ELSEVIER

Alzheimer's & Dementia 9 (2013) S45–S55

Alzheimer's & Dementia

Rationale for use of the Clinical Dementia Rating Sum of Boxes as a primary outcome measure for Alzheimer's disease clinical trials

Jesse M. Cedarbaum[a,*], Mark Jaros[b], Chito Hernandez[b], Nicola Coley[c,d], Sandrine Andrieu[c,d,e], Michael Grundman[b,g], Bruno Vellas[c,d,f]; and the Alzheimer's Disease Neuroimaging Initiative

[a]*Cytokinetics, Inc., South San Francisco, CA, USA*
[b]*Elan Pharmaceuticals, South San Francisco, CA, USA*
[c]*INSERM U1027, University of Toulouse III, Toulouse, France*
[d]*University of Toulouse III, Toulouse, France*
[e]*Department of Epidemiology and Public Health, CHU Toulouse, Toulouse, France*
[f]*Department of Geriatric Medicine, Alzheimer Centre, CHU Toulouse, Gerontopole, Toulouse, France*
[g]*Global R&D Partners, LLC., San Diego, CA, USA*

Abstract

Background: We used the database of the Alzheimer's Disease Neuroimaging Initiative (ADNI) to explore the psychometric properties of the Clinical Dementia Rating Sum of Boxes (CDR-SB) to consider its utility as an outcome measure for clinical trials in early and mild, as well as later, stages of Alzheimer's disease (AD).

Methods: We assessed internal consistency, structural validity, convergent validity, and 2-year internal and external responsiveness of the CDR-SB using data from 382 subjects with early or mild AD at entry into the ADNI study.

Results: The CDR-SB assesses both cognitive and functional domains of AD disability. Mean scores declined nearly linearly; CDR-SB cognitive and functional subsums contributed equally to total scores at both very mild (early) and mild stages of the disease.

Conclusions: The CDR-SB has psychometric properties that make it attractive as a primary outcome measure that comprehensively assesses both cognitive and functional disability in AD patients. It may prove particularly useful for studies in early, predementia stages of AD.

Keywords: Alzheimer's disease; Mild cognitive impairment; Clinical trials; Outcome measures; ADNI

1. Introduction

Recent advances in the understanding of Alzheimer's disease (AD) are leading to a paradigm shift in the way we conceive of the pathological and clinical evolution of the disorder. The recognition that AD represents a continuous process that passes through a presymptomatic phase and a stage of "mild" cognitive impairment (MCI), with early cognitive but little or no evident functional impairment [1], has led to a proposed revision of the research diagnostic criteria for AD [2] that incorporates both clinical and biomarker evidence of disease, enabling diagnosis of AD at its very early stages.

Typically, clinical AD appears to become evident first as a syndrome of amnestic MCI, in which cognitive impairment is largely confined to deficits in memory and complex activities of daily living (ADLs) [3,4]. Functional, behavioral, and social impairments inexorably emerge as the disorder segues into what we clinically recognize as dementia of the Alzheimer type [5].

Clinical trials in subjects with MCI have used "conversion" from the MCI syndrome to "Probable Alzheimer's Disease" (the stage of clinically defined dementia) using Diagnostic and Statistical Manual of Mental Disorders - IV (DSM-IV) or National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related

The authors have no conflicts of interest to report.

*Corresponding author. Tel.: 650-624-3031; Fax: 650-624-3225.
E-mail address: jcedarbaum@cytokinetics.com

1552-5260/$ - see front matter
doi:10.1016/j.jalz.2011.11.002

Disorders Association (NINCDS–ADRDA) criteria as the primary outcome measure [6,7]. The use of "conversion" to "AD" or "dementia" is, however, problematic as a clinical trial outcome for at least four reasons:

1. Rates of "conversion" occurring within the feasible time frame of a therapeutic clinical trial are low; therefore, large numbers of subjects are required. Current criteria for enrichment of patient populations based on clinical and/or biomarker criteria in an attempt to select for subjects with AD pathology, as opposed to other causes of cognitive impairment, still yield study populations in which a minority of patients "convert" within a reasonable time frame.
2. In clinical trials practice, adjudication of "conversion" requires centralized committee review of both quantitative data and subjective reports, a cumbersome and highly subjective practice, such that clinical study results using "conversion" as an outcome will be difficult to translate into clinical practice.
3. Rates of disease progression and "conversion" are subject to interstudy variation based on the population (a) screened for and (b) actually enrolled. Thus, subtle differences in the design or implementation of inclusion criteria have resulted in vastly different rates of "conversion to AD" in recent clinical trials [6–8]; use of a continuous end point that tracks disease progress throughout the disease course may facilitate more uniform outcomes across clinical trials.
4. Because patients entering clinical trials may progress to what is recognized as mild-to-moderate AD, clinical outcome measures that have sensitivity to clinical change throughout the course of AD would enable consistent use of the same outcome measure as these patients progress.

Therefore, a more desirable metric would be a continuous comprehensive outcome measure that is capable of tracking the progression of the disease from the MCI stage through mild, moderate, and severe dementia and to monitor progression of cognitive and functional change in AD clinical trials. However, currently available cognitive assessment instruments, such as the Alzheimer's Disease Assessment Scale, cognitive subscale (ADAS-cog) [9], lack sensitivity in the early, milder stages of AD [10] and exhibit a high degree of variability in rates of change [11], necessitating enrollment of large numbers of subjects in clinical trials. Similarly, currently available and validated functional rating scales show little change in early stages of AD [1,6–8].

The use of a single comprehensive instrument that integrates the assessment of both manifestations of the primary disease activity (loss of cognitive function) and consequent loss of functional abilities is common clinical practice in the study of potential treatments for chronic neurological disorders, most notably the Unified Parkinson Disease Rating Scale [12] for Parkinson disease, which has been used as a primary outcome measure both when the aim of the clinical trial is to assess symptomatic benefits of a new intervention, or to slow or delay disease progression [13].

