

The ADAS-Cog revisited: Novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials

Nandini Raghavan^{a,*}, Mahesh N. Samtani^a, Michael Farnum^b, Eric Yang^b, Gerald Novak^c, Michael Grundman^d, Vaibhav Narayan^c, Allitia DiBernardo^c; for the Alzheimer's Disease Neuroimaging Initiative[†]

^aJanssen Pharmaceutical Companies of Johnson & Johnson, Raritan, NJ

^bJanssen Pharmaceutical Companies of Johnson & Johnson, Springhouse, PA

^cJanssen Pharmaceutical Companies of Johnson & Johnson, Titusville, NJ

^dGlobal R&D Partners, San Diego, CA

Abstract

Background: The Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) has been used widely as a cognitive end point in Alzheimer's Disease (AD) clinical trials. Efforts to treat AD pathology at earlier stages have also used ADAS-Cog, but failure in these trials can be difficult to interpret because the scale has well-known ceiling effects that limit its use in mild cognitive impairment (MCI) and early AD. A wealth of data exists in ADAS-Cog from both historical trials and contemporary longitudinal natural history studies that can provide insights about parts of the scale that may be better suited for MCI and early AD trials.

Methods: Using Alzheimer's Disease Neuroimaging Initiative study data, we identified the most informative cognitive measures from the ADAS-Cog and other available scales. We used cross-sectional analyses to characterize trajectories of ADAS-Cog and its individual subscales, as well as other cognitive, functional, or global measures across disease stages. Informative measures were identified based on standardized mean of 2-year change from baseline and were combined into novel composite endpoints. We assessed performance of the novel endpoints based on sample size requirements for a 2-year clinical trial. A bootstrap validation procedure was also undertaken to assess the reproducibility of the standardized mean changes of the selected measures and the corresponding composites.

Results: All proposed novel endpoints have improved standardized mean changes and thus improved statistical power compared with the ADAS-Cog 11. Further improvements were achieved by using cognitive–functional composites. Combining the novel composites with an enrichment strategy based on cerebral spinal fluid beta-amyloid ($A\beta_{1-42}$) in a 2-year trial yielded gains in power of 20% to 40% over ADAS-Cog 11, regardless of the novel measure considered.

Conclusion: An empirical, data-driven approach with existing instruments was used to derive novel composite scales based on ADAS-Cog 11 with improved performance characteristics for MCI and early AD clinical trials. Together with patient enrichment based on $A\beta_{1-42}$ pathology, these modified endpoints may allow more efficient clinical trials in these populations and can be assessed without modifying current test administration procedures in ongoing trials.

© 2013 The Alzheimer's Association. All rights reserved.

The authors have no conflicts of interest to report.

N.R., M.N.S., M.F., E.Y., G.N., V.N., and A.D. are employees of Johnson & Johnson and hold stock options in Johnson & Johnson. M.G. is a consultant to Johnson & Johnson.

[†]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators in the ADNI contributed to

the design and implementation of ADNI or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Authorship_List.pdf.

*Corresponding author. Tel.: 908-218-6424; Fax: 908-927-3782.

E-mail address: nraghava@its.jnj.com

Keywords: Alzheimer's disease; MCI; Mild cognitive impairment; ADAS-Cog; Composite endpoints; Novel endpoints; Clinical trials; Power; Sample size; Ceiling effects

1. Introduction

Alzheimer's disease (AD) clinical research has entered a new era of therapeutics that aim to modify underlying disease pathology rather than ameliorate symptoms [1–3]. Among the challenges faced by disease-modifying strategies targeting early disease stages is uncertainty over appropriate endpoints for early AD and mild cognitive impairment (MCI) trials. The standard research tool for cognitive assessment in clinical trials is the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) [4], which provides a single score based on arbitrary weightings of performance on test items relevant to AD. In its original configuration, the ADAS-Cog assessed learning and memory, language, and spatial cognition, but lacked coverage of executive function; an expanded version was developed in an attempt to address this concern [5].

Although the ADAS-Cog has been used successfully in trials of symptomatic treatments for mild-to-moderate AD [6], certain features of the ADAS-Cog limit its use in earlier stages. There is a nonlinear relationship between disease severity and rate of decline, with the fastest rate of decline seen in patients with moderate AD [7,8], and slower rates of decline in patients with mild AD or MCI [9]. Up to half of ADAS-Cog subscales demonstrate ceiling effects in subjects with mild or moderate AD [10], which makes these items likely to be uninformative in earlier stages of the disease. Limited coverage of early cognitive deficits compounds the insensitivity of ADAS-Cog to mild progression. Last, variability in ADAS-Cog scores can exceed the annual rate of change in clinical trials [11,12], with variability in uninformative subscales potentially obscuring changes on items that can actually track mild deficits.

Another endpoint-related challenge for disease-modifying trials is the precedent set in symptomatic antidementia trials requiring statistical significance on separate primary endpoints of cognition and function [13,14]. This requirement translates into larger sample sizes needed to achieve the same study power compared with a single primary endpoint. Composite cognitive–functional endpoints optimized for earlier stages of disease could therefore provide an alternative to the high hurdle of co-primary endpoints and have been proposed as a methodological advance that may help early AD trials succeed [15].

We hypothesized that by eliminating less informative items from ADAS-Cog and substituting more responsive measures of cognition or function, we could improve sensitivity to change, reduce variability and develop a single scale optimized for MCI and early AD trials. We derived new cognitive and cognitive–functional composite scales based on ADAS-Cog and other existing instruments, and compared their power with the ADAS-Cog in these populations. We also

compared the efficiency of the novel and composite endpoints with and without biomarker enrichment based on the presence of amyloid beta ($A\beta$) pathology in the MCI cohort. Last, we used a bootstrap procedure to assess the reproducibility of the standardized mean changes.

2. Methods

2.1. Data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [16]. Analyses were performed on data downloaded from the ADNI web portal [17] using a data cut-off of November 15, 2010. Of 819 subjects in the data set, 798 subjects were included in the analyses; 21 subjects who reverted from AD to MCI were excluded. Five subjects not labeled as converters from MCI to AD in the original download were relabeled as converters for the analysis to resolve discrepancies between conversion status and current diagnosis. Conversion status was extracted from the ADNI fields for clinical category. The final analysis set at baseline included 229 normal elderly control subjects (NECs), 212 MCI-nonconverters (MCI-NCs), 165 MCI-converters (MCI-Cs), and 192 subjects with AD. Of the subjects with MCI at baseline (142 subjects), 64 MCI-NCs (30.2%) and 78 MCI-Cs (47.3%) had cerebrospinal fluid (CSF) $A\beta_{1-42} \leq 192$ pg/mL, the cutoff we chose to follow for enrichment based on published cutoffs that discriminate between MCI converters and nonconverters as defined in ADNI [18].

2.1.1. Subjects with longitudinal follow-ups

One hundred ninety-eight NECs, 138 MCI-NCs, 139 MCI-Cs, and 131 subjects with AD had data for visits at months 12 and 24; 168 NECs, 99 MCI-NCs, and 103 MCI-Cs had data for month 36. The study protocol did not follow subjects with AD beyond 2 years. We created 15 cross-sectional cohorts by clinical category and visit month utilizing subjects with data for visits at baseline and months 12, 24, and 36: NEC, months 0, 12, 24, and 36; MCI-NC, months 0, 12, 24, and 36; MCI-C, months 0, 12, 24, and 36; and AD, months 0, 12, and 24.

