ELSEVIER

CrossMark

Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 10 (2018) 153-160

Alzheimer's & Dementia

Cognitive & Behavioral Assessment

A novel cognitive-functional composite measure to detect changes in early Alzheimer's disease: Test–retest reliability and feasibility

Roos J. Jutten[a,*], John Harrison[a,b,c], Philippe R. Lee Meeuw Kjoe[a], Esther M. Opmeer[d], Niki S. M. Schoonenboom[e], Frank Jan de Jong[f], Craig W. Ritchie[g], Philip Scheltens[a], Sietske A. M. Sikkes[a,h]

[a]*Alzheimer Center, Department of Neurology, VU University Medical Center, Amsterdam Neuroscience, Amsterdam, The Netherlands*
[b]*Metis Cognition Ltd, Park House, Kilmington Common, Wiltshire, United Kingdom*
[c]*Institute of Psychiatry, Psychology & Neuroscience, King's College, London, United Kingdom*
[d]*Department of Neurosciences, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands*
[e]*Department of Neurology, Spaarne Gasthuis, Haarlem, The Netherlands*
[f]*Department of Neurology, Erasmus Medical Center, Rotterdam, The Netherlands*
[g]*Centre for Dementia Prevention, University of Edinburgh, Edinburgh, United Kingdom*
[h]*Department of Epidemiology & Biostatistics, VU University Medical Center, Amsterdam, The Netherlands*

Abstract

Introduction: To improve the detection of changes in Alzheimer's disease (AD), we designed the cognitive-functional composite (CFC). As a first validation step, we investigated its test–retest reliability and feasibility of use.

Methods: We performed a test–retest study with 2–3 weeks between assessments, including patients with mild cognitive impairment (MCI) or mild AD dementia and cognitively healthy participants. We calculated intraclass correlation coefficients (ICCs) type absolute agreement for all CFC measures and compared baseline and retest scores using paired-samples *t*-tests. We evaluated feasibility by interviewing participants.

Results: Forty-three patients (40% female, mean age = 69.9) and 30 controls (50% female, mean age = 65) were included. Subtest intraclass correlation coefficients ranged from .70 to .96. We found negligible improvements after retesting on only two subtests. Overall, patients perceived the administration of the CFC as feasible.

Discussion: The CFC is a stable and feasible measure in MCI and mild AD dementia, and thereby meets important quality metrics for clinically meaningful outcome measures.

Keywords: Activities of daily living; Alzheimer's disease; Cognition; Feasibility; Test–retest reliability; Outcome measures

1. Introduction

Neurodegenerative diseases leading to dementia are characterized by progressive cognitive decline and increasing interference in daily function [1]. Alzheimer's disease (AD), which is the main cause of dementia worldwide, is a continuum starting with a preclinical phase in which pathology develops but clinical symptoms are still absent [2]. It can be accompanied by subjective complaints as the first signal of the disease [3,4]. Cognitive deficits become more prominent in the prodromal phase of mild cognitive impairment (MCI) [5] and are ultimately severe enough to interfere with daily life in the dementia stage [6]. Measuring cognition and everyday functioning across the AD continuum is pivotal for monitoring clinical progression and evaluating both symptomatic relief and disease-modifying therapies. Currently used cognitive and

*Corresponding author. Tel.: +31 20 4448527.
E-mail address: r.jutten@vumc.nl

https://doi.org/10.1016/j.dadm.2017.12.002

functional measures have shown to be of insufficient quality for these purposes, due to their insensitivity to clinically meaningful changes over time [7]. For example, widely used tests such as the Mini–Mental State Examination (MMSE) [8] and the Alzheimer's Disease Assessment Scale–cognitive subscale (ADAS-Cog) [9] have been shown to be inappropriate for use in the design and evaluation of clinical trials targeted at MCI and mild AD [10–13]. As to the measurement of everyday functioning, recent reviews pointed out that most commonly used questionnaires are only of limited use to detect the early functional decline [14,15].

Consequently, many researchers have expressed the need for an improved measure that is capable of detecting clinically meaningful changes in MCI and mild AD [16,17]. To this end, the "Capturing Changes in Cognition" (Catch-Cog) project was initiated [18]. Based on preparatory work [19,20] and input from patients and experts, we designed a novel composite assessment combining measures of cognition and function: the "cognitive-functional composite" (CFC) [18]. The CFC comprises existing cognitive tests focusing on the domains that are known to be vulnerable to decline in incipient AD, specifically episodic memory (EM), working memory (WM), and executive functioning (EF) [17]. To amplify its clinical relevance, the CFC encompasses an everyday functioning questionnaire focusing on instrumental activities of daily living (IADL) [20,21]. IADL are activities that require the use of multiple cognitive processes and include activities such as cooking, driving, and managing finances. Difficulties in IADL performance are among the earliest clinical symptoms in MCI and mild AD dementia [15,22].

