ELSEVIER

Alzheimer's & Dementia 5 (2009) 324–339

Alzheimer's & Dementia

Perspectives

Scales as outcome measures for Alzheimer's disease

Ronald Black[a], Barry Greenberg[b], J. Michael Ryan[a], Holly Posner[c], Jeffrey Seeburger[d], Joan Amatniek[e], Malca Resnick[f], Richard Mohs[g], David S. Miller[h], Daniel Saumier[i,j], Maria C. Carrillo[k,*], Yaakov Stern[l]

[a]*Wyeth Research, Collegeville, PA, USA*
[b]*Toronto Western Research Institute, Toronto, Ontario, Canada*
[c]*Eisai Medical Research, Inc., New York, NY, USA*
[d]*Merck Research Laboratories, North Wales, PA, USA*
[e]*Ortho-McNeil Neurologics, Inc., Raritan, NJ, USA*
[f]*Forest Laboratories, Inc., New York, NY, USA*
[g]*Eli Lilly & Co., Indianapolis, IN, USA*
[h]*United BioSource Corporation, Wayne, PA, USA*
[i]*Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada*
[j]*BELLUS Health, Inc., Laval, Quebec, Canada*
[k]*Alzheimer's Association, Chicago, IL, USA*
[l]*Department of Clinical Neuropsychology, Taub Institute, Columbia University, New York, NY, USA*

Abstract The assessment of patient outcomes in clinical trials of new therapeutics for Alzheimer's disease (AD) continues to evolve. In addition to assessing drugs for symptomatic relief, an increasing number of trials are focusing on potential disease-modifying agents. Moreover, participants with AD are being studied earlier in their course of disease. As a result, the limitations of current outcome measures have become more apparent, as has the need for better instruments. In recognition of the need to review and possibly revise current assessment measures, the Alzheimer's Association, in cooperation with industry leaders and academic investigators, convened a Research Roundtable meeting devoted to scales as outcome measures for AD clinical trials. The meeting included a discussion of methodological issues in the use of scales in AD clinical trials, including cross-cultural issues. Specific topics related to the use of cognitive, functional, global, and neuropsychiatric scales were also presented. Speakers also addressed academic and industry initiatives for pooling data from untreated and placebo-treated patients in clinical trials. A number of regulatory topics were also discussed with agency representatives. Panel discussions highlighted areas of controversy, in an effort to gain consensus on various topics.

Keywords: Neuropsychological assessment; Outcome measure; Cognitive testing; Functional scale; Global scale

1. Introduction

Before 1984, consensus measures did not exist for diagnosing or assessing the progression of Alzheimer's disease (AD). Clinical trials were heterogeneous, inclusion criteria were vague and various, and outcomes were idiosyncratic. For example, diagnostic inclusion criteria comprised vague entities such as organic brain syndrome, senile cognitive decline, or organic psychosyndrome. Outcomes included miscellaneous clinician rating scales and various neuropsychological subscales. In the early 1980s, there were attempts to arrive at a consensus on criteria and measures. The National Institute of Neurological and Communicative Disorders and Stroke—Alzheimer's Disease and Related Disorders Criteria for Alzheimer's Disease (also known as the McKhann criteria) were proposed and immediately applied as inclusion criteria in dementia trials [1]. Early clinical trial work with physostigmine in healthy participants and in participants with AD led to the development of the Alzheimer's Disease

*Corresponding author. Tel.: 312-335-5722; Fax: 866-741-3716.
E-mail address: Maria.Carrillo@alz.org

1552-5260/09/$ – see front matter
doi:10.1016/j.jalz.2009.05.667

Assessment Scale (ADAS) as a cognitive-assessment instrument and outcome, specifically for clinical trials [2,3]. Other neuropsychological assessments were also developed for trials at this time.

Early experiences in AD clinical trials led a United States Food and Drug Administration (FDA) advisory panel in 1989 to recommend that AD clinical trials be at least 3 months, preferably 6 months, in duration and use a standard cognitive-assessment instrument and a clinician's global assessment as primary outcomes. As a result, the vast majority of AD registration clinical trials have been 6 months long, and used the Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-Cog) [3] as the primary cognitive outcome. Over time, 6 months were considered insufficient, and trials were lengthened to 12 months. Most of these trials targeted participants with mild to moderate levels of AD severity.