2.1.2. Measures evaluated

In this article, we considered several cognitive, functional, and global measures that are available in ADNI as described in the *ADNI General Procedures Manual* [19]. We evaluated ADAS-Cog 11, ADAS-Cog 13, and the individual ADAS-Cog subscales: Word Recall (Q1), Commands (Q2),

Construction (Q3), Delayed Word Recall (Q4), Naming (Q5), Ideational Praxis (Q6), Orientation (Q7), Word Recognition (Q8), Recall Instructions (Q9), Spoken Language (Q10), Word Finding Difficulty (Q11), Comprehension (Q12), and Number Cancellation (Q14). In addition, the following scales were also considered: Mini-Mental State Examination (MMSE); Boston Naming Test (BNT); Clock Drawing and Clock Copying Tests; Wechsler Adult Intelligence Scale-R Digit Symbol Substitution Test (Digit Symbol); Digit Span Backward and Forward Tests (Digit Backward, Digit Forward); Logical Memory Immediate and Delayed Recall Tests (LM Immed, LM Delayed); Trail Making Tests A and B (Trails A and Trails B); Auditory-Verbal Learning Tests Delayed, Immediate, with Interference and Recognition (AVLT Delayed, AVLT Immed, AVLT Inter, AVLT Recog); Clinical Dementia Rating-Sum of Boxes (CDR-SB); Functional Assessment Questionnaire (FAQ); and Category Fluency Tests Animals and Vegetables (Category Animals, Category Vegetables).

2.2. Analysis

We used the R statistical computing platform [20], version 2.13.0, and libraries available there.

2.2.1. Longitudinal characterization of ADAS-Cog and other scales using cross-sectional cohorts

For the 15 cross-sectional cohorts defined in Section 2.1, values of ADAS-Cog 11, ADAS-Cog 13, all the ADAS-Cog subscales, as well as all the other measures listed in Section 2.1.1 were extracted (Supplemental Table 1). Summary statistics and boxplots for each subscale and cohort were generated to identify ceiling effects, floor effects, and change

across clinical categories (e.g., MCI to AD). Ceiling effects occur when the measure cannot distinguish among the best performers (i.e., scores that plateau at 0% errors or 100% correct). We define a ceiling effect for a measure if at least 10% of the subjects in at least one cohort achieve perfect scores. Conversely, floor effects occur when the measure cannot distinguish among the poorest performers (0% correct or 100% errors). For this article, a floor effect is defined if at least 10% of the subjects in at least one cohort get no test items correct. We describe change across clinical categories as the difference in the median percent scores of the NEC and the AD cohorts at baseline.

2.2.2. Statistical development and assessment of composite scores

Standardized measures (z scores) were calculated for all cognitive and functional scales and subscales of the ADAS-Cog described in Section 2.1.1 across all subjects in all diagnostic categories by subtracting the baseline mean and dividing by the baseline standard deviation.

2.2.2.1. Composite component selection

Standardized mean 2-year change from baseline was used to select informative subscales and measures for target disease-stage cohorts—AD, MCI, or MCI with Aβ₁₋₄₂ pathology—using a threshold of CSF Aβ₁₋₄₂ ≤ 192 pg/mL (MCI-Aβ). We focused on the MCI and the MCI with Aβ₁₋₄₂ pathology cohorts for this article. Measures with standardized mean changes exceeding a threshold of 0.4 for MCI were selected to develop composite scores. We selected this cutoff based on comparison with ADAS-Cog, which has an effect size of 0.54 (see Section 3.2.1), reasoning that subscales and tests with standardized effects sizes well below this

Table 1

Descriptive statistics for the three cohorts: mild cognitive impairment (MCI), MCI with Aβ₁₋₄₂ pathology (MCI-Aβ), and Alzheimer's disease (AD)

Characteristic	MCI	% Missing MCI	MCI-Aβ	% Missing		
				MCI-Aβ	AD	
n	377	NA	142	NA	192	NA
Age, mean (SD)	74.95 (7.37)	0	74.21 (7.3)	0	75.51 (7.48)	0
Sex, % female	0.36 (0.48)	0	0.37 (0.49)	0	0.47 (0.5)	0
Education, mean (SD)	15.68 (3.08)	0	15.74 (2.99)	0	14.69 (3.13)	0
APOE ε4, % 0 alleles	0.54 (0.5)	0	0.65 (0.48)	0	0.66 (0.48)	0
MMSE, mean (SD)	27.01 (1.78)	0	26.81 (1.78)	0	23.34 (2.07)	0
ADAS-Cog 11, mean (SD)	11.71 (4.39)	0	12.61 (4.57)	0	18.64 (6.31)	1.04
CDR-global, fraction of 0, 0.5, and 1	0: 1/377, 0.5: 376/377	0	0.5: 142/142	0	0.5: 99/192, 1: 93/192	0
CDR-SB, mean (SD)	1.62 (0.89)	0	1.64 (0.9)	0	4.3 (1.63)	0
CC1 (ADAS-3, AVLT Immed, MMSE), mean (SD)	0.15 (1.64)	0	0.49 (1.55)	0	3.15 (1.68)	1.04
CC2 (ADAS-3, CDR-SB-Cog), mean (SD)	0.07 (0.94)	0	0.22 (0.95)	0	2.38 (1.28)	0
CFC1 (ADAS-3, AVLT Immed, MMSE, FAQ), mean (SD)	-0.11 (1.02)	<1	0.06 (0.98)	1.41	2.4 (1.42)	1.04
CFC2 (ADAS-3, CDR-SB-Cog, FAQ), mean (SD)	-0.13 (1)	<1	0 (1.01)	1.41	2.48 (1.51)	0
CFC3(CDR-SB-Cog, FAQ), mean (SD)	-0.17 (1)	<1	-0.1 (0.98)	1.41	2.51 (1.58)	0

Abbreviations: ADAS-3, the sum of Word Recall (Q1), Delayed Word Recall (Q4), and Orientation (Q7) in ADAS-Cog; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive; APOE, apolipoprotein E; AVLT Immed, Auditory-Verbal Learning Test Immediate; CC1, cognitive composite 1; CC2, cognitive composite 2; CDR-SB-Cog, Clinical Dementia Rating-Sum of Boxes cognitive component; CFC1, cognitive-functional component 1; CFC2, cognitive-functional component 2; CFC3, cognitive-functional component 3; FAQ, Functional Assessment Questionnaire; MMSE, Mini-Mental State Examination; NA, not applicable; SD, standard deviation.

size would not yield sufficient gains for inclusion in novel composites. For the component selection, only subjects with longitudinal follow-up visits were considered. In addition, differential sensitivity of ADAS-Cog 11 subscales to changes in MCI compared with AD was assessed by evaluating change at year 1 and year 2 from baseline for both groups.

2.2.2.2. Composite evaluation

We assessed the performance of the composite measures and compared them with existing measures by calculating the power to detect a hypothesized 25% treatment effect [21,22] in a clinical trial. For this, we built statistical models of disease progression for each measure using linear mixed effects models with random slopes and intercepts to base power calculations for a 2-year, two-arm, parallel design clinical trial. Power calculations were done for subjects with (i) early AD, (ii) MCI, and (iii) MCI-A β at samples sizes ranging from 100 to 1000 per treatment arm. The disease progression model and the power calculations used any available data from subjects for visits

from baseline up to and including month 24 for the 192 AD subjects, for the 377 subjects with MCI, and for the 142 subjects with MCI-A β pathology.

2.2.2.3. Validation of component and composite scores

To assess whether the selected components of the composites could be selected reproducibly using this procedure in similar populations, we performed a bootstrap validation of the entire composite component selection process. Furthermore, we assessed whether these components, and the corresponding composites, demonstrated similar standardized mean changes and, consequently, similar power for clinical trials across the bootstrap samples. This was implemented as follows: A total of 1000 bootstrap samples of subjects were selected with replacement from the population of subjects with MCI in ADNI who had visits recorded for up to 24 months. For each bootstrap sample, for each measure, standardized mean 2-year changes from baseline were calculated. The empirical distribution of these standardized mean changes for each measure based on the 1000 bootstrap samples was determined.

