Original Research Article

Dementia and Geriatric Cognitive Disorders

Dement Geriatr Cogn Disord 2009;28:63–69
DOI: 10.1159/000230709

Accepted: May 25, 2009
Published online: July 30, 2009

How Well Do the ADAS-cog and its Subscales Measure Cognitive Dysfunction in Alzheimer's Disease?

Jared F. Benge[a, c] Steve Balsis[e] Lisa Geraci[e] Paul J. Massman[b, d] Rachelle S. Doody[b]

[a]Michael E. DeBakey Veteran's Affairs Medical Center, and [b]Alzheimer's Disease and Memory Disorders Center, Department of Neurology and [c]The Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, and [d]Department of Psychology, University of Houston, Houston, Tex., and [e]Department of Psychology, Texas A & M University, College Station, Tex., USA

Key Words

Alzheimer's Disease Assessment Scale · Alzheimer's disease · Dementia · Item response theory · Psychometric assessment · Neuropsychological assessment

Abstract

Background/Aims: The Alzheimer's Disease Assessment Scale-cognitive (ADAS-cog) is regularly used to assess cognitive dysfunction in Alzheimer's disease (AD) clinical trials. Yet, little is known about how the instrument and its subscales measure cognition across the spectrum of AD. The current investigation used item response theory (IRT) analyses to assess the measurement properties of the ADAS-cog across the range of cognitive dysfunction in AD. ***Methods:*** We used IRT-based analyses to establish the relationship between cognitive dysfunction and the probability of obtaining observed scores on each subscale and the test as a whole. Data were obtained from 1,087 patients with AD and amnestic mild cognitive impairment. ***Results:*** Results showed that the ADAS-cog and its subscales provide maximum information at moderate levels of cognitive dysfunction. Raw score differences toward the lower and higher ends of the scale corresponded to large differences in cognitive dysfunction, whereas raw score differences toward the middle of the scale corresponded to smaller differences. ***Conclusions:*** The utility of the ADAS-cog and its subscales is optimal in the moderate range of cognitive dysfunction, but raw score differences in that region correspond to relatively small differences in cognitive dysfunction. Implications for tracking and staging dementia and for clinical trials are discussed.

The Alzheimer's Disease Assessment Scale-cognitive (ADAS-cog) [1] is the standard assessment tool used to measure cognitive dysfunction in clinical trials of Alzheimer's disease (AD) medications [2, 3]. It consists of 11 subscales that are designed to assess various cognitive abilities, including those associated with memory, language and praxis. Problems in these areas of cognitive functioning are considered hallmarks of AD [1]. However, little is known about how well the subscales perform across the entire spectrum of cognitive dysfunction in AD. A thorough understanding of the measure's psychometric properties across a broad range of cognitive dys-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

1420–8008/09/0281–0063$26.00/0

Accessible online at:
www.karger.com/dem

Jared F. Benge, PhD
Mental Health Care Line, Michael E. DeBakey VA Medical Center
2002 Holcombe Blvd.
Houston, TX 77030 (USA)
Tel. +1 713 791 1414, ext. 4048, E-Mail jared.benge@va.gov

function is critical to make optimal use of this instrument in clinical trials. As the development of therapeutics shifts to interventions earlier in the disease course [4], corresponding to milder cognitive dysfunction, it is especially important to know which subscales provide the most reliable information in this range of dementia.

In addition to knowing how the subscales function, it is important to understand how the test as a whole functions. Such knowledge can inform the interpretation of results from clinical trials. A Cochrane review of AD pharmaceutical trials relied, in part, on ADAS-cog change scores when evaluating the efficacy of dementia treatments [3]. This review noted that groups receiving cholinesterase inhibitors obtained an average of 2.7 fewer ADAS-cog error points. But the question remains as to what these 2.7 points mean in terms of the underlying cognitive dysfunction, and whether a 2.7-point difference means the same thing in terms of underlying cognitive dysfunction across the scale.