Recently, with the increasing interest in disease modification and the enrollment of participants at milder stages of AD, AD clinical trials have been lengthened to 18 months. Most trials still rely on the Mini-Mental State Examination (MMSE) [4] and Clinical Dementia Rating (CDR) [5] for staging, the ADAS-Cog for cognitive outcomes, the Alzheimer's Disease Cooperative Study-Activity of Daily Living (ADCS-ADL) [6] or the Disability Assessment for Dementia (DAD) [7] for activities of daily living, the CDR or Clinician's Global Impression of Change (CGIC) [8] for global clinical measures, and the Neuropsychiatric Inventory (NPI) [9] for assessing behavior. However, the move toward 18-month trials presents some significant technical issues, such as subject retention and how to handle the increased number of dropouts [10].

Current outcome measures may vary greatly in the linearity of decline over these longer trials, and in their relative sensitivity to change across different strata of disease severity. This raises questions about the scales used in trials to measure progression, and about what should be considered a meaningful difference in the ADAS-Cog and other measures. There is also a perception that control groups are not deteriorating as rapidly as they did in the 1980s and early 1990s on the scales now used in trials, although no clear evidence exists for this.

One issue with current measures, including the ADAS-Cog, involves the broad distribution of cognitive test scores at baseline, even within a narrowly defined group such as mild AD. Moreover, within-subject change is highly variable, with considerable overlap in scores and standard deviations between placebo and treated groups. Overall, the clinical decline in patients in placebo groups may be relatively small compared with the variability in patients, even in 18-month trials, so it may be difficult to detect a drug treatment effect if one exists. Irizarry et al., examining individual scores over 6, 12, or 18 months, reported considerable participant variation, wherein some deteriorated and some actually showed improvement while on a placebo [11]. This variability seems to increase over time, suggesting that for longer trials, the data may produce greater deviations. In addition to between-subject differences, ADAS-Cog variability is compounded by site-to-site and country-to-country differences.

2. Methodological issues in clinical trials

Cognitive scales are essential for AD clinical trials because decline in cognition is the defining symptom. For this reason, cognitive tests are generally given in phase 2 trials, and are a regulatory requirement in phase 3.

2.1. Measurement properties of cognitive tests

A good cognitive test or test battery for AD trials should sample all major cognitive functions affected by AD, should be sensitive over a range of impairment levels, reliable, and valid, should have minimal floor and ceiling effects, should be sensitive to longitudinal changes with minimal practice (learning) effects, and should provide a composite measure of overall performance. The test must also work in the real world, and cannot overtax study participants. Information on practice effects is important, and the availability of equivalent forms for repeated measurements is necessary [12].

Perhaps what is most required in any cognitive test is content validity, or the extent to which the test actually measures what it is intended to measure. Validity cannot be achieved without good interrater reliability. Sensitivity over a range of cognitive-ability levels is becoming increasingly important as trials in cognitively normal people and people with mild dementia become more common. There is a need to improve psychometric properties in this regard. Test bias is often not adequately addressed, and can be problematic, especially when transferring tests to other languages or cultures. Test bias, because of changing psychometric raters during trials, is also a potential issue. It becomes increasingly difficult to maintain rater consistency as trials become increasingly longer.

The ADAS-Cog has been the gold standard for cognitive assessment in clinical trials, but has some limitations. It does not adequately measure certain domains, including delayed memory, attention, and executive function. The Alzheimer's Disease Cooperative Study (ADCS) has improved upon this by adding new tests that address those components [13]. In addition, floor effects make it less useful for longitudinal studies with severe AD. Instead, the Severe Impairment Battery or Modified Ordinal Scales of Psychological Development are often used [14,15]. At the other end of the spectrum, more sensitive cognitive tests are needed for participants with mild cognitive impairment (MCI). Primary prevention trials may require a different set of tests to detect very small changes in memory, typically the first domain affected, at the normal end of the spectrum. Measurement scales that can be conducted at home or over the phone would also be advantageous.

An important factor in the implementation of cognitive scales is that the numbers generated by the scale in question are measurements of the central dependant variable (in the context of assessing an AD patient, the variable would represent functional or physical states of the brain) on which clinical decisions are based. Poorly chosen rating scales can