ISCTM Autumn Conference | September 16th - 17th, 2026 | Philadelphia, PA

Model-Based Detection of Rater Drift in CDR-SoB Using Centralized Monitoring Across Two Early Alzheimer's Disease Trials

Kott, Alan, MUDr.[1]; Wang, Xingmei, MS[2]; Miller, David S., MD, MA[2]
[1] Signant Health, Prague, Czech Republic | [2] Signant Health, Blue Bell, PA

WHAT IS THE METHODOLOGICAL ISSUE BEING ADDRESSED?

- In the current analysis we explore the utility of centralized statistical approaches to detect rater drift in CDR-SoB ratings.

INTRODUCTION

- Ensuring data quality in early Alzheimer's disease (AD) trials requires reliable detection of rater-related biases, including drift in scoring over time.
- Although rater drift is well recognized, scalable methods to detect it across large, multicenter programs remain limited.
- We evaluated centralized statistical approaches to identify temporal changes in rater behavior using Clinical Dementia Rating–Sum of Boxes (CDR-SoB) data from two large early AD trials.

METHODS

- CDR-SoB data from 772 raters across two independent studies were analyzed (N = 48,188 CDR assessments).
- For each study, linear mixed-effects models were fitted to characterize rater-specific scoring trajectories over time, allowing random intercepts and slopes for each rater while adjusting for subject-level disease progression.
- Rater-specific drift estimates were derived as Best Linear Unbiased Predictors (BLUPs) of the time slope.
- Positive slopes indicate increasing severity ratings over time, while negative slopes indicate decreasing severity.
- To identify atypical drift patterns, four complementary methods were applied: z-scores ($|z| > 2.5$), median absolute deviation ($|\text{slope} - \text{median}| > 2.5\times\text{MAD}$), Tukey's fences (outside 1.5×IQR), and statistical significance ($p < 0.05$).
- Raters were classified as exhibiting drift if at least three of the four methods were concordant.

RESULTS

- Using this framework, 12/408 (3%) raters in Study 1 and 9/364 (2%) raters in Study 2 were identified as exhibiting drift.
- The most pronounced cases showed slopes exceeding ±0.003 CDR-SoB points per day, corresponding to approximately ±1 point or more per year, a magnitude that could potentially impact study outcomes. (Figures 1 and 2)

CONCLUSION

- These findings demonstrate that rater drift in CDR-SoB can be systematically detected using scalable, model-based centralized monitoring.
- The observed patterns suggest that drift is not purely random but reflects consistent shifts in scoring behavior among a small subset of raters.
- Such drift may affect endpoint sensitivity and bias treatment effect estimates, particularly in early-stage populations where signal detection is challenging.
- Integrating routine drift detection into ongoing data monitoring workflows alongside data analytics and endpoint programs inclusive of audio/video reviews of individual assessments could support earlier identification of atypical raters and enable targeted retraining.
- Future work should assess the prospective implementation of these methods and their impact on trial integrity and statistical power.

FIGURE 1. BEST LINEAR UNBIASED PREDICTORS (BLUPS) OF THE TIME SLOPE FOR STUDY 1 WITH HIGHLIGHTED RATERS IDENTIFIED AS EXHIBITING DRIFT