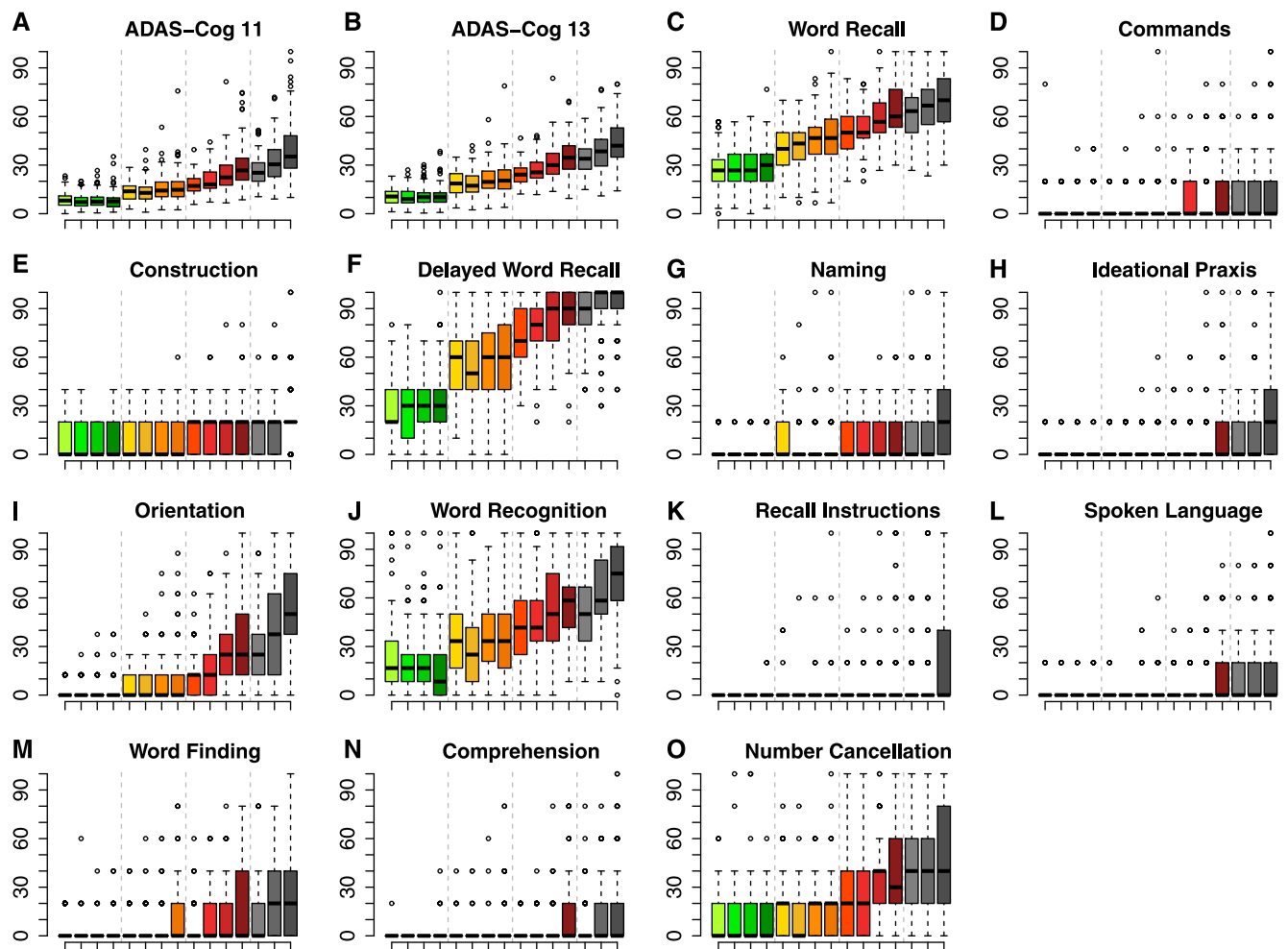


Fig. 1. (A–O) Cross-sectional boxplots of the Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) omnibus scales and subscales. From left to right the x-axis in each plot represents: Normal elderly control subject (NEC) cohorts (green) corresponding to NEC 0, 12, 24, and 36 months; mild cognitive impairment nonconverter (MCI-NC) cohorts (yellow) corresponding to MCI-NC 0, 12, 24, and 36 months; MCI converter (MCI-C) cohorts (red) corresponding to MCI-C 0, 12, 24, and 36 months; and Alzheimer's disease (AD) cohorts (gray) corresponding to AD 0, 12, and 24 months are shown in this order. The y-axis represents the number of errors as a percent of the maximum possible errors.

3. Results

3.1. Longitudinal characterization of ADAS-Cog and other scales using cross-sectional cohorts

Table 1 shows the baseline characteristics of the AD, MCI, and MCI- $A\beta$ groups for MMSE, global CDR, CDR-SB, ADAS-Cog 11, age, gender, and *APOE* $\epsilon 4$ status as well as the novel composite scales, CC1, CC2, CFC1, CFC2 and CFC3. There is no significant difference in age across the three groups. The AD group has more women and has slightly less education than the MCI and MCI- $A\beta$ groups. Both the AD and MCI- $A\beta$ groups have more *APOE* $\epsilon 4$ carriers than the MCI group. As expected, the AD group showed more severity at baseline in all measures.

Fig. 1 shows boxplots of the trajectories across the 15 cohorts for ADAS-Cog 11 and ADAS-Cog 13, as well as each ADAS-Cog subscale. The y-axis represents errors as a percentage of the maximum possible score and the x-axis represents the 15 cohorts (from left to right: NEC 0, 12, 24, 36 through AD 0, 12, and 24). In each plot, for each disease stage (NEC to AD), the medians of the boxplots trace the course of the particular subscale over the duration observed. Juxtaposed to each other, the median trace suggests the disease trajectory for the given subscale across the disease spectrum observable in ADNI. The analysis identified several ADAS-Cog subscales that exhibit ceiling effects in virtually every cohort: Commands (Q2), Construction (Q3), Naming (Q5), Praxis (Q6), Recall Instructions (Q9), Language

