

Assessing Within-Site Participant Similarity as a Marker of Recruitment Bias and Data Integrity in Alzheimer's Disease Clinical Trials

Kott, Alan, MUDr.¹; Wang, Xingmei, MS²; Miller, David S., MD, MA²

¹Signant Health, Prague, Czech Republic; ²Signant Health, Blue Bell, PA USA

INTRODCUTION

- Study participants recruited within a given clinical site are generally expected to be more similar than those recruited across different sites.
- While excessive within-site similarity may reflect local demographic or procedural consistency, it may also indicate site-specific bias in recruitment and assessment practices, or in rare instances, data tampering.
- This retrospective analysis evaluated whether participants in Alzheimer's disease (AD) trials show greater similarity within sites than between sites, and whether regional differences exist between the United States (US) and the rest of the world (ROW).

METHODS

- Screening and Baseline data were extracted from six AD clinical trials.
- Each participant was represented as a multidimensional vector comprising standardized inclusionary and efficacy variables.
- Pairwise Euclidean distances quantified similarity between participants.
- Comparisons were categorized as within-site or between-site.
- For each trial, permutation testing determined whether within-site participants were significantly more similar than between-site participants.
- In trials including both US and ROW data, Euclidean distances were also compared to examine regional

RESULTS

- The pooled dataset included 9,797 participants across six trials.
- In all trials, within-site participants were significantly closer in Euclidean space than participants from different sites ($p < 0.001$). (Figure 1)
- Among the five trials that included both US and ROW data, within-site distances were significantly smaller in US sites for three trials and numerically smaller (nonsignificant) in two.

DISCUSSION

- Participants from the same site were consistently more similar to one another than to participants from other sites, consistent with site-specific recruitment and assessment patterns.
- While moderate within-site similarity is expected, excessively high similarity may reflect rater bias, homogenous recruitment, or potential data quality issues.
- The greater within-site similarity observed in the US suggests regional variability that may arise from local recruitment practices or data integrity factors increasingly recognized in the US trial landscape.
- Euclidean distance limitations—its assumption of independence and equal scaling across variables—were acknowledged, as Mahalanobis distance could not be applied reliably due to small site-level sample sizes.
- Detecting unusual within-site similarity patterns may offer a valuable signal for monitoring data integrity and ensuring valid interpretation of multicenter trial outcomes.

FIGURE 1. DIFFERENCE IN EUCLIDEAN DISTANCES BETWEEN WITHIN- AND BETWEEN-SITE PARTICIPANTS.

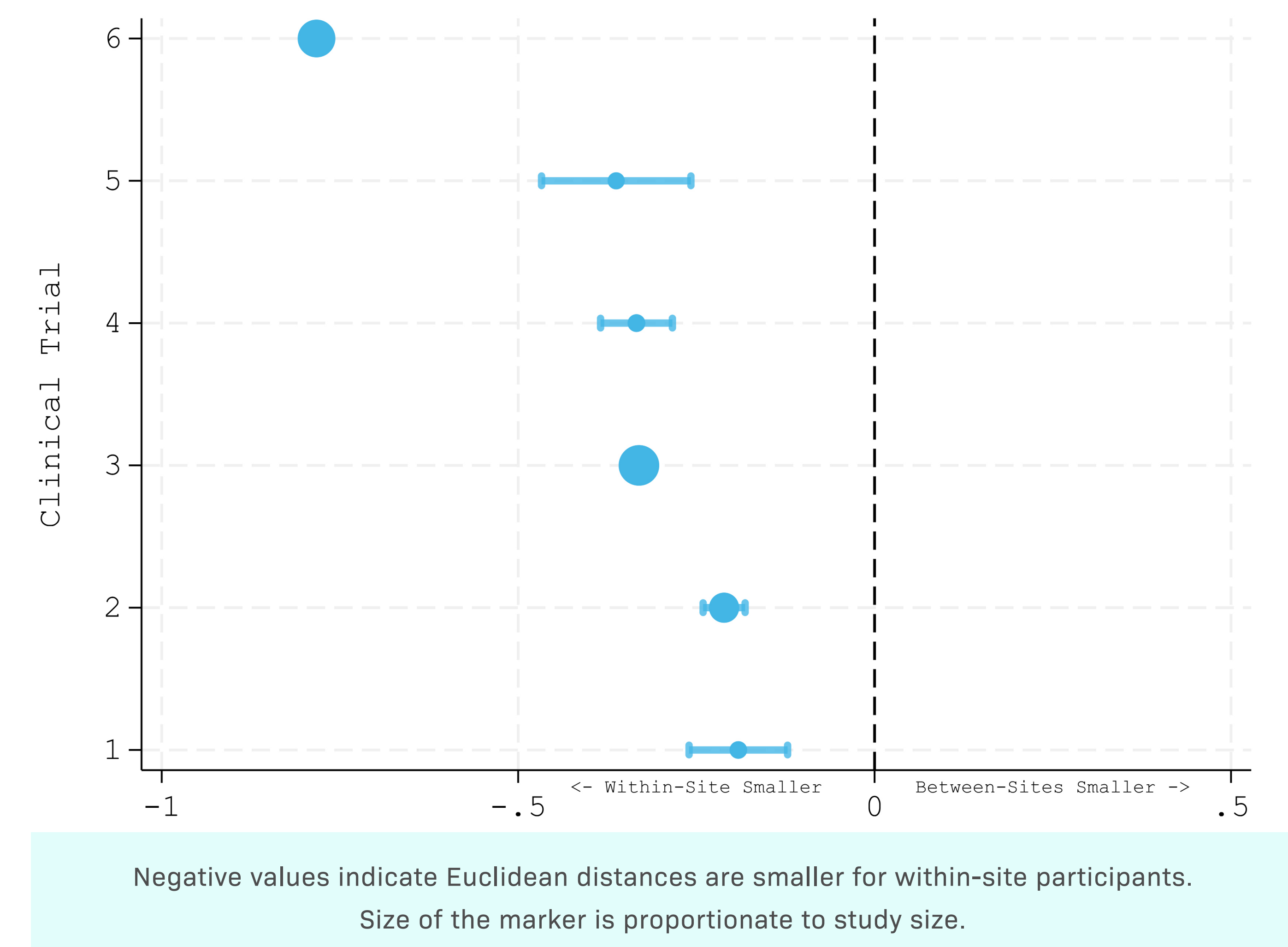


FIGURE 2. DIFFERENCE IN WITHIN-SITE EUCLIDEAN DISTANCES BETWEEN THE ROW AND US PARTICIPANTS.

