Alzheimer's & Dementia 11 (2015) 404-414

Development of a straightforward and sensitive scale for MCI and early AD clinical trials

Yifan Huang[a,*], Kaori Ito[a], Clare B. Billing, Jr.,[a,b], Richard J. Anziano[a], for the Alzheimer's Disease Neuroimaging Initiative

[a]*Pfizer, Inc., Groton, CT, USA*
[b]*SystaMedic, Inc., Groton, CT, USA*

Abstract

Background: Although the Clinical Dementia Rating Scale-Sum of Boxes score (CDR-SB) is a widely accepted and commonly used global scale, validated clinical endpoints of cognitive changes are unavailable in the predementia stages of Alzheimer's disease (AD), and a new clinical assessment with reliability and sensitivity is needed in the mild cognitive impairment (MCI) population.

Methods: Using Alzheimer's Disease Neuroimaging Initiative (ADNI)-1/GO data, signal-to-noise ratios (SNRs) were calculated to quantify the sensitivity of a measure for detecting disease progression and hypothetical treatment effects. All possible combinations of selected sensitive measures were assessed for developing composite scores. The analyses were performed in the MCI population and subpopulations enriched by apolipoprotein E4 (*APOE* ε4), hippocampal volume, and cerebrospinal fluid β-amyloid.

Results: The best composite score was "Word Recall + Delayed Word Recall + Orientation + CDR-SB + FAQ", more sensitive than 13-item Alzheimer's Disease Assessment Scale-cognitive subscale or CDR-SB.

Conclusion: The proposed composite score derived from the existing clinical endpoints demonstrated higher sensitivity in the MCI population and is easy to implement and standardize across studies.

Keywords: Alzheimer's disease; Mild cognitive impairment; Alzheimer's Disease Assessment Scale-cognitive subscale; Composite endpoints; Clinical trials; Signal-to-noise ratio

1. Introduction

Alzheimer's disease (AD) is a continuously growing affliction worldwide and the fifth leading cause of death for individuals age 65 and older; however, no disease-modifying treatment is currently available. By 2025, the number of people age 65 and older with AD is estimated to reach 7.1 million—a 40% increase from the 5 million aged 65 and older currently affected in the United States [1].

Recent research suggests that AD begins many years before the development of symptoms such as memory loss or behavioral changes, and that new technologies have the potential to identify brain changes that precede the development of symptoms. In 2011, the National Institute of Aging (NIA) and the Alzheimer's Association proposed new criteria and guidelines for diagnosing AD [2–5], which identified three stages of AD. The new diagnostic criteria require the presence of biological evidence demonstrated by imaging or cerebrospinal fluid (CSF) biomarkers. The new guidelines also recognize the importance of identifying and treating patients at the early stage before symptoms develop. If AD can be detected earlier, then there may be a better chance for disease-modifying treatment to become successful [6].

The 11-item Alzheimer's Disease Assessment Scale-cognitive subscale ($ADAS\text{-}cog_{11}$) has been the most

*Corresponding author. Tel.: 860-686-2616.
E-mail address: yifan.huang@pfizer.com

1552-5260/$ - see front matter
http://dx.doi.org/10.1016/j.jalz.2014.03.008

commonly used outcome measure of cognitive function in antidementia clinical trials and has been an effective outcome in approval of symptomatic agents in mild to moderate AD patients. However, ADAS-cog$_{11}$ is not sensitive enough to detect changes in predementia stages because of significant ceiling effects. Although the Clinical Dementia Rating Scale-Sum of Boxes score (CDR-SB) is a widely accepted and commonly used global scale, there are no generally accepted, validated clinical endpoints of cognitive changes for use in therapeutic trials in the predementia stages of AD.

Given the current state of clinical research in AD, a new clinical assessment scale would ideally address the following criteria: (1) being reliable and valid in patients with mild cognitive impairment (MCI) due to AD because most current therapeutic interventions are directed at AD pathology; (2) being sensitive to clinical decline and potential treatment effects that might only manifest as a reduction in the rate of decline because most current therapeutic interventions are directed at disease modification; (3) maintaining good reliability, validity, and sensitivity once patients have progressed to mild AD dementia because most current clinical trials extend over very long durations (e.g., 18–24 months); (4) including cognitive and functional assessments so that it could be used as a single primary composite efficacy measure in line with the recent U.S. Food and Drug Administration (FDA) draft guidance [7] for clinical trials in early stages of AD.

There currently exist several precompetitive initiatives and consortia that are pursuing new assessments to address these criteria. In addition, several pharmaceutical companies have ongoing efforts toward the development of sensitive clinical endpoints for trials in MCI and early AD [8–13]. In this report we present our work in developing easy-to-use composite endpoints that are sensitive to disease progression and to hypothetical treatment effects of disease-modifying therapies. We quantified the sensitivity using signal-to-noise ratios (SNRs) that measure the strength of the desired signal for detecting the hypothetical treatment effect relative to the level of the variability. The SNRs we proposed are in essence the standardized mean changes, which are useful measures of effect size in designing the clinical trials.

We examined the SNRs for various cognitive, functional, and global measures, including their individual components. Composite scores were derived based on the most sensitive measures, with a similar hypothesis as in some of the previous work [12] that by combining the most informative scales we could create a composite with stronger signal and reduced variability for the early AD population. In addition to the MCI population, subpopulations enriched by apolipoprotein E (*APOE*) status, baseline hippocampal volume, and baseline CSF β-amyloid (Aβ) were also analyzed.

We compared the SNRs of the derived composite scales to the existing measures such as ADAS-cog 11-item or 13-item total score and CDR-SB. We also examined the similarity between the measures selected to derive our composites and those selected for the previously reported composite scores and to the items selected based on the Item Response Theory (IRT) [14,15] by other researchers.

2. Methods

2.1. Data

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). Analyses were performed on the data downloaded on June 18, 2013 from the ADNI web portal (http://adni.loni.ucla.edu/data-samples/access-data/). All data points available for MCI subjects enrolled in ADNI-1 and followed through ADNI-GO were included in this analysis. Detailed ADNI protocol information can be found at www.adni-info.org.