Until recently, a thorough understanding of this measure's scales and the measure as a whole remained elusive. Recent advances in the application of relatively new statistical machinery, such as item response theory (IRT), now provide a framework for analyzing this measure in psychometrically sophisticated ways. According to IRT, items on a test are related to a latent construct [5] (in the case of the current study, cognitive dysfunction). As the levels of the latent construct change, the probability that a person will receive a particular score on the subscales (and the test as a whole) also changes. This feature of IRT analyses allows investigators to determine the extent to which individual subscale scores or entire instruments are associated with the latent construct of interest – critical for interpreting change associated with clinical drug trials. Until recently, IRT has been applied to the ADAS-cog only in terms of establishing how the scale meets assumptions for a specific type of IRT modeling, known as Rasch Analysis [6]. In this study, the authors showed that the ADAS-cog, as it is currently scored, did a relatively poor job across the spectrum of cognitive dysfunction at meeting the assumptions of that particular statistical model, but found that modifications in weighted scores could improve the instrument's performance. In contrast, the goal of this paper is to apply IRT techniques to analyze how the subscales of the ADAS and the measure as a whole function across the spectrum of AD severity. Among other implications, these analyses will help to build a framework from which we can better interpret changes in raw scores that occur during clinical trials.

Methods

Participants

Participants were 1,087 patients from the Baylor College of Medicine Alzheimer's Disease and Memory Disorders Clinic (ADMDC) who were recruited and evaluated according to previously published procedures [7]. Patients were selected for the analyses if they had a diagnosis of probable AD (96% of sample) or amnestic mild cognitive impairment (MCI) (4% of sample) and had completed the ADAS-cog (with fewer than 4 missing subscale scores) during their most recent visit to the clinic. The decision to add individuals with amnestic MCI to the analyses was to make sure the sample represented the full spectrum of cognitive impairment that could be observed in AD, including mild or preclinical cases. Diagnostic determinations were made at a clinical consensus meeting using relevant diagnostic guidelines (for example, NINCDS-ADRDA criteria [8] for the diagnosis of AD and Petersen et al. [9] criteria for MCI). The mean age of the participants was 75.0 years (SD = 8.1) and the mean education level was 13.7 years (SD = 3.5). Of the participants, 66.6% were female and 92.5% were White. Participants had a mean ADAS-cog score of 31.2 (SD = 16.5).

Measures

The ADAS-cog measures several domains of cognition including recall of a 10-item word list, recognition memory, orientation, naming, language comprehension, expressive language and praxis [1]. Test responses are scored using summed error points, where 0 represents no errors and 70 represents errors on all items.

Procedure

Patients at the ADMDC underwent full neurological and neuropsychological evaluations at baseline and yearly as described elsewhere [7]. Diagnoses were established by clinical consensus conferences among the staff. The consensus meetings included a full review of the patients' medical history and psychometric test scores as well as a review of an informant report. Most patients received yearly follow-up neuropsychological evaluations, medication management, and neurological examinations.

As part this longitudinal protocol, patients at the ADMDC were administered the ADAS-cog. The current data analyses were conducted using the patients' most recent ADAS-cog scores. By selecting the patients' most recent ADAS-cog scores, we sampled a breadth of cognitive dysfunction, as some scores came from patients seen only for their first visit to the clinic (more mild cases) and other scores were taken from later follow-up visits (more severe cases). It is important to note that the ADAS-cog was originally designed to measure moderate AD. Despite this fact, the measure is commonly used in therapeutic trials that focus on milder stages of the disease [10]. Thus, in this study we found it appropriate to include a broad sampling of patients in order to evaluate its utility at all levels of cognitive dysfunction.

Data Analysis

The main data analyses were conducted within an IRT framework. When working with IRT models, it is important to determine that the latent construct measured by all the items is statistically unidimensional (meaning that the scale measures one latent construct) [5]. An important distinction can be made about

unidimensionality as a conceptual and statistical construct. While it is clear that different test items measure different domains of cognitive function, statistical unidimensionality reflects on the tendency of items to measure at least in part a cohesive factor across tasks. We assessed for unidimensionality using the 2 different factor analytic techniques described below.

