Assessing cognition and daily function in early dementia using the Cognitive-Functional Composite: findings from the Catch-Cog study cohort

Background: The Cognitive-Functional Composite (CFC) was designed to improve the measurement of clinically relevant changes in predementia and early dementia stages. We have previously demonstrated its good test-retest reliability and feasibility of use. The current study aimed to evaluate several quality aspects of the CFC, including construct validity, clinical relevance and suitability for the target population.

Methods: Baseline data of the Capturing Changes in Cognition study was used: an international, prospective cohort-study including participants with subjective cognitive decline (SCD), mild cognitive impairment (MCI), Alzheimer’s disease (AD) dementia, and dementia with Lewy Bodies (DLB). The CFC comprises seven existing cognitive tests focusing on memory and executive functions (EF), and the informant-based Amsterdam Instrumental Activities of Daily Living Questionnaire (A-IADL-Q). Construct validity and clinical relevance were assessed by 1) confirmatory factor analyses (CFA) using all CFC subtests; and 2) linear regression analyses relating the CFC score (independent) to reference measures of disease severity (dependent), correcting for age, sex and education. To assess suitability for the target population, we compared score distributions of the CFC to those of traditional tests (Alzheimer’s Disease Assessment Scale – Cognitive subscale, Alzheimer’s Disease Cooperative Study - Activities of Daily Living scale and Clinical Dementia Rating scale).

Results: A total of 184 participants were included (age 71.8±8.4; 42% female; n=14 SCD, n=80 MCI, n=78 AD and n=12 DLB). CFA showed that the hypothesized three-factor model (memory, EF and IADL) had adequate fit (CFI=.931, RMSEA=.091, SRMR=.06). Moreover, worse CFC performance was associated with more cognitive decline as reported by the informant (β=.61, p<.001), poorer quality of life (β=.51 , p<.001), higher caregiver burden (β=.51, p<.001), more apathy (β =-.36, p<.001) and less cortical volume (β=.34, p=.02). Whilst correlations between the CFC and traditional measures were moderate-to-strong (ranging from -.65 to .83, all p<.001), histograms showed floor and ceiling effects for the traditional tests as compared to the CFC.

Conclusions: Our findings illustrate that the CFC has good construct validity, captures clinically relevant aspects of disease severity and shows no range restrictions in scoring. It therefore provides a more useful outcome measure than traditional tests to evaluate cognition and function in MCI and mild AD.
1. Background

Alzheimer’s disease (AD) is the leading cause of dementia worldwide, and has been the target of clinical trials and intervention studies for many years [1]. In the past decade, the research field has shifted towards earlier clinical stages of dementia and to pre-dementia stages such as Mild Cognitive Impairment (MCI) [2]. Remarkably, the selection of cognitive and functional outcome measures to evaluate treatment effects has not been adapted to the shift in treatment target populations. Measures originally designed for mild to severe dementia, such as the Alzheimer’s Disease Assessment Scale – Cognitive subscale (ADAS-Cog [3]), are still widely used as primary endpoints in MCI and early AD dementia trials. Several studies have shown that those older, traditional measures are insensitive to change over time in MCI and mild dementia [4, 5], as they focus on cognitive domains and everyday activities that are unaffected in those disease stages [5-7]. This limits their clinical relevance in the pre-dementia window and leads to range restrictions in scoring (i.e. floor and ceiling effects) [8, 9]. Hence, researchers and regulatory agencies have expressed the urgent need for a sensitive measure that is capable of detecting clinically relevant changes at early clinical stages of AD [10-14]. The same holds for dementia with Lewy bodies (DLB), the second most common cause of dementia, of which both pathology as well as clinical manifestations show considerable overlap with AD [15, 16].

The Capturing Changes in Cognition (Catch-Cog) project was initiated to fulfill the need for a sensitive, clinically relevant outcome measure for use in MCI and mild dementia. We designed a novel Cognitive-Functional Composite (CFC) measure in our expert working group, basing our selections on previously published work and input from MCI and dementia patients and caregivers [17]. The resulting CFC consists of a short cognitive test battery focusing on memory and executive functioning (EF) [18], as these are the cognitive domains that have been shown to decline in pre-dementia and early stages of dementia [14]. The rationale of specific cognitive tests included in this battery has been described in more detail elsewhere [17]. Briefly, the selection was based on empirical evidence on their sensitivity to change, as reflected by the absence of floor and ceiling effects in MCI and mild AD [9, 18]. To amplify its clinical relevance, we augmented the cognitive test battery with a previously validated everyday functioning measure: the Amsterdam Instrumental Activities of Daily Living Questionnaire (A-IADL-Q) [19]. The A-IADL-Q assesses problems in cognitively complex everyday activities such as cooking, managing finances and using technological devices [20, 21]. In item response theory analyses, these activities were found to be the activities most sensitive to cognitive decline [22]. The A-IADL-Q was previously demonstrated to be sensitive to decline in dementia, as well as able to capture difficulties in IADL...
functioning in MCI and subjective cognitive decline (SCD) [22, 23].