2.2. Subject population

Three hundred and ninety-seven MCI subjects were included in this analysis. Among them, 312 (78.6%) subjects had their hippocampal volume measured at baseline and were considered for the subgroup analysis by baseline hippocampal volume. The 198 (49.9%) subjects with baseline CSF A$\beta_{1\text{-}42}$ levels were considered for the subgroup analysis using recently proposed CSF cutoffs [16]. All 397 subjects had *APOE* information collected at baseline and were considered for the subgroup analysis by *APOE* status. These 397 subjects were followed every 6–12 months from baseline to beyond 3 years. At the time of the data download, 357, 302, and 260 subjects had clinical data available for the 12-, 24-, and 36-month visits, respectively.

2.3. Clinical measures

In this analysis, we considered all of the cognitive, functional, and behavioral assessments that are available in ADNI-1 and ADNI-GO, including ADAS-cog, Mini-Mental State Examination (MMSE), CDR, Functional Assessment Questionnaire (FAQ), Clock Drawing Test, Rey Auditory Verbal Learning Test (AVLT), Logical Memory Test, Digit Span Test, Category Fluency Test, Trail Making Test, Digit Symbol Substitution Test, and Boston Naming Test. Total scores and individual item scores were evaluated.

2.4. Analysis methods

To quantify the sensitivity of a clinical measure, we examined SNRs associated with two different definitions of hypothetical treatment effect. The SNRs proposed are in essence the standardized mean changes, which are the average changes from baseline standardized by the variability of the measures. These quantities are associated with the effect sizes used in designing the clinical trials and can directly influence the sample size and power. The applicability of these SNRs in a clinical trial setting relies

on the assumptions that link the disease progression to the hypothetical treatment effects. One key assumption is that by using an endpoint that is sensitive to measuring disease progression in the untreated control group of a clinical trial, we will be able to sensitively compare that group with a treated group that has hypothetically slow or no progression.

The first hypothetical treatment effect was defined by assuming that the disease-modifying therapy can slow down the disease progression at a certain time point by a certain percentage (for example, 50% at 2 years). Given this assumption, a clinical scale or endpoint that is sensitive to the progression of the disease is also sensitive to the hypothetical treatment effect. Therefore, the following SNR is used in the assessment of sensitivity for this hypothetical treatment effect. Let δ_t be the mean change from baseline at time t (for example, $t = 1, 2, 3$ years). We define the SNR at time t as

$$SNR_t = |\delta_t| / SD(|\delta_t|),$$

where $SD(|\delta_t|)$ is the standard deviation of δ_t. SNR_t reflects the relative strength of the clinical measures to detect disease progression over time and thus to detect the first hypothetical treatment effect. The effect size used in designing a trial is proportional to SNR_t. For example, if a treatment is assumed to slow down the progression by 50% at 2 years, the effect size of this treatment can be calculated as SNR_2*50%. A more sensitive endpoint with a larger SNR_2 will result in a larger effect size for the treatment and thus a smaller sample size when designing the clinical trial.

The second hypothetical treatment effect was defined by assuming that the disease-modifying therapy can delay the disease progression by a certain period of time. For this hypothesis, we focused on the scenario in which the treatment can delay the progression of the disease by 1 year. Under this hypothesis, MCI subjects with the treatment at 2 years should have the same disease assessments as those without the treatment at 1 year. For the ADNI data, if we randomly divide the MCI population into two groups, then the difference in disease assessments between Group 1 at 1 year and Group 2 at 2 years represents the hypothetical treatment effect. Therefore, a clinical measure that is sensitive to this difference is sensitive to the hypothetical treatment effect. To estimate the hypothetical treatment effect and quantify the sensitivity, the following SNR is used. Let $\Delta = |\delta_2| - |\delta_1|$, where δ_1 is the mean change from baseline at 1 year in Group 1 and δ_2 is the mean change from baseline at 2 years in Group 2. We define the SNR for detecting the effect of 1-year delay in disease progression as

$$SNR_{2-1} = \Delta / SD(\Delta),$$

where $SD(\Delta) = \sqrt{var(|\delta_1|) + var(|\delta_2|)}$ because the two groups are independent. To obtain the estimate for this SNR, we applied a resampling approach as follows. For each iteration, we randomly divided the MCI subjects in ADNI into two groups with equal allocation and then calculated the $SNR_{2\text{-}1}$. We repeated this process many times (we used 10,000 iterations in the analysis). The point estimate was obtained by averaging the $SNR^b_{2\text{-}1}$, $b = 1, 2, \ldots, B$, where B is the number of total iterations. Similar to SNR_t, a more sensitive endpoint with a larger $SNR_{2\text{-}1}$ will result in a larger effect size for the same treatment.

For each clinical measure (including the individual components) described in Section 2.3, we computed SNR_1, SNR_2, and SNR_3 (SNR at 1, 2, and 3 years, respectively) and the $SNR_{2\text{-}1}$ (assuming a hypothetical treatment effect of 1-year delay in disease progression) on the basis of the ADNI data in the MCI population described in Section 2.2. Larger SNRs indicate more sensitive measures. We mainly focused on SNR_2 and $SNR_{2\text{-}1}$ because our goal was to identify the sensitive measures for early-phase clinical trials that typically examine the endpoint at 24 months; however, SNR_1 and SNR_3 were also calculated to evaluate the consistency in sensitivity over time. Candidate measures with high SNR values were selected for developing composite scores. We developed composite endpoints by combining these candidate measures. Simple sum of the individual scores from each candidates were used to generate the composite scores. For each composite, the SNRs were computed to compare the performance or sensitivity in the target population described in Section 2.2.

We also assessed the development of composite scores in the enriched MCI populations. Three subpopulations were examined: (1) *APOE* ε4 allele carriers; (2) subjects with low hippocampal volume (i.e., baseline volume ≤ 6700 mm^3, which is approximately the median value in the data set); (3) Aβ-positive subjects (i.e., baseline CSF $A\beta_{1-42} \leq 192$ pg/mL [16]). These subpopulations have been considered to have higher risk of disease progression.

3. Results

3.1. Baseline characteristics

Among the 397 MCI subjects, 212 carried one or two *APOE* ε4 alleles, 191 had a baseline hippocampal volume smaller than the cutoff of 6700 mm^3, and 147 had a baseline CSF $A\beta_{1-42}$ level less than 192 pg/mL. The baseline characteristics of the overall, biomarker-enriched, and biomarker-negative MCI populations are summarized in Table 1. Our focus was on the overall and biomarker-enriched MCI populations, and the biomarker-negative MCI populations are included in Table 1 for completeness. There was no clinically meaningful difference in age, MMSE score, CDR-SB score, or AVLT Trial 1–5 total score among the overall and enriched MCI populations. The ADAS-cog and FAQ total scores were higher in the enriched populations. The enriched population with a low hippocampal volume had more women than the other populations. All three enriched populations had more *APOE* ε4 carriers than the overall MCI population in this analysis.

