ELSEVIER

Alzheimer's & Dementia 9 (2013) S56–S60

Alzheimer's & Dementia

Perspectives

Establishing the psychometric underpinning of cognition measures for clinical trials of Alzheimer's disease and its precursors: A new approach

Holly B. Posner[a], Stefan Cano[b], Maria C. Carrillo[c,*], Ola Selnes[d], Yaakov Stern[e], Ronald G. Thomas[f], John Zajicek[a], Jeremy Hobart[a]; for the Alzheimer's Disease Neuroimaging Initiative[†]

[a]*Pfizer Inc, New York, NY, USA*
[b]*Clinical Neurology Research Group, Peninsula College of Medicine and Dentistry, Plymouth, UK*
[c]*Alzheimer's Association, Chicago, IL, USA*
[d]*Johns Hopkins University School of Medicine, Baltimore, MD, USA*
[e]*Columbia University, New York, NY, USA*
[f]*University of California, San Diego, CA, USA*

Abstract The Alzheimer's disease (AD) Cognitive Behavior Section (ADAS-Cog) is the most commonly used cognitive test in clinical trials of AD. Recent trials have focused on people earlier in the course of disease; however, there are concerns about using the ADAS-Cog at this crucial stage. Using data from the Alzheimer's disease Neuroimaging Initiative study, we used a range of traditional psychometric tests to evaluate those concerns. This issue of *Alzheimer's & Dementia* includes two articles that evaluate the ADAS-Cog. These articles report evaluations using two psychometric approaches: traditional methods and new methods. In this review, we provide accompanying background information to this program of research.

Keywords: Alzheimer's disease; Cognition; Psychometric; MCI; ADNI; Neuropsychology; ADAS-Cog

1. Introduction

Cognitive impairment is the most prominent and clinically relevant feature of Alzheimer's disease (AD). Thus, assessment of cognition is critical in evaluating the efficacy of new therapies. For the past 20 years, the Alzheimer's Disease Assessment Scale Cognitive Behavior Section, widely known as the ADAS-Cog, has been the most widely used cognitive coprimary outcome measure in clinical trials of AD treatments [1]. Developed in the 1980s based on well-developed conceptual and neuropsychological underpinnings [2], the ADAS-Cog has since been used in >170 clinical trials, including those that led to the regulatory approval of all currently marketed therapies for mild to moderate AD. Yet, despite the widespread use of this scale, questions remain about its performance as an instrument of measurement among individuals with milder forms of AD. Because there is now widespread consensus in the field that the disease must be treated in its earliest stages, it is particularly important that instruments demonstrate the ability to detect subtle changes in cognition early in the disease process [3]. This is particularly important for evaluating cognition in clinical trials or for use as a primary end point in intervention studies. Although other neuropsychological tests such as the Wechsler Adult Intelligence Scale, the Wechsler Memory Scale, and the California

The authors have no conflicts of interest to report.

[†]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators in the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this article report. A complete listing of ADNI investigators can be found at http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

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

1552-5260/$ - see front matter © 2013 The Alzheimer's Association. All rights reserved.
http://dx.doi.org/10.1016/j.jalz.2012.10.008

Verbal Learning Test have been used extensively in assessing cognition in mild AD in clinical practice and in epidemiological and exploratory interventional settings, these tools have not been used in large phase 3 studies as primary outcomes.

This issue of *Alzheimer's & Dementia* includes two articles that evaluating the ADAS-Cog [4,5]. These articles report evaluations using two psychometric approaches: traditional methods and new methods. In this review, we provide some accompanying background information to this program of research. The goal of this program of research was to rethink our understanding of how neuropsychological assessments function as tools for measuring cognition in a manner suited to the evaluation of interventions in clinical trials. Beyond useful interpretation of scores for diagnostic purposes or adding the number of correct or incorrect responses, the questions and answers in neuropsychological assessments can be ordered from most likely to be answered correctly (easiest) to least likely to be answered correctly (most difficult). Also, the individual questions and cognitive constructs in the ADAS-Cog and other assessments, such as the Paired Associate Learning test [6], the Auditory Verbal Learning Test [7], or the Boston Naming Test [8] can be considered each a nidus for expansion in difficulty level, either toward harder or easier questions, in a neuropsychologically meaningful way for people with AD or its precursors. During the initial stages of this program of research, we conducted an extensive evaluation of the ADAS-Cog in two stages using traditional and modern psychometric methods [4,5], providing a framework for further analyses that could include additional cognitive assessments, such as the Boston Naming Test, Trails A & B, and Logical Memory to fill in identified gaps.