Combining these cognitive and functional measures into the CFC summarizes cognitive and everyday performance and may thereby provide a single clinically relevant score. Both the Food and Drug Association (FDA) and European Medical Agency (EMA) encourage the use of such composite endpoints to evaluate novel drug therapies and interventions [12, 13]. These agencies also stipulate that composite measures should be 1) carefully designed, 2) validated in an independent prospective cohort-study, and 3) ‘bear some relevance to existing tools for which historical experience exists’ [13]. In the Catch-Cog study, we have sought to meet these criteria by performing an extensive construct validation of the CFC. In a previous report, we demonstrated that the CFC exhibits high test-retest reliability and good feasibility of use [24], which are pivotal prerequisites for a reliable and valid outcome measure [25]. Having demonstrated this, we embarked on a longitudinal construct validation in an independent prospective cohort across the clinical spectrum from SCD to mild dementia. The main aim of this study is to validate the CFC in MCI and mild AD dementia stages, but we will also explore whether the CFC could be of use in individuals with SCD and DLB.

In the current study we performed a psychometric evaluation of the CFC using baseline data of the Catch-Cog study. We aimed to evaluate several quality aspects of the CFC, including the construct validity, clinical relevance and suitability for the target population. Therefore, we investigated the CFC’s factor structure and compared the CFC score with reference measures of disease severity such as informant-reports and global cortical atrophy [17]. We also examined CFC score distributions in direct comparison to currently used tests, including the ADAS-Cog [3], the Clinical Dementia Rating scale Sum of Boxes (CDR-SB) [26], the Alzheimer's Disease Cooperative Study - Activities of Daily Living scale (ADCS-ADL) [27] and the Alzheimer's Disease Composite Score (ADCOMS) procedure [28].

2. Methods

2.1. Study design and participants
In this cross-sectional study, we employed baseline data from the Catch-Cog study, which is an international, multicenter, prospective cohort-study. Participants (N=184) were recruited via the 1) Alzheimer Center Amsterdam, Amsterdam UMC, location VU University Medical Center, the Netherlands (n=102); 2) Alzheimer Center Erasmus Medical Center (EMC, n=14), Rotterdam, the Netherlands; 3) University Medical Center Groningen (UMCG, n=39), The Netherlands; and 4) the Centre for Dementia Prevention, Edinburgh, Scotland.
We recruited participants who met the research criteria for SCD [29], the clinical criteria for MCI due to AD [30], probable AD dementia [31] or DLB dementia [16]. Other inclusion criteria were: 1) Mini-Mental State Examination (MMSE) score ≥ 18; 2) age ≥ 50; and 3) availability of a study partner who was able and willing to participate. Exclusion criteria were 1) presence of any other neurological disorder; 2) presence of a major psychiatric disorder such as severe personality disorder or depression (Geriatric Depression Scale score ≥ 6 [32]); 3) current abuse of alcohol and/or drugs; 4) simultaneously participating in a clinical trial.

Before inclusion, participants had undergone a standard diagnostic work-up in their study center, including at least medical history, neurological examination and cognitive assessment. Structural brain imaging was available for a subset of the study cohort. Diagnoses were performed during a multidisciplinary consensus meeting, containing at least a neurologist or psychiatrist and with neuropsychology input. In the UMCG, SCD and MCI participants were also recruited via advertisements in local newspapers. After responding to this advertisement, eligible participants were screened by a neuropsychologist and neurologist to investigate whether they met the criteria for SCD or MCI [29, 30].

The Medical-Ethical Committee of the VU University Medical Center approved the study for all Dutch centers. The South East Scotland Research Ethic Committee approved the study for the Scottish site. All participants and study partners provided written and oral informed consent.