Table 1
Baseline characteristics of MCI populations

Characteristics	All MCI	*APOE* ε4		Hippocampal volume		Aβ	
		Carrier	Noncarrier	Low*	High	Positive†	Negative
n	397	212	185	191	121	147	51
Age	74.7 (7.40)	73.9 (6.72)	75.8 (8.0)	75.8 (6.60)	71.3 (7.52)	74.5 (7.04)	74.4 (8.65)
Female	141 (35.5%)	79 (37.3%)	62 (33.5%)	86 (45.0%)	29 (24.0%)	52 (35.4%)	14 (27.5%)
APOE ε4 (alleles)							
0	185 (46.6%)	0 (0%)	185 (100%)	69 (36.1%)	65 (53.7%)	53 (36.1%)	39 (%)
1	165 (41.6%)	165 (77.8%)	0 (0%)	90 (47.1%)	48 (39.7%)	73 (49.7%)	12 (%)
2	47 (11.8%)	47 (22.2%)	0 (0%)	32 (16.8%)	8 (6.6%)	21 (14.3%)	0 (0%)
MMSE	27.0 (1.78)	26.9 (1.79)	27.1 (1.76)	26.7 (1.72)	27.4 (1.81)	26.8 (1.79)	27.3 (1.76)
ADAS-cog$_{11}$	11.5 (4.42)	12.3 (4.30)	10.5 (4.4)	12.6 (4.48)	9.9 (4.11)	12.2 (4.55)	10.1 (4.37)
ADAS-cog$_{13}$	18.6 (6.27)	19.9 (5.94)	17.2 (6.4)	20.3 (6.06)	16.0 (6.02)	19.9 (6.06)	16.1 (6.35)
CDR-SB	1.6 (0.86)	1.7 (0.90)	1.5 (0.86)	1.6 (0.90)	1.5 (0.74)	1.6 (0.92)	1.3 (0.72)
FAQ	3.8 (4.47)	4.3 (4.73)	3.3 (4.10)	4.4 (4.69)	2.9 (4.03)	4.3 (4.40)	2.4 (4.25)
AVLT (five-item total)	30.7 (9.03)	29.4 (8.24)	32.2 (9.7)	29.4 (8.26)	32.7 (10.28)	29.1 (8.13)	33.2 (9.31)

Abbreviations: MCI, mild cognitive impairment; *APOE*, apolipoprotein E; Aβ, β-amyloid; MMSE, Mini-Mental State Examination; ADAS-cog$_{11}$, 11-item Alzheimer's Disease Assessment Scale-cognitive subscale; ADAS-cog$_{13}$, 13-item Alzheimer's Disease Assessment Scale-cognitive subscale; CDR-SB, Clinical Dementia Rating-Sum of Boxes; FAQ, Functional Assessment Questionnaire; AVLT, Auditory Verbal Learning Test; CSF, cerebrospinal fluid.

NOTE. Mean (SD) or *n* (%).

*Baseline hippocampal volume $\leq$ 6700 mm^3.

†Baseline CSF A$\beta_{1-42} \leq$ 192 pg/mL.

3.2. Measures selected for development of composite scores

The SNR_1, SNR_2, SNR_3, and SNR_{2-1} were computed for all individual items and are summarized in Table 2 for the MCI population. For the enriched populations, Table 3 presents the SNR_2 and SNR_{2-1}, assuming that a 24-month study is of primary interest.

Because we were trying to develop composites that are more sensitive than the ADAS-cog or CDR-SB and we were focusing on studies with primary endpoints at 18–24 months, we selected candidate measures with SNR_2 benchmarked against ADAS-cog. A threshold of 0.45, 0.50, and 0.55 was used for the overall MCI population, the *APOE*-enriched population, and the other two biomarker-enriched populations, respectively. Clinical measures with SNR_2 larger than the thresholds were included in the development of composite scores. There were two exceptions: the Delayed Word Recall and the other AVLT Trial 1–5 total score. Their SNR_2 in some populations was just below the threshold value, but we included them in the composite score development because episodic memory impairment is a core feature of AD and amnestic MCI and previous research studies have shown that episodic memory tests were useful for identifying MCI patients with a high likelihood of progressing to AD dementia [17,18].

The measures selected as candidates for the composite score were ADAS-cog$_{11}$ total score, 13-item Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog$_{13}$) total score, Word Recall (Q1), Delayed Word Recall (Q4), Orientation (Q7), MMSE, CDR-SB, FAQ total score, and AVLT Trial 1–5 total score (bold font, Table 2). The same measures also showed high sensitivity in the enriched populations (Table 3). These measures were considered in the development of composite scores.

Some subitems of CDR and FAQ are as sensitive as the total score, but we considered only the total score for the composite because illnesses such as dementia may not always progress uniformly in all subdomains of CDR and FAQ, and the total score is not difficult to obtain.

3.3. Composite scores

After examining the SNRs of all of the possible combinations of the candidate measures, the composite scores by combining ADAS-cog individual items (Word Recall, Delayed Word Recall, and Orientation) with CDR-SB and FAQ were identified as being the most sensitive to the disease progression and hypothetical treatment effects. The SNR values for the top-ranked composites and other combinations of interest are presented in Table 4 and Fig. 1. The following observations are similar across the four populations (the MCI population and the three enriched subpopulations) that we examined.

The composite score derived by summing up the scores of Word Recall (Q1), Delayed Word Recall (Q4), Orientation (Q7), CDR-SB, and FAQ was the most sensitive, with much higher sensitivity than the ADAS-cog$_{13}$ total score. CDR-SB alone was more sensitive than the ADAS-cog$_{13}$ total score, but our proposed composite improved the sensitivity even more.