The present study reports on the first validation step of the novel CFC, in which we focused on its stability and feasibility. First, we investigated test–retest reliability of the CFC subtests. Second, we examined the influence of potential practice effects on the cognitive parts. Practice effects are improvements in cognitive test performance that may result from repeated exposure [23]. They are a potential threat for longitudinal cognitive assessment, as they can result in either underestimation of actual cognitive decline or overestimation of real treatment effects [24]. It is therefore important to explore the presence of practice effects on novel outcome measures designed for longitudinal use, such as the CFC. Previous studies on the presence of practice effects on cognitive tests in individuals with and without cognitive impairment have shown contrasting results [23–27]. Consequently, we explored potential practice effects on the CFC subtests separately for individuals with MCI or mild AD dementia and cognitively healthy individuals. Third, we computed an overall CFC score including all subtests. We investigated whether this score was influenced by age or education, and we examined the stability of this score in both groups. Finally, we evaluated feasibility of the CFC, with a focus on patients' experiences with respect to its administration time, modality, and perceived burden.

2. Methods

2.1. Study design

This study is a multicenter, observational, prospective cohort study, conducted at three Dutch sites and one United Kingdom (UK) site. We used a test–retest design with 2–3 weeks between assessments. Data were collected between November 2015 and November 2016. The Medical Ethical Committee of the VU University Medical Center (VUmc) approved the study for all Dutch centers. The South East Scotland Research Ethics Committee approved the study for the UK site. All participants and study partners gave written informed consent.

2.2. Participants

We included patients with a clinical diagnosis of MCI or probable AD dementia (n = 48) and cognitively healthy participants (n = 30) with their study partners. Patients were recruited via the VUmc Alzheimer Center Amsterdam, the Spaarne Gasthuis Haarlem, the Alzheimer Center Rotterdam, the University Medical Center Groningen, and the Centre for Dementia Prevention at the University of Edinburgh. Before inclusion in the present study, all patients had undergone a screening assessment including medical history, neurological, and neuropsychological examination in their center. Diagnoses were made in a multidisciplinary meeting containing at least a neurologist, psychiatrist or geriatrician and with neuropsychology input. MCI and probable AD were diagnosed according to the corresponding National Institute of Aging-Alzheimer's Association core clinical criteria [5,6]. Biomarkers (structural brain imaging or cerebrospinal fluid) were available for most but not all patients and were used to increase or decrease the likelihood of AD according to McKhann et al [6]. If not available, we relied on the clinical diagnosis of MCI and AD. Other inclusion criteria were as follows: (1) MMSE score $\geq$ 18 [8]; (2) age $\geq$ 50; and (3) availability of a study partner (i.e. a spouse, relative, or close friend) who was capable and willing to participate. Exclusion criteria were as follows: (1) neurological or psychiatric diagnoses other than AD (Geriatric Depression Scale score $\geq$ 6 [28]); (2) current or a history of substance abuse; or (3) participation in a clinical trial during the timeframe of the present study.

Cognitively healthy participants originated from an existing healthy volunteer database from the VUmc Alzheimer Center. Eligible participants had (1) an age $\geq$ 50; (2) neuropsychological test results within age- and education-adjusted norms; and (3) an available study partner.

2.3. Materials

The CFC was designed by the Catch-Cog working group [18]. It comprises seven existing, validated cognitive tests. Three EM tests originate from the ADAS-Cog [9]. These include the following: (1) Word Recognition, for which a participant has to learn a list of 12 words and identify these words when mixed among 12 other distracter words (one point for each incorrect response, score range 0–12); (2) Orientation, containing eight questions regarding the participant's orientation to person, place, and time (one point for each incorrect response, score range 0–8); and (3) Word Recall, where the participant is given three trials to learn a list of 10 high-imagery nouns (total score entails the average number of words not recalled across the three trials, score range 0–10). Other subtests originate from the neuropsychological test battery (NTB) and address different aspects of EF [29–31]. These include the following: (4) the Controlled Oral Word Association Test (COWAT), assessing the participant's phonological fluency skills using the letters D-A-T (F-A-S in the UK) and a total time of 60 seconds per letter (one point for each correct nonrepeated word); and (5) the Category Fluency Test, examining the participant's semantic fluency by generating as many animals within 60 seconds (one point for each correct unique animal). First results show this combination of three ADAS-Cog and two NTB tests to be a reliable composite measure in MCI and mild AD, with good internal consistency (Cronbach's alpha = .80) and good test–retest reliability at 4 (r = .89), 12 (r = .85), 18 (r = .84), and 24 weeks (r = .84) [19]. To cover the WM and EF domains more broadly, the following tests were additionally included: (6) the Wechsler Memory Scale Digit Span backward [32], demanding a participant to reproduce sequences of digits of increasing length in the reversed order (score range 0–12); and (7) the Wechsler Digit Symbol Substitution Test (DSST) [33], a timed EF test for which participants have to substitute as many digits by unique geometric symbols within 90 seconds (one point for each correct substituted symbol).