### 2.2. The Cognitive-Functional Composite

**Cognitive component.** The cognitive test battery of the CFC included the three ADAS-Cog memory subscales Word Recognition, Word Recall and Orientation [3]; the Controlled Oral Word Association Test (COWAT); Category Fluency Test (CFT); Digit Span Backward (DSB) and Digit Symbol Substitution Test (DSST) [33]. During the Word Recognition test, the participant is required 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). During Word Recall, the participant is given three trials to learn a list of ten high-imagery nouns (total score entails the average number of words not recalled across the three trials, score range 0-10). The Orientation subtest includes eight questions regarding the participant's orientation to person, place and time (one point for each incorrect response, score range 0-8); and (3). The COWAT assesses the participant’s phonemic fluency skills using the letters D-A-T in The Netherlands or F-A-S in English and a total time of 60 seconds per letter (one point for each correct non-repeated word). The CFT examines the participant’s semantic fluency by requiring them to generate as many exemplars of the category animals within 60 seconds (one point for each correct unique animal). The DSB requires the participant to reproduce sequences...
of digits of increasing length in the reversed order (score range 0-12). The DSST is a timed EF test during which participants have to substitute as many digits by unique geometric symbols within 90 seconds (one point for each correct substituted symbol).

Functional component. The functional component consisted of the short version of the A-IADL-Q [22]. The A-IADL-Q is a computerized, informant-based questionnaire covering a broad range of complex IADL [20]. The short version consists of 30 items covering household, administration, work, computer use, leisure time, appliances, and transport 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]. This results in a latent trait score (z-score), reflecting one's level of IADL functioning, with higher scores indicating better IADL functioning [22].

CFC scoring. To create CFC scores, the directionality of the three ADAS-Cog subtest scores were reversed, so that higher scores reflected better performance. Subsequently, all cognitive subtest scores were transformed into z-scores with total group mean and standard deviation (SD) as reference values. The cognitive composite was computed as a weighted z-score of all seven cognitive subtests, whereas the functional component score was the A-IADL-Q score. The overall CFC composite score was computed as a weighted z-score of the cognitive composite and A-IADL-Q, with higher scores indicating better performance.

2.3. Reference measures

Traditional tests of cognition and function. Traditional tests to compare the CFC with included the MMSE, ADAS-Cog-13, ADCS-ADL, and CDR-SB. The MMSE is a global cognitive screening test, with total score ranging from 0 to 30 and higher scores reflecting better performance [35]. The ADAS-Cog-13 yields a measure of cognitive performance by combining ratings of 13 subtests (e.g. word lists recognition and recall, constructional praxis, object and finger naming). Total scores range from 0 to 85, with higher scores indicating more severe impairment [3]. The ADCS-ADL assesses functional abilities affected in mild-to-moderate AD. For 23 different basic and instrumental activities, the levels of performance and independency during the past four weeks were rated by the study partner. Total scores range from 0 (non-performance or need for extensive help) to 78 (independent performance) [27]. The CDR has been developed for the staging of dementia severity. The participant's cognitive and functional performance is rated in six areas: memory, orientation, judgment & problem solving, community affairs, home & hobbies, and personal care. Each area is rated as 0 ('healthy'), 0.5 ('questionable dementia'), 1 ('mild dementia'), 2 ('moderate dementia'), or 3
‘severe dementia’). Adding the rating of all boxes results in a total CDR-SB score ranging from 0 to 18, with higher scores reflecting severe dementia [26, 36]. The ADCOMS is a recently designed, statistically derived composite scoring procedure, consisting of two MMSE items (‘Orientation to time’ and ‘Copy design’), four ADAS-Cog subtests (Delayed Word Recall, Orientation, Word Recognition and Word Recall) and all six CDR-SB subscores. All items are differentially weighted yielding a score ranging from 0 to 1.27 with higher scores implying greater impairment [28].

Reference measures of disease severity. Informant reports of disease severity included the Cognitive Function Instrument (CFI) study partner version [37], Quality of Life in Alzheimer’s Disease (QoL-AD) [38], the short version of Zarit Burden Inventory (ZBI-12) [39] and the Apathy Evaluation Scale (AES) [40]. The CFI includes 14 items on decline in day-to-day cognitive and functional abilities compared to one year ago. Response options include ‘yes’ (0), ‘no’ (1) or ‘maybe’ (0.5), with total scores ranging from 0 to 14. The QoL-AD consist of 13 items, rated on a 4-point scale. Total scores range from 13 to 52, with higher scores reflecting better quality of life. The ZBI is one of the most commonly used instruments for assessing aspects of caregiver burden [39]. Each item was rated on a 5-point scale, with total scores ranging from 0 to 60 and higher scores suggesting greater caregiver burden. The AES consists of 18 statements about the participant’s thoughts, feelings, and activity, which are rated on a 4-point scale. Total scores range from 0 to 72, with higher scores indicating more severe apathy.