CDR-SB or FAQ seemed to play an important role in enhancing the sensitivity of composite scores because the composite of summing Word Recall, Delayed Word Recall, and Orientation scores without adding CDR-SB or FAQ had much lower sensitivity compared with the top-ranked

Table 2
SNRs for clinical measures in MCI population

Assessment	Description	SNR_1	SNR_2	SNR_3	$SNR_{2\text{-}1}$
ADAS-cog	**ADAS-cog$_{11}$ total score**	**0.20**	**0.52**	**0.60**	**0.29**
	ADAS-cog$_{13}$ items total score	**0.28**	**0.56**	**0.65**	**0.29**
	Word Recall (Q1)	**0.27**	**0.47**	**0.55**	**0.18**
	Commands (Q2)	0.07	0.09	0.25	0.06
	Construction (Q3)	0.04	0.04	0.18	0.05
	Delayed Word Recall (Q4)	**0.27**	**0.42**	**0.50**	**0.12**
	Naming (Q5)	0.03	0.10	0.16	0.07
	Ideational Praxis (Q6)	0.05	0.12	0.29	0.07
	Orientation (Q7)	**0.25**	**0.50**	**0.64**	**0.23**
	Word Recognition (Q8)	0.03	0.26	0.30	0.15
	Recall Instructions (Q9)	0.03	0.10	0.24	0.07
	Spoken Language (Q10)	0.13	0.25	0.24	0.13
	Word Finding (Q11)	0.19	0.23	0.37	0.06
	Comprehension (Q12)	0.12	0.16	0.32	0.07
	Number Cancellation (Q14)	0.09	0.35	0.31	0.32
MMSE	**Score**	**0.26**	**0.49**	**0.61**	**0.25**
CDR	**Sum of Boxes Score**	**0.55**	**0.74**	**0.76**	**0.35**
	Memory	0.29	0.51	0.61	0.28
	Orientation	0.45	0.65	0.68	0.27
	Judgment and Problem Solving	0.30	0.55	0.61	0.25
	Community Affairs	0.44	0.68	0.76	0.31
	Home and Hobbies	0.40	0.57	0.69	0.20
	Personal Care	0.08	0.27	0.33	0.18
FAQ	**Total Score**	**0.47**	**0.73**	**0.88**	**0.34**
	Financial	0.35	0.57	0.71	0.21
	Forms	0.33	0.57	0.68	0.22
	Shopping	0.25	0.51	0.73	0.25
	Game	0.20	0.45	0.56	0.23
	Beverage	0.15	0.21	0.40	0.09
	Meal	0.30	0.48	0.65	0.22
	Event Tracking	0.21	0.36	0.51	0.14
	TV	0.15	0.40	0.46	0.22
	Remembering	0.19	0.43	0.55	0.18
	Travel	0.30	0.57	0.71	0.26
Clock Drawing Test	Drawing Administration	0.10	0.15	0.28	0.07
	Copying Administration	0.05	0.09	0.21	0.06
AVLT	**Trial 1–5 Total Score**	**0.20**	**0.38**	**0.60**	**0.16**
	Trial 1 Administration	0.04	0.15	0.40	0.09
	Trial 2 Administration	0.08	0.26	0.36	0.14
	Trial 3 Administration	0.19	0.35	0.56	0.13
	Trial 4 Administration	0.13	0.29	0.49	0.12
	Trial 5 Administration	0.21	0.34	0.55	0.11
	Trial 6 Administration	0.06	0.19	0.40	0.11
	List B Administration	0.10	0.24	0.25	0.10
	Delayed Administration	0.19	0.25	0.36	0.07
	Recognition Administration	0.16	0.31	0.43	0.13
Logical Memory Test	Immediate Recall	0.13	0.01	0.13	0.11
	Delayed Recall	0.20	0.15	0.10	0.06
Digit Span Test	Forward	0.06	0.16	0.16	0.09
	Backward	0.16	0.18	0.20	0.05
Category Fluency Test	Animal Total	0.15	0.30	0.47	0.12
	Vegetable Total	0.20	0.33	0.56	0.13
Trail Making Test	Part A	0.07	0.22	0.34	0.13
	Part B	0.16	0.25	0.36	0.10
Digit Symbol Substitution Test	Score	0.10	0.25	0.50	0.14
Boston Naming Test	Score	0.08	0.19	0.41	0.09

Abbreviations: SNR, signal-to-noise ratio; ADAS-cog, Alzheimer's Disease Assessment Scale-cognitive subscale; ADAS-cog$_{11}$, 11-item ADAS-cog; ADAS-cog$_{13}$, 13-item ADAS-cog; CDR-SB, Clinical Dementia Rating-Sum of Boxes; FAQ, Functional Assessment Questionnaire; AVLT, Auditory Verbal Learning Test; SNR_1, SNR_2, and SNR_3, SNR at 1, 2, and 3 years, respectively; $SNR_{2\text{-}1}$, SNR assuming a hypothetical treatment effect of 1-year delay in disease progression.

NOTE. Bold indicates selected candidates.