IRT analyses were conducted using Multilog software [11]. Using observed data from each subscale, the probability of receiving a particular score on a subscale at each level of the latent dimension, theta (θ), was predicted by fitting Samejima's [12] graded model to the observed data; θ can be thought of as similar to a factor in factor analysis, i.e., the observed variables are combined to give an indication of an unobserved construct that each of the items has in common. Thus, the ADAS-cog subscores, thought to measure an unobserved, unidimensional factor, are fitted to the model to allow for a general estimate of the latent variable. Parameters that define the response functions for each possible score were estimated for each subscale. In IRT, these parameters can be and were used to calculate the response probability for each possible score at each of the 801 values of θ from –4.00 to 4.00 (–4.00, –3.99, –3.98, … 3.98, 3.99, 4.00) expressed in standard deviation (SD) units with mean M = 0.0, SD = 1.0. These values were subsequently multiplied by the raw value for each possible score, and summed to generate the 'item' or subscale function. To create domain functions, the relevant subscale level response functions were summed at θ from –4.00 (almost no cognitive dysfunction) to 4.00 (severe cognitive dysfunction). To create the test function, all of these domain level response functions were similarly combined. In addition, information estimates were generated by these analyses to determine which levels of cognitive dysfunction were measured best by the ADAS-cog (see [11] for details on creating these functions).

Note that these analyses require the grouping of seemingly disparate cognitive functions such as constructional and ideational praxis. In fact, these are distinct neuropsychological constructs. Their grouping in the current study is not related to whether they are distinct neuropsychological constructs but is related to their covariation in this demented sample.

Results

To test for adequate unidimensionality of the ADAS-cog for IRT analyses, we first conducted an exploratory factor analysis using SPSS v. 16 [13], excluding individuals with missing data. Results of this analysis suggested a large first factor (eigenvalue = 6.88, explaining 62.48% of the variance) with no other factor reaching an eigenvalue >1 and a clear break on a screen plot, indicating an internal structure suggestive of unidimensionality, a finding in keeping with the results of other evaluations of ADAS-cog factor structure [14]. Kline [15] suggests that a ratio between the first and second eigenvalues >3.00 is a good indicator of unidimensionality. The ratio for the present analysis was 7.73, suggesting very strong unidimensionality. To further confirm the unidimensionality of this measure, we conducted a confirmatory factor analysis using MPLUS software [16]. Once again, we found indications of excellent fit of a one-factor model. The Tucker-Lewis incremental fit index [17] (TLI = 0.98) showed excellent fit as did the comparative fit index [18] (CFI = 0.92). Given these converging indicators of strong unidimensionality and IRT's relative robustness to any minor deviation from unidimensionality, we were confident that we had adequate fit for subsequent IRT analyses.

ADAS-cog subscales were grouped into 3 domains: praxis (ideational praxis and construction), language (word reading, speech comprehension, commands, aural comprehension, and naming) and memory (word recall, word recognition, orientation, and recall of the test instructions). Curves that represent these domains are depicted in figure 1. The domain curves were rescaled for the purpose of visual inspection to account for differences in the total number of points possible in each domain, allowing for domain performance comparisons across the spectrum of cognitive dysfunction (x-axis). In conducting a visual analysis of the domain curves, one looks for where the slope of the lines are steepest, with a steeper curve indicating more discriminative power at the given range of θ. Thus in figure 1, the memory domain curve has its steepest slope between –2 and 1 SDs of cognitive dysfunction, while the praxis and language domain curves share a similarly shaped and sloped curve from approximately between –1 and 2 SDs of cognitive dysfunction. This finding suggests that the memory domain is maximally discriminative at relatively milder stages of the disease when compared with the praxis and language domains, which functioned quite similarly, indicating that one did not provide unique discriminative data in comparison to the other.

