Lanabecestat: Central Monitoring of Rater Performance and Error Characteristics of Efficacy Assessments in the AMARANTH Study

Alette M. Wessels, PhD[1], Lisle R. Kingery, PhD[2], Edward I. Bartolic, PhD[2], Laura E. Nickell, BBus[1], Jamie A. Mullen, MD[3], John R. Sims, MD[1]

[1]Eli Lilly and Company, Indianapolis, IN, USA; [2]Cogstate, New Haven, USA; [3]AstraZeneca, Gaithersburg, MD, USA

BACKGROUND

- Lanabecestat is a brain-permeable, orally administered inhibitor of human beta-site amyloid (Aβ) precursor protein-cleaving enzyme 1 (BACE1)[1]
- Lanabecestat reduces Aβ production and was under investigation as a potential disease-modifying treatment for AD
- The Phase 2/3 AMARANTH study evaluated the potential effects of lanabecestat as a disease-modifying treatment for Early AD by reducing Aβ production via BACE1 inhibition (NCT02245737)
- Rater errors are common on outcome scales used in AD clinical trials; central monitoring using form and audio reviews has been shown to identify and potentially reduce rater errors[2-4]

Abbreviations: AD, Alzheimer's disease; BACE1, beta-site amyloid precursor protein-cleaving enzyme 1; MCI, mild cognitive impairment

OBJECTIVES

- To describe the central monitoring (CM) methodology for the Phase 2/3 AMARANTH study (NCT02245737)
- To summarize key findings regarding rater performance on the efficacy assessments

METHODS

- **Efficacy Assessments**
  - Primary Efficacy Assessment: The 13-item version of the Alzheimer's Disease Assessment Scale- Cognitive subscale (ADAS-Cog$_{13}$)
  - Secondary Efficacy Assessments: Alzheimer's Disease Cooperative Study—Activities of Daily Living Inventory (ADCS-ADL), the Functional Activities Questionnaire (FAQ), the Clinical Dementia Rating Scale (CDR) and the Mini-Mental State Examination (MMSE)
- Raters were qualified to rate in the study based on educational background, clinical and scale administration experience
- Training included voiced-over didactic presentations for all scales. Additionally, ADAS-Cog certification included a video-scoring exercise and an audio-recorded practice administration with a mock subject that was reviewed by a local expert neuropsychologist
- CDR raters had to successfully complete and receive certification through Washington University's online CDR training program

Central Monitoring (CM) Methodology

- All scales were administered using paper worksheets.
- Rater performance was centrally monitored in a standardized manner using a Scale Review Form (SRF) created for each scale. Local expert reviewers relied on the SRF to promote detailed and standardized reviews. The SRF classifies rater errors in the following categories:
  - **Administration error** – deviation from a standardized test administration or interviewing procedure
  - **Recording error** – recorded subject response on the test form for an item did not accurately reflect the subject's actual spoken response or behavior, was not captured in the manner specified during training, or a required field was left blank on the form
  - **Scoring error** – item or summary score associated with a subject's obtained response was incorrect
- It is important to note that the SRF was structured so that experts assessed the three error types as independently as possible. For example, if an error in administration occurred for a particular item or subtest, there was no assumption that the resulting score for that item or subtest was incorrect as a result. The accuracy of scoring was determined by what was observable by the expert, based on the manner in which the item or subtest was actually administered by the rater
- CM sampling strategy included:
  - **Form review**– Review of scale worksheets only
  - **Audio review**– Review of scale worksheets along with audio recording of the scale administration
- Follow-up with raters by email and/or teleconference occurred if concerns about administration and/or scoring practices were observed in central monitoring

RESULTS

- Data available for analyses were from 1463 raters from 15 countries (US, Canada, Europe, Asia, Australia)
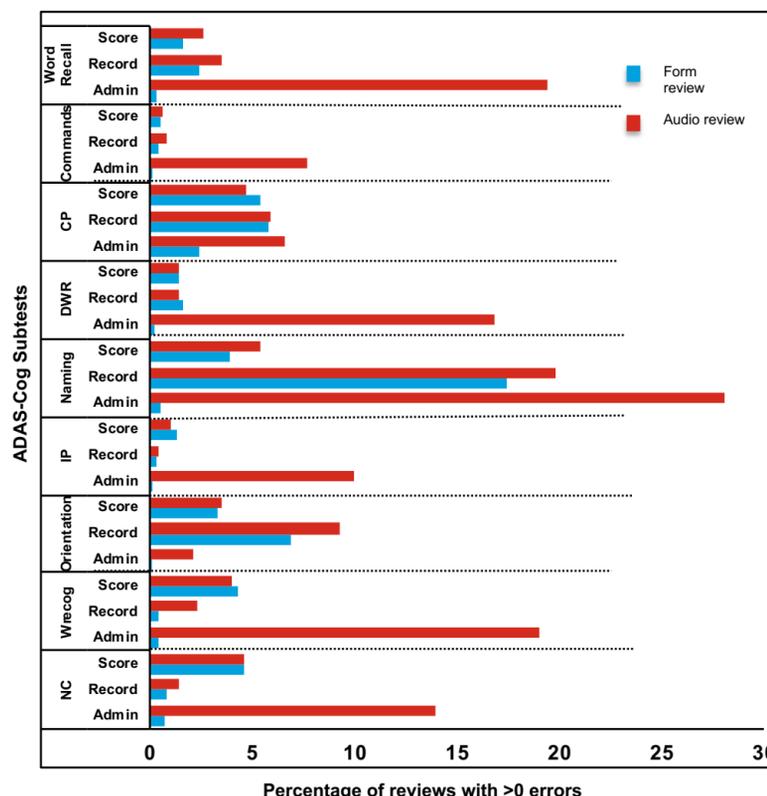
Raters (Category)	Number of Raters (N)
Clinical raters	504
ADAS-Cog raters	453
MMSE raters	506