Table 3
SNRs for clinical measures in enriched populations

Assessment	Description	*APOE* ε4 carrier		Low hippocampal volume*		Aβ positive†	
		SNR_2	$SNR_{2\text{-}1}$	SNR_2	$SNR_{2\text{-}1}$	SNR_2	$SNR_{2\text{-}1}$
ADAS-cog	**ADAS-cog$_{11}$ total score**	**0.56**	**0.34**	**0.67**	**0.30**	**0.68**	**0.39**
	ADAS-cog$_{13}$ total score	**0.60**	**0.32**	**0.74**	**0.32**	**0.73**	**0.38**
	Word Recall (Q1)	**0.56**	**0.22**	**0.61**	**0.23**	**0.55**	**0.22**
	Commands (Q2)	0.03	0.08	0.06	0.08	0.10	0.09
	Construction (Q3)	0.07	0.07	0.05	0.11	0.01	0.10
	Delayed Word Recall (Q4)	**0.48**	**0.14**	**0.60**	**0.20**	**0.55**	**0.17**
	Naming (Q5)	0.14	0.11	0.08	0.08	0.20	0.16
	Ideational Praxis (Q6)	0.18	0.12	0.13	0.10	0.07	0.09
	Orientation (Q7)	**0.50**	**0.24**	**0.62**	**0.30**	**0.64**	**0.28**
	Word Recognition (Q8)	0.21	0.16	0.30	0.13	0.39	0.22
	Recall Instructions (Q9)	0.12	0.10	0.13	0.08	0.15	0.13
	Spoken Language (Q10)	0.24	0.13	0.28	0.15	0.27	0.12
	Word Finding (Q11)	0.36	0.12	0.29	0.10	0.31	0.12
	Comprehension (Q12)	0.19	0.08	0.24	0.08	0.15	0.11
	Number Cancellation (Q14)	0.42	0.33	0.43	0.34	0.51	0.37
MMSE	**Score**	**0.63**	**0.29**	**0.66**	**0.30**	**0.62**	**0.24**
CDR	**Sum of Boxes Score**	**0.87**	**0.41**	**0.95**	**0.44**	**0.93**	**0.41**
	Memory	0.61	0.29	0.76	0.39	0.63	0.28
	Orientation	0.80	0.34	0.82	0.34	0.81	0.36
	Judgment and Problem Solving	0.63	0.29	0.73	0.31	0.69	0.26
	Community Affairs	0.82	0.39	0.80	0.37	0.93	0.39
	Home and Hobbies	0.67	0.23	0.72	0.24	0.72	0.22
	Personal Care	0.24	0.16	0.34	0.21	0.25	0.19
FAQ	**Total Score**	**0.86**	**0.40**	**0.90**	**0.46**	**0.91**	**0.48**
	Financial	0.63	0.24	0.66	0.25	0.70	0.29
	Forms	0.63	0.29	0.67	0.30	0.63	0.28
	Shopping	0.58	0.28	0.67	0.34	0.64	0.34
	Game	0.50	0.26	0.51	0.24	0.58	0.39
	Beverage	0.28	0.13	0.30	0.15	0.24	0.13
	Meal	0.58	0.26	0.63	0.31	0.61	0.29
	Event Tracking	0.45	0.14	0.47	0.22	0.51	0.21
	TV	0.49	0.23	0.53	0.31	0.42	0.27
	Remembering	0.51	0.17	0.58	0.32	0.47	0.20
	Travel	0.69	0.33	0.73	0.37	0.73	0.37
Clock Drawing Test	Drawing Administration	0.07	0.08	0.17	0.10	0.13	0.11
	Copying Administration	0.07	0.08	0.09	0.08	0.02	0.09
AVLT	**Trial 1–5 Total Score**	**0.56**	**0.25**	**0.55**	**0.27**	**0.50**	**0.23**
	Trial 1 Administration	0.30	0.16	0.27	0.18	0.23	0.15
	Trial 2 Administration	0.35	0.18	0.35	0.19	0.24	0.14
	Trial 3 Administration	0.48	0.19	0.49	0.22	0.48	0.20
	Trial 4 Administration	0.44	0.14	0.37	0.17	0.43	0.19
	Trial 5 Administration	0.47	0.21	0.43	0.18	0.40	0.17
	Trial 6 Administration	0.38	0.20	0.40	0.17	0.36	0.19
	List B Administration	0.32	0.09	0.31	0.11	0.26	0.12
	Delayed Administration	0.38	0.11	0.46	0.16	0.37	0.14
	Recognition Administration	0.47	0.18	0.40	0.16	0.35	0.13
Logical Memory Test	Immediate Recall	0.23	0.13	0.21	0.14	0.21	0.15
	Delayed Recall	0.14	0.11	0.19	0.14	0.13	0.10
Digit Span Test	Forward	0.14	0.11	0.33	0.17	0.24	0.09
	Backward	0.18	0.07	0.13	0.08	0.32	0.09
Category Fluency Test	Animal Total	0.39	0.17	0.56	0.20	0.44	0.17
	Vegetable Total	0.45	0.23	0.34	0.16	0.44	0.23
Trail Making Test	Part A	0.26	0.13	0.28	0.15	0.31	0.18
	Part B	0.41	0.21	0.38	0.14	0.46	0.21
Digit Symbol Substitution Test	Score	0.37	0.19	0.34	0.16	0.41	0.18
Boston Naming Test	Score	0.31	0.10	0.27	0.11	0.31	0.15

Abbreviations: SNR, signal-to-noise ratio; ADAS-cog$_{13}$, 13-item ADAS-cog; CDR-SB, Clinical Dementia Rating-Sum of Boxes; FAQ, Functional Assessment Questionnaire; SNR_1, SNR_2, and SNR_3, SNR at 1, 2, and 3 years, respectively; $SNR_{2\text{-}1}$, SNR assuming a hypothetical treatment effect of 1-year delay in disease progression; CSF, cerebrospinal fluid; Aβ, β-amyloid.

NOTE. Bold indicates selected candidates.

*Baseline hippocampal volume $\leq$ 6700 mm^3.

†Baseline CSF A$\beta_{1-42} \leq$ 192 pg/mL.

Table 4
SNRs for composite scores

Composite	All MCI				*APOE* ε4 carrier				Low hippocampal volume*				Aβ-positive†			
	SNR_1	SNR_2	SNR_3	$SNR_{2\text{-}1}$	SNR_1	SNR_2	SNR_3	$SNR_{2\text{-}1}$	SNR_1	SNR_2	SNR_3	$SNR_{2\text{-}1}$	SNR_1	SNR_2	SNR_3	$SNR_{2\text{-}1}$
Top rank combination vs. some candidate measures																
Q1+Q4+Q7+CDR-SB + FAQ	0.62	0.82	0.93	0.37	0.75	0.93	1.12	0.42	0.70	1.04	1.25	0.51	0.74	1.04	1.24	0.51
CDR-SB	0.55	0.74	0.76	0.35	0.67	0.87	0.96	0.41	0.70	0.95	1.01	0.44	0.72	0.93	0.92	0.41
ADAS-cog_{13}	0.28	0.56	0.65	0.29	0.28	0.60	0.74	0.32	0.41	0.74	0.86	0.32	0.34	0.73	0.81	0.38
Other combinations of interest																
Q1+Q4+Q7+CDR-SB	0.52	0.74	0.82	0.31	0.63	0.83	1.02	0.35	0.69	1.01	1.16	0.44	0.68	0.96	1.12	0.51
Q1+Q4+Q7+FAQ	0.57	0.80	0.92	0.36	0.70	0.91	1.09	0.41	0.63	1.01	1.21	0.50	0.68	1.02	1.25	0.40
Q1+Q4+Q7	0.38	0.61	0.71	0.23	0.45	0.68	0.90	0.25	0.49	0.83	1.00	0.33	0.49	0.78	1.02	0.30

Abbreviations: Aβ, β-amyloid; SNR, signal-to-noise ratio; ADAS-cog_{13}, 13-item ADAS-cog; CDR-SB, Clinical Dementia Rating-Sum of Boxes; FAQ, Functional Assessment Questionnaire; SNR_1 and SNR_2, SNR at 1 and 2 years, respectively; $SNR_{2\text{-}1}$, SNR assuming a hypothetical treatment effect of 1-year delay in disease progression; CSF, cerebrospinal fluid.

*Baseline hippocampal volume $\leq$ 6700 mm^3.