The functional part of the CFC consists of the Amsterdam IADL Questionnaire (A-IADL-Q): a computerized, informant-based questionnaire covering a broad range of cognitive IADL [20]. The A-IADL-Q consists of 70 items, which can be divided into eight subcategories: household, administration, work, computer use, leisure time, appliances, transport, and other activities. For each item, difficulty in performance is rated on a 5-point Likert scale (ranging from "no difficulty in performing this task" to "no longer able to perform this task"). Scoring is based on item response theory, a paradigm linking item responses to an underlying latent trait [34]. For the A-IADL-Q, this latent trait reflects IADL impairment. Total A-IADL-Q scores are normally distributed (M = 50, standard deviation [SD] = 10), with lower scores reflecting more IADL impairment. Previous work showed that the A-IADL-Q has excellent psychometric qualities: factor analysis supported unidimensionality, high internal consistency (reliability coefficient: .97), and good test–retest reliability (κ values $>$ .60 for 87.9% of the items) [20]. A validation study showed medium to high correlations with other measures of cognition and everyday functioning, suggesting good construct validity [21]. In addition, a longitudinal validation study demonstrated that the A-IADL-Q was sensitive to changes over time [35].

2.3.1. Feasibility interview

We evaluated the overall feasibility of the cognitive component by interviewing a subsample of the patient group (n = 15). These were all patients originating from the VUmc Alzheimer Center. The interviews contained open, neutrally formulated questions about the patient's general impression of the assessment. Subsequently, patients were asked to share their thoughts on specific aspects, such as total administration time, perceived burden and difficulty level, and quality of the test materials. If they addressed other issues that were important for them, we invited them to elaborate on these topics as well.

2.4. Procedures

Assessments took place at either the hospital or the participant's home, depending on the participant's preference. A trained rater assessed the cognitive tests according to standardized instructions, in the following order: Word Recognition, Orientation, COWAT, Category Fluency Test, Digit Span backward DSST, and Word Recall. In total, this took approximately 20–25 minutes. In the meantime, the study partner completed the A-IADL-Q independently on a tablet computer in a separate room, which took about 20–25 minutes as well. Baseline and retest assessments had similar content for participants and study partners, except that parallel forms of the wordlists for Word Recognition and Word Recall were used. Feasibility interviews were held directly after the retest assessment. Before the start of the interview, we mentioned that the test was still in development, and we explicitly invited participants to mention aspects for improvement if they had any.

2.5. Statistical analyses

Statistical analyses were performed using SPSS, version 22.0. Statistical significance was set at $P < .05$. Demographic differences were investigated using independent-samples t-tests or chi-square tests as appropriate. For the subtest analyses, we used the raw test scores for the cognitive measures and the total A-IADL-Q score, which were all unadjusted for age and education. We investigated test–retest reliability for all individual subtests using intraclass correlation coefficients (ICCs), using a two-way random model and type absolute agreement. ICCs $>$ .70 were considered acceptable [36]. For subtests with lower ICCs, we applied the Bland–Altman method to explore systematic differences

by (1) visual inspection; and (2) linear regression analyses with difference between baseline and retesting as dependent variable and the average score of baseline and retesting as independent variable [37]. To investigate potential practice effects, we compared baseline and retest scores for all cognitive subtests in the separate groups using paired-samples *t*-tests or Wilcoxon signed-rank test as nonparametric alternative. Subsequently, we calculated Cohen's *d* effect-sizes for change scores. For subtests showing a significant change score, we performed multiple linear regression analyses with change score as dependent variable, and age and education as independent variables, which were stepwise entered into the model based on strength of their correlation with dependent variable.

To create an overall CFC score, we first reversed the ADAS-Cog subtest scores so that higher scores reflected better performance. Subsequently, all subtest baseline and retest scores were converted into Z-scores with baseline means and SDs of the total group as reference values, and added up to an overall CFC score. We investigated whether baseline CFC scores differed between patients and controls using an independent-samples *t*-test. We investigated possible associations with age or education by computing Pearson's correlation coefficients or Spearman's rho if appropriate. We examined stability of the CFC scores using the same test–retest analyses as described previously.

