Can Screen to Baseline MMSE Variability in AD Clinical Trials Affect Primary Outcome Measurement?

David S. Miller, M.D., M.A1., Yanhui Xu, Ph.D.1, Priscilla Samuelson, R.N.1, David Henley, M.D.2, Gopalan Sethuraman, Ph.D.2, Karen Sundell, B.S.2

1Bracket, Wayne, PA, USA; 2Eli Lilly & Co., Indianapolis, IN, USA
**BACKGROUND**

In Alzheimer’s disease (AD) clinical trials, where other potential causes of a subject’s cognitive impairment must be ruled out, it is paramount that only subjects who meet the defined inclusion/exclusion and specified severity criteria be enrolled. Frequently, the Mini-Mental State Exam (MMSE), a brief cognitive assessment with a maximum score of 30 points, is used to determine the subject’s cognitive function to use for further study subjects within a trial. The MMSE entry criteria are often required to be met only at the screening visit. Previous studies have shown that 17% of subjects on five occasions, and consistently similar, multi-national AD clinical trials had MMSE score changes from Screen to Baseline that would have excluded them from the trial, should they have been required to meet MMSE criteria at Baseline as well [2]. In those studies, raters whose experience with the scales fell below a predetermined level received additional enriched training. This curriculum has previously been shown to enable these raters to both certify to participate, and perform similarly to those more experienced colleagues regarding the need for remediation in study. Additionally, an endpoint reliability program was implemented to ensure that raters were adhering to scale administration and study conventions at select visits. When errors were detected by calibrated raters, the raters in question were remediated and the scoring was corrected. Such a program has been shown to significantly decrease the error rate on subsequent assessments [3].

**OBJECTIVES:**

1. To determine if the degree of MMSE change from Screen to Baseline influenced the ability to separate responders to active drug versus placebo.
2. To determine how raters who received enriched training performed compared to those with more experienced colleagues with regard to their ability to separate responders to active drug from placebo.

**METHODS**

Data from 552 subjects in 2 of the 4 AD trials previously discussed [2] were available for analysis. MMSE scores at both Screen (when inclusion criteria were established) and Baseline visits were assessed for the degree of change, and a sensitivity analysis of change in the different AD programs. The degree of MMSE score change from Screen to Baseline was similar to that showed worsening. (Fig 1) This is consistent with what we had previously demonstrated in the combined analysis of the 2 separate, but methodologically similar, multi-national AD programs. The degree of MMSE score change from Screen to Baseline significantly correlated with a rater’s ability to detect a difference on the ADAS-Cog (r-coefficient) between patients on drug compared to those on placebo. When we compared patients who changed >3 points on the MMSE from Screen to Baseline visits with those who changed <3, only those who changed <3 points separated from placebo at 52 weeks (p=0.019 for ADAS-Cog (11), p=0.043 for ADAS-Cog (14), compared by LSMean from mixed model). (Table 1) We further compared raters who went through the additional Enriched training curriculum with those who did not regarding their performances in study. When we compared subjects who had a change of >2 points on the MMSE from Screen to Baseline visits with those whose change was <2, only when the change was <2 points did active drug separate from placebo. (p=0.019 for ADAS-Cog (11), p=0.043 for ADAS-Cog (14), compared by LSMean from mixed model).

**RESULTS**

468 (86.4%) subjects changed >2 points on the MMSE from Screen to Baseline (including 180 (31.25%) whose score did not change), while 114 (19.6%) changed 3 or more points. The distribution of score changes that showed improvement on the MMSE from Screen to Baseline was similar to that that showed worsening. (Fig 1). This is consistent with what we had previously demonstrated in the combined analysis of the 2 separate, but methodologically similar, multi-national AD programs. The degree of MMSE score change from Screen to Baseline significantly correlated with a rater’s ability to detect a difference on the ADAS-Cog (r-coefficient) between patients on drug compared to those on placebo. When we compared patients who changed >3 points on the MMSE from Screen to Baseline visits with those who changed <3, only those who changed <3 points separated from placebo at 52 weeks (p=0.019 for ADAS-Cog (11), p=0.043 for ADAS-Cog (14), compared by LSMean from mixed model). (Table 1) We further compared raters who went through the additional Enriched training curriculum with those who did not regarding their performances in study. When we compared subjects who had a change of >2 points on the MMSE from Screen to Baseline visits with those whose change was <2, only when the change was <2 points did active drug separate from placebo. (p=0.019 for ADAS-Cog (11), p=0.043 for ADAS-Cog (14), compared by LSMean from mixed model).

**DISCUSSION**

One potential explanation for the degree of MMSE change seen in these studies is the practice effect. However, this is unlikely, given the fact that the distribution of scores that “improved” on the MMSE from screen to baseline was similar to those that “worsened.” This pattern was consistent with what we had previously demonstrated in the combined analysis of the 2 separate, but methodologically similar, multi-national AD programs [2]. Reviewing the MMSE scores at both Screen and Baseline visits would uncover and correct any scoring errors. Doing this would also identify unusual score swings and alert raters to the need to seek an explanation for this (eg – change in the patient’s medication and/or medical condition) and possibly preclude this subject’s enrollment in the study at that time. Regarding the ADAS-Cog, neither group of raters (those requiring enriched training and those who did not) was able to separate active drug from placebo at 52 weeks. That was a trend for those on drug to be rated as doing worse than those on placebo is not surprising, as subjects in this study did tend to do worse on drug at 18 months. As noted above, raters who required and received enriched training performed similarly to their more experienced colleagues over the two time points assessed in this study. Any concern about whether the same rater administered the MMSE at each of the time points is obviated by the fact that the endpoint reliability program was in effect for both of those visits. In an error was identified, the rater was contacted, remediated and the error was corrected. Additionally, the endpoint reliability program employed was able to substantially decrease the scale administration and scoring errors seen by both groups of raters.

**REFERENCES**