Magnetic resonance (MR) images were acquired locally at each center in 3T scanners. A minimum acceptable protocol was approved and then optimized at each site due to scanner differences (see Appendix). Images were checked for quality by an experienced rater. Volumetric measurements were processed on 3D T1-weighted (3DT1) images with Statistical Parametric Mapping 12 (SPM12) software (Wellcome Trust Centre for Neuroimaging, University College London, UK) running in MATLAB 2011a (MathWorks Inc., Natick, MA, USA). Prior to processing, the origin in each scan was manually set to the anterior commissure. Scans were segmented into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Total GM (i.e., the sum of all GM voxels) and total intracranial volume (TIV, i.e. the sum of GM, WM, and CSF volumes) were derived from the segmented images in native space (units in liter). Cortical volume was defined as the total GM volume normalized for head size, divided by TIV.

2.4. Procedures
Study visits took place at the hospital or the participant’s home, depending on the participant’s preference. A trained rater assessed the cognitive tests according to standardized instructions, starting with the MMSE and followed by the cognitive part of the CFC (Word Recognition, Orientation, CFT, COWAT, DSST, DSB, Word
Recall), and the remaining ADAS-Cog-13 tests. In the meantime, the study partner completed the A-IADL-Q, ZBI and QoL-AD independently on an iPad. Subsequently, the participant completed the QoL-AD on the iPad with assistance from the rater. Finally, the rater completed the ADCS-ADL and CDR interview with the study partner. The total duration of a complete assessment was approximately 90 minutes. A shortened protocol was used in the SCD and DLB participants, as it was not our purpose compare the CFC to the traditional tests that were not designed to assess progression in these groups. Therefore SCD and DLB participants only underwent the MMSE and cognitive battery of the CFC whilst their study partner completed the A-IADL-Q.

**MRI procedures.** MR scans acquired less than 6 months prior to the study visit were available for a subset of the study cohort. These included at least 3D-T1, T2-weighted imaging (T2) and 3D- Fluid Attenuated Inversion Recovery (FLAIR). Participants without a recent MRI scan available but who agreed to undergo a structural MRI scan, were also scanned at 3T with the same structural sequences which took about 30 minutes.

### 2.5. Statistical analyses

Statistical analyses were performed using SPSS version 22.0 (IBM Corp., Armonk, NY) and R Studio (R Core Team, 2018). Statistical significance level was set at p-value <.05, unless otherwise indicated. Demographic and clinical differences between groups were investigated using Chi-square tests, one-way analyses of variance (ANOVA) followed by Hochberg’s post-hoc tests, and independent t-tests for measures only available for the MCI and AD group.

**Construct validity and clinical relevance.** We performed confirmatory factor analyses (CFA) including all CFC subtests to investigate the CFC’s underlying factor structure. We evaluated a single-factor, two-factor (memory and EF) and three-factor (memory, EF and IADL) model. In the two-factor model, the memory factor included the Word Recognition, Orientation and Word Recall tests and the EF factor the CFT, COWAT, DSST, DSB and A-IADL-Q. The three-factor model had a similar memory and EF factor, except that the A-IADL-Q was excluded from the EF factor and included as separate factor. We compared these models using Chi-Square tests and by evaluating their Comparative Fit Index (CFI), Root Mean Squares of Error Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) indices, with CFI ≥.90, RMSEA <.08 and SRMR <.08 considered as adequate fit [41]. We hypothesized that the three factor model would fit, based on preparatory work on the cognitive component showing two underlying factors [42], and A-IADL-Q reflecting one underlying factor [20]. As sensitivity analysis, all aforementioned CFA model evaluations were repeated in a restricted sample of MCI and mild AD participants, as this was the primary target population of the CFC.
Next, we investigated differences in CFC scores across diagnostic groups using ANOVA followed by Hochberg's post-hoc tests, to examine whether scores would decrease from SCD to dementia. We assessed the association between the CFC variables and reference measures of disease severity, by performing linear regression analyses for each reference measure (CFI, QoL-AD, ZBI-12, and AES, as dependent) and CFC score, age, sex and education as independents. To evaluate the added clinical value of the A-IADL-Q, we also investigated a second model including the cognitive component score and A-IADL-Q score as separate independents. The association between CFC score and grey matter volume was assessed with a linear regression analysis correcting for age, sex, years of education and scanner type. We computed Pearson's correlation coefficients to investigate the relation between CFC scores and traditional cognitive and functional measures.

Quality for the target population. As the CFC was initially designed for MCI and mild AD dementia, the comparison analyses with traditional measures and ADCOMS were performed using the MCI and AD groups only. Using this sample, we compared histograms of score distributions of the CFC, traditional tests and ADCOMS to inspect range restrictions in scoring. To allow for appropriate comparisons between the CFC components and traditional tests, the histograms for the ADAS-Cog, ADCS-ADL and CDR-SB score distributions were based on the standardized scores. Additionally, we reported original score ranges and distribution parameters (percentiles, skewness and kurtosis) for all tests.