The Clinical Dementia Rating (CDR) [14] is a well-validated instrument that has been in use for more than 20 years in clinical trials in AD and MCI. The CDR assesses three domains of cognition (memory, orientation, judgment/problem solving) and three domains of function (community affairs, home/hobbies, personal care) using structured interviews of both the study subject and a companion/informant carried out by a trained rater and scored using a standard methodology. The scores for the six domains (range from 0 to 3) tested can be summed (CDR Sum of Boxes or CDR-SB); an algorithm is used for integrating the information obtained into an overall score, termed here the "CDR Global" score [15]. The CDR includes structured discussions with the subject and informant. It has an advantage for trials lasting a year or more in that it does not require the rater to remember remote details of the subject's baseline performance or to make an assessment of the subject's clinical change from baseline. We [16] have recently proposed using the CDR-SB as the sole primary outcome measure for clinical trials in early Alzheimer's disease (eAD).

Tractenberg et al [17] performed a factor analysis of the CDR-SB in conjunction with other measures that separately rated cognitive function (Mini-Mental State Examination [MMSE]), functional abilities (Alzheimer's Disease Cooperative Study Activities of Daily Living [ADCS-ADL] scale), and behavior (Behavior Rating Scale for Dementia) in a sample of 242 subjects, at 27 sites in the United States, with probable AD. The authors created two CDR subscores—a "cognitive" subsum comprising the sum of the memory, orientation, and judgment/problem-solving box scores of the CDR, and a "functional" subsum, which combined the scores for community activities, personal care, and home/hobbies boxes. In this model, the 12-month change in the CDR cognitive subsum loaded onto a factor with only the MMSE, whereas the change in functional subsum loaded onto a factor with only the change in ADCS-ADL score. The correlation between the change in the total CDR-SB and change in the MMSE was −0.46, between CDR-SB and ADAS-cog was 0.42, and between the CDR-SB and ADCS-ADL was 0.50, indicating a "modest" degree of association between the change in the total CDR-SB score with decline in both cognitive and functional measures.

To further validate the suitability of the CDR-SB as a sole primary outcome measure for AD (and particularly for eAD) clinical trials, we have taken advantage of publicly available data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study conducted in North America. ADNI, a consortium of universities and medical centers in the United States and Canada, was established to develop standardized imaging techniques and biomarker procedures in normal subjects, subjects with MCI, and subjects with mild AD [1]. The major goals of ADNI are to develop improved methods that will lead to uniform standards for acquiring longitudinal

multisite magnetic resonance imaging and positron emission tomography data on patients with AD, patients with MCI, and elderly control subjects; to develop an accessible data repository that describes longitudinal changes in brain structure and metabolism while acquiring, in parallel, clinical, cognitive, and biochemical data; to develop methods that will maximize statistical power to determine treatment effects in clinical trials; and to test a series of hypotheses based on clinical and biomarker data. Data from ADNI are publicly available from its Web site (www.loni.ucla.edu/ADNI) in near-real time. The baseline clinical characteristics of the population enrolled in ADNI have been reported by Petersen et al [1].

We used this database to confirm and extend the findings of Trachtenberg et al regarding (1) the internal structure of the CDR-SB as containing separate cognitive and functional domains, (2) the correlation of disease progression as assessed by the CDR-SB with external cognitive and functional measures, and (3) relative progression of cognitive and functional subsums of the CDR-SB originally defined by Trachtenberg et al [17], as well as their relative contribution to the overall CDR-SB over time. This work was planned as a companion study to that of Coley et al [18] that explored the psychometric properties of the CDR-SB using data from REAL.FR, a study that followed disease progression in more than 600 French AD patients. Because neither database was large enough to support independent test and replication data sets, we planned, to the extent possible, given the limitations created by differences between the REAL.FR and ADNI populations and study designs, to apply similar methods of analysis in both studies.

2. Methods

We evaluated the CDR-SB as an outcome measure for AD clinical trials. We were particularly interested in its utility as a primary outcome measure for subjects with early (predementia) or "prodromal" [2] and mild AD.

2.1. Patient data sets

The population of analysis was drawn from the ADNI database (downloaded on January 14, 2010). The general inclusion and exclusion criteria for the ADNI study are described by Petersen et al [1]. The ADNI study was approved individually by the institutional review boards of all the participating institutions. Informed written consent was obtained from all participants at each site.

The ADNI study enrolled subjects who were English- or Spanish-speaking males or females ranging in age between 55 and 90 years (inclusive). All were required to have a study partner to provide an independent evaluation of functioning. Subjects were enrolled in the following three categories: normal control subjects, subjects with MCI, or subjects with mild AD. The MMSE score range was 24 to 30 for the normal subjects and subjects with MCI and 20 to 26 for those with AD; all are inclusive. For subjects with MCI, the CDR global score had to be 0.5, with the memory box score being 0.5 or greater; for subjects with AD, the CDR global score had to be 0.5 or 1. Subjects had to meet a minimum memory criterion, using delayed recall of one paragraph from the Logical Memory II subscale of the Wechsler Memory Scale–Revised (maximum score of 25). Cutoff scores, described by Petersen et al [1], were based on education level. In addition, subjects with MCI could not qualify for the diagnosis of dementia, whereas subjects in the AD group met the NINCDS–ADRDA criteria for probable AD. The use of acetylcholinesterase inhibitors and memantine was permitted if the dose had been stable for 4 weeks before screening for subjects with MCI and AD, but patients were excluded if they were on psychoactive medications that were believed to possibly affect cognitive function.

Subjects enrolled in the ADNI MCI cohort by definition did not meet NINCDS–ADRDA diagnostic criteria for AD. This group was expected to include some subjects who might either remain as stable MCI or progress to non-AD dementias. We, therefore, separately analyzed data from ADNI subjects who were either enrolled in the "Alzheimer's Disease" cohort (those meeting the aforementioned criteria for a diagnosis of dementia of the Alzheimer type and a subset of individuals we termed "Early" AD [eAD]). eAD was defined as a CDR global score equal to 0.5 and scores of $\geq$4 on the ADAS-cog and of $\geq$3 on the Functional Activities Questionnaire (FAQ), at entry. Because of the way the ADNI study recruited and classified subjects, subjects were found in both the ADNI "MCI" and "AD" populations. Our definition of eAD was based on a previous analysis of the ADNI database, in which it was found that these cognitive and functional criteria identified an MCI subgroup more likely to have underlying AD based on cerebrospinal fluid biomarker criteria and more likely to progress on clinical scales than the general MCI population recruited to ADNI (data on file, Elan Pharmaceuticals, Inc.). The eAD sample in the present study, which consists of 280 subjects, also has a degree of impairment roughly matching the REAL.FR population studied by Coley et al [18], with CDR global scores of 0.5. All available data were used in the analyses.