Widespread use of rating scales in clinical trials has meant that these sorts of instruments now play key roles in crucial decisions about patient care, health policy, and the direction of research. Confidence and evidence that scales are fit for this responsibility are crucial [9]. The extent to which cognitive tests and scales, such as the ADAS-Cog, are robust clinically and scientifically has been examined previously using psychometric methods [10].

Our overarching goal in this stage of the research was to provide guidance for the potential improvement, in measurement terms, of an existing scale that has been used as the basis of approval in previous clinical trials—in other words, the ADAS-Cog. In particular, we aimed to examine the potential to capture more completely the cognitive performance of people with mild AD and its precursors. We engaged the full range of key stakeholders in this effort, including experts in the fields of AD, neuropsychology, and psychometrics; pharmaceutical companies; and the advocacy community. The project was funded by the Alzheimer's Association and another nonprofit organization.

2. Background

The ADAS was developed in 1984 as an instrument to assess longitudinally the severity of both cognitive and noncognitive dysfunction in AD patients. The cognitive–behavior section (subscale), known as ADAS-Cog, is comprised of 11 components (word recall, word recognition, constructional praxis, orientation, naming objects and fingers, commands, ideational praxis, remembering test instruction, spoken language, word finding, and comprehension), which represent various cognitive domains: memory; language; ability to orient oneself to time, place, and person; construction of simple designs and planning; and performing simple behaviors in pursuit of a basic, predefined goal. At the time of its construction, psychometric techniques that are considered standard today were not yet widely applied, and not in this case. Since then, the ADAS-Cog has undergone limited psychometric evaluation (eg, [11,12]). These studies supported its reliability and validity at the scale level. Other evaluations, such as scaling assumptions (ie, evidence to support summing component scores for a single total score) and targeting [13] have been little studied, especially when considering the ADAS-Cog as a set of subscales.

The issue of targeting is particularly important with regard to milder levels of AD. We previously conducted such a psychometric analysis of the ADAS-Cog among individuals with mild to moderate AD using data from three clinical trials of donepezil (Aricept) [14]. This study found satisfactory performance at the scale level, but noted that most ADAS-Cog components were too simple for many of the patients in this study. Thus, the scale failed to measure the range of cognitive performance adequately in these subjects. There have been some attempts to improve the ADAS-Cog by expanding the number of cognitive domains tested [15]. For example, a decline in episodic memory may be among the earliest manifestation of disease [16], suggesting that adding tests of delayed recall and digit cancellation might expand sensitivity in the early stages of disease. Other studies have suggested that assessments for mild dementia may require additional neuropsychological tests [17].

3. The ADAS-Cog in ADNI

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the U.S. Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Determination of

sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians in developing new treatments and monitoring their effectiveness, as well as lessening the time and cost of clinical trials.

The principal investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California–San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from >50 sites across the United States and Canada. The initial goal of ADNI was to recruit 800 adults, age 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be monitored for 3 years, 400 people with MCI to be monitored for 3 years, and 200 people with early AD to be monitored for 2 years. For up-to-date information, see www.adni-info.org.

ADNI was established in 2003 to assess various disease markers across the AD continuum, with a focus on the transition from MCI to AD. In the original ADNI study (ADNI-1), subjects were evaluated over a 2- or 3-year period using clinical, neuropsychological, neuropsychiatric, imaging, and fluid biomarker studies. ADNI data are freely available to investigators via the publically available ADNI database. Among the tests conducted as part of the cognitive assessment is the ADAS-Cog, which was collected across five time points in the ADNI database. Thus, these data provided an ideal opportunity to examine the measurement properties underpinning the ADAS-Cog.