3. Results

3.1. Descriptive characteristics
The total sample (N=184) had a mean age of 71.8 ± 8.4 years, 42% were female, and mean years of education was 13.6 ± 3.8. The majority had a diagnosis of MCI (n=80) or AD dementia (n=78). Table 1 presents the demographic and clinical characteristics separately for each diagnostic group. Groups differed in terms of age (F= 2.99, p =.033), but there were no significant differences regarding sex and education. MMSE scores were lower for dementia (AD: 24.0±3.3; DLB: 24.8±3.1) compared to MCI (26.7±2.3) and SCD (29.3±1.2). The AD group also performed worse on the ADAS-Cog (mean difference 6.3 points, p<.001), ADCS-ADL (mean difference 3.8 points, p=.01) and CDR-SB (mean difference 2.4 points, p<.001) when compared to the MCI group. Study partners of AD participants reported worse CFI scores (mean difference 2 points, p=.001), lower quality of life scores (mean difference 2.2 points, p=.026), higher caregiver burden (mean difference 3.7 points, p=.004) and higher apathy levels (mean difference 3.4 points, p=.039) than study partners of MCI participants.
Table 1. Descriptive characteristics and test scores separately for each diagnostic group.

<table>
<thead>
<tr>
<th></th>
<th>SCD (n=14)</th>
<th>MCI (n=80)</th>
<th>AD (n=78)</th>
<th>DLB (n=12)</th>
<th>P</th>
<th>Post-hoc between group comparisons*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>68.1 (6.6)</td>
<td>73.6 (8.0)</td>
<td>71.3 (9.1)</td>
<td>68.3 (6.1)</td>
<td>.033</td>
<td>Non-significant</td>
</tr>
<tr>
<td>Female (%)</td>
<td>8 (57.1%)</td>
<td>29 (36.3%)</td>
<td>39 (50%)</td>
<td>2 (16.7%)</td>
<td>.060</td>
<td>N.A.</td>
</tr>
<tr>
<td>Education</td>
<td>15.2 (4.8)</td>
<td>13.7 (3.6)</td>
<td>13.2 (4.0)</td>
<td>13.6 (3.3)</td>
<td>.36</td>
<td>N.A.</td>
</tr>
<tr>
<td>MMSE</td>
<td>29.3 (1.2)</td>
<td>26.7 (2.3)</td>
<td>24.0 (3.3)</td>
<td>24.4 (2.9)</td>
<td>&lt;.001</td>
<td>SCD &gt; MCI &gt; AD; SCD &gt; DLB</td>
</tr>
<tr>
<td>Cognitive composite</td>
<td>.88 (.50)</td>
<td>.19 (.50)</td>
<td>-.31 (.61)</td>
<td>-.35 (.71)</td>
<td>&lt;.001</td>
<td>SCD &gt; MCI &gt; AD, SCD &gt; DLB, MCI &gt; DLB</td>
</tr>
<tr>
<td>A-IADL-Q</td>
<td>.91 (.79)</td>
<td>.39 (.73)</td>
<td>-.48 (.84)</td>
<td>-.65 (.90)</td>
<td>&lt;.001</td>
<td>SCD &gt; AD; MCI &gt; AD; SCD &gt; DLB; MCI &gt; DLB</td>
</tr>
<tr>
<td>CFC score</td>
<td>.89 (.57)</td>
<td>.29 (.51)</td>
<td>-.39 (.61)</td>
<td>-.51 (.75)</td>
<td>&lt;.001</td>
<td>SCD &gt; MCI &gt; AD; SCD &gt; DLB; MCI &gt; DLB</td>
</tr>
<tr>
<td>Reference measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td>N.A.</td>
<td>6.9 (3.2)</td>
<td>4.9 (2.8)</td>
<td>N.A.</td>
<td>&lt;.001</td>
<td>N.A.</td>
</tr>
<tr>
<td>QoL-AD</td>
<td>N.A.</td>
<td>33.8 (5.3)</td>
<td>31.6 (5.2)</td>
<td>N.A.</td>
<td>.026</td>
<td>N.A.</td>
</tr>
<tr>
<td>ZBI-12</td>
<td>N.A.</td>
<td>10.7 (7.7)</td>
<td>14.4 (7.9)</td>
<td>N.A.</td>
<td>.004</td>
<td>N.A.</td>
</tr>
<tr>
<td>AES</td>
<td>N.A.</td>
<td>39.9 (10.5)</td>
<td>43.3 (10.0)</td>
<td>N.A.</td>
<td>.053</td>
<td>N.A.</td>
</tr>
<tr>
<td>ADAS-Cog</td>
<td>N.A.</td>
<td>22.2 (6.8)</td>
<td>28.5 (7.6)</td>
<td>N.A.</td>
<td>&lt;.001</td>
<td>N.A.</td>
</tr>
<tr>
<td>ADCS-ADL</td>
<td>N.A.</td>
<td>67.5 (7.8)</td>
<td>63.7 (8.8)</td>
<td>N.A.</td>
<td>.010</td>
<td>N.A.</td>
</tr>
<tr>
<td>CDR-SB</td>
<td>N.A.</td>
<td>2.7 (1.9)</td>
<td>5.1 (2.2)</td>
<td>N.A.</td>
<td>&lt;.001</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