†Baseline CSF $A\beta_{1\text{–}42} \leq$ 192 pg/mL.

composite. However, there could be some overlap between CDR-SB and FAQ because dropping one of them from the top-ranked composite had a smaller effect on sensitivity. In addition, CDR-SB was more sensitive than the composite of summing Word Recall, Delayed Word Recall, and Orientation scores.

The sensitivity of the measures has a direct effect on the sample size of a clinical trial. For the hypothetical treatment effect of slowing down progression by 50% in a 2-year MCI trial (type I error = 0.05, power = 80%), a total sample size of 402, 231, and 189 subjects would be required for ADAS-cog_{13}, CDR-SB, and the proposed composite score, respectively. Therefore, a 53% or 18% saving in sample size could be achieved by using the composite score compared with ADAS-cog_{13} or CDR-SB alone.

The composite score using a subset of ADAS-cog individual items (e.g., summing Word Recall, Delayed Word Recall, and Orientation scores) had comparable sensitivity to the ADAS-cog total score. As the authors noted before [19], some of the ADAS-cog individual items with low sensitivity could be simply adding noise to the detection of disease progression in the MCI population, and a subset of the individual items may be used as a more efficient measure for early AD.

As in each of the clinical measures, the composite scores had higher SNR values in the enriched populations than the MCI population (Table 4, Fig. 1), indicating a faster progression of the disease in these high-risk populations, especially in those with low hippocampal volume or those who were Aβ-positive at baseline.

4. Discussion

In this analysis, SNRs, which are essentially standardized mean changes, were proposed to quantify the relative strength of a measure or instrument to detect disease progression and hypothetical treatment effects. ADNI-1 and ADNI-GO clinical data at 12, 24, and 36 months were

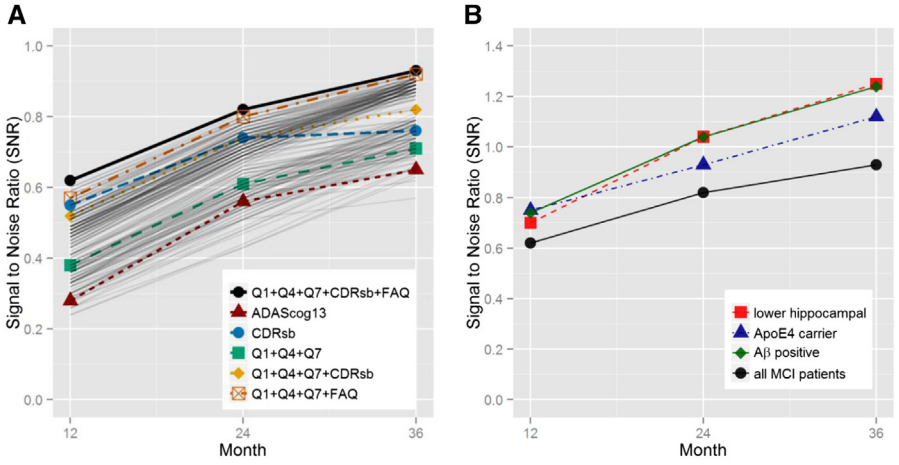

Fig. 1. Time course of signal-to-noise ratio (SNR). (A) The SNRs at 1, 2, and 3 years for all composite scores. The lines without corresponding legends (gray color) are all of the other possible combinations assessed for composite development. (B) Comparison between apolipoprotein ε4 (*APOE* ε4) carrier group, low hippocampal volume group, and cerebrospinal fluid (CSF) β-amyloid (Aβ)-positive group vs. all mild cognitive impairment (MCI) subjects for the best combination (Q1+Q4+Q7+CDR-SB+FAQ). CDR-SB, Clinical Dementia Rating-Sum of Boxes; FAQ, Functional Assessment Questionnaire.

used to identify sensitive instruments and best combinations of the selected instruments in MCI subjects. We did not include ADNI-2 data because only limited data up to 24 months were available with ADNI-2 subjects as of June 18, 2013. The instruments examined in this analysis include ADAS-cog (ADAS-cog$_{11}$, ADAS-cog$_{13}$, and subitems), MMSE, CDR (CDR-SB and subitems), FAQ (total score and subitems), Clock Drawing Test, AVLT (total and subitems), Logical Memory Test, Digit Span Test, Category Fluency Test, Trail Making Test, Digit Symbol Substitution Test, Boston Naming Test (Table 2), and all possible combinations of the most sensitive instruments selected from this list. Highly overlapping combinations in the same domain or category (e.g., ADAS-cog$_{13}$ + Delayed Word Recall in which Delayed Word Recall is a component of ADAS-cog$_{13}$) were excluded from the analysis. The best combination or composite score is Word Recall + Delayed Word Recall + Orientation + CDR-SB + FAQ, in which Word Recall, Delayed Word Recall, and Orientation are from ADAS-cog (Table 3). This proposed composite endpoint includes cognitive, functional, and global instruments; therefore, it is also in line with the recent draft FDA guidance for clinical trials in the early stage of AD [7]. It has higher sensitivity than ADAS-cog$_{13}$ or CDR-SB (Table 4, Fig. 1), indicating improved statistical power for MCI trials.

Our proposed composite contains similar elements as those identified in previously reported analyses. Logovinsky and colleagues [13] proposed "ADCOMS," a composite score derived from ADAS-cog subitems (Delayed Word Recall, Orientation, Word Recognition, and Word Finding), MMSE subitems (orientation and construction praxis), and all of the CDR subitems. The authors applied a linear regression model to data from MCI subjects in ADNI-1 and placebo groups of three clinical trials (NCT00293176, NCT00403520, and ADCS-MCI). The selected items were combined using weighting factors to maximize the sensitivity of the composite. Nandini and colleagues [12] proposed two composite scores on the basis of ADNI-1 clinical data. The first one is "TriAD," a cognitive endpoint combining Word Recall, Delayed Word Recall, Orientation, and CDR cognitive components. The second one is "TriAD-G," a cognitive-functional endpoint that adds FAQ to TriAD. In their analysis, sensitive measures were identified based on standardized mean changes at 2 years and were combined into composite scores. The most sensitive composites were selected based on their performance in terms of sample size reduction for 2-year clinical trials. Comparing to the composite scores proposed by other researchers using the ADNI data, all of the composites included some items from the memory domain of ADAS-cog, and all items from the memory domain of CDR. Most of the other composites did not use the functional assessment FAQ whereas ours and "TriAD-G" by Raghavan and colleagues [12] used FAQ to improve sensitivity. None of the studies that examined all available clinical scales in ADNI has identified Executive Function measures (e.g., Trail Making, Digit Symbol, Digit Span, Category Fluency tests) in the development of sensitive composites. In summary, although each researcher used different statistical methodologies and sometimes different data sets, the cognitive measures in the memory domain were always selected, and adding some global and/or functional assessments could improve the performance of the composite scores.