Figure 1. Prevalence of Errors Across Scales
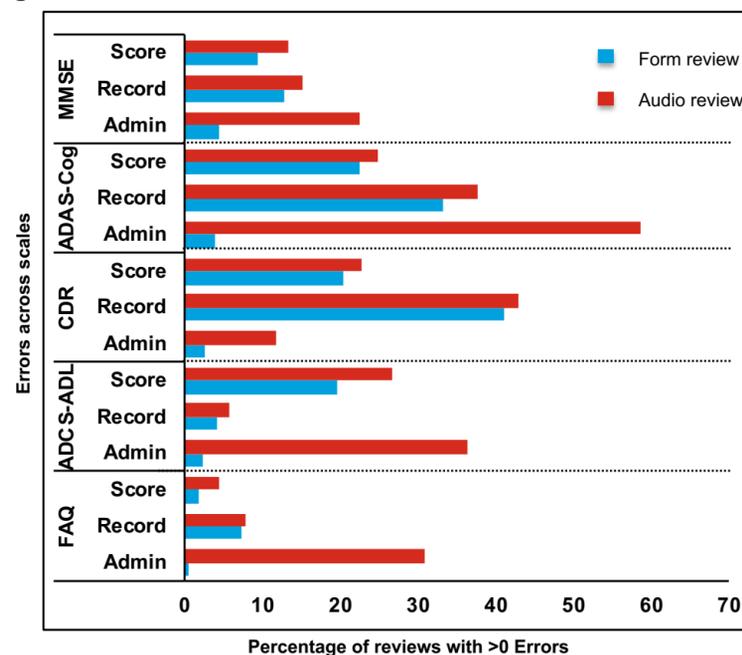
Errors: Admin=administration; Record=recording; Score=scoring

Figure 2. Error Prevalence for ADAS-Cog$_{13}$ Subtests

Errors: Admin=administration; Record=recording; Score=scoring

Scale (Category)	Number of Reviews Completed	
	Form only	Audio
MMSE	4676	1798
ADAS-Cog$_{13}$	4639	2473
CDR	3991	2591
ADCS-ADL	2751	2405
FAQ	2958	2205

Clinical raters: Raters responsible for CDR, FAQ and ADCS-ADL assessments

Abbreviations: ADAS-Cog, Alzheimer's Disease Assessment Scale—Cognitive subscale; CDR, Clinical Dementia Rating; FAQ, Functional Activities Questionnaire; MMSE, Mini-Mental State Examination;

Discussion

- Standardized review of audio recorded scale assessments showed deviations from standard administration guidelines in 25-55% of the scale administrations
- Highest prevalence of errors was observed in ADAS-Cog$_{13}$ scale due to the higher number of items on the SRF and the length and complexity of the scale
- Scoring errors and recording errors were detected equally as often by form review and audio review
- Consistent with previous research, relatively higher rates of error were observed on the Naming and Word Recognition subtests of the ADAS-Cog$_{13}$ scale, as well as Word recall, Delayed word recall and Number cancellation[5,6]
- Recording errors were most common on Naming and Orientation
- Scoring errors were most common on Constructional Praxis, Naming, Word Recognition, and Number Cancellation
- Data demonstrated that Form only reviews are not sufficient for identifying all types of raters' errors when administering clinical outcome assessments, particularly ADAS-Cog$_{13}$. While a small number of administration errors can be deduced by a reviewer in a form only review, based on the way subject responses are recorded on the source document, the vast majority of administration errors are only evident by hearing the interaction between the rater and the subject via audio review

CONCLUSIONS

- CM programs that feature audio reviews of outcome assessments promote greater standardization of scale administration within and across raters and improve the overall quality of study data

References

1. Eketjall S et al. *J Alzheimers Dis.* 2016;50(4):1109-1123
2. Meyer S.M et al. 2016
3. Randolph C et al. 2015
4. Feaster H.T et al. AAIC. 2017
5. Veroff, A. *Alzheimer's & Dementia.* 2012;8(4):538.
6. Atkins A.S et al. 2015. http://www.neurocogtrials.com