*Based on Hochberg’s post-hoc tests

Abbreviations: ADAS-Cog = Alzheimer’s disease assessment scale - Cognitive subscale; ADCS-ADL = Alzheimer’s Disease Cooperation Study – Activity of Daily Living; AES = Apathy Evaluation Scale; A-IADL-Q = Amsterdam IADL Questionnaire; CDR-SB = Clinical Dementia Rating Sum of Boxes; CFC = Cognitive-Functional Composite; CFI = Cognitive Functioning Instrument; MMSE = Mini-Mental State Examination; QoL = Quality of life; ZBI = Zarit Burden Inventory.
3.2. Construct validity and clinical relevance

CFA showed that the hypothesized three-factor model including memory, EF and IADL had adequate fit (CFI = .931, RMSEA = .091 (90% CI = [.058 – .124] and SRMR = .06), although the RMSEA index did not reach the pre-defined cut-off (<.08). The three-factor model (Figure 1) described the data better than the single or two-factor models (Table 2). Similar results were found after repeating the analyses in a restricted sample of MCI and mild AD participants (CFI = .918, RMSEA = .090 (90% CI = [.053 – .122] and SRMR = .06).

As expected, overall CFC scores decreased concomitantly to progression across the spectrum from SCD (.89 ± .57), to MCI (.29 ± .51) to AD- or DLB-dementia (AD: -.39 ± .61, DLB -.52 ± .75), with significant differences between all groups except the two dementia groups (AD and DLB). A similar pattern was found for the cognitive composite score. A-IADL-Q scores were significantly lower for AD as compared to SCD and MCI, as well as for DLB compared to MCI and SCD (Table 1). Figure 2 visualizes the decreased scores across the diagnostic groups, with the cognitive composite divided into a memory and EF score according to the CFA results. It also shows that the CFC score is similar for the two dementia groups, and that in AD this score is driven by the memory factor rather than the EF factor, whereas the opposite is observed in DLB.

Lower CFC scores were associated with worse cognitive functioning as reported by the informant (corrected β = .61, p<.001), quality of life (corrected β = .51, p<.001), caregiver burden (corrected β = -.51, p<.001), and apathy level (corrected β = -.36, p<.001). Regression models including the cognitive composite and A-IADL-Q as separate scores demonstrated the added clinical value of the A-IADL-Q (Table 3). We found moderate-to-strong associations between the cognitive composite and ADAS-Cog (r = -.83; 95% CI = [-.87 – -.77]); the A-IADL-Q and ADCS-ADL (r = .65; 95% CI = [.54 – .73]); and CFC score and CDR-SB (r = -.69; 95% CI = [-.77 – -.56]).

Brain MR images were available for 70 participants (n=7 SCD; n=27 MCI and n=36 AD). Linear regression analyses showed a significant association between normalized GM volume and CFC score (corrected β = .34, p=.01, Figure 3), indicating that worse performance on the CFC was related to less cortical volume independently of age, sex, education and scanner type.

3.2. Construct validity and clinical relevance

CFA showed that the hypothesized three-factor model including memory, EF and IADL had adequate fit (CFI = .931, RMSEA = .091 (90% CI = [.058 – .124] and SRMR = .06), although the RMSEA index did not reach the pre-defined cut-off (<.08). The three-factor model (Figure 1) described the data better than the single or two-factor models (Table 2). Similar results were found after repeating the
analyses in a restricted sample of MCI and mild AD participants (CFI = .918, RMSEA = .090 (90% CI = [.053 –.122] and SRMR = .06).

As expected, overall CFC scores decreased concomitantly to progression across the spectrum from SCD (.89±.57), to MCI (.29±.51) to AD- or DLB-dementia (AD: -.39±.61, DLB -.52±.75), with significant differences between all groups except the two dementia groups (AD and DLB). A similar pattern was found for the cognitive composite score. A-IADL-Q scores were significantly lower for AD as compared to SCD and MCI, as well as for DLB compared to MCI and SCD (Table 1). Figure 2 visualizes the decreased scores across the diagnostic groups, with the cognitive composite divided into a memory and EF score according to the CFA results. It also shows that the CFC score is similar for the two dementia groups, and that in AD this score is driven by the memory factor rather than the EF factor, whereas the opposite is observed in DLB.