In addition to generating domain functions, IRT analyses can be used to generate another function known as information (see test information function in fig. 1). The information function shows how well a particular instrument captures the latent phenomenon, in this case cognitive dysfunction. A measure has optimal information where the information curve peaks. Examination of figure 1 shows that the ADAS-cog has the highest level of information in the more moderate levels of cognitive dysfunction (the middle panel of fig. 1, which demarcates approximately between –1.0 and 1.75 SDs of cognitive dysfunction). This finding suggests that, when considered as a whole, the test does a relatively good job discriminating among the different degrees of cognitive dysfunction in this range. Notice that in this range, the memory, language, and praxis measures have meaningful slopes. The slopes of the domain curves can be thought

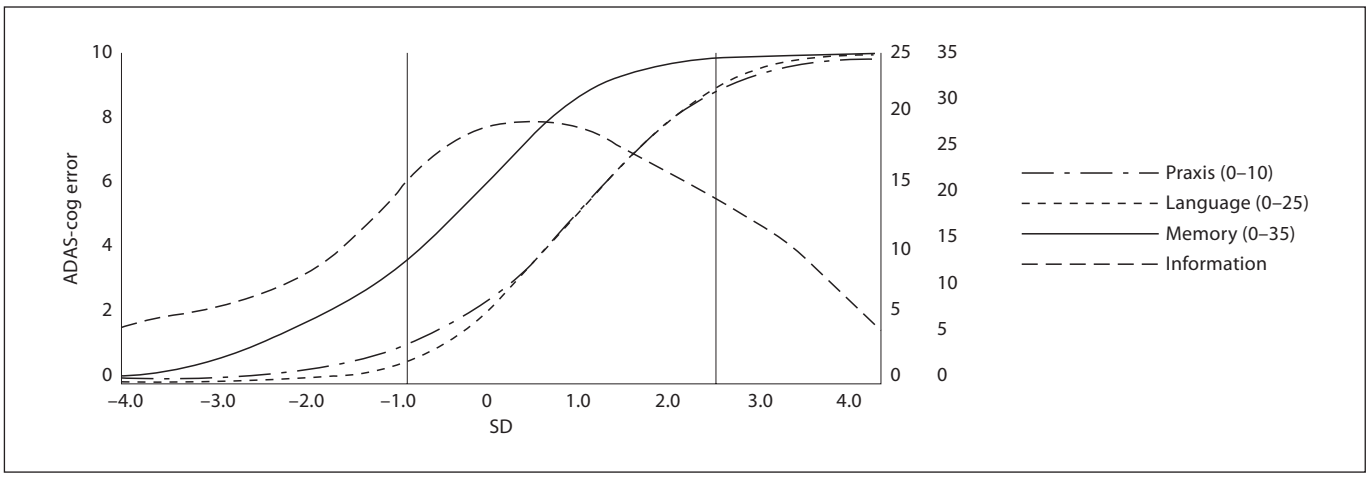

Fig. 1. ADAS-cog domain scale curves and test information curve. Curves were rescaled for the purpose of visual inspection. See text for details.

of as component parts of the information curve. Where the slopes are steepest, the measure does the best job at discriminating between gradations of cognitive dysfunction, and the information curve is highest. Notice that in this moderate range where the information curve is highest, a unit change in cognitive dysfunction (x-axis) indicates a greater change in errors (y-axis) than a unit change at lesser or greater levels of cognitive dysfunction. This pattern suggests that in this range, the scale can discriminate best between smaller gradations of cognitive dysfunction. The left panel (which demarcates approximately between –4.0 and –1.0 SDs, i.e., more mild levels of impairment) depicts a different pattern of results. There are no domains with meaningful slopes here. As a result, the information function remains low, suggesting that the domains and the measure as a whole do a relatively poor job discriminating among the different degrees of cognitive dysfunction when dementia severity is mild. The same is true at the more severe levels of cognitive dysfunction, shown in the rightmost panel of figure 1. Taken together, results from these analyses show that the subscales of the ADAS-cog, when considered together, are maximally useful for measuring the moderate stages of cognitive dysfunction in AD.

