**CONCLUSION**

- Education level is not predictive of in-study rating performance. While experience with the scale appears to influence scoring errors, these associations were weak.

- These findings are consistent with previous literature showing that even the most experienced raters make scoring errors that affect clinical trial outcomes, and underscore the need for central oversight to ensure quality control in AD trials.  

**REFERENCES**


---

**BACKGROUND**

- Clinician-administered endpoints are typically conducted by multiple raters per site across a large number of global trial sites in late-phase Alzheimer’s disease (AD) trials. Raters and trial sites vary widely with regard to their level education as well as experience with the scales. Inconsistent clinical outcome assessments have a potential to lead to scoring variability and contribute to high placebo response rates and inconclusive results.

- While it is assumed individuals with higher levels of education and experience with a scale will rate more accurately and reliably, there is currently little data to support this notion. In the present study, we examined scoring errors as a function of rater education and experience using widely administered scales in AD trials.

**METHODS**

- Three clinical endpoints (ADAS-Cog, CDR, and MMSE) from ongoing, double-blind, placebo-controlled, multinational trials in prodromal AD were combined for analysis.

- A total of 16,267 assessments were completed by 813 raters across 232 sites.

- Raters reported information on highest degree obtained and years of experience with the scales as part of their pre-qualification survey.

- Scoring errors in each scale were identified via review of audio recordings and worksheets by a cohort of expert calibrated clinicians.

**RESULTS**

- Table 1 shows the percentages of reviews with discrepancies as a function of education level. Rater education levels included Associate’s, Bachelor’s, Master’s, M.D., or Ph.D. The percentages of reviews with one or two or more errors did not differ based on education level for all three scales, suggesting that education level did not influence scoring performance.

- Raters with higher scale experience appeared to make fewer scoring errors (Figure 1). There was, however, a high degree of variability in the years of experience raters had with the scales (ranges 0-30), and the correlations between scale experience and scoring errors were weak (ADAS-Cog, r = 0.162, CDR, r = 0.088 and MMSE, r = 0.199).

- There were no clear advantages observed based on raters’ experience with the AD population (Table 2).

**REFERENCES**