3. Results

3.1. Demographic characteristics

Five participants (all patients, mean age 69.2, n = 1 female) withdrew from the study before their retest visit due to personal reasons and were therefore excluded from the analyses. They did not differ from the other patients with respect to demographics. The remaining patients (n = 43, 40% female) had a mean age of 69.9 (SD = 7.4), a mean MMSE score of 25.3 (SD = 3.3), and received 14.3 (SD = 5.1) years of education. Controls (n = 30, 50% female) were slightly younger (M = 65, SD = 7.1, P = .006) and marginally more highly educated (M = 16.9, SD = 4.0, P = .022) than patients (see Table 1).

Table 1
Demographic characteristics of participants and their study partners, separate for the patient and control group

Characteristics	Patients (n = 43)	Controls (n = 30)	*P* value
Participants			
Female (%)	17 (40)	15 (50)	.376†
Age (SD)	69.9 (7.4)	65.0 (7.1)	.006*
Years of education (SD)	14.3 (5.1)	16.9 (4.0)	.022*
MMSE (SD)	25.2 (3.2)	NA	
Study partners			
Female (%)	28 (65)	19 (63)	.876†
Age (SD)	66.9 (9.8)	62.5 (10.2)	.070*
Relationship partner (%)	37 (86)	23 (77)	
Relationship > 10 years (%)	42 (98)	30 (100)	

Abbreviations: SD, standard deviation; MMSE, Mini–Mental State Examination; NA, not assessed.
*Tested using independent *t*-test.
†Tested using chi-square test.

3.2. Test–retest reliability

Table 2 presents the ICCs for all subtests, which ranged from .70 to .96. Word Recognition was the only subtest with a borderline ICC of .70. The corresponding Bland–Altman plot showed an equal number of data points above and below the mean difference line, revealing no systematic differences between the two measurements (Fig. 1). This was supported by the regression analysis that showed no significant linear relationship among the data points (β = .104, t = 1.407, P = .164). When looking at the distribution of data points for the distinct groups, the Bland–Altman plot showed that the control group had many similar scores and low between-subject variance which probably influenced the ICC.

3.2.1. Practice effects

Table 2 also shows the ranges, baseline, and retest scores of the individual subtests. In the patient group, we only found a significant higher score after retesting for the DSST (observed change = +2.4, SD = 5.2, P = .005). In the control group, we found a significant higher score after retesting for the DSST (observed change = +2.8, SD = 5.5, P = .011) and COWAT (observed change = +2.5, SD = 7.8, P = .012). The corresponding effect-sizes of these changes scores fell in the low to medium range (Table 2). Regression analyses showed that COWAT and DSST change scores were not confounded by age or education: all regression coefficients were nonsignificant (Table 3).

3.2.2. CFC score

Figure 2 displays the distribution of baseline CFC scores for patients and controls. Mean score significantly differed between groups (mean difference = 9.7, SD = .8, t = 12.12, P < .001). We found no associations with age (patients: r = .09, P = .573; controls: r = −.18, P = .35) or education (patients: r = .14, P = .378; controls: r = −.01, P = .946) and baseline CFC score. The last row in Table 2 presents the mean CFC scores at baseline and retesting. We found no significantly different CFC score at retest for patients, whereas in controls, the CFC score was slightly higher after retesting (observed change = +.9, SD = 1.5, P = .006).

3.3. Feasibility

Patients' general impressions of the CFC were mainly positive (e.g. described as "interesting", "enjoyable", and "relevant"), although some patients described the test as "challenging" or "confronting". Some specific subtests were frequently defined as "difficult", particularly the tests involving word lists. The majority stated that they did not

Table 2
Subtest characteristics: ICC's, score ranges, and mean scores in the patient and control group

	Total group									
			Patients				Controls			
Measure	ICCs	Score range	Baseline (SD)	Retest (SD)	*P* value	Cohen's d	Baseline (SD)	Retest (SD)	*P* value	Cohen's d
ADAS-Cog Word Recognition	.70	0–12	4.6 (2.8)	4.8 (3.1)	.684†	.07	1.2 (1.3)	0.7 (1.1)	.097†	.40
ADAS-Cog Orientation	.72	0–4	1.1 (1.3)	1.2 (1.3)	.728†	.08	0.0 (0.2)	0.1 (0.3)	.564†	.12
COWAT (DAT/FAS, 60 sec)	.85	4–69	31.7 (13.3)	33.7 (14.4)	.067*	.15	42.9 (10.6)	45.4 (8.7)	.012†	.26
CFT (animals, 60 sec)	.83	4–35	13.7 (5.7)	14.4 (5.7)	.260*	.12	24.2 (5.0)	25.3 (4.7)	.185*	.23
DSST (90 sec)	.93	3–73	31.1 (11.7)	33.5 (11.4)	.005*	.21	52.4 (9.6)	55.2 (11.0)	.011*	.28
DSB	.72	2–12	5.5 (1.5)	5.5 (2.0)	.561†	.04	6.7 (1.6)	7.2 (2.0)	.081*	.28
ADAS-Cog Word Recall	.89	0–9	5.6 (1.5)	5.8 (1.4)	.192†	.14	2.5 (1.22)	2.7 (1.2)	.141†	.17
A-IADL-Q	.96	22.8–76.0	49.7 (10.4)	49.8 (10.4)	.830*	NA	71.2 (4.8)	71.8 (3.7)	.432*	NA
Overall CFC score	.95	−14.9 to 12.36	−3.9 (4.3)	−3.7 (4.3)	.412*	NA	5.7 (2.4)	6.6 (2.7)	.006*	NA