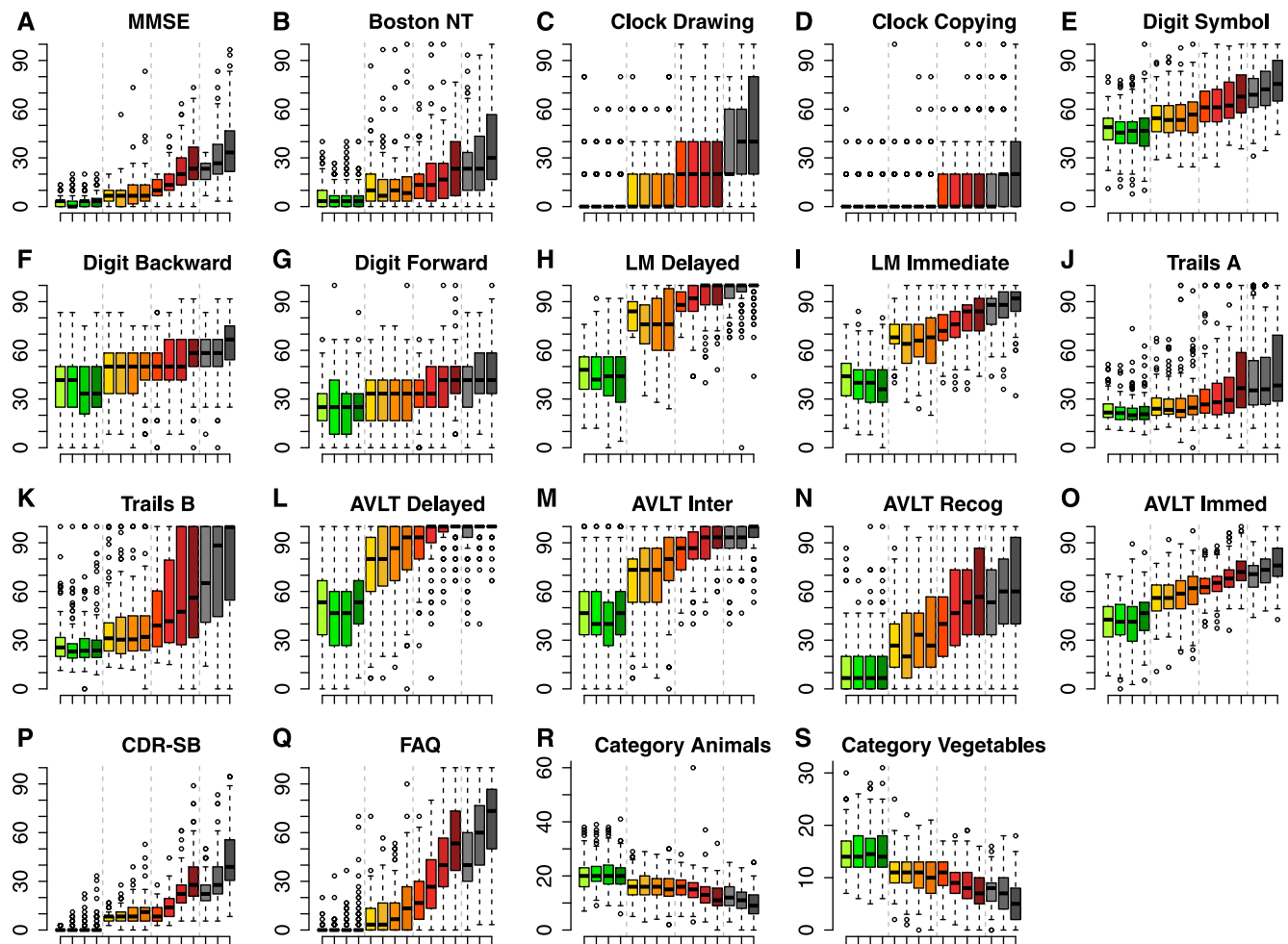


Fig. 2. (A–S) Cross-sectional boxplots of cognitive–functional measures (not including the Alzheimer’s Disease Assessment Scale–Cognitive [ADAS-Cog]). From left to right the x-axis in each plot represents: Normal elderly control subject (NEC) cohorts (green) corresponding to NEC 0, 12, 24, and 36 months; mild cognitive impairment nonconverter (MCI-NC) cohorts (yellow) corresponding to MCI-NC 0, 12, 24, and 36 months; MCI converter (MCI-C) cohorts (red) corresponding to MCI-C 0, 12, 24, and 36 months; and Alzheimer’s disease (AD) cohorts (gray) corresponding to AD 0, 12, and 24 months are shown in this order. The y-axis represents the number of errors as a percent of the maximum possible errors for comparability with the ADAS-Cog subscales. For Trails A and Trails B, the y-axis represents the percentage of maximum allowed time taken for the test. For Clinical Dementia Rating–Sum of Boxes (CDR-SB) and Functional Assessment Questionnaire (FAQ), the y-axis represents the score as a percentage of the maximum score. For Category Fluency Animals and Vegetables, the y-axis represents the number of responses, because there is no maximum set. AVLT Delayed, Auditory–Verbal Learning Test Delayed; AVLT Immed, Auditory–Verbal Learning Test Immediate; AVLT Inter, Auditory–Verbal Learning Test with Interference; AVLT Recog, Auditory–Verbal Learning Test with Recognition; LM, Logical Memory; MMSE, Mini-Mental State Examination; NT, Naming Test.

(Q10), Word Finding Difficulty (Q11), and Comprehension (Q12). Orientation (Q7) and Number Cancellation (Q14) also exhibit ceiling effects to a lesser degree. Delayed Word Recall (Q4) and Word Recognition (Q8) exhibit modest ceiling effects for the NEC cohorts, although this is not visible in the plots. ADAS-Cog 11, ADAS-Cog 13, and Word Recall (Q1) do not exhibit ceiling effects for any cohort. Delayed Word Recall (Q4), Word Recognition (Q8), and Number Cancellation (Q14) exhibit floor effects in the AD cohort. Delayed Word Recall (Q4) shows the largest amount of change across clinical categories, followed by Word Recall (Q1), Word Recognition (Q8), and Number Cancellation (Q14). A summary of the findings in the boxplots is given in Supplemental Table 1.

To understand more completely the different sensitivities of ADAS-Cog subscales to progression in MCI compared with AD, we evaluated the changes from baseline for each subscale at months 12 and 24 (Supplemental Fig. 1). A similar pattern and magnitude of change was seen between MCI and AD for several subscales: Word Recall (Q1), Delayed Word Recall (Q4), Orientation (Q7), Word Recognition (Q8), Recall Instructions (Q9), and Language (Q10). Subjects with MCI changed less than subjects with AD on Commands (Q2), Construction (Q3), Naming (Q5), Praxis (Q6), Word Finding Difficulty (Q11), and Comprehension (Q12), suggesting these items may make ADAS-Cog 11 less sensitive to changes in cognitive function in subjects with MCI.

We performed the same set of analyses on MMSE, BNT, Clock Drawing and Copying, Digit Symbol, Digit Span, LM

Delayed and Immed, Trails A and B, the AVLT components, CDR-SB, FAQ, and the category tests. Fig. 2 shows the resulting boxplots. Ceiling effects in the MCI group are seen in MMSE, BNT, Clock Copying and Drawing Tests, AVLT Recog, and FAQ. Floor effects, especially for the AD group, are seen in LM Delayed, AVLT Delayed, and Trails B. Less severe floor effects are seen in LM Immed, AVLT Inter, AVLT Recog, and Trails A. However, with the exception of Trails A, tests that exhibit floor effects also exhibit, in general, the largest changes across disease stages.

3.2. Statistical development and assessment of composite scores

3.2.1. Composite component selection

We plotted the mean/standard deviation (SD) versus SD of 2-year changes from baseline for the MCI cohort, for all items tested (Fig. 3). ADAS-Cog 11 and ADAS-Cog 13 are shown for comparison. Subscales of the ADAS-Cog identified in Fig. 1 as having ceiling effects—Commands (Q2), Construction (Q3), Naming (Q5), Praxis (Q6), Recall Instructions (Q9), Language (Q10), Word Finding Difficulty (Q11), Comprehension (Q12)—fall in the lower right region in Fig. 3, indicating large variability and correspondingly small effect sizes. ADAS-Cog subscales that exceeded our target standardized mean changes threshold of 0.4 for the MCI cohort were Word Recall (Q1), Delayed Word Recall (Q4), and Orientation (Q7). They had small standardized 2-year changes and relatively low variability, except for