3.1. Overview of analysis

This psychometric analysis examined the legitimacy of producing a total ADAS-Cog score, which is obtained by summing its 11 items. We also examined the individual items of the ADAS-Cog, which themselves are made up of multiple subcomponents of the total score. Key psychometric properties, including reliability and validity, were assessed both at the scale level (total score) and the component level, and using both traditional [4] and modern (Rasch analysis) psychometric methods [5]. In simple terms, and explained more here [5] and elsewhere [9,10,18], the main advantage of Rasch analysis [19] was that it allowed us to diagnose specific issues regarding performance of the ADAS-Cog and to identify potential areas of improvement. In addition, Rasch analysis uses mathematical models to evaluate the legitimacy of summing items [20].

3.2. What we learned from traditional psychometric analysis

As demonstrated by Cano and colleagues [4] using traditional psychometric analysis, at the scale level and for individuals with milder AD, the ADAS-Cog performed adequately in terms of data completeness, scaling assumptions, reliability, and validity. However, in the mild AD and MCI subsamples, analysis of individual components showed large ceiling effects for 8 of 11 components in mild AD patients and 9 of 11 components in MCI patients, indicating that the components are too easy for most subjects with a diagnosis of dementia. Thus, these measures failed to detect subtle changes in cognitive performance, particularly among individuals with milder impairments. Taken together, these analyses suggest that when used in clinical trials in asymptomatic or MCI populations, the total score may provide a misleading answer (ie, not sufficiently sensitive) to the question of whether cognitive performance has been affected by a treatment.

3.3. What we learned from Rasch analysis

Using the Rasch measurement methods, our analysis [5] of the ADNI data confirmed that the 11 components tested in the ADAS-Cog map a continuum of cognitive impairment with good stability across time points and diagnosis and without bias. However, the scale targeted a range of cognition that was significantly worse than that of the subjects in this study, with this mismatch particularly strong among the more mildly affected patients. In addition, the response categories were not working as intended; they were unable to detect the range of responses and thus unable to discriminate subtle changes in cognition. Component-level analysis further showed that many components were too easy even for people with moderate AD.

The advantage of the Rasch analysis is that it identifies explicitly how the scale can be improved. If the scale is envisioned as a ruler, by moving the ruler down toward harder questions that represent better cognition, it would be more likely not only to detect milder levels of impairment but also changes in cognition over time. Within individual components, response categories will need to be redefined.

In a separate analysis [21], we also compared the 11-item with the 13-item ADAS-Cog using Rasch methods to determine whether adding new measures would improve performance. This analysis showed that although this expanded the neuropsychological profile, there were no substantial improvements in terms of measurement performance or reduction of errors. Neuropsychological and measurement properties of cognitive assessments are both independent and related considerations for evaluation.

4. Next steps

We next plan to analyze the full ADNI-1 item-level data for all the cognitive assessments by traditional and Rasch psychometric methodologies and build a Rasch-based algorithm that will convert individual patient-level scores on specific cognitive tests into a robust, overall cognitive score. The justification for using Rasch measurement over non-Rasch–based methods to achieve this goal includes the ability to examine key psychometric properties, including parameter separation, statistical sufficiency, and invariance.

These are essential to generate stable linear measurements. These advantages are further discussed by Hobart and associates [5] and are described extensively elsewhere [9,10,18]. Concurrently, we plan to work with the U.S. Food and Drug Administration and other regulatory authorities to be able to use the Rasch–based algorithm in future clinical trials.

5. Conclusions

The Rasch analysis we conducted not only identified weaknesses in the ADAS-Cog, but, more important, also points to a scientifically explicit path toward improvement. This improvement will take neuropsychologically sound measures of cognition and turn them into instruments that can measure change in cognitive performance over time. Such rulers are needed to assess more completely the drugs being evaluated in clinical trials. Indeed, there are concerns that the poor sensitivity of the ADAS-Cog to detect change in milder patients may have contributed to the disappointing results of recent trials. A full replication of this algorithm is needed to confirm these conclusions.

It is important to note that this approach does not propose a new method of assessing cognition, but rather a more psychometrically sound interpretation of the performance of individuals on the range of cognitive tests that comprise the assessment of cognitive performance in AD and its precursors. Moreover, this approach will allow the new methodology to be tested side by side in the same study with the tabulation of traditional ADAS-Cog scores.