Abbreviations: ADAS-Cog, Alzheimer's Disease Assessment Scale–cognitive subscale; ICC, intraclass correlation coefficient; SD, standard deviation; COWAT, Controlled Oral Word Association Test; CFT, Category Fluency Test; DSST, Digit Symbol Substitution Test; DSB, Digit Span backward; A-IADL-Q, Amsterdam Instrumental Activities of Daily Living Questionnaire; CFC, cognitive-functional composite; NA, not assessed for this score.

*Tested using paired-samples *t*-test.

†Tested using Wilcoxon rank-sign test.

perceive the test administration as burdensome. In general, the total duration (20–25 minutes) was experienced as acceptable. Patients' descriptions of their ability to concentrate on the entire test ranged from "reasonable" to "very good". Test materials were described as "clear" and "very readable".

4. Discussion

We addressed the stability and feasibility of the CFC by performing a test–retest study and collecting input from participants. Our findings demonstrated that all CFC components had moderate to high test–retest reliability and were only limitedly affected by practice effects. We showed that the overall CFC score was not influenced by age or education and provided a stable measure especially for patients. Finally, qualitative interviews indicated that patients perceived the administration of the CFC overall as feasible. The present study expands on previous work on the cognitive composite [19] and A-IADL-Q [20], which form the underlying basis of the CFC. Our findings with respect to test–retest reliability of the CFC components are largely in line with these former validation studies.

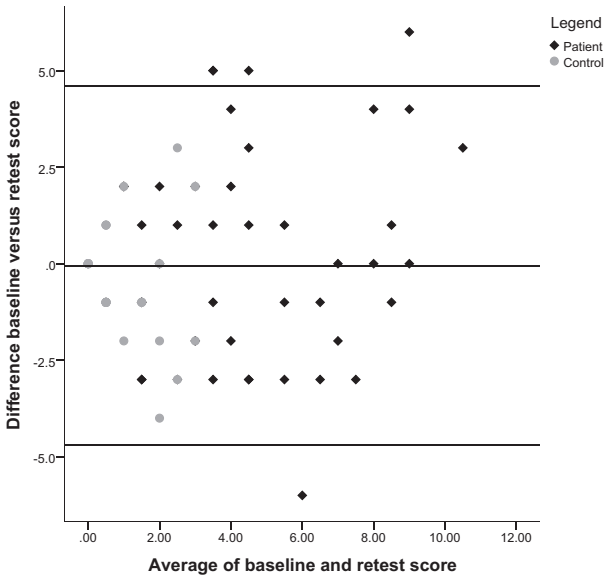

Fig. 1. Bland–Altman plot for subtest ADAS-Cog Word Recognition. Abbreviation: ADAS-Cog, Alzheimer's Disease Assessment Scale–cognitive subscale.

Our findings regarding practice effects in the patient group are similar to previous research, whereas our findings in the control group were somewhat different compared with other studies. In patients, we found limited presence of practice effects, which was also reported earlier [23,26]. We only observed a significant improvement on the DSST; however, the clinical meaningfulness of the extent of this change (only two points) seems negligible. A slight improvement on the DSST in patients with AD seems to be a robust finding because other studies observed this as well [24]. In the cognitively healthy group, we only observed small practice effects on the DSST and COWAT. This is somewhat different from previous studies demonstrating larger practice effects on a wide variety of cognitive tests in cognitively healthy adults, for example on word list tasks, category fluency, and digit span backward [23,24]. A possible explanation for the current findings is that most tests showed ceiling effects in baseline scores of the control group, making it by definition impossible to further improve these scores and to observe a potential practice effect. Furthermore, the use of parallel versions for the word lists tasks may have limited the magnitude of practice effects in our study, in both patients and controls [27].