The attributes of the CDR that we explored were as follows:

- *Internal reliability*, using Cronbach's α, a measure of consistency among individual items in a scale. Values of at least 0.70 can be seen as an acceptable reliability coefficient [19].
- *Factor analyses* to assess the internal structure of the CDR-SB and determine the extent to which it concurrently evaluates a patient's decline in both the cognitive and functional domains. Baseline, 2-year, and 2-year change from baseline scores for each of the six individual CDR box scores were subjected to a factor analysis

using principal components with promax rotation. Promax rotation was chosen to allow correlation between the factors.

- *Convergent ("face") validity* by calculating Spearman correlations between the CDR-SB and its individual components, as well as other cognitive and functional tests, including the MMSE [20], the original 11-item ADAS-cog [9], as well as a 13-item version (which includes delayed word recall and number cancellation tasks) [21], and the FAQ [22] at baseline and follow-up visits. Subjects enrolled in the ADNI MCI cohort had evaluations performed at baseline and every 6 months for 2 years; ADNI "AD" subjects did not undergo the month 18 evaluation.
- *Internal responsiveness*, which characterizes the ability of a measure to change over time, using standardized effect size (mean change from baseline divided by the standard deviation of baseline scores) and standardized response mean (mean change from baseline divided by the standard deviation of change from baseline) statistics calculated on 2-year change from baseline scores. Higher effect sizes or standardized response means indicate better responsiveness—values of 0.20, 0.50, and 0.80 or greater for either measure are considered to represent small, moderate, and large responsiveness, respectively [23].
- *External responsiveness*, which examines the extent to which changes in a measure over time relate to corresponding changes in a reference measure [23]. Spearman correlation coefficients were calculated to evaluate correlations between 2-year changes in the CDR-SB and 2-year changes in the other cognitive and functional measures. We also explored the contribution of each component "box" scores of the CDR to the total score and to change in the total score over time.

All analyses were performed using SAS version 9.1 (SAS Institute, Cary, NC). Baseline characteristics are presented using descriptive statistics.

3. Results

The demographic characteristics of the study population are depicted in Table 1. The mean age of the 382 subjects considered in this analysis was approximately 75 years; 227 (59%) were male and 155 (41%) were female. The median duration of education was 16 years. Almost 25% were college graduates and 31.7% had some education beyond college, 17.5% had completed high school as their greatest level of education, and only 3% had an eighth-grade education or less.

Of the 382 subjects, 280 (177 males and 103 females) met the definition of eAD and 188 met the criteria for AD dementia. Because of the way the ADNI study participants were recruited (2:1 ratio of MCI to mild AD subjects), and because the overall study population excluded subjects with a global CDR score greater than 1, subjects with CDR scores of 0.5, representing "very mild" or "questionable" AD [15], are overrepresented in the patient population compared with mild AD subjects.

Table 1
Demographic characteristics

Variable	Mean	SD	Median	Minimum	Maximum
All subjects (N = 382)					
Age	74.76	7.34	75.0	54.0	91.0
Years of education	15.25	3.14	16.0	4.0	20.0
ADAS-cog 11*	15.62	6.23	14.5	4.0	42.7
ADAS-cog 13†,§	24.67	8.11	24.0	8.0	54.7
FAQ	10.17	6.38	9.0	0.0	30.0
MMSE	25.13	2.63	25.0	18.0	30.0
CDR-SB	3.15	1.76	3.0	0.5	9.0
CDR cognitive subsum	1.99	0.94	2.0	0.5	6.0
CDR functional subsum	1.17	1.00	1.0	0.0	5.0
eAD subjects (N = 280)					
Age	74.5	7.1	75.0	54.0	91.0
Years of education	15.5	3.2	16.0	4.0	20.0
ADAS-cog 11	14.0	5.2	13.3	4.0	34.7
ADAS-cog 13‡,§	22.6	7.1	22.0	8.0	47.7
FAQ	8.3	4.6	7.0	3.0	29.0
MMSE	25.9	2.3	26.0	20.0	30.0
CDR-SB	2.4	1.0	2.3	0.5	5.0
CDR cognitive subsum	1.6	0.7	1.5	0.5	4.0
CDR functional subsum	0.8	0.6	1.0	0.0	2.5
AD subjects (N = 188)					
Age	74.9	7.6	75.0	54.0	91.0
Years of education	14.7	3.1	15.0	4.0	20.0
ADAS-cog 11*	18.7	6.3	15.0	7.7	42.7
ADAS-cog 13†,§	29.0	7.7	28.8	12.7	54.7
FAQ	13.1	6.8	12.0	0	30.0
MMSE	23.3	2.0	23.0	18.0	27.0
CDR-SB	4.4	1.6	4.5	1.0	9.0
CDR cognitive subsum	2.6	0.8	2.5	1.0	6.0
CDR functional subsum	1.7	1.0	1.5	0	5.0

Abbreviations: CDR-SB, Clinical Dementia Rating Sum of Boxes; ADAS-cog, Alzheimer's Disease Assessment Scale, cognitive subscale; FAQ, Functional Activities Questionnaire; MMSE, Mini-Mental State Examination; SD, standard deviation; eAD, early Alzheimer's disease; AD, Alzheimer's disease.

*Two subjects had missing baseline values.

†Eight subjects had missing baseline values.

‡Three subjects had missing baseline values.

§ADAS-cog 13 includes the 11-item ADAS-cog plus a delayed recall task and the Digit Symbol Substitution Test.

Not surprisingly, the age and level of education of the 280 eAD subjects were reflective of the overall group. Baseline test scores revealed somewhat less cognitive and functional impairment in this group; baseline CDR-SB scores averaged 2.4.