The approach we took in conducting this study—building a partnership in precompetitive space with academia, industry (through the Foundation for the National Institutes of Health consortium), nonprofit organizations, advocacy groups such as the Alzheimer's Association, and organizations such as the International Society for CNS Clinical Trials and Methodology and the Critical Path Institute—is essential to advance the development of psychometrically sound assessments for use in MCI and AD clinical trials. Only through partnerships such as this will the field move forward collectively toward the goal of finding effective treatments for AD and other dementias.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Amorfix Life Sciences Ltd; AstraZeneca; Bayer HealthCare; BioClinica, Inc; Biogen Idec, Inc; Bristol-Myers SquibbCompany; Eisai, Inc; Elan Pharmaceuticals, Inc; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development, LLC; Medpace, Inc; Merck & Co, Inc; Meso Scale Diagnostics, LLC; Novartis Pharmaceuticals Corporation; Pfizer, Inc; Servier; Synarc, Inc; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private-sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by National Institutes of Health grants P30 AG010129 and K01 AG030514. We thank Lisa J. Bain for editorial assistance with this manuscript.

References

[1] Mohs RC, Rosen WG, Davis KL. The Alzheimer's disease assessment scale: an instrument for assessing treatment efficacy. Psychopharmacol Bull 1983;19:448–50.
[2] Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. Am J Psychiatry 1984;141:1356–64.
[3] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007; 6:734–46.
[4] Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R, et al. Putting the Alzheimer's cognitive test to the test I: Traditional psychometric methods. Alzheimers Dement 2012;9:S4–9.
[5] Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R, et al. Putting the Alzheimer's cognitive test to the test I: Rasch Measurement Theory. Alzheimers Dement 2013;9:S10–20.
[6] Inglis J. A paired-associate learning test for use with elderly psychiatric patients. J Ment Sci 1959;105:440–3.
[7] Rey A. L'examen clinique en psychologie. Paris: Presses Universitaires de France; 1964.
[8] Kaplan EF, Goodglass H, Weintraub S. Boston Naming Test. Philadelphia, PA: Lea & Febiger; 1983.
[9] Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. Lancet Neurol 2007;6:1094–105.
[10] Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009;13:1–177.
[11] Doraiswamy PM, Kaiser L, Bieber F, Garman RL. The Alzheimer's Disease Assessment Scale: evaluation of psychometric properties and patterns of cognitive decline in multicenter clinical trials of mild to moderate Alzheimer's disease. Alzheimer Dis Assoc Disord 2001; 15:174–83.
[12] Weyer G, Erzigkeit H, Kanowski S, Ihl R, Hadler D. Alzheimer's Disease Assessment Scale: reliability and validity in a multicenter clinical trial. Int Psychogeriatr 1997;9:123–38.
[13] Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002;11:193–205.
[14] Cano SJ, Posner HB, Moline ML, Hurt SW, Swartz J, Hsu T, et al. The ADAS-Cog in Alzheimer's disease clinical trials: psychometric evaluation of the sum and its parts. J Neurol Neurosurg Psychiatry 2010; 81:1363–8.

[15] Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope: the Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord 1997;11:S13–21.

[16] Salmon DP. Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease. In: Pardon MC, Bondi MW, editors. Behavioral Neurobiology of Aging. Volume 10: Springer Berlin Heidelberg; 2012:187–212.

[17] Wouters H, Appels B, van der Flier WM, van Campen J, Klein M, Zwinderman AH, et al. Improving the accuracy and precision of cognitive testing in mild dementia. J Int Neuropsychol Soc 2012; 18:314–22.

[18] Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoecon Outcomes Res 2011;11:571–85.

[19] Rasch G. On general laws and the meaning of measurement in psychology. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV. Berkeley, CA: University Of California Press; 1961;4:321–33.

[20] Cano SJ, Barrett LE, Zajicek JP, Hobart JC. Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. Mult Scler 2010;17:214–22.

[21] Hobart JC, Posner HB, Aisen P, Selnes O, Stern Y, Thomas R, et al. The ADAS-Cog's performance as a measure: lessons from the ADNI study. Part 3: do the scale modifications add value? Neurology 2009;72:A92.