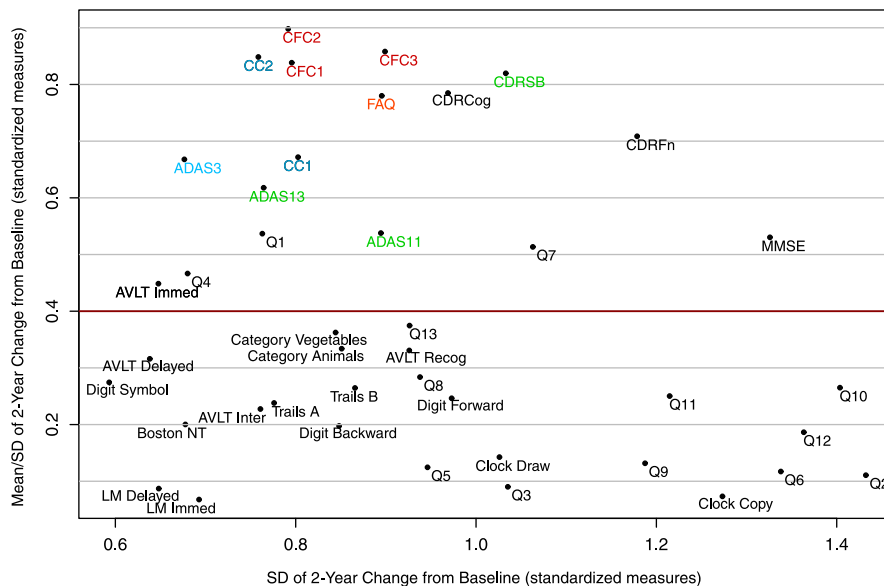


Fig. 3. Plot of mean/standard deviation (SD) versus SD of 2-year change from baseline for the mild cognitive impairment (MCI) cohort. The plot is based on standardized values of all measures. Standardized values are obtained by subtracting the baseline mean and dividing by baseline SD of the measure, where mean and SD are taken across all subjects. For brevity and clarity, Alzheimer's Disease Assessment Scale-Cognitive subscales are represented as Q1 through Q14 (there is no Q13) as follows: Q1, Word Recall; Q2, Commands; Q3, Construction; Q4, Delayed Word Recall; Q5, Naming; Q6, Ideational Praxis; Q7, Orientation; Q8, Word Recognition; Q9, Recall Instructions; Q10, Spoken Language; Q11, Word-Finding Difficulty; Q12, Comprehension; and Q14, Number Cancellation; CDRSB, Clinical Dementia Rating—Sum of Boxes; CDRCog, CDRSB-Cognitive Component; CDRFn, CDRSB-Functional Component; Digit Symbol, Digit Symbol Substitution Test; Digit Forward, Digit Span Forward; Digit Backward, Digit Span Backward; FAQ, Functional Assessment Questionnaire; Category Animals, Category Fluency Animals; Category Vegetables, Category Fluency Vegetables; AVLT, Auditory Verbal Learning Test; LM, Logical Memory; MMSE, Mini-Mental State Examination; NT, Naming Test.

Orientation. Additional cognitive measures that met the selection threshold were AVLT-Immed, which had the smallest standardized 2-year change of any item selected but the lowest variability as well, and MMSE, which had a relatively large 2-year change but high variability, which diminished its standardized 2-year change (compared with ADAS-Cog 11, Word Recall, and Orientation). The FAQ and CDR-SB both had greater 2-year changes than ADAS-Cog 11 or ADAS-Cog 13. Variability of 2-year changes was greater for CDR-SB than FAQ, ADAS-Cog 11, or ADAS-Cog 13.

By eliminating subscales with poor standardized mean changes, we developed the following six composites using standardized scores for selection. ADAS-3 is the sum of Word Recall (Q1), Delayed Word Recall (Q4), and Orientation (Q7) in ADAS-Cog. Cognitive composite 1 (CC1) combines ADAS-3, AVLT-Immed, and MMSE taking directionality of change into account. Cognitive composite CC2 is composed of ADAS-3 and the cognitive portion of CDR-SB. Cognitive–functional composite CFC1 combines CC1 with FAQ, CFC2 combines CC2 with FAQ, and CFC3 combines the cognitive portion of CDR-SB with FAQ.

Our novel cognitive–functional composites—CFC1, CFC2, and CFC3—have the largest standardized 2-year changes of any item (Fig. 3), in part driven by smaller variability than FAQ, CDR-SB, or ADAS-Cog 11, although the variability of CFC3 was slightly greater than ADAS-Cog 13. Our reduced ADAS-3 had the smallest variability and smallest standardized 2-year change of all the novel composites. CC1 had a greater 2-year change than ADAS-Cog 11 or

ADAS-Cog 13, with variability similar to ADAS-Cog 13 and less than ADAS-Cog 11. The novel cognitive composite CC2, which combines ADAS-3 and the cognitive portion of CDR-SB, takes advantage of the low variability of the former and the high standardized change of the latter, resulting in a cognitive composite comparable with the cognitive–functional composites.

3.2.2. Composite evaluation

Fig. 4 shows plots of the statistical power (y-axis) as a function of sample size requirements (x-axis), for a hypothesized 25% treatment effect for each novel composite as well as for ADAS-Cog 11, ADAS-Cog 13, and CDR-SB, for the three cohorts (AD, MCI, and MCI- $A\beta$). Although our focus is on MCI, we compared the utility of the novel composites for AD also.

Table 2 summarizes the 2-year rate of change on ADAS-Cog 11, ADAS-Cog 13, CDR-SB, FAQ, and the novel composites for the MCI, MCI- $A\beta$, and AD populations. Also shown is the number needed to treat to observe a 25% treatment effect with 80% power.

In subjects with MCI, clear advantages for the novel composites are seen over ADAS-Cog 11 and ADAS-Cog 13. With a target 25% treatment effect, ADAS-Cog 11 requires 772 patients per arm for 80% power and ADAS-Cog 13 requires 582 per arm. This level of power is achieved by CDR-SB and the novel cognitive–functional composites with far fewer patients, less than 400 per arm. Indeed, it is possible to achieve 80% power with as few as 375 patients per arm using CDR-SB or

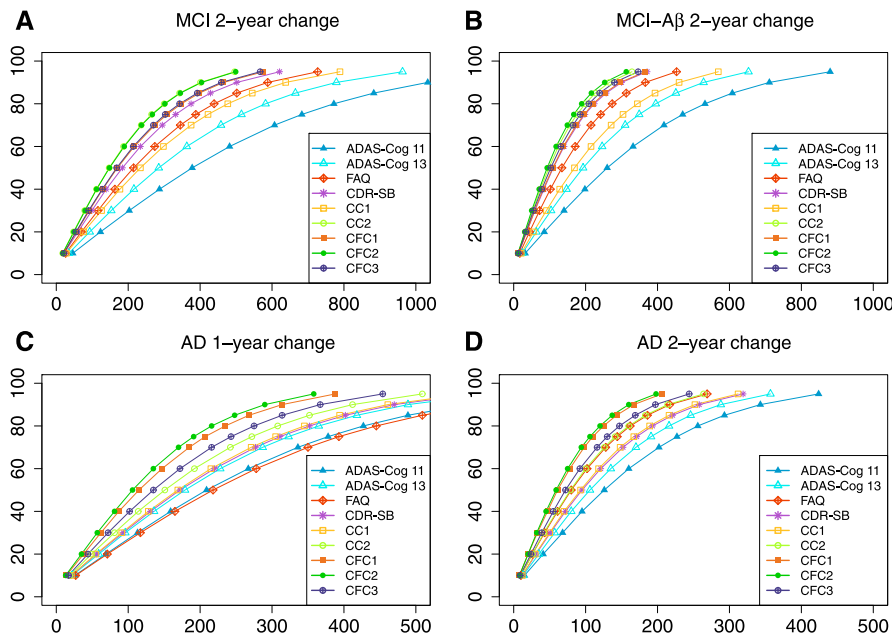


Fig. 4. Plots of power (between 0% and 100% on the y-axis) versus sample size requirements per treatment arm (x-axis). Power calculations are based on a linear mixed-effects model for each outcome measure, with random slopes and intercepts, for a two-arm parallel design clinical trial and assuming a 25% treatment effect relative to placebo. (A) Mild cognitive impairment (MCI) 2-year change. (B) MCI with $A\beta_{1-42}$ pathology (MCI- $A\beta$) 2-year change. (C) Alzheimer's disease (AD) 1-year change. (D) AD 2-year change. Each figure shows power plots for Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) 11, ADAS-Cog 13, Functional Assessment Questionnaire (FAQ), Clinical Dementia Rating–Sum of Boxes (CDR-SB), Cognitive composites (CC1 and CC2), and Cognitive–functional composites (CFC1, CFC2, and CFC3).