To more fully understand the reason for this trend, an analysis of all 11 subscales is necessary. Figure 2 displays the results from all 11 subscales grouped by domain. Note that they have been rescaled so they could be depicted together when necessary to again allow differential comparison of performance across the abscissa. Using the same visual interpretative strategy employed for the domain curves, we see that of all the subscales, the recall subscale slopes upward earlier than the other subscales. Thus, performance on this memory test is the best indicator of the disease at mild levels of cognitive impairment.

Figure 2 also depicts the individual subscales that constitute the praxis and language domains. There are 2 important trends to note in this figure. One, the curves maximally discriminate within the moderate levels of cognitive dysfunction. Two, the curves for these subscales overlap with one another to a large extent, suggesting that in terms of measuring underlying cognitive dysfunction, performance on these tests may be largely redundant. In other words, performance in any one of these tests indicates roughly the same information about the level of a patient's cognitive dysfunction.

All subscales can be combined to form a test characteristic curve (see fig. 3). In figure 3, the relationship between all possible observed raw scores and underlying cognitive dysfunction is plotted between –4.00 and 4.00 SDs. Thus, reading the figure from left to right represents the changes in cognitive dysfunction from very mild to very severe. There are several noteworthy trends in this figure. First, as cognitive dysfunction increases, so do the expected raw scores on the ADAS-cog. This finding is intuitive as higher scores on the ADAS-cog represent errors and an increase in cognitive errors should be related to cognitive dysfunction. For example, at 0.24 SDs of cognitive dysfunction (in this relatively demented sample), a