Strengths of this study include the use of a prospective cohort to investigate the CFC's test–retest reliability. An independent validation of promising measures outside of the

Table 3
Regression coefficients for age and education on change scores for the DSST and COWAT

	Patients			Controls		
Measure	β	*t*	*P* value	β	*t*	*P* value
DSST						
Education	−.238	−1.511	.139	.206	1.086	.287
Age	−.090	−.573	.570	.150	.790	.436
COWAT						
Education	NA			.122	.642	.526
Age	NA			.123	.649	.522

Abbreviations: DSST, Digit Symbol Substitution Test; COWAT, Controlled Oral Word Association Test; NA, not applicable for this group.

larger, retrospective data sets in which they were designed is a key requirement for any proposed new measure. In addition, the involvement of participants to evaluate feasibility of our measurement is a unique and important aspect of this study. An important advantage of this qualitative method is that it allows participants to address topics that we had not considered on forehand [38]. Our semi-structured interview allowed us to ask more in-depth questions on topics that appear to be relevant, which contributes to a more comprehensive understanding of the participant's thoughts and beliefs [39]. Ultimately, acceptability by the target population is highly important as it may contribute to future implementation of the CFC.

There are however some limitations that should be considered. First, we were not able to investigate the CFC separately in the MCI and mild AD dementia group due to our relatively small sample size, which was based on recommendations for test–retest studies [40]. Validating the CFC in a larger cohort, including both MCI and mild AD patients, is however needed and will be the scope of a future study addressing the longitudinal psychometric properties of the CFC [18]. This larger cohort will also allow us to explore whether the CFC score further improves if the components are weighted differently [41]. Second, the etiology of AD was not confirmed for all patients with MCI because some of them had no biomarkers available. In these cases, we relied on the clinical diagnosis, and we excluded patients with other neurological or psychiatric symptoms to ensure that MCI was not attributable to these factors. Third, our control and patient group were not equally matched with respect to age and education. However, we showed that practice effects were not confounded by education and age nor was our CFC score, indicating that this has probably not influenced our results. Finally, our study involved only one repeated measurement, whereas other studies on practice effects included more measurements over a larger time period [23,25,26]. However, these studies have shown that practice effects seem to be most prominent after the first repeated measurement. Our finding that the CFC only shows negligible practice effects after testing the second time around is thus promising for any additional repeated assessments.

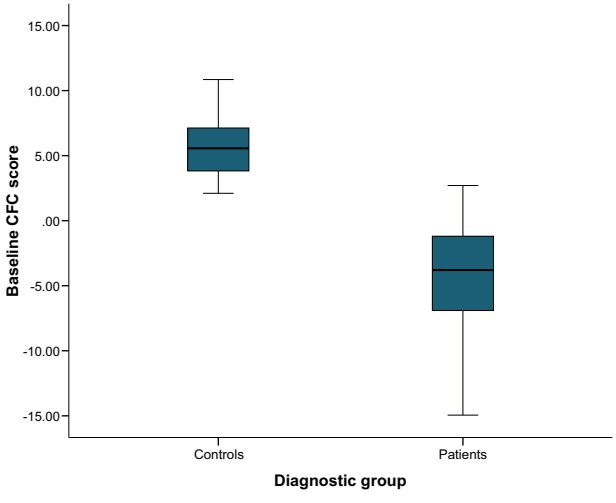

Fig. 2. Boxplots displaying baseline CFC scores separately for patients and controls. Abbreviation: CFC, cognitive-functional composite.

In conclusion, we demonstrated that the CFC is a stable and feasible measure in MCI and mild AD dementia. It thereby meets important quality metrics for clinically meaningful outcome measures. These findings are promising for the next step in our validation plan, which is to determine the CFC's sensitivity to clinically relevant changes over time. The ultimate goal of the Catch-Cog study on the CFC is to improve longitudinal clinical measurement in early AD. The CFC is a concise measure with a relatively short administration time, while being comprehensive enough to have the potential to capture clinically meaningful changes over time. Especially, the combination of cognitive tests and an everyday functioning measure will advance the clinical relevance of the composite score [42]. The CFC will be more time efficient than traditional measures and provide a more meaningful and sensitive outcome measures for monitoring progression and evaluating future clinical trials. This will eventually contribute to the still ongoing quest for a disease-modifying or even preventive (drug) therapy against AD.

Acknowledgments

We would like to thank Sahar Bahrami, Kate Forsyth, Sophie Leijdesdorff, Raymond den Bakker, and Ralph Vreeswijk for their help with data collection and recruitment. In addition, we would like to acknowledge Stichting Buytentwist for their contributions.
The present study is supported by a grant from ZonMw Memorabel (grant number 733050205). Research of the VUmc Alzheimer Center is a part of the neurodegeneration research program of the Amsterdam Neuroscience institute. The VUmc Alzheimer Center is supported by Alzheimer Nederland and Stichting VUmc Fonds. Part of this article has been presented at the 2017 Alzheimer's Association International Conference.