Table 2. Fit statistics for confirmatory factor analysis models.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Chi-Square</th>
<th>df</th>
<th>P</th>
<th>CFI</th>
<th>RMSEA [90%CI]</th>
<th>SRMR</th>
<th>Comparison with Model 1</th>
<th>Comparison with Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>62.952</td>
<td>20</td>
<td>&lt;.001</td>
<td>.891</td>
<td>.108 [.079 -.139]</td>
<td>.070</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Model 2</td>
<td>52.543</td>
<td>19</td>
<td>&lt;.001</td>
<td>.915</td>
<td>.098 [.067 -.130]</td>
<td>.064</td>
<td>ChiSq(1)= 10.409</td>
<td>N.A.</td>
</tr>
<tr>
<td>Model 3</td>
<td>45.269</td>
<td>18</td>
<td>&lt;.001</td>
<td>.931</td>
<td>.091 [.058 -.124]</td>
<td>.060</td>
<td>Chisq(2)= 17.683</td>
<td>Chisq(1)= 15.873</td>
</tr>
</tbody>
</table>

Model 1: Single factor based on all eight CFC subtests
Model 2: Two-factor model: MEM + EF
Model 3: Three-factor model: MEM + EF + IADL

114 Chapter 3.3
3.3. Quality for the target population

Histograms including the total MCI and AD sample (n=158) showed expected floor and ceiling effects in scoring for all the traditional measures (Figure 4). Range restrictions were especially apparent for the ADAS-Cog and ADCS-ADL (Table 4). For example for the ADCS-ADL, 35% of the participant scores were at the maximum end of the scale (between 70 and 78). By comparison, all CFC components showed normal distributions without range restrictions in scoring. Figure 5 displays a direct comparison between the ADCOMS and CFC score distributions, separately for the MCI and mild AD dementia group. Despite strong correlations between the ADCOMS and CFC scores (r = -.76; 95% CI = [-.82 - -.68]), it can be seen that the ADCOMS is more influenced by ceiling-effects in scoring, particularly in the MCI subgroup.

**Figure 1.** Path diagrams showing the three-factor structure of the CFC, including the covariance between domains and variables.
**Figure 2.** Boxplots displaying scores on the CFC subcomponents (memory, EF, and IADL factor) and the overall CFC score, separately for each diagnostic group.

**Figure 3.** Scatterplot displaying the correlation between the CFC score and grey matter volume (corrected for total intracranial volume).
Figure 4. Score distributions of the CFC components (top) and traditional tests (bottom) in combined sample of MCI and mild AD subjects (n=158). Scores are standardized using the total sample mean and standard deviation as reference values.
Table 3. Beta coefficients obtained from linear regression analyses relating CFC components to reference measures of disease severity.

<table>
<thead>
<tr>
<th></th>
<th>CFI</th>
<th>QoL-AD</th>
<th>ZBI-12</th>
<th>AES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta</td>
<td>95% CI</td>
<td>P</td>
<td>Beta</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFC score</td>
<td>.61</td>
<td>.48</td>
<td>.74</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age</td>
<td>.14</td>
<td>.02</td>
<td>.27</td>
<td>.028</td>
</tr>
<tr>
<td>Sex</td>
<td>.21</td>
<td>.08</td>
<td>.34</td>
<td>.002</td>
</tr>
<tr>
<td>Education</td>
<td>-.04</td>
<td>-.17</td>
<td>.09</td>
<td>.578</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>composite</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-IADL-Q</td>
<td>.72</td>
<td>.58</td>
<td>.86</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age</td>
<td>.11</td>
<td>-.01</td>
<td>.23</td>
<td>.066</td>
</tr>
<tr>
<td>Sex</td>
<td>.18</td>
<td>.06</td>
<td>.30</td>
<td>.004</td>
</tr>
<tr>
<td>Education</td>
<td>.04</td>
<td>-.09</td>
<td>.16</td>
<td>.552</td>
</tr>
</tbody>
</table>

Abbreviations: AES = Apathy Evaluation Scale; A-IADL-Q = Amsterdam IADL Questionnaire; CFC = Cognitive-Functional Composite; CFI = Cognitive Functioning Instrument; QoL = Quality of life; ZBI = Zarit Burden Inventory.
Table 4. Score ranges, percentiles and distribution parameters for the traditional test scores and CFC scores.

