ARE CLINICAL ENDPOINTS OF AD TRIALS CLINICALLY MEANINGFUL?

Michael T. Ropacki & Michael Krams
Janssen Research & Development, LLC

Since 1980 there have been over 200 failed Alzheimer's disease (AD) drug candidates despite great efforts by industry and academia to develop new AD therapies. There are many lessons to be learned from failed drug programs. In the case of potential disease modifying therapies, a consensus emerged that investigations should move earlier in the disease course before the accumulation of beta amyloid and other related deleterious effects have manifested. Given the multitude of failed AD programs, questions have also been raised regarding the appropriateness of the clinical measures employed and endpoints selected, as well as their clinical meaningfulness.

The current criteria for the clinical diagnosis of AD include memory and other cognitive impairments that result in functional decline and interfere with activities of daily living (ADLs). More recent criteria developed for research and clinical trial enrollment include laboratory and neuroimaging biomarkers. Several measures frequently employed in AD clinical trials such as the Folstein Mini-Mental State Examination (MMSE), Clinical Dementia Rating Scale (CDR), and Alzheimer's Disease Assessment Scale - Cognitive subscale (ADAS-Cog) are used to screen, diagnose, and/or stage AD. Preclinical AD corresponds to a CDR score of 0 or MMSE scores 27-30; mild AD to CDR scores of 0.5-1 or MMSE scores 21-26; moderate AD to CDR scores of 2 or MMSE scores 10-20; and severe AD to a CDR score of 3 or MMSE scores <10 (Jicha 2010).

Since the 1980's the most commonly employed primary endpoint for AD clinical trials is the ADAS-Cog. The ADAS-Cog is a suitable endpoint for clinical trials of AD patients at specific stages of disease. In early moderate AD (MMSE = 16-20), the ADAS-Cog is sensitive to cognitive change. At this stage the ADAS-Cog is also correlated with clinically meaningful measures that assess ADLs. The main advantage of the ADAS-Cog is utilization of a single global score (which provides improved reliability and statistical power), but this is also a major weakness as it is not clinically meaningful. Therefore, this measure is not used in clinical practice.

The ADAS-Cog is relatively insensitive in the mild, late moderate and severe stages of AD. Therefore, 18-month clinical trials in mild-to-moderate AD (defined in clinical trials as a MMSE = 16-26) with stratified baseline enrollment are at a significant disadvantage. Specifically, at baseline half of the patients are in an AD stage where this measure is relatively insensitive (mild AD: MMSE = 21-26), whereas the other half (early moderate AD: MMSE = 16-20) are within this measure's sensitive range. Over the course of the 18-month trial, patients are expected to decline approximately 6 points on the MMSE. By study completion, many of the mild AD participants originally in the insensitive range have progressed into the sensitive range. Likewise, many of the early moderate AD participants who began the trial in the sensitive range have progressed and are no longer in it. Overall, when selecting measures and clinical endpoints investigators must consider the AD stage at baseline, study duration, expected AD progression over the trial, evidence supporting the measure's psychometric properties and sensitivity in this population, as well as clinical meaningfulness of the results.

When conducting studies with a preclinical AD/early MCI (eMCI) population, sensitivity and clinical meaningfulness issues are compounded because, in theory, there are only finite differences between healthy and affected individuals. Moreover, you would not expect this population to have impaired ADLs (e.g., toileting) measured by commonly used instruments like the Disability Assessment of Dementia (DAD). To assist in this regard the United States Food and Drug Administration (FDA) recently published draft guidance for industry consideration when developing drugs for the treatment of early AD. This guidance discusses employing clinical measures that combine assessment of cognition and function, composite scales/scores and isolated neuropsychological measures. The challenge of moving earlier in

the disease course is to find neuropsychological measures that are psychometrically validated, sensitive across the range of disease, devoid of floor and ceiling effects, not prone to practice effects, and applicable cross-culturally. In addition, these measures need to be feasible to implement, easy to administer and score, have low patient burden and good compliance, and provide clinically meaningful data.

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) and Neuropsychological Assessment Battery (NAB) are two measures that satisfy these criteria. The RBANS is a concise, 20-25 minute comprehensive cognitive measure originally developed for cognitive assessment, detection and characterization of dementia in the elderly. There is a large body of research supporting the RBANS's psychometrics, sensitivity and capability to diagnose MCI/early AD, as well as its ability to predict longitudinal functional status, driving capacity, and basic ADLs in this population. The measure was also shown to correlate with CDR in MCI, as well as biomarkers such as anti-RAGE and anti-Aβ IgG. The RBANS has been linguistically translated, empirically validated in over 30 languages and has four alternative forms to control for practice effects.

The NAB is a comprehensive, modular battery of neuropsychological measures assessing attention, language, memory, visuospatial and executive functions that has excellent psychometric properties and extensive normative and validation data. This battery was normed on a large group (N = 1,448) of healthy normal subjects excluded at baseline for conditions known to interfere with cognition (e.g., neurologic or psychiatric conditions and substance abuse), as well as other impairments that would interfere with test performance (e.g., uncorrected hearing or vision loss, physical impairments affecting upper extremity motor performance). A large proportion (n = 841) of the normative sample was 60-97 years old. NAB modules and alternative forms were co-developed, normed and validated, thereby limiting differences between forms and controlling for practice effects. Each module also contains objective, ecologically valid 'Daily Living Tests' that obviate the need for a subjective caregiver ADL assessment of ADLs. For instance, the Memory module examines memory for medication instructions, names, addresses and telephone numbers, while the Spatial module has a map reading test.

In summary, a significant number of failed AD clinical trials have raised questions around the appropriate stage of AD to study, clinical measures and endpoints employed, and clinical meaningfulness. The most commonly utilized measures in AD research are not sensitive in preclinical AD/eMCI and, therefore, unlikely to yield statistically significant or clinically meaningful information. In the future, tailoring our methods and selecting more sensitive clinical measures and endpoints to the population under investigation should increase our chances of obtaining statistically significant and clinically meaningful information for regulators, physicians, patients and their families.