Figure 1 shows the distribution of global CDR and individual CDR box scores of the population at baseline for the eAD and AD study populations. As might be expected in a population recruited to represent MCI as defined by Petersen criteria (Petersen et al, 2001), subjects we identified with eAD did not score above 1 in the Judgment and Problem-Solving, Home and Hobbies, and Community Affairs box scores, although one subject did report a score above 1 in the Personal Care box.

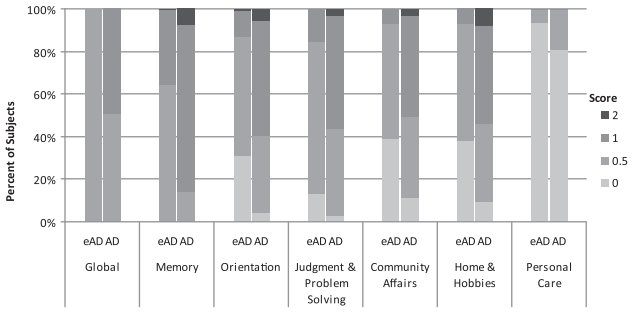

Fig. 1. Distribution of Clinical Dementia Rating (CDR) global scores and individual CDR box scores for the early Alzheimer's disease (eAD) and Alzheimer's disease (AD) populations at baseline.

3.1. Internal reliability

Values of Cronbach's α for the full CDR-SB in the overall study population were 0.90 at screening and 0.92 at month 24, indicating a high degree of reliability. Deleting either the cognitive or the functional subsum reduced the value of Cronbach's α, but not below 0.85 at either time point. Cronbach's α for the CDR-SB change to month 24 was 0.90. Again, deletion of the individual subsums slightly reduced the α coefficient. For the eAD subgroup, values of Cronbach's α were 0.85 at baseline, 0.91 at month 24, and 0.90 for change to month 24, and for AD subjects, the values of Cronbach's α were 0.87 at screening, 0.92 at month 24, and 0.90 for change to month 24.

3.2. Factor analysis

Table 2 depicts the factor structure of the CDR-SB at screening and month 24, as well as the change between the two time points. Confirming the results of Trachtenberg et al [17], a rotated factor analysis revealed that the CDR-SB can be described as having two factors—a "Cognitive" factor, consisting of the Memory, Orientation, and Judgment and Problem-Solving box scores; and a "Functional" factor, consisting of the Community Affairs, Home and Hobbies, and Personal Care boxes, although the ADNI population, which is weighted toward less impaired subjects, showed overlap in certain domains where functional independence may be retained when the data are examined longitudinally. Interestingly, the separation of the CDR-SB into two distinct factors is most evident when examining change over time. This observation may relate to the way the CDR interview itself is structured, as an emphasis in postbaseline assessments is placed on change in performance rather than performance level itself (J. Morris, personal communication). The loading of CDR box scores onto the two factors was not as clear in the AD group, probably because of its smaller size.

3.3. Convergent validity

As shown in Table 3, at baseline, Spearman coefficients for the correlation of the total CDR-SB score with other cognitive and functional tests across all subjects ranged from 0.53 for the ADAS-cog to 0.66 for the FAQ. CDR-SB total score correlated somewhat better with the FAQ than with the ADAS-cog and MMSE. The CDR cognitive subsum correlated modestly well with the ADAS-cog, MMSE, and FAQ. However, the functional subsum correlated more strongly with the functional (i.e., FAQ) than the cognitive external measures. All correlations were statistically significant (see legend to Table 3).

Table 2
Factor loadings for 2-factor solutions after promax rotation for baseline, 2-year, and 2-year change scores

	All subjects		eAD subjects		AD subjects	
Variable	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Screening						
Memory	0.77		0.72			0.64
Orientation	0.74		0.68			0.59
Judgment and problem solving	0.44		0.35			0.37
Community affairs	0.38	0.47	0.40	0.35	0.56	
Home and hobbies		0.56	0.33	0.42	0.65	
Personal care		0.46			0.53	
Two-year scores						
Memory	0.75		0.79			0.78
Orientation	0.76		0.83			0.77
Judgment and problem solving	0.39	0.49	0.51		0.55	
Community affairs	0.42	0.49	0.44	0.47	0.47	0.39
Home and hobbies		0.71		0.70	0.74	
Personal care		0.63		0.62	0.69	
Two-year change scores						
Memory	0.50		0.72		0.31	0.47
Orientation	0.52		0.68			0.60
Judgment and problem solving	0.47		0.58			0.57
Community affairs		0.48	0.43	0.37	0.53	
Home and hobbies		0.56		0.66	0.55	
Personal care		0.37		0.60	0.47	

Table 3
Correlations between baseline scores (Spearman correlation coefficients)

Variable	CDR-SB	CDR cog. subsum	CDR fun. subsum	ADAS-cog 11	ADAS-cog 13	FAQ	MMSE
All subjects							
CDR-SB	1						
CDR cog. subsum	0.91	1					
CDR fun. subsum	0.91	0.68	1				
ADAS-cog 11	0.53	0.55	0.42	1			
ADAS-cog 13	0.57	0.59	0.45	0.96	1		
FAQ	0.66	0.63	0.58	0.46	0.49	1	
MMSE	−0.57	−0.63	−0.42	−0.57	−0.60	−0.45	1
eAD subjects							
CDR-SB	1						
CDR cog. subsum	0.85	1					
CDR fun. subsum	0.84	0.46	1				
ADAS-cog 11	0.37	0.43	0.20	1			
ADAS-cog 13	0.41	0.47	0.24	0.95	1		
FAQ	0.57	0.54	0.42	0.30	0.33	1	
MMSE	−0.37	−0.48	−0.14	−0.46	−0.49	−0.31	1
AD subjects							
CDR-SB	1						
CDR cog. subsum	0.82	1					
CDR fun. subsum	0.90	0.52	1				
ADAS-cog 11	0.42	0.40	0.36	1			
ADAS-cog 13	0.43	0.42	0.36	0.97	1		
FAQ	0.62	0.51	0.55	0.48	0.48	1	
MMSE	−0.30	−0.36	−0.17	−0.49	−0.46	−0.33	1

Abbreviations: CDR cog. subsum, CDR cognitive subsum; CDR fun. subsum, CDR functional subsum.
NOTE. All correlations have $P < .05$.
All subjects: all are <0.0001, eAD: all are <0.02, AD: all are <0.02.

