generates the same relationship from both a successful drug (donanemab) and a clearly unsuccessful drug (solanezumab), so we must conclude that the method can be causally misleading – this analytic approach does not require the drug to have any effect to argue that drug effects are explained by a candidate surrogate

IPD from RCTs remain our best source of hard evidence for finding reliable surrogate outcomes. We must simply use existing causally-interpretable methods that preserve randomization if we aim to identify clinically reliable surrogates. ^{2, 3}

References:

¹Lu M et al. *JAMA Neurol.* 2025;82(12):1251–1256. doi:10.1001/jamaneurol.2025.3869

²Ackley SF et al. *Alzheimers Dement (N Y).* 2026;12(1):e70205. Published 2026 Feb 14. doi:10.1002/trc2.70205

³Flanders MD et al. *JAMA Neurol.* Published online May 18, 2026. doi:10.1001/jamaneurol.2026.1240

This work was supported by NIH/NIA: R00AG073454

Data Source: A4 Study & A4/LEARN Data Package: Additional information and data is available at <https://www.a4studydata.org/>. A4/LEARN was funded by the National Institute on Aging/NIH, Eli Lilly and Co., Alzheimer's Association, GHR Foundation, additional philanthropic donors.

Scan the QR code for our recent publication in *JAMA Neurology*, containing more details and contact info