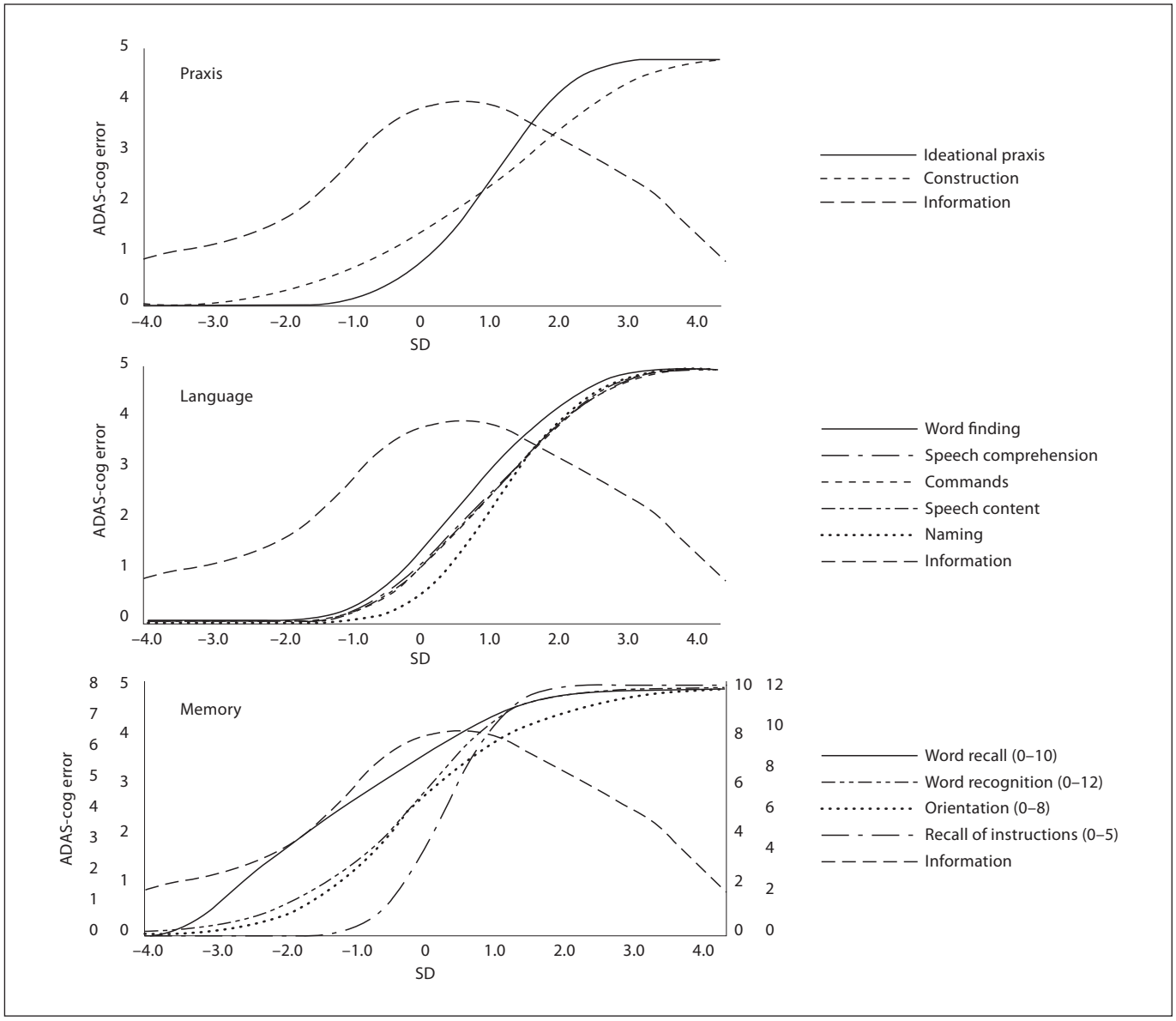

Fig. 2. ADAS-cog subscale curves and test information curve. Memory curves were rescaled for the purpose of visual inspection. See text for details.

score of 36 is expected, whereas at 1.07 SDs of cognitive dysfunction, a score of 52 is expected. Second, there are portions of the trace line that are relatively flat and other portions that are relatively steep. These differing slopes indicate that for some ranges of raw scores (toward the lower and higher extremes) at approximately <–1 SD or >2 SD, underlying cognitive dysfunction changes quite dramatically, but for other similar-sized ranges of raw scores (in the middle of the score range), cognitive dysfunction changes very little.

To aid other investigators in evaluating patient raw scores in terms of underlying cognitive dysfunction in relation to our sample, θ values (SDs of cognitive dysfunction) corresponding to each raw score value of the ADAS-cog are available in table 1. Evaluation of figure 2 and table 1 demonstrates that at milder stages of the disease (ADAS-cog raw scores <14), even small changes in raw score are associated with relatively large changes in cognitive dysfunction (θ). A similar pattern is observed in higher ADAS-cog raw scores. In the moderate range of

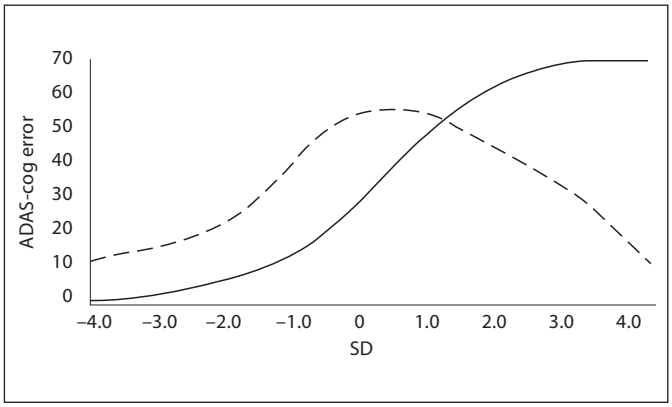

Fig. 3. ADAS-cog test characteristic curve and test information curve. See text for details.