3.4. Internal responsiveness

CDR-SB total scores worsened approximately linearly over the 2 years of observation. The rate of deterioration was slightly higher in subjects with starting CDR global scores of >0.5 than in subjects with starting CDR global scores of 0.5 (eAD). As shown in Fig. 2, the cognitive and functional subsums contributed approximately equally to the total score at all time points assessed.

Table 4 shows the 2-year change in CDR-SB and other cognitive and functional tests. Of note, both the effect size and the standardized response mean for the CDR-SB are the largest of all the measures used, suggesting that the CDR-SB has the potential to require smaller sample sizes when used as a clinical study outcome, compared with the ADAS-cog (11- or 13-item versions), the MMSE, or the FAQ. The difference in effect size between the CDR-SB and ADAS-cog is greatest in the eAD population, suggesting that this measure might be particularly suitable for studies of disease-modifying agents that might delay progression of AD to the dementia stage. All correlations were statistically significant (see legend to Table 4).

3.5. External responsiveness

As shown in Table 5, the correlation between the change in CDR-SB and change in the functional measure, the FAQ (0.58), appears to be somewhat greater than that between the CDR-SB and the MMSE (0.49). William's formula was used to compare the correlation coefficients, yielding a nonsignificant P equal to .11 [24]. The CDR functional subsum correlated more strongly with the FAQ than with either the MMSE or ADAS-cog. The change in the CDR cognitive subsum correlated similarly with the MMSE (0.47) and FAQ (0.51) but less strongly with either the 11-item or the 13-item version of the ADAS-cog (0.32 and 0.34, respectively). The lower correlations of the CDR with the ADAS-cog may reflect the lack of correlation of the ADAS-cog with impairments that are noticeable to clinicians and caregivers in general. It is also possible that the restricted range of the scales in the defined analysis populations reduced the observed correlations [25].

4. Discussion

The CDR-SB assesses both cognitive and functional aspects of AD. Scores on the individual cognitive and functional subsums correlate well with corresponding individual domain measures across the spectrum of disease severity from the earliest, prodromal or "early" stages of AD through the mild/moderate stages of the disease. These properties make it well suited to serve as a comprehensive primary outcome measure for a study that will enroll subjects with eAD and follow them to more advanced stages.

Inter-rater reliability validity and stability of CDR-SB scores have been established [26,27]; the scale has been shown to be sensitive to within-subject clinically meaningful

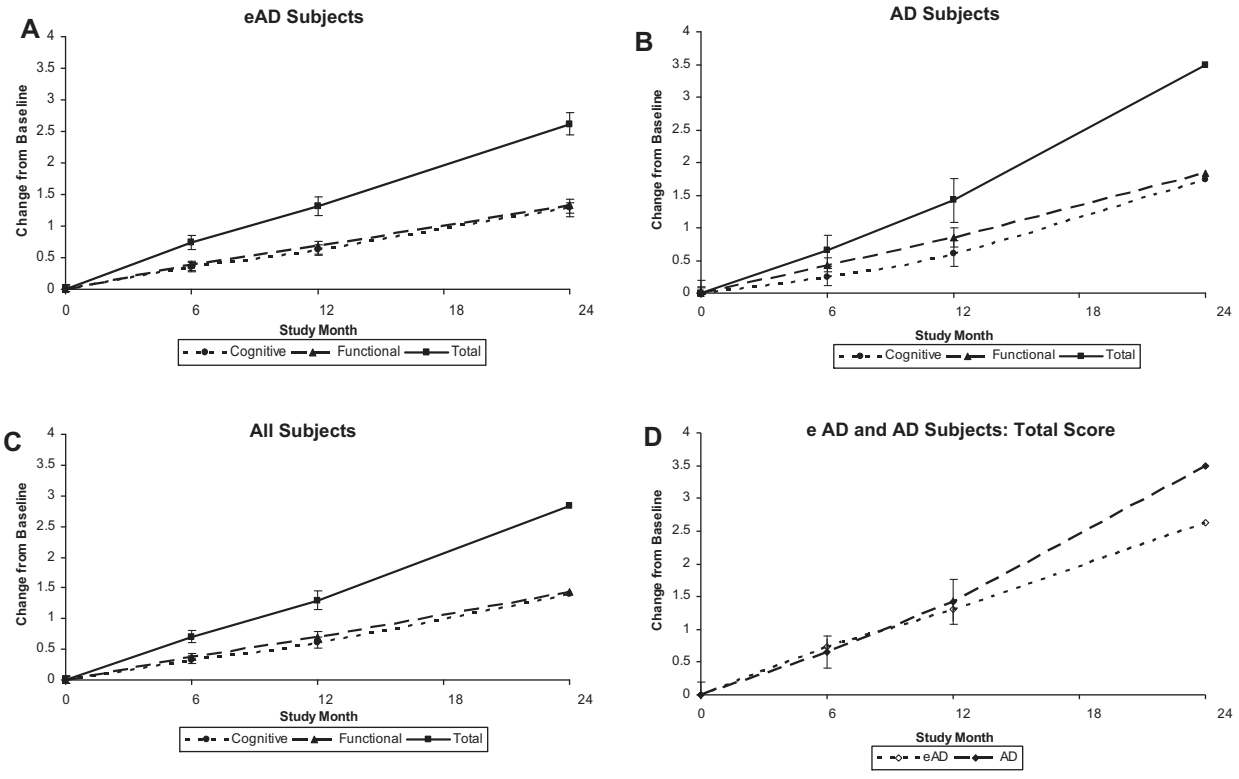

Fig. 2. Two-year progression of total Clinical Dementia Rating Sum of Boxes (CDR-SB) scores, cognitive subsums, and functional subsums in all, eAD, and AD subjects. CDR cognitive subsums (circles, dotted lines), functional subsums (triangles, dashed lines), and total CDR-SB scores (squares, solid lines) for the entire study population. (A) "eAD" subjects (CDR global = 0.5, Alzheimer's Disease Assessment Scale, cognitive subscale: ≥4, and Functional Activities Questionnaire: ≥3), (B) AD subjects, (C) all subjects, (D) comparison of progression rate for total CDR-SB score for eAD (open symbols, dashed line) and AD (filled symbols, solid line) subjects. Note that rate of progression in the second year in AD subjects appears more rapid than in eAD subjects, and that progression rate in eAD subjects is similar to progression rate for AD during the first year of observation. Symbols represent mean ± SEM.