Table 2
Mean, Standard Deviation and Estimated Sample Sizes for Reference Measures and Novel Composites for the MCI, MCI-A β and AD Cohorts

Measure	MCI, mean (SD)	MCI-A β , mean (SD)	AD, mean (SD)	MCI, n	MCI-A β , n	AD, N
ADAS-Cog 11	0.48 (0.89)	0.59 (0.95)	1.47 (1.4)	772	532	256
ADAS-Cog 13	0.47 (0.76)	0.58 (0.8)	1.10 (0.94)	582	395	216
CDR-SB	0.85 (1.03)	0.99 (1.00)	1.86 (1.62)	375	224	193
FAQ	0.70 (0.90)	0.91 (0.92)	1.05 (0.88)	439	274	162
CC1 (ADAS-3, AVLT Immed, MMSE)	0.54 (0.80)	0.66 (0.82)	0.98 (0.90)	477	344	189
CC2 (ADAS-3, CDR-SB-Cog)	0.64 (0.76)	0.77 (0.74)	1.09 (0.85)	300	199	160
CFC1 (CC1, FAQ)	0.67 (0.80)	0.84 (0.79)	1.10 (0.83)	348	221	125
CFC2 (CC2, FAQ)	0.71 (0.79)	0.88 (0.79)	1.14 (0.79)	302	189	120
CFC3 (CDR-SB-Cog, FAQ)	0.77 (0.90)	0.95 (0.90)	1.37 (1.03)	343	209	147

Abbreviations: AD, Alzheimer's disease; ADAS-3, the sum of Word Recall (Q1), Delayed Word Recall (Q4), and Orientation (Q7) in ADAS-Cog; ADAS-Cog, Alzheimer's Disease Assessment Scale-Cognitive; AVLT Immed, Auditory-Verbal Learning Test Immediate; CC1, cognitive composite 1; CC2, cognitive composite 2; CDR-SB, Clinical Dementia Rating-Sum of Boxes; CDR-SB-Cog, Clinical Dementia Rating-Sum of Boxes cognitive component; CFC1, cognitive-functional component 1; CFC2, cognitive-functional component 2; CFC3, cognitive-functional component 3; FAQ, Functional Assessment Questionnaire; MCI, mild cognitive impairment; MCI-A β , mild cognitive impairment amyloid beta; MMSE, Mini-Mental State Examination; SD, standard deviation.

NOTE. Changes are given in standardized units for ease of comparison. In columns 1 through 3, 2-year changes (mean/SD) for ADAS-Cog 11, ADAS-13, CDR-SB, FAQ, and novel composites for MCI, cerebrospinal fluid-enriched MCI (MCI-A β), and AD cohorts are given. In columns 4 through 6, sample sizes (n) for a target 25% treatment effect and 80% power in a 2-year trial are given for the MCI, MCI-A β , and AD cohorts. Sample size calculations are based on a linear mixed-effects model for each outcome measure, with random slopes and intercepts, for a two-arm parallel design clinical trial and assuming a 25% treatment effect relative to placebo.

302 patients per arm using the best cognitive-functional composite CFC2, given a 25% target treatment effect.

Additional gains in power are seen with the MCI-A β population for all scales, including ADAS-Cog 11. For a target treatment effect of 25%, it would be possible to achieve 80% using CDR-SB with approximately 224 patients per arm, compared with 189 for CFC2, whereas this level of power would require 395 patients per arm for ADAS-Cog 13 and about 532 per arm for ADAS-Cog 11.

3.2.3. Validation of component and composite scores

Supplemental Fig. 2 shows boxplots of the empirical distributions of standardized mean changes for each measure based on 1000 bootstrap samples from the subjects with MCI in ADNI with visits of at least 2 years. The empirical distributions for the measures selected from the original population exceeded the threshold of 0.4 significantly: Word Recall (99% exceeded 0.4), Delayed Word Recall (88% exceeded 0.4), Orientation (98% exceeding 0.4), MMSE (99% exceeding 0.4), AVLT Immed (80% exceeded 0.4), FAQ (100% exceeded 0.4), and CDR-SB (100% exceeded 0.4). Conversely, the empirical distributions for measures that had poor standardized mean changes in the original population largely fell well below the threshold. Thus, this demonstrates that the standardized mean changes of the measures that were selected, as well as their composites, are reproducible in populations similar to the original population.

4. Discussion

We present several novel composite scores for early AD developed by (i) improving the ADAS-Cog for MCI populations by removing uninformative subscales, (ii) supplementing with other sensitive cognitive measures for MCI, and (iii) supplementing with a functional measure to produce a single composite end point that could be used in clinical trials in lieu of

co-primary endpoints. We evaluated the resulting composite scores as outcome measures for 2-year clinical trials involving patients with MCI or MCI with A β pathology (based on a threshold of CSF A $\beta_{1-42} \leq 192$ pg/mL), and for comparison with AD. Linear mixed-effects disease progression models were used to develop statistical power calculations for each novel and for several existing measures, and these show increased power for the novel composites compared with ADAS-Cog 11 or ADAS-Cog 13. The most significant gains in power were seen with our novel cognitive-functional endpoints. With a target 25% treatment effect, ADAS-Cog 11 would require 772 subjects with MCI per arm to achieve 80% power, whereas our novel cognitive-functional composites CFC1, CFC2, and CFC3 all require less than 350 subjects per arm.

Ceiling effects for subscales removed from the ADAS-Cog were seen in three clinical trials of donepezil in AD [8,23,24], in which perfect scores were achieved for subjects with AD in 55% to 82% of trials. Conversely, subscales retained in the novel composites have face validity for early AD. Recall is well-known as an early symptom of disease that quickly reaches "floor" levels, whereas Orientation has previously been noted in a clinical trial as the single subscale in ADAS-Cog to show significant decline in mild AD [25]. The composition of ADAS-3, which forms the core of our novel composites (with the exception of CFC3, which uses the cognitive portion of CDR-SB), is also consistent with findings from the MCI trial of donepezil and vitamin E, in which 81% of the errors among the MCI group on ADAS-Cog items stemmed from word list recall or word recognition deficits [24], suggesting that these items should be included in any composite designed for MCI trials.

Cognitive composite CC1 achieved modestly improved effect size and power over ADAS-3 by incorporating MMSE and AVLT-Immed. Although AVLT-Immed is a sensitive measure of episodic memory deficits [26], and its inclusion is perhaps

unsurprising, MMSE is a less obvious scale to increase the accuracy of a cognitive composite given its well-known variability, which is also reflected in Fig. 3. MMSE, however, is a fairly robust disease severity marker and rates of decline on functional [27] as well as cognitive [28,29] scales have been shown to vary with baseline MMSE scores. Because performance on ADAS-Cog, and by extension its components, is highly correlated with MMSE, the presence of MMSE in the cognitive composite may reflect the correlation of this scale with other assessments of AD. One concern regulators and clinical “trialists” may have about using measures like ADAS-3 as outcome measures is that their performance is driven by too few domains. Including MMSE mitigates against this type of concern without much loss in effect size.

Cognitive composite CC2 combines ADAS-3 and the cognitive portion of CDR-SB into the best-performing cognitive endpoint, which exceeded even some of the cognitive–functional composites in standardized mean change and power. Additional gains over the cognitive portion of CDR-SB alone may be a result of the increased specificity of ADAS-3 items for domains affected in MCI.