Table 1. The relationship between ADAS-cog raw scores and cognitive dysfunction in SD units

ADAS raw score	Theta	ADAS raw score	Theta	ADAS raw score	Theta
0	<–4.0	24	–0.39	48	0.82
1	–3.39	25	–0.34	49	0.87
2	–3.04	26	–0.28	50	0.93
3	–2.77	27	–0.23	51	0.99
4	–2.54	28	–0.18	52	1.05
5	–2.32	29	–0.13	53	1.12
6	–2.12	30	–0.08	54	1.18
7	–1.94	31	–0.03	55	1.25
8	–1.78	32	0.02	56	1.32
9	–1.63	33	0.07	57	1.39
10	–1.50	34	0.12	58	1.47
11	–1.38	35	0.17	59	1.55
12	–1.27	36	0.22	60	1.64
13	–1.17	37	0.26	61	1.73
14	–1.07	38	0.31	62	1.82
15	–0.99	39	0.36	63	1.93
16	–0.91	40	0.41	64	2.04
17	–0.83	41	0.46	65	2.17
18	–0.76	42	0.51	66	2.30
19	–0.69	43	0.56	67	2.47
20	–0.63	44	0.61	68	2.68
21	–0.56	45	0.66	69	3.01
22	–0.50	46	0.71	70	>4.00
23	–0.45	47	0.76		

dysfunction, between approximately –1.0 and 2.0 SDs of cognitive dysfunction, each ADAS-cog error point represents an incrementally smaller change in underlying cognitive dysfunction.

Discussion

The current analyses were conducted to examine how well each subscale of the ADAS-cog and the test as a whole measure cognitive dysfunction at varying levels of dementia severity. Results indicated that the ADAS-cog subscales measure AD-related cognitive dysfunction best at moderate levels of impairment. Of the 3 main cognitive domains designated a priori (memory, language, and praxis), performance on the memory subscales discriminated best at lower levels of dysfunction. Of the memory subscales, difficulty on the word recall test was sensitive to milder levels of impairment, whereas difficulty recalling the test instructions was not affected until the more severe stages of cognitive dysfunction.

The finding that memory, the defining cognitive feature of AD, is impaired at relatively mild stages of cognitive dysfunction is not surprising. The ADAS-cog, however, does not have other subscales that might be sensitive to relatively early indications of cognitive decline. The fact that there are no other measures that are sensitive to early stages of cognitive decline may hinder the ability of this instrument to detect the disease in its earliest stages [19]. In addition, while a delayed recall memory trial is frequently administered with the ADAS-cog, its findings are not included in the standard 70-point error scoring of this measure. As difficulties with delayed recall are the cardinal feature of MCI [20], its inclusion in studies of the more mild preclinical stages of the disease (as has more recently been done in clinical trials in MCI [21]) may be particularly useful in improving the psychometric properties of the instrument at this early stage of the disease. Also, it should be noted that adding sensitive measures of other domains (beyond memory) that might change early in the course of AD may also improve the psychometric properties of the instrument at this end of the cognitive dysfunction spectrum.

The other 2 domains tapped by the ADAS-cog, language and praxis, provide somewhat redundant information about the moderate stages of cognitive dysfunction. While these domains are undoubtedly of clinical utility in assessing AD, their utility for measuring cognitive dysfunction as a global entity may be redundant. Turning towards more severe cognitive dysfunction, including simple repetition tasks (tasks that would still have meaningful variability even in late stages of the disease) could help discriminate among severe levels of cognitive dys-

function to increase the amount of information at this end of the spectrum.

When the test was examined as a whole, there were 2 key findings. First, the magnitude of cognitive dysfunction represented by each point on the ADAS-cog was not equal across the scale. This can be most clearly demonstrated by examining table 1 and considering the roughly 3-point treatment effect of cholinesterase inhibitors. A 3-point difference in scores can represent an actual difference in underlying cognitive dysfunction of anywhere from 0.72 SDs (from a raw score of 5 to a raw score of 2) to 0.14 SDs of cognitive dysfunction (from a raw score of 37 to a raw score of 34). Thus, a 3-point difference between a subject's scores may represent a different magnitude of change in cognitive dysfunction at different points on the scale.