changes. Standardized training and certification protocols are available ([26]; http://alzheimer.wustl.edu/cdr/default.htm), and the scale has been translated into and used in multiple languages and across cultures ([15,26]; http://alzheimer.wustl.edu/cdr/PDFs/Translations/). CDR scores also correlated well with pathological stages of AD at autopsy in a small number of carefully studied AD patients [28]. Given the presence of memory impairment,

Table 4
Two-year change in CDR-SB and other cognitive and functional tests

Variable	Mean 2-year change	SD 2-year change	Effect size	Standardized response mean
All subjects (N = 227)				
CDR-SB	−2.51	2.37	−1.42	−1.06
ADAS-cog 11	−5.67	7.04	−0.91	−0.81
ADAS-cog 13	−7.16	7.82	−0.88	−0.92
FAQ	−5.86	6.37	−0.92	−0.92
MMSE	−2.91	3.80	−1.11	−0.77
eAD subjects (N = 172)				
CDR-SB	−2.39	2.32	−2.30	−1.03
ADAS-cog 11	−4.85	6.62	−0.94	−0.73
ADAS-cog 13	−6.45	7.48	−0.91	−0.86
FAQ	−5.57	6.42	−1.20	−0.87
MMSE	−2.74	3.63	−1.19	−0.76
AD subjects (N = 99)				
CDR-SB	−3.01	2.69	−1.86	−1.12
ADAS-cog 11	−8.23	7.51	−1.30	−1.10
ADAS-cog 13	−9.47	8.35	−1.24	−1.13
FAQ	−6.71	6.26	−0.98	−1.07
MMSE	−3.26	4.46	−1.60	−0.73

NOTE. Only includes subjects with both baseline and 2-year data.
All change $P < .0001$.

Table 5
Correlations between 2-year change from baseline scores (Spearman correlation coefficients)

Variable	CDR-SB	CDR cog. subsum	CDR fun. subsum	ADAS-cog 11	ADAS-cog 13	FAQ	MMSE
All subjects (N = 227)							
CDR-SB	1						
CDR cog. Subsume	0.89	1					
CDR fun. Subsume	0.92	0.65	1				
ADAS-cog 11	0.38	0.32	0.39	1			
ADAS-cog 13	0.40	0.34	0.40	0.95	1		
FAQ	0.58	0.51	0.56	0.31	0.34	1	
MMSE	0.49	0.47	0.43	0.42	0.43	0.44	1
eAD subjects (N = 172)							
CDR-SB	1						
CDR cog. Subsume	0.88	1					
CDR fun. Subsume	0.92	0.65	1				
ADAS-cog 11	0.33	0.30	0.30	1			
ADAS-cog 13	0.36	0.32	0.34	0.94	1		
FAQ	0.70	0.59	0.68	0.34	0.37	1	
MMSE	0.51	0.50	0.42	0.40	0.41	0.44	1
AD subjects (N = 99)							
CDR-SB	1						
CDR cog. Subsume	0.89	1					
CDR fun. Subsume	0.92	0.64	1				
ADAS-cog 11	0.49	0.39	0.50	1			
ADAS-cog 13	0.50	0.41	0.52	0.96	1		
FAQ	0.46	0.43	0.42	0.33	0.37	1	
MMSE	0.49	0.43	0.45	0.52	0.55	0.49	1

NOTE. Only includes subjects with both baseline and 2-year data.
P values for all correlations shown are $<.05$.
All subjects: all are <0.0001, eAD: all are <0.0001, AD: all are <0.0008.

which is required at study entry, progressive dysfunction in any one or more of five domains can drive the assessment of overall decline.

In the present study, we have explored the internal reliability, factor structure, face validity, and internal and external responsiveness of the CDR-SB in the ADNI database. We found total CDR-SB scores and scores on its individual subsum domains to correlate with independent cognitive and functional measures. However, correlation of the cognitive subsum with ADAS-cog was weaker than might be expected (but still statistically significant). This observation may relate in part to the relatively mild level of overall cognitive dysfunction in the ADNI population we studied, in which global CDR scores ranged only from 0.5 to 2, with the vast majority being in the 0.5 and 1 ranges. The ADAS-cog is relatively less sensitive to milder degrees of cognitive dysfunction [11, 29, 30].

The CDR showed better responsiveness to change over the 2-year period of observation in the ADNI study compared with other cognitive or functional scales. The standard effect size (mean change from baseline divided by the standard deviation of baseline scores) is larger for the CDR-SB than for any of the other standard outcome measures used in AD clinical trials, suggesting that sample sizes for clinical trials using the CDR-SB as a primary outcome measure may be smaller than those for studies that rely on the ADAS-cog, and a functional measure [18]. However, it should be noted that the ability to generate large standard effect sizes (and hence to require smaller sample sizes when used as an outcome measure in clinical trials) is only one factor involved in the selection of an instrument as a key outcome measure. The clinical relevance of a scale and its ability to faithfully track disease progression in an interpretable and meaningful fashion are also of critical importance.

Importantly, factor analysis substantiates consideration of the CDR-SB as separately assessing both cognitive impairment and functional disability, and the individual subsums each account for about half of the test score change when the CDR is assessed repeatedly over time. In the present study, the cognitive domain of the CDR-SB appeared to associate as strongly with functional as with cognitive outcome measures, whereas in the companion study by Coley et al [18], convergent validity of the cognitive subsum appeared to be greater. We attribute this discrepancy to the differing nature of the two populations. The REAL.FR population was selected on the basis of the presence of dementia (however mild) and included subjects with a wider range of CDR scores at baseline, whereas the ADNI study recruited only subjects with CDR global scores of 0 to 1 at baseline. The level of formal education was also much lower in the French population, which was a geographically based sample of patients attending memory center clinics, than in the North American study, which recruited interested volunteers to specialized academic research centers.