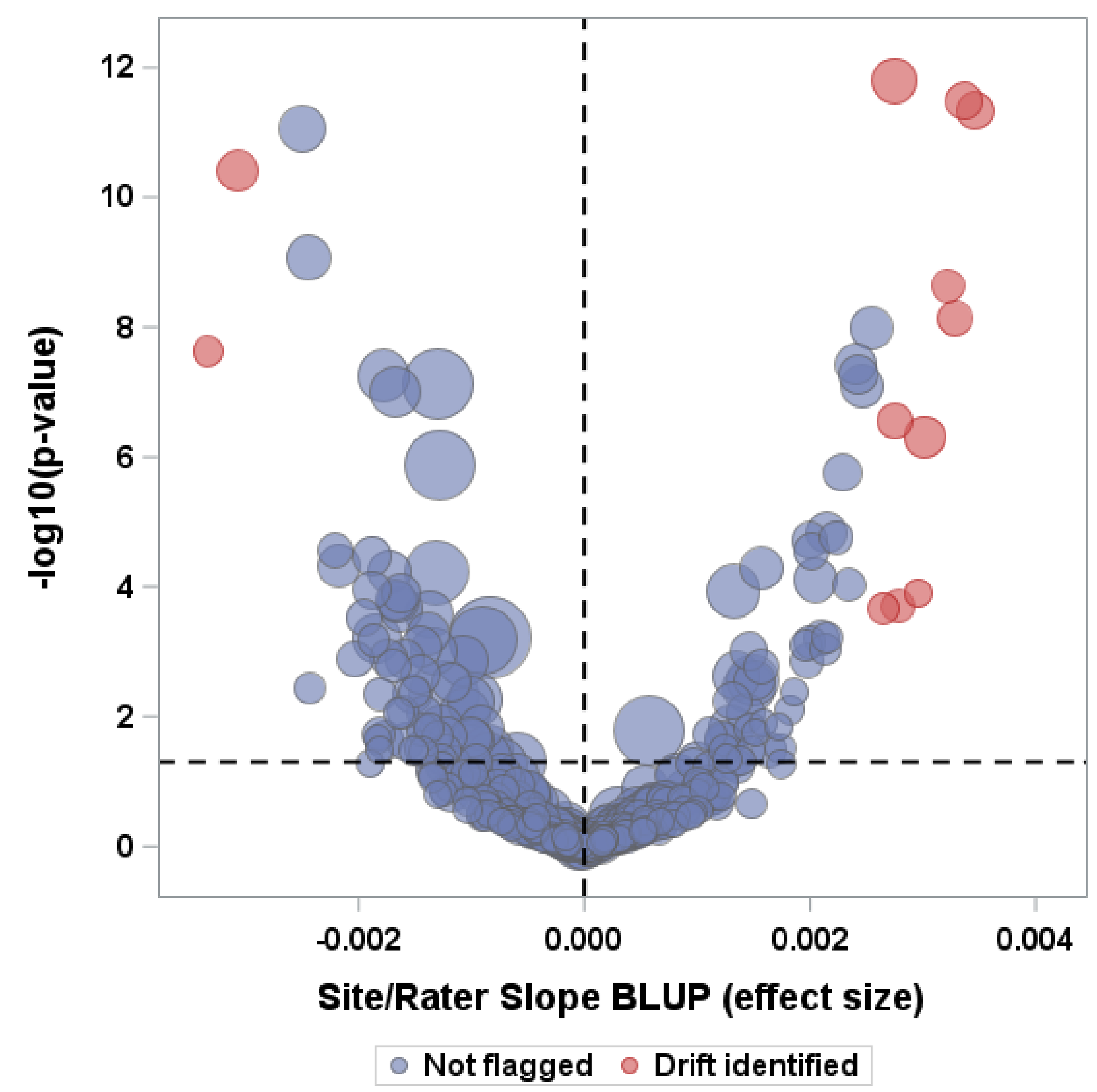
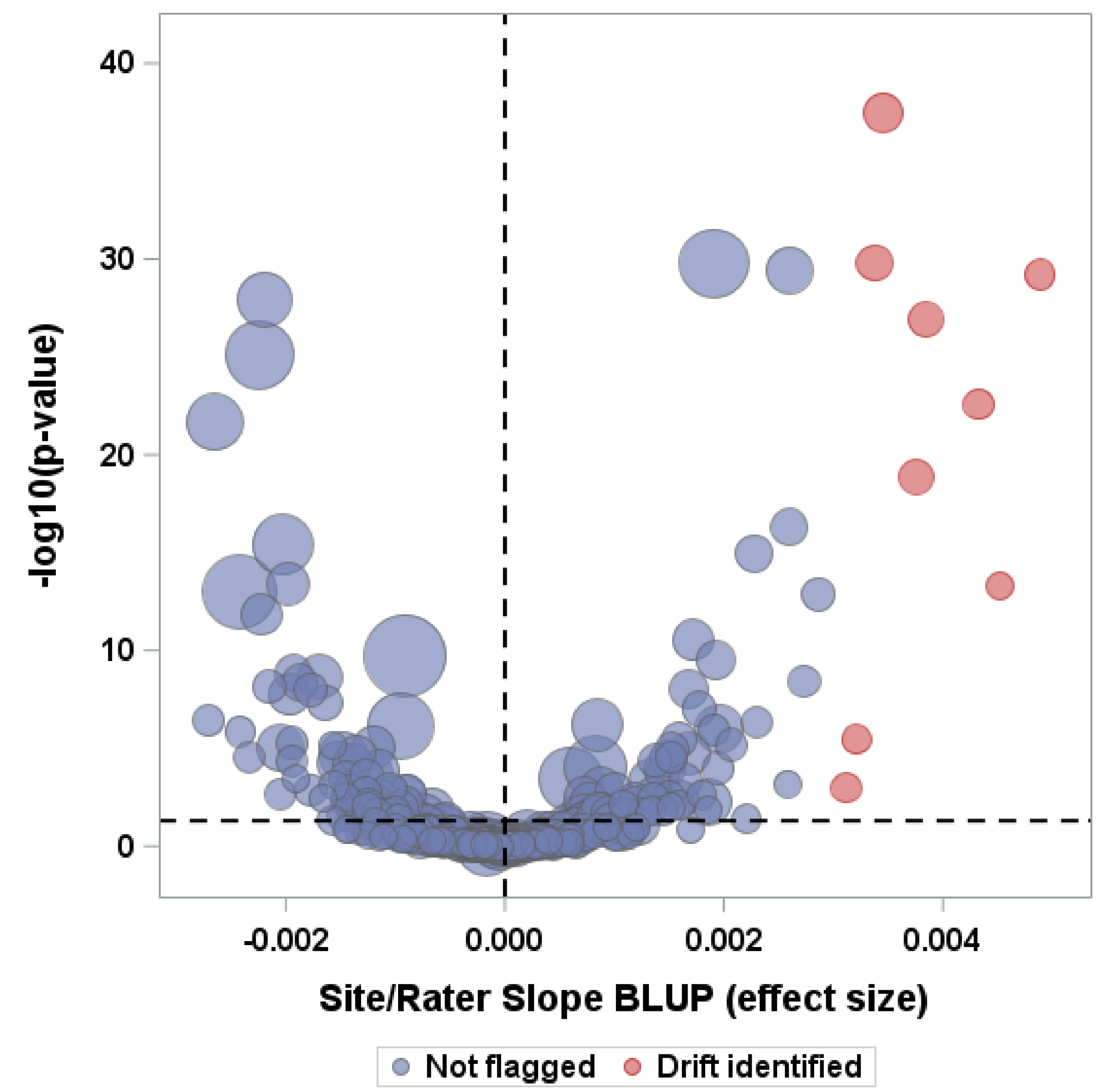

-log10(p-value)
12
10
8
6
4
2
0
-0.002
0.000
0.002
0.004
Site/Rater Slope BLUP (effect size)
Not flagged
Drift identified

FIGURE 2. BEST LINEAR UNBIASED PREDICTORS (BLUPS) OF THE TIME SLOPE FOR STUDY 2 WITH HIGHLIGHTED RATERS IDENTIFIED AS EXHIBITING DRIFT


-log10(p-value)
40
30
20
10
0
-0.002
0.000
0.002
0.004
Site/Rater Slope BLUP (effect size)
Not flagged
Drift identified

All authors are Signant Health employees.

SIGNANT HEALTH
THE EVIDENCE GENERATION COMPANY®