In addition, the rank of ADAS-cog subitems that was based on the SNRs was consistent with the rank that was based on the Fisher's information used in the IRT analysis (Table 5). Ueckert and colleagues [14,15] calculated the information content for each subitem using data from the ADNI and the Coalition Against Major Diseases (CAMD) databases (2744 patients in total). The top three items (Word Recall, Delayed Word Recall, and Orientation) are the same between our analysis and the IRT analysis. These

Table 5
Rank of sensitive measure selected from SNR analysis compared with IRT analysis [14,15]

	SNR analysis			IRT analysis		
Rank	Component	24 mo + 36 mo	% Total	Component	Information	% Total
1	**Orientation (Q7)**	1.14	15.3	**Delayed Word Recall**	4.82	30.1
2	**Word Recall (Q1)**	1.02	13.7	**Word Recall**	4.10	25.6
3	**Delayed Word Recall (Q4)**	0.92	12.4	**Orientation**	2.02	12.6
4	Number Cancellation (Q14)	0.66	8.9	Word Recognition	1.91	12.0
5	Word Finding (Q11)	0.6	8.1	Naming Objects and Fingers	1.10	6.9
6	Word Recognition (Q8)	0.56	7.5	Number Cancellation	0.40	2.5
7	Spoken Language (Q10)	0.49	6.6	Construction	0.34	2.1
8	Comprehension (Q12)	0.48	6.5	Word Finding	0.29	1.8
9	Ideational Praxis (Q6)	0.41	5.5	Remembering	0.25	1.5
10	Commands (Q2)	0.34	4.6	Comprehension	0.21	1.3
11	Recall Instructions (Q9)	0.34	4.6	Ideational Praxis	0.20	1.3
12	Naming (Q5)	0.26	3.5	Spoken Language	0.13	0.8
13	Construction (Q3)	0.22	3.0	Commands	0.12	0.8

Abbreviations: SNR, signal-to-noise ratio; IRT, Item Response Theory.
NOTE. Bold indicates ADAS-cog individual items selected for the development of composite scores.

consistent results indicate that the selected cognitive items (Word Recall, Delayed Word Recall, and Orientation) could be the most sensitive items from ADAS-cog and would be the best candidates among the ADAS-cog subitems for constructing composite endpoints.

Biomarkers have been playing an important role in AD drug development. Dubois and colleagues [20,21] proposed a revision of the diagnosis of AD on the basis of the recent advance of technology and understanding of pathology, widening the range of its categories to encompass predementia and dementia phases. The new criteria include at least one "abnormal" biomarker among structural neuroimaging with magnetic resonance imaging, molecular neuroimaging with positron emission tomography, and CSF analysis of Aβ or tau proteins. Some other researchers have examined the subpopulation enriched by baseline CSF $A\beta_{1\text{-}42}$ in their composite score analysis [12]. We examined three subpopulations with amyloid burden measured by CSF $A\beta_{1\text{-}42}$, low hippocampal volume, and positive *APOE* ε4 status, respectively. The sensitive measures found in the enriched populations were consistent with the MCI population, and the same composite score was identified for the overall and enriched populations. The SNR values were higher in the enriched populations than the overall population, suggesting that an enrichment strategy may further improve the statistical power in clinical trials.

Among different types of composite endpoints (e.g., patient-level total score, index score derived from multiple item scores, and time to first event derived from multiple categories of events), our analysis was about the development of a total score from multiple items. In addition to our method, there are several other analytical approaches for constructing total scores as composite endpoints. The partial least squares (PLS) regression used by Logovinsky and colleagues [13] is one of them, in which a linear regression model was used to identify and combine sensitive measures. Rasch Measurement Theory (RMT) used in Hobart and colleagues [22] and IRT used in Ueckert and colleagues [14,15] can generate information-based ratings for individual items according to their responsiveness in the population. Our analysis found consistent results as those based on IRT (Table 5). Partitioning trees (e.g., Llano and colleagues [23]) and other machine learning models may be used to generate composite scores for discriminating between populations.

There are in general two types of strategies to improve the sensitivity of a composite score: finding more optimal weights for the individual items or adding more content (or items) to broaden the scope of domain coverage. Our work was essentially an extension of the latter. Not only did we add sensitive scales, we also eliminated the items that were just adding noise. Although some other researchers have attempted to develop more sensitive composite scores using weighting factors (e.g., model coefficients) derived from the data set used in the analysis [13], our goal at this point was to develop a sensitive measure in a clinical trial setting that is easy to use and standardize; therefore, data-driven weighing systems were not considered. As described in the *Methods*, the sensitive measures for the development of composite scores were selected simply based on sensitivity at various time points during the course of the disease, and our analysis was not dependent on parametric models, which makes the developed composite score less prone to assumption errors or bias. In addition, we applied a simple sum to the components of the composite score because data-dependent weighting factors may require more data sets to test their reliability before standardizing the composite score for clinical use. Therefore, our derived composite that was based on a simple sum of existing instruments may be a better choice for clinical trial endpoints because a simple sum is easy to standardize and implement across studies and to explain and interpret for clinicians.

There are some limitations of and potential improvements to our analysis. Firstly, although ADNI is a quality study and its population is close to those who would be enrolled into clinical trials, it is an observational study without treatment intervention; therefore, the generalizability of the proposed composite endpoint, including its underlying assumptions for testing hypothetical treatment effects, has yet to be determined in real clinical trials. Secondly, our search for sensitive measures was restricted to the clinical assessments administered in ADNI. Other clinical measures or different versions of the same measures that were not used by ADNI may have great sensitivity to disease progression in the MCI, especially in an epidemiologically selected real-life population. Analyzing data sets with a larger pool of sensitive assessments may result in even more sensitive composite scores. Lastly, the nominal visit time and the actual visit time in ADNI did not always coincide with each other. We have used the nominal visits in the ADNI data sets because most subjects had their actual visits occur near the times that were close to the nominal visit times. For the nominal visits at 12, 24, and 36 months, the actual visit time from baseline had an interquartile range of 11.9–12.2 months, 23.9–24.5 months, and 35.9–36.6 months, respectively. However, there were some subjects whose actual visit times deviated from the nominal visit times. The actual visit time from baseline for the nominal 12-, 24-, and 36-month visits ranged from 11.0–23.9 months, 21.4–36.4 months, and 29.7–54.1 months, respectively. Refinement of the analysis using the actual visit times may improve the accuracy in modeling the disease progression over time and can be explored in future research. Another future research direction can be the assessment of validity and reliability of the proposed composite measure. Our analysis focused on the identification of sensitive measures and development of more sensitive composites. However, validity is also an important aspect of any measure and needs careful

theoretical considerations and comprehensive data analyses. Validity and reliability of this sensitive composite endpoint may be assessed when more data are available (e.g., when the ADNI-2 data are more mature, or some clinical trial data are available).