Although ideally the CDR-SB could be used as a primary outcome measure throughout the course of AD from the

"Very Mild" or "Early" stage of the disease, there is a floor effect in the "Personal Care" box score, which is frequently rated as 0 in subjects with early cognitive change. However, accrual of disability in this domain is reflected in an increased contribution of this box score as the disease progresses. Likewise, in early disease stages, the Judgment and Problem-Solving as well as the Community Affairs box scores appear to load equally onto the cognitive and functional domains, perhaps because functional (ADL) limitations are less prominent than they are later in the disease course.

4.1. Comparison with similar analyses performed on the REAL.FR database

Our findings generally agree with those derived from a similar analysis of the REAL.FR database [18]. The REAL.FR study is a multicenter prospective cohort study of 686 community-dwelling AD patients that was carried out in 16 university medical centers in France, beginning in 2000 [31]. REAL.FR was similar in design to the clinical component of ADNI. The ADNI population we studied was somewhat younger, had more years of formal education, and was less severely affected by AD than the REAL.FR population analyzed by Coley et al [18]. Separation of the CDR-SB into cognitive and functional factors was more clear-cut in the REAL.FR study population, perhaps because ADNI contained (by design) a higher proportion of subjects classed as MCI at study entry, who had minimal impairment in ADLs and, therefore, had scores of 0 on the CDR Personal Care box. Additional factors contributing to the overlap of factor groupings in the ADNI subject population may have included the (relatively, for a factor analysis) small number of subjects; the ordinal nature of the CDR scale with a limited number of steps (five); the small number of scale items to sort (six), possibly some content overlap between the variables; and the restricted range of disease severity in the population available for study [25].

The greater degree of correlation between changes in CDR scores and the functional outcome measure, FAQ, compared with the ADAS-cog in the present study may at first seem to be of concern regarding the convergent validity of the CDR. However, in the study by Coley et al [18], the CDR cognitive subsum did correlate more strongly with the ADAS-cog than with the functional measure used in REAL.FR. Although the reason for this is unclear at present, it may be because of demographic differences between the two populations, as well as the high level of formal education in the ADNI participants. On the other hand, the relatively stronger correlation of functional assessments with the cognitive subsum of the CDR, which we observed in the ADNI population, may be a function of the nature of the functional measures used. Two functional rating scales were used in the REAL.FR study—an inventory of basic ADLs, such as feeding, dressing, and bathing [32], and an assessment of instrumental activities of daily living (IADLs), including the ability to shop, handle finances, and use public transportation [33]. The FAQ [22], which was the functional outcome measure in ADNI, focuses on IADLs that have a substantial cognitive component, and therefore it resembles the IADL scale used in REAL.FR, while not assessing basic ADL function at all. It is, therefore, interesting to note that the correlation between the cognitive subsum of the CDR-SB and the IADL scale in the overall REAL.FR study population is similar to the correlation between the CDR-SB cognitive subsum and FAQ scores in the ADNI population. Also, the correlation between the cognitive subsum and the ADAS-cog was better in the more impaired subjects in the REAL.FR baseline analysis [18]; therefore, it is perhaps not surprising to find a lower correlation in the ADNI subjects, who were, on average, less impaired than the REAL.FR population. The FAQ also contains within it items that may overlap with domains of the CDR cognitive subsum, such as handling of financial and business affairs, as well as memory item, querying patients' ability for "Remembering appointments, family occasions, holidays, medications," which could also increase the correlation with the cognitive subsum. Thus, the observed degree of correlation between the CDR-SB cognitive subsum and the FAQ, especially in higher-functioning patients, such as those enrolled in the ADNI cohort, may not be entirely unexpected.

One of the limitations of the present study is that the ADNI data set we used (N = 382) was smaller than the REAL.FR data set (N = 667). We, therefore, explored two somewhat overlapping subgroups: (1) an eAD subgroup, diagnosed with MCI or AD, who were given a CDR global score of 0.5 and met minimal thresholds of cognitive and functional impairment on the ADAS-cog and FAQ; and (2) a population diagnosed with AD by ADNI investigators, according to conventional criteria, and with global CDR scores at baseline of 0.5 or greater. The CDR-SB behaved similarly in both subgroups. Thus, taken together, the results of our parallel analyses of the performance of the CDR-SB in ADNI and REAL.FR, two natural history populations, lend further validity to the potential to use of the CDR-SB as a sole primary outcome measure for AD clinical trials across the spectrum of disease severity.

Use of the CDR-SB as a sole outcome measure for clinical trials in early or very mild AD has some limitations. It is a lengthy instrument to administer. Raters administering the instrument must be well trained and must adhere closely to the semistructured interview format [26]. Unlike other rating scales and instruments, it does not include numerically scored patient performance measures, but the Memory and Orientation boxes do assess changes in patient performance in objective memory tasks. Most importantly, the accuracy of the information obtained is highly dependent on the reliability of the study subjects' study partner/caregiver/informant, who has regular access to and observes the patient regularly, and the consistency of the availability of a single informant throughout the duration of the clinical trial.

However, inter-rater reliability studies have demonstrated high levels of intra- and inter-rater reliability in multicenter clinical trials [27]. Language, socioeconomic factors, and cultural factors may also promote variability in CDR-SB scores. It is, therefore, interesting to note that we came to similar conclusions in analyses of study populations in the ADNI study, which was carried out in the United States and Canada, where the education level is high and the CDR is administered in either English or Spanish, as well as in France, where the population was older, education levels lower, the range of test scores greater, and the instrument was administered in French [18].

Acknowledgments

The authors thank Dr. Janet Wittes of Statistics Collaborative, Inc. for her thoughtful comments on the manuscript, and Drs. Aleksandra Pastrak, John Choe, and Tony Cruz of Transition Therapeutics for their contributions in the development of the criteria used to define the eAD population within the ADNI data set.