The Amsterdam IADL Questionnaire© is free for use in all public health and not-for-profit agencies and can be obtained from the authors following a simple registration.

In the past 2 years, J.H. has received honoraria and paid consultancy from 23andMe, Abbvie, A2Q, Amgen, Anavex, Astellas, AstraZeneca, Avraham, Axon, Axovant, Biogen Idec, Boehringer Ingelheim, Bracket, Catenion, Cognitive Therapeutics, CRF Health, DeNDRoN, EnVivo Pharma, Enzymotec, ePharmaSolutions, Eisai, Eli Lilly, Forum Pharma, Fresh Forward, GfHEu, Heptares, Janssen AI, Johnson & Johnson, Kaasa Health, Kyowa Hakko Kirin, Lundbeck, MedAvante, Merck, MyCognition, Mind Agilis, Neurocog, Neurim, Neuroscios, Neurotrack, Novartis, Nutricia, Orion Pharma, Pharmanet/i3, Pfizer, Prana Biotech, PriceSpective, Probiodrug, Prophase, Prostrakan, Regeneron, Reviva, Roche, Sanofi, Servier, Shire, Takeda, TCG, TransTech Pharma, and Velacor.

S.A.M.S. is supported by grants from JPND and ZonMw and has provided consultancy services in the past 2 years for Nutricia and Takeda. All funds were paid to her institution.

The authors have declared that no conflict of interest exists.

RESEARCH IN CONTEXT

1. Systematic review: We aimed to assess test–retest reliability, potential to be influenced by practice effects, and feasibility of use of our recently designed cognitive-functional composite (CFC). We searched PubMed for publications reporting on practice effects on cognitive tests that are included in the CFC.
2. Interpretation: We found moderate to high test–retest reliability for all CFC subtests. We only found negligible practice effects on one subtest in patients with mild cognitive impairment or mild dementia due to Alzheimer's disease. The administration of the CFC was perceived as feasible. We demonstrated that the CFC is a stable and feasible measure and thereby meets important quality metrics for clinically meaningful outcome measures to assess changes over time.
3. Future directions: These findings are promising for the next step in our validation plan: we will perform a longitudinal construct validation study to determine the sensitivity of the CFC to clinically relevant changes over time.

References

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). Arlington, VA: American Psychiatric Pub; 2013.

[2] Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 2010;9:119–28.

[3] Jessen F, Amariglio RE, van Boxtel M, Breteler M, Ceccaldi M, Chételat G, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement 2014;10:844–52.

[4] Sperling R, Mormino E, Johnson K. The evolution of preclinical Alzheimer's disease: implications for prevention trials. Neuron 2014; 84:608–22.

[5] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC. 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.

[6] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH. 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.

[7] Karin A, Hannesdottir K, Jaeger J, Annas P, Segerdahl M, Karlsson P, et al. Psychometric evaluation of ADAS-Cog and NTB for measuring drug response. Acta Neurol Scand 2014;129:114–22.

[8] 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.

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

[10] Chapman KR, Bing-Canar H, Alosco ML, Steinberg EG, Martin B, Chaisson C, et al. Mini Mental State Examination and Logical Memory scores for entry into Alzheimer's disease trials. Alzheimers Res Ther 2016;8:9.

[11] Spencer RJ, Wendell CR, Giggey PP, Katzel LI, Lefkowitz DM, Siegel EL, et al. Psychometric limitations of the mini-mental state examination among nondemented older adults: an evaluation of neurocognitive and magnetic resonance imaging correlates. Exp Aging Res 2013;39:382–97.

[12] Cano SJ, Posner HB, Moline ML, Hurt SW, Swartz J, Hsu T, et al. The ADAS-cog in Alzheimer's disease clinical trials: psychometric evaluation of the sum and its parts. J Neurol Neurosurg Psychiatry 2010; 81:1363–8.

[13] Grochowalski JH, Liu Y, Siedlecki KL. Examining the reliability of ADAS-Cog change scores. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2016;23:513–29.

[14] Kaur N, Belchior P, Gelinas I, Bier N. Critical appraisal of questionnaires to assess functional impairment in individuals with mild cognitive impairment. Int Psychogeriatr 2016;28:1425–39.

[15] Jekel K, Damian M, Wattmo C, Hausner L, Bullock R, Connelly PJ, et al. Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review. Alzheimers Res Ther 2015;7:17.

[16] Snyder PJ, Kahle-Wrobleski K, Brannan S, Miller DS, Schindler RJ, DeSanti S, et al. Assessing cognition and function in Alzheimer's disease clinical trials: do we have the right tools? Alzheimers Dement 2014;10:853–60.