The regulatory requirement for co-primary endpoints in AD trials has, in general [30], mandated statistical significance on each co-primary endpoint at the studywise false-positive rate. This typically requires larger trials to achieve the same study power compared with a single primary endpoint. An alternative approach is to use a single composite end point, such as the CDR-SB, to evaluate a primary cognitive deficit and its clinical relevance simultaneously. The CDR-SB has been proposed for early AD trials [31] and it performed very well in our power calculations. Indeed, the cognitive component of CDR-SB outperformed all existing cognitive measures in the analysis of 2-year change, supporting its current use in MCI and prodromal AD trials. It is interesting to observe, however, that the effect size of the CDR-SB score in the MCI population is driven more by its cognitive component than its functional component, suggesting that the functional part of this composite could be optimized further for the mild functional deficits seen in this population. We noted that FAQ captured more functional deficits than the functional portion of CDR-SB in subjects with MCI, prompting us to use it as the functional core of all our cognitive–functional composites. CDR-SB and all three novel cognitive–functional composites (CFC1, CFC2 and CFC3) demonstrated adequate standardized mean changes for the MCI and MCI- $A\beta$ cohorts, suggesting that any would be appropriate for MCI or prodromal AD clinical trials. However, the variability of the novel composites was less than that of CDR-SB. Another potential advantage of most of the novel composites relates to their independence from the clinical expertise of the rater. Clinician-rated scales such as CDR-SB, which perform well in ADNI, may be less reliable in clinical trial settings compared with scales that rely on quantitative assessments or on patient-reported outcomes.

A principal concern with cognitive–functional composites is that the cognitive component of a composite may drive

statistical significance, without movement on the functional component. In all our novel cognitive–functional composites, the standardized mean change in functional component matched that in the cognitive components, in contrast to CDR-SB. Furthermore, FAQ shows moderately good correlations with the cognitive components of all the composites (0.51 with CC1 in CFC1, 0.62 with CC2 in CFC2, and 0.65 with the cognitive portion of CDR-SB in CFC3), suggesting reasonably well-balanced contributions from the cognitive and functional components. Previous work suggests a close relationship between changes on measures of function and cognition in AD [32–34]. The relationship is less well characterized for MCI, in which the expected deficits are different and milder. However, including measures with similar standardized mean changes and reasonable correlations should allay concerns over independent cognitive effects alone driving change in a cognitive–functional composite.

Here we focused on the statistical development of novel composites based on quantitative assessments to select components of preexisting measures for improved efficiency in clinical trials, followed by a statistical evaluation of their performance. Although we did not undertake a comprehensive psychometric assessment in this work, we assessed features relevant to a psychometric approach in the course of this work or as supplemental analyses.

Internal responsiveness, representing the ability of a measure to change over time, was measured using standardized response means, calculated as the ratio of the mean change from baseline to the SD of the change from baseline for a given group [35]. Standardized response means of the novel composites (this corresponds to the standardized mean changes in Fig. 3) suggest the internal responsiveness for these measures was moderate to large, comparable with CDR-SB and superior to ADAS-Cog 11 or ADAS-Cog 13. Floor and ceiling effects for the individual components have been described previously. Factor analysis, with principal factors extraction with promax rotation, a type of non-orthogonal rotation, on the baseline and 2-year change from baseline values of the components of each composite, was done to explore the structural validity of the novel composites. The results are presented in Supplemental Table 2. Unlike findings in Coley et al. [36] on a different cohort, the cognitive and functional components of our composites show mixed loadings between the two factors for both the MCI cohorts and the AD cohort. Interestingly, the components of CDR-SB (results not shown) for these cohorts in ADNI also show mixed loadings between the two factors, suggesting that this may reflect cohort-related differences between the ADNI cohorts and the cohort considered in Coley et al. [36]. Convergent validity was assessed using Spearman's correlations at baseline between the composites and the components, as well as reference measures MMSE, ADAS-Cog 11, CDR-SB, and FAQ in the MCI and AD cohorts (Supplemental Table 3). External responsiveness was assessed in part using Spearman's correlations of 2-year

changes from baseline between the composites and reference measures MMSE, ADAS-Cog 11, CDR-SB, and FAQ. In both cases, moderate to high correlations (0.46–0.82) were noted (Supplemental Table 3), although some exceptions were noted for ADAS-Cog 11 and MMSE with functional composite CFC3, and for CDR-SB and FAQ with cognitive composite CC1 in the MCI cohorts. Overall, these exploratory analyses suggest adequate, if preliminary, validity of our novel composites, and support their further development and psychometric characterization.

In these analyses, we used raw scores standardized against ADNI population baseline values, as described in Section 2.2.2, rather than normative scores used typically for cognitive measures in psychometric literature. Our interest is primarily in longitudinal change from baseline, and because subjects act as their own control, the need for normative scores is diminished. Furthermore, our approach to standardization allows for comparison across all measures—cognitive and functional—and potentially biomarker endpoints as well, whereas this is not possible given the limited number of normative datasets that exist for only cognitive endpoints.

The approach presented here can be applied to data from proof-of-concept or Phase II trials to facilitate the selection of optimal endpoints to take forward into later stages of testing. Furthermore, our proposed composites can be analyzed as primary endpoints without altering current test administration procedures or materials. That is to say, any of the novel composites can be calculated from scores of ADAS-Cog or other endpoints administered intact. Interpretability of the findings, as well as clinical meaningfulness, can also be addressed further by specifying key components of the composites and clinical measures as secondary endpoints.

We confirmed that a prespecified MCI sample enrichment strategy, using a threshold of CSF A $\beta_{1-42} \leq 192$ pg/mL, can be used in conjunction with our novel endpoints to enhance trial design. Donohue et al. [21] also show gains for ADAS-Cog 11 (about 40%) and CDR-SB (about 70%) with an MCI-A β cohort using simulations. Using an enrichment strategy based on the A β pathology yields a 20% to 40% increase in power, regardless of the outcome measure considered.

Several potential limitations of this work pertain to our methodology. Our method does not penalize redundant or correlated items. Two-year changes in ADAS-3 correlate moderately with those in MMSE ($r = -0.52$) and AVLT Immed ($r = -0.47$), and with the cognitive component of CDR-SB ($r = 0.49$). Although such correlations could produce increased variability, interrogating the same deficit through multiple instruments—some quantitative and some clinical—should provide better overall clinical accuracy in assessing that deficit, especially given that these instruments have large variability in general. Variance reduction techniques could be used to gain additional statistical efficiencies in the composites; however, this could come at the cost of lost interpretability. We undertook a bootstrap validation to assess the reproducibility of the standardized mean changes of the selected measures as well as their composites.

This assessment demonstrated clearly the robustness of both, in populations similar to the ADNI MCI population. Understanding the limits of scales is also of interest. The scales perform moderately well into mild AD, but the performance in moderate AD, for which they are not adapted, is not known. Similarly, the performance for early MCI has not been assessed rigorously because subjects with early MCI were not available in ADNI at the time of writing.

We plan next to assess the impact of covariates on each of the proposed endpoints and undertake analyses based on Item Response Theory (IRT) models to understand the clinical significance of unit changes for the novel scales and robustness of the extreme values for each scale. Incorporation of other biomarkers into the composites may also increase power by reducing variability (data not shown), suggesting that development of a composite scale that captures cognitive, functional, and a biomarker with pathological relevance may be possible. Such a global composite might further advance clinical trial methodology and provide an important link to the development of surrogate endpoints.

In conclusion, use of novel cognitive composites based on a reduced ADAS-3 results in significant gains in power over ADAS-Cog 11 and ADAS-Cog 13 for MCI trials. Larger gains are achieved by using composites of cognition and function, the latter based on FAQ, which is more sensitive to mild functional deficits in MCI. Last, a dual strategy of enrichment combined with improved endpoints would thus result in cumulatively improved power for MCI trials.

Acknowledgments

The authors thank Harry Chen, of J&J Medical China, and Chihshan (Sandy) Lei, of Johnson & Johnson PRD, for generating analysis-ready files from the Alzheimer's Disease Neuroimaging Initiative (ADNI) data. The authors also thank Victor Lobanov, Rudi Verbeeck, and Tim Schultz of the Informatics Center of Excellence. Johnson & Johnson PRD for curation and data management.