5. Conclusion

A simple composite score is derived from the existing clinical endpoints that have been widely used in mild to moderate AD clinical studies. The derived composite score is sensitive to disease progression in the MCI population and to detecting hypothetical disease-modifying treatment effects tested in MCI trials, and it is easy to use and standardize across studies. The components of the composite overlap with those identified in previously proposed composite scores as well as the IRT analysis.

Acknowledgments

The authors acknowledge members of the AD clinical development team at Pfizer for their valuable input.

Data used in preparation of this article were obtained from the ADNI database (adni.loni.ucla.edu). As such, the investigators within 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 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

Data collection and sharing for this project was funded by ADNI (National Institutes of Health Grant U01 AG024904). ADNI is funded by the NIA, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec, Inc.; Bristol-Myers Squibb Company; 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.; IXICO, Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development, LLC; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer, Inc.; Piramal Imaging; Servier; Synarc, Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research provides 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 NeuroImaging at the University of California–Los Angeles.

This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

RESEARCH IN CONTEXT

1. Systematic review: The authors identified and evaluated the scientific question and relevant knowledge through channels including participation in AD working initiatives (e.g., CAMD AD), interaction with health regulators, and literature review using online sources (e.g., PubMed). The accumulated knowledge indicated a great need for a reliable and sensitive clinical endpoint for the MCI or early AD trials.
2. Interpretation: We proposed a composite score derived from the existing clinical endpoints that demonstrated higher sensitivity in the MCI population, and it is easy to implement and standardize across studies.
3. Future directions: The proposed composite score can be used as a clinical endpoint to design more efficient MCI trials to screen potential disease-modifying treatments for early signs of efficacy. Validation of this composite score may be conducted when the ADNI-2 data are more mature or some clinical trial data are available.

References

[1] Alzheimer's Disease Facts and Figures. Alzheimer's & Dementia 2013;9:208–45.
[2] Jack CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, et al. Introduction to the recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:257–62.
[3] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:263–9.
[4] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox N, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:270–9.
[5] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:280–92.
[6] Sperling RA, Jack CR, Aisen PS. Testing the right target and right drug at the right stage. Sci Transl Med 2011;3:1–5.
[7] FDA Guidance for Industry. Alzheimer's disease: Developing drugs for the treatment of early stage disease. Washington, DC: U.S. Food and Drug Administration; 2013.

[8] Raghavan N, Samtani MN, Farnum M, Yang E, Lobanov V, Novak G, et al. Optimizing the ADAS-Cog for MCI and early AD. Presented at: FDA-Industry Statistics Workshop; September 19-21, 2011; Washington, DC.
[9] Raghavan N, Samtani MN, Farnum M, Yang E, Lobanov V, Novak G, et al. Optimizing the ADAS-cog for MCI and early AD. Presented at: Clinical Trials on Alzheimer's Disease; Nov 3-5, 2011; San Diego, CA.
[10] Hendrix S, Logovinsky V, Perdomo C, Wang J, Satlin A. Introducing a new tool for optimizing responsiveness to decline in early AD. Presented at: Alzheimer's Association International Conference; July 14-19, 2012; Vancouver, Canada.
[11] Hendrix S, Logovinsky V, Perdomo C, Wang J, Satlin A. A new tool for optimizing responsiveness to decline in early AD. Presented at: Clinical Trials on Alzheimer's Disease; October 29-31, 2012; Monte Carlo, Monaco.
[12] Raghavan N, Samtani MN, Farnum M, Yang E, Novak G, Grundman M, et al. Alzheimer's Disease Neuroimaging Initiative. The ADAS-cog revisited: Novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. Alzheimers Dement 2013;9(1 suppl):S21–31.
[13] Logovinsky V, Hendrix, S, Perdomo C, Wang J, Satlin A. New composite score demonstrates sensitivity to disease progression and treatment effects. Presented at: Alzheimer's and Parkinson's Disease Congress; March 6-10, 2013; Florence, Italy.
[14] Ueckert S, Plan EL, Ito K, Karlsson MO, Corrigan B, Hooker AC. Application of Item Response Theory to ADAS-cog scores modelling in Alzheimer's disease. 2012. Available at: http://www.page-meeting.org/?abstract=2318. Accessed: May 15, 2014.]
[15] Ueckert S, Plan EL, Ito K, Karlsson MO, Corrigan B, Hooker AC, the Alzheimer's Disease Neuroimaging Initiative. Improved utilization of ADAS-cog assessment data through item response theory based pharmacometric modeling. Pharm Res 2014;31:2152–65.
[16] 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.
[17] Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, et al., Alzheimer's Disease Neuroimaging Initiative. Comparing predictors of conversion and decline in mild cognitive impairment. Neurology 2010;75:230–8.
[18] Rizk-Jackson A, Insel P, Petersen R, Aisen P, Jack C, Weiner M. Early indications of future cognitive decline: stable versus declining controls. PLoS One 2013;8:e74062.
[19] Huang Y. Simplified ADAS-cog: A more efficient assessment for early Alzheimer's disease. Alzheimers Dement 2011;7:S803.
[20] 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.
[21] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol 2010;9:1118–27.
[22] Hobart J, Cano S, Posner H, Selnes O, Stern Y, Thomas R, et al., Alzheimer's Disease Neuroimaging Initiative. Putting the Alzheimer's cognitive test to the test II: Rasch Measurement Theory. Alzheimers Dement 2013;9(1 suppl):S10–20.
[23] Llano DA, Laforet G, Devanarayan V. Derivation of a new ADAS-Cog composite using tree-based multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. Alzheimer Dis Assoc Disord 2011;25:73–84.