[17] Vellas B, Andrieu S, Sampaio C, Coley N, Wilcock G. Endpoints for trials in Alzheimer's disease: a European task force consensus. Lancet Neurol 2008;7:436–50.

[18] Jutten RJ, Harrison J, de Jong FJ, Aleman A, Ritchie CW, Scheltens P, et al. A composite measure of cognitive and functional progression in Alzheimer's disease: Design of the Capturing Changes in Cognition study. Alzheimer's & Dementia: Translational Research & Clinical Interventions 2017;3:130–8.

[19] Harrison J, Dgetluck N, Gawryl M, Moebius H, Hilt D. Validation of a novel cognitive composite assessment for mild and prodromal Alzheimer's disease. Alzheimers Dement 2013;9:P661.

[20] Sikkes SA, de Lange-de Klerk ES, Pijnenburg YA, Gillissen F, Romkes R, Knol DL, et al. A new informant-based questionnaire for instrumental activities of daily living in dementia. Alzheimers Dement 2012;8:536–43.
[21] Sikkes SA, Knol DL, Pijnenburg YA, de Lange-de Klerk ES, Uitdehaag BM, Scheltens P. Validation of the Amsterdam IADL Questionnaire(c), a new tool to measure instrumental activities of daily living in dementia. Neuroepidemiology 2013; 41:35–41.
[22] Jutten RJ, Peeters CF, Leijdesdorff SM, Visser PJ, Maier AB, Terwee CB, et al. Detecting functional decline from normal aging to dementia: development and validation of a short version of the Amsterdam IADL Questionnaire. Alzheimers Dement (Amst) 2017; 31:26–35.
[23] Calamia M, Markon K, Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. Clin Neuropsychol 2012;26:543–70.
[24] Goldberg TE, Harvey PD, Wesnes KA, Snyder PJ, Schneider LS. Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer's disease randomized controlled trials. Alzheimers Dement (Amst) 2015;1:103–11.
[25] Hassenstab J, Ruvolo D, Jasielec M, Xiong C, Grant E, Morris JC. Absence of practice effects in preclinical Alzheimer's disease. Neuropsychology 2015;29:940–8.
[26] Machulda MM, Pankratz VS, Christianson TJ, Ivnik RJ, Mielke MM, Roberts RO, et al. Practice effects and longitudinal cognitive change in normal aging vs. incident mild cognitive impairment and dementia in the Mayo Clinic Study of Aging. Clin Neuropsychol 2013; 27:1247–64.
[27] Beglinger LJ, Gaydos B, Tangphao-Daniels O, Duff K, Kareken DA, Crawford J, et al. Practice effects and the use of alternate forms in serial neuropsychological testing. Arch Clin Neuropsychol 2005; 20:517–29.
[28] Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1983;17:37–49.
[29] Harrison J, Minassian SL, Jenkins L, Black RS, Koller M, Grundman M. A neuropsychological test battery for use in Alzheimer disease clinical trials. Arch Neurol 2007;64:1323–9.
[30] Harrison J, Rentz DM, McLaughlin T, Niecko T, Gregg KM, Black RS, et al. Cognition in MCI and Alzheimer's disease: baseline data from a longitudinal study of the NTB. Clin Neuropsychol 2014;28:252–68.
[31] Lezak MD. Neuropsychological assessment. USA: Oxford University Press; 2004.
[32] Wechsler D. Wechsler memory scale–fourth edition (WMS-IV). San Antonio, TX: Pearson; 2009.
[33] Wechsler D. Wechsler adult intelligence scale–Fourth Edition (WAIS–IV), 22. San Antonio, TX: NCS Pearson; 2008. p.498.
[34] Embretson SE, Reise SP. Item response theory. New York, NY: Psychology Press; 2013.
[35] Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SA. The sensitivity to change over time of the Amsterdam IADL Questionnaire((c)). Alzheimers Dement 2015;11:1231–40.
[36] Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Conditioning Res 2005; 19:231–40.
[37] Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;327:307–10.
[38] Ostlund U, Kidd L, Wengstrom Y, Rowa-Dewar N. Combining qualitative and quantitative research within mixed method research designs: a methodological review. Int J Nurs Stud 2011;48:369–83.
[39] Green J, Thorogood N. Qualitative methods for health research. London, United Kingdom: Sage; 2013.
[40] de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. New York: Cambridge University Press; 2011.
[41] Edland SD, Ard MC, Li W, Jiang L. Design of pilot studies to inform the construction of composite outcome measures. Alzheimers Dement (N Y) 2017;3:213–8.
[42] Royall DR, Lauterbach EC, Kaufer D, Malloy P, Coburn KL, Black KJ. The cognitive correlates of functional status: a review from the Committee on Research of the American Neuropsychiatric Association. J Neuropsychiatry Clin Neurosci 2007;19:249–65.