Data collection and sharing for this project was funded by ADNI (National Institutes of Health [NIH] 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; AstraZeneca AB; Bayer Schering Pharma AG; Bristol-Myers Squibb; Eisai Global Clinical Development; Elan Corporation; Eli Lilly and Company; F. Hoffman-La Roche; GE Healthcare; Genentech; GlaxoSmithKline; Innogenetics; Johnson and Johnson; Medpace, Inc.; Merck and Company, Inc.; Novartis AG; Pfizer, Inc.; Schering-Plough; Synarc, Inc.; as well as the nonprofit partners the Alzheimer's Association and Alzheimer's Drug Discovery Foundation, with participation from the U.S. Food and Drug Administration. Private-sector contributions to ADNI 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 NeuroImaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514, and the Dana Foundation.

References

- [1] Salloway S, Sperling R, Gilman S, Fox NC, Blennow K, Raskind M, et al. A phase 2 multiple ascending dose trial of bapineuzumab in mild to moderate Alzheimer disease. *Neurology* 2009;73:2061–70.
- [2] Siemers ER, Friedrich S, Dean RA, Gonzales CR, Farlow MR, Paul SM, et al. Safety and changes in plasma and cerebrospinal fluid amyloid β after a single administration of an amyloid β monoclonal antibody in subjects with Alzheimer disease. *Clin Neuropharmacol* 2010;33:67–73.
- [3] Ostrowitzki S, Deptula D, Thurfjell L, Barkhof F, Bohrmann B, Brooks DJ, et al. Mechanism of amyloid removal in patients with Alzheimer disease treated with gantenerumab. *Arch Neurol* 2011; 69:198–207.
- [4] Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry* 1984;141:1356–64.
- [5] 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.
- [6] Birks J. Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database Syst Rev* 2006;25:CD005593.
- [7] Samtani MN, Farnum M, Lobanov V, Yang E, Raghavan N, DiBernardo A, et al. An improved model for disease progression in subjects from Alzheimer's Disease Neuroimaging Initiative. *J Clin Pharmacol* 2012;52:629–44.
- [8] 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.
- [9] Samtani MN, Raghavan N, Shi Y, Novak G, Farnum M, Lobanov V, et al. Disease progression model in MCI subjects from Alzheimer's Disease Neuroimaging Initiative: CSF biomarkers predict population subtypes. *Br J Clin Pharmacol*. 2012. doi: 10.1111/j.1365-2125.2012.04308.x.
- [10] 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.
- [11] Aisen PS, Schafer KA, Grundman M, Pfeiffer E, Sano M, Davis KL, et al. Effects of rofecoxib or naproxen vs. placebo on Alzheimer disease progression: a randomized controlled trial. *JAMA* 2003;289:2819–26.
- [12] Gilman S, Koller M, Black RS, Jenkins L, Griffith SG, Fox NC, et al. Clinical effects of Abeta immunization (AN1792) in patients with AD in an interrupted trial. *Neurology* 2005;64:1553–62.
- [13] Leber P. Guidelines for the clinical evaluation of antidementia drugs. First draft. Food and Drug Administration, 1990.
- [14] European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on medicinal products for the treatment of Alzheimer's disease and other dementias. In: European Medicines Agency (EMA) Pre-Authorisation Evaluation of Medicines for Human Use. Doc. ref. CPMP/EWP/553/95 Rev. 1. London, UK: 2008.
- [15] Grundman M, Petersen RC, Bennett DA, Feldman HH, Salloway S, Visser PJ, et al. Alzheimer's Association Research roundtable meeting on mild cognitive impairment: what have we learned? *Alzheimer Dement* 2006;2:220–33.
- [16] Available at: The Alzheimer's Disease Neuroimaging Initiative (ADNI), <http://adni.loni.ucla.edu/about-data-samples>. Accessed November 10, 2010.
- [17] Available at: The Alzheimer's Disease Neuroimaging Initiative (ADNI), <http://adni.loni.ucla.edu/about-data-samples>. Accessed November 10, 2010.
- [18] Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's Disease Neuroimaging Initiative subjects. *Ann Neurol* 2009;65:403–13.
- [19] ADNI General Procedures Manual. Available at: http://adni.loni.ucla.edu/wp-content/uploads/2010/09/ADNI_GeneralProceduresManual.pdf. Accessed November 4, 2010.
- [20] R Development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Available at: <http://www.R-project.org>. Accessed April 13, 2011.
- [21] Donohue MC, Gamst AC, Thomas RG, Beckett L, Petersen RC, Weiner MW, et al. The relative efficiency of time-to-threshold and rate of change in longitudinal data. *Contemp Clin Trials* 2011;32:685–93.
- [22] Donohue MC, Gamst AC, Aisen PS. Correspondence about "Requiring an amyloid biomarker for prodromal Alzheimer's disease or mild cognitive impairment does not lead to more efficient clinical trials.". *Alzheimer Dement* 2011;7:245–6.
- [23] Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, et al. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med* 2005;352:2379–88.
- [24] Grundman M, Petersen RC, Ferris SH, Thomas RG, Aisen PS, Bennett DA, et al. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Arch Neurol* 2004;61:59–66.
- [25] Sevigny JJ, Peng Y, Liu L, Lines CR. Item analysis of ADAS-Cog: effect of baseline cognitive impairment in a clinical AD trial. *Am J Alzheimer Dis Other Dement* 2010;25:119–24.
- [26] Schoenberg MR, Dawson KA, Duff K, Patton D, Scott JG, Adams RL. Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. *Arch Clin Neuropsychol* 2006;21:693–703.
- [27] Stern Y, Jacobs DM. Preliminary findings from the predictors study: utility of clinical signs for predicting disease course. *Alzheimer Dis Assoc Disord* 1995;9:S14–8.
- [28] Doody RS, Massman P, Dunn JK. A method for estimating progression rates in Alzheimer disease. *Arch Neurol* 2001;58:449–54.
- [29] Doody R, Pavlik V, Massman P, Rountree S, Darby E, Chan W. Predicting progression in Alzheimer's disease. *Alzheimer Res Ther* 2010;2:2.
- [30] Offen W, Chuang Stein C, Dmitrienko A, Littman G, Maca J, Meyerson L, et al. Multiple co-primary endpoints: medical and statistical solutions: a report from the Multiple Endpoints Expert Team of the Pharmaceutical Research and Manufacturers of America. *Drug Info J* 2007;41:31–46.
- [31] Aisen PS, Andrieu S, Sampaio C, Carrillo M, Khachaturian ZS, Dubois B, et al. Report of the task force on designing clinical trials in early (predementia) AD. *Neurology* 2011;76:280–6.
- [32] Cullum CM, Saine K, Davis-Chan L, Martin-Cook K, Gray KF, Weiner M, et al. A performance-based instrument to assess functional capacity in dementia: the Texas Functional Living Scale. *Neuropsychiatr Neuropsychol Behav Neurol* 2001;14:103–8.
- [33] Tractenberg RE, Weiner MF, Patterson MB, Teri L, Thal LJ. Emergent psychopathology in AD patients over 12 months associated with functional, not cognitive, changes. *J Geriatr Psychiatry Neurol* 2002;15:110–7.
- [34] Tractenberg RE, Weiner MF, Cummings JL, Patterson MB, Thal LJ. Independence of changes in behavior from cognition and function in community-dwelling persons with Alzheimer's disease: a factor analytic approach. *J Neuropsychiatry Clin Neurosci* 2005;17:51–60.
- [35] Husted JR, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;53:459–68.
- [36] Coley N, Andrieu S, Jaros M, Weiner M, Cedarbaum J, Vellas B. Suitability of the Clinical Dementia Rating–Sum of Boxes as a single primary end point for Alzheimer's disease trials. *Alzheimer Dement* 2011;7:602–6102.