ADNI: Data used in the preparation of this article were obtained from the ADNI database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Manuscript_Citations.pdf). The principal investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco, CA. Subjects have been recruited from more than 50 sites across the United States and Canada. The Foundation for the National Institutes of Health (www.fnih.org) coordinates the private sector participation in the ADNI public–private partnership that was begun by the National Institute on Aging and supported by the National Institutes of Health. Corporate contributions have been provided to the Foundation for the National Institutes of Health by Abbott, AstraZeneca AB, Bayer Schering PharmaAG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson & Johnson, Eli Lilly and Co., Merck & Co. Inc., Novartis AG, Pfizer Inc., F. Hoffmann-LaRoche, Schering-Plough, Synarc Inc., and Wyeth, as well as nonprofit partners, the Alzheimer's Association and the Institute for the Study of Aging.

References

[1] Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology 2010;74:201–9.

[2] 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.

[3] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Gangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999;56:303–8.

[4] Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L. Mild cognitive impairment represents early-stage Alzheimer disease. Arch Neurol 2001;58:397–405.

[5] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Sadlan E. Clinical Diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;24:939–44.

[6] Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, et al. Alzheimer's Disease Cooperative Study Group. Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med 2005;352:2379–88.

[7] Feldman HH, Ferris S, Winblad B, Sfikas N, Mancione L, He Y, et al. Effect of rivastigmine on delay to diagnosis of Alzheimer's disease from mild cognitive impairment: the InDDEx study. Lancet Neurol 2007;6:501–12.

[8] Winblad B, Gauthier S, Scinto L, Feldman H, Wilcock GK, Truyen L, Mayorga AJ, Wang D, Brashear HR, Nye JS, GAL-INT-11/18 Study Group. Safety and efficacy of galantamine in subjects with mild cognitive impairment. Neurology 2008;70:2024–35.

[9] Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. Am J Psychiatry 1984;141:1356–64.

[10] Benge J, Balsis S, Massman PJ, Doody RS. How well to the ADAS-cog and its subscales measure cognitive dysfunction in Alzheimer's disease? Dementia Geriatr Cog Disord 2009;28:63–9.

[11] Doraiswamy PM, Kaiser L, Bieber F, Garman RL. The Alzheimer's Disease Assessment Scale: evaluation of psychometric properties and patterns of cognitive decline in multicenter clinical trials of mild to moderate Alzheimer's disease. Alzheimer Dis Assoc Disord 2001;15:174–83.

[12] Fahn S, Elton R. UPDRS Program Members. Unified Parkinson's Disease Rating Scale. In: Fahn S, Marsden C, Calne D, Goldstein M, eds. Recent Developments in Parkinson's Disease. Florham Park. NJ: Macmillan Healthcare Information; 1987. p. 153–64.

[13] Olanow CW, Rascol O, Hauser R, Feigin PD, Jankovic J, Lang A, et al. A double-blind, delayed-start trial of rasagiline in Parkinson's disease. N Engl J Med 2009;361:1268–78.

[14] Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982;140:566–72.

[15] Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993;43:2412–4.

[16] Cedarbaum JM, Crans G, Grundman M. Seeing with new eyes: finding a path to early intervention trials in Alzheimer's Disease. J Nutr Health Aging 2010;14:306–9.

[17] 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.

[18] Coley N, Andrieu S, Jaros M, Weiner M, Cedarbaum J, Vellas B. Suitability of the Clinical Dementia Rating-Sum of Boxes (CDR-SB) as a single primary endpoint for Alzheimer's disease trials. Alzheimers Dement 2011;7. 602.e2–10.e2.

[19] Nunnally J, Bernstein I. Psychometric theory. 3rd ed. New York, NY: McGraw-Hill; 1994.

[20] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98.

[21] Aisen PS, Petersen RC, Donohue MC, Gamst A, Raman R, Thomas RG, et al. Alzheimer's Disease Neuroimaging Initiative. Clinical core of the Alzheimer's disease neuroimaging initiative: progress and plans. Azheimers Dement 2009;6:239–46.

[22] Pfeffer RI, Kuroasaki TT, Harrah CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol 1963;37:323–9.

[23] Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol 2000;53:459–68.
[24] Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull 1980;87:245–51.
[25] Ghiselli EE. Theory of psychologoical measurement. New York, NY: McGraw-Hill Book Company; 1964. p. 357–69.
[26] Morris JC, Ernesto C, Schafer K, Coats M, Leon S, Sano M, et al. Clinical dementia rating training and reliability in multicenter studies: the Alzheimer's Disease Cooperative Study experience. Neurology 1997; 47:1508–10.
[27] Rockwood K, Strang D, MacKnight C, Downer R, Morris JC. Interrater reliability of the Clinical Dementia Rating in a multicenter trial. J Am Geriatr Soc 2000;48:558–9.
[28] Morris JC, McKeel DW, Sorandt M, Rubin EH, Price JL, Grant MJ, et al. Very mild Alzheimer's disease: informant-based clinical, psychometric and pathologic distinction from normal aging. Neurology 1991;41. 463–48.
[29] Cano SJ, Posner HB, Moline ML, Hurt SW, Swartz J, Hsu T, Hobart JC. The ADAS-cog in Alzheimer's disease clinical trials: psychometric evaluation of the sum and its parts. Neurol Neurosurg Psychiatry 2010;81:1363–8.
[30] Benge JF, Balsis S, Geraci L, Massman PJ, Doody RS. How well do the ADAS-cog and its subscales measure cognitive dysfunction in Alzheimer's disease? Dement Geriatr Cogn Disord 2009; 28:63–9.
[31] Gillette-Guyonnet S, Nourhashemi F, Andrieu S, Cantet C, Micas M, Ousset PJ, Vellas B, REAL.FR Group. The REAL.FR research program on Alzheimer's disease and its management: methods and preliminary results. J Nutr Health Aging 2003;7:91–6.
[32] Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The Index of ADL: a standardized measure of biological and psychosocial function. JAMA 1963;185:914–9.
[33] Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969; 9:179–86.