A strength of the findings derived from the current IRT analyses is that they rest on parameter estimates that typically can be considered sample invariant [5], meaning that the information presented (specifically in table 1) can be applied to other studies after appropriate transformation. Clinicians and researchers alike can use this information to help equate and evaluate the magnitude of scores in their dementia samples to determine the influence of cognitive dysfunction on subsequent outcomes and response to treatment.

Acknowledgements

This study was supported by the Mitchell Foundation (J.F.B., S.B., P.J.M.) and the Cain Foundation (R.S.D.).

References

1 Rosen WG, Mohs, RC, Davis KL: A new rating scale for Alzheimer's disease. Am J Psychiatry 1984;141:1356–1364.

2 Rockwood K, Fay S, Gorman M, Carver D, Graham J: The clinical meaningfulness of ADAS-cog changes in Alzheimer's disease patients treated with donepezil in an open-label trial. BMC Neurol 2007;7:26.

3 Birks J: Cholinesterase inhibitors for Alzheimer's disease. Cochrane Database Syst Rev 2006;1:CD005593.

4 Hackinski V: Shifts in thinking about dementia. JAMA 2008;300:2172–2173.

5 Hambleton RK, Swaminathan H, Rogers HJ: Fundamentals of Item Response Theory. Newbury Park, Sage Publications, 1991.

6 Wouters H, van Gool W, Schmand B, Lindeboom R: Revising the ADAS-cog for a more accurate assessment of cognitive impairment. Alzheimer Dis Assoc Disord 2008;22:236–244.

7 Doody RS, Pavlik V, Massman P, Kenan M, Yeh S, Powell S, Cooke N, Dyer C, Demirovic J, Waring S, Chan W: Changing patient characteristics and survival experience in an Alzheimer's disease center patient cohort. Dement Geriatr Cogn Disord 2005;20:198–208.

8 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM: Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. Neurology 1984;34:939–944.

9 Petersen RC, Doody R, Kurz A, et al: Current concepts in mild cognitive impairment. Arch Neurol 2001;58:1985–1992.

10 Jelic V, Kivipelto M, Winblad B: Clinical trials in mild cognitive impairment: lessons for the future. J Neurol Neurosurg Psychiatry 2006;77:429–438.

11 Thissen D: Multilog User's Guide, Version 6.3. Scientific Software, 1991.

12 Samejima F: Estimation of latent ability using a response pattern of graded scores. Psychometrika Monographs 1969, No. 17.

13 SPSS for Windows, Release 13.0. SPSS Inc., 2004.

14 Kim YS, Nibbelink DW, Overall JE: Factor structure and reliability of the Alzheimer's Disease Assessment Scale in a multicenter trial with linopirdine. J Geriatr Psychiatry Neurol 1994;7:74–83.

15 Kline RB: Principles and practice of structural equation modeling. New York, The Guilford Press, 2004.

16 Muthén LK, Muthén BO: MPLUS user's guide (ed 5). Los Angeles, Muthén & Muthén, 2007.

17 Tucker LR, Lewis C: A reliability coefficient for maximum likelihood factor analysis. Psychometrika 1973;38:1–10.

18 Bentler PM: Comparative fit indexes in structural models. Psychol Bull 1990;107:238–246.

19 Harrison JC, Minassian SL, Jenkins L, et al: A neuropsychological test battery for use in Alzheimer disease clinical trials. Arch Neurol 2007;64:1323–1329.

20 Petersen RC: Mild underlying cognitive dysfunction as a diagnostic entity. J Intern Med 2004;256:183–194.

21 Doody RS, Ferris SH, Salloway S, et al: Donepezil treatment of patients with MCI: a 48-week randomized placebo-controlled trial. Neurology 2009;72:1555–1561.