<table>
<thead>
<tr>
<th>Score range</th>
<th>Percentiles 5 10 25 50 75 90 95</th>
<th>Distribution Skewness Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>Current sample</td>
<td>5 10 25 50 75 90 95</td>
</tr>
<tr>
<td>ADAS-Cog</td>
<td>0 - 80</td>
<td>6 - 47</td>
</tr>
<tr>
<td>ADCS-ADL</td>
<td>0 - 78</td>
<td>39 - 78</td>
</tr>
<tr>
<td>CDR-SB</td>
<td>0 - 18</td>
<td>0 - 12</td>
</tr>
<tr>
<td>ADCOMS</td>
<td>0 - 1.47</td>
<td>.05 - 1.21</td>
</tr>
<tr>
<td>A-IADL-Q</td>
<td>N.A.</td>
<td>-2.01 - 1.95</td>
</tr>
<tr>
<td>Cognitive Composite</td>
<td>N.A.</td>
<td>-1.44 - 1.33</td>
</tr>
<tr>
<td>CFC</td>
<td>N.A.</td>
<td>-1.51 - 1.37</td>
</tr>
</tbody>
</table>

4. Discussion

We performed a psychometric evaluation of the novel CFC. Factor analyses confirmed the underlying structure of the CFC, reflecting the domains memory, EF and IADL. The associations that we found between the CFC score and reference measures of disease severity further supported the construct validity and clinical relevance of the CFC. We also demonstrated that the CFC scores yielded fewer range restrictions in scoring as compared to traditional tests of cognition and function, indicating a better quality for the target population.

Construct validity, clinical relevance and suitability for the target population are important quality aspects for clinical outcome measures [43]. Construct validity refers to whether an instrument measures what it intends to measure [25]. As a first step to assess this, we performed a CFA which confirmed that our novel combination of tests can be described by a memory, EF and IADL component. Interestingly, we observed that the two cognition factors contributed differentially to the overall CFC score as seen in AD versus DLB, with the AD group

![Figure 5](image-url). Score distributions of the CFC (top) and ADCOMS (bottom), separately for MCI (left) and mild AD dementia (right).
scoring worse on the memory component compared to the EF component and the DLB group vice versa. Notably, this is in line with the clinical pictures of both diseases with memory problems as most prominent symptoms of early AD as opposed to more predominant EF problems in early DLB [15]. These results further confirm that the CFC score is an adequate reflection of the dimensionality of the construct to be measured.

Another aspect of construct validity is testing whether scores on an instrument are associated with scores on instruments that measure a similar construct [25]. We found moderate-to-strong associations with the traditional measures of cognition and function, which supports the construct validity of the CFC. It should, however, be noted that cognitive composite included three ADAS-Cog measures, which probably accounts for the strong association we found between these measures. We also demonstrated that the CFC is associated with other clinically relevant measures of disease severity, such as cognitive decline as reported by the informant, quality of life, caregiver burden and apathy. Additionally, the association between CFC score and GM volume, which has shown to be a good biomarker of neurodegeneration in AD [44, 45], illustrates that the CFC assesses a construct that is related to the underlying disease process. Taken together, these associations suggest that several clinically relevant aspects of the disease and its severity are captured by the CFC.

Quality for the target population was evaluated by inspecting range restrictions in scoring in our MCI and early dementia sample. This population largely corresponds with "Stage 3 patients" as described in the recently proposed NIA-AA clinical staging scheme and aligned FDA guidance [12, 46]. We found that both the traditional tests and the recently designed ADCOMS showed ceiling effects in scoring in this stage, indicating that these participants showed a high level of functioning. These ceiling effects hamper the measurement of changes and especially the measurement of potential improvement due to treatment. In contrast, the score distributions of the CFC component were normally distributed, showing potential to indicate both decline and improvement with respect to baseline measurements. Therefore, these cross-sectional results support that the CFC is a promising measure with which to assess changes over time without exhibiting the limitations of traditionally used tests.

Several other endeavors have been undertaken to design and validate composite measures that are more appropriate to assess changes in earlier clinical stages of AD. These composites range from purely statistically-driven [28, 47] to more theoretically-based such as the Preclinical Alzheimer's Cognitive Composite (PACC) [48] and the Alzheimer Prevention Initiative Cognitive Composite (APCC) [49]. The PACC and APCC focus on preclinical stages of AD and do not include a measure of daily function. However, measuring everyday
functioning is highly relevant in the MCI and dementia stages, as evolving IADL problems are an important clinical hallmark in the transitional stage from MCI to dementia and predict future decline in dementia [7]. The ADCOMS procedure includes a functional measure (CDR-SB), and the fact that previously changes were detected in a clinical trial [28] supports the idea that adding a functional measure advances a cognitive outcome measure. However, the ADCOMS selection has been largely driven by statistical considerations rather than its content, and therefore its clinical relevance is as yet uncertain. It is our view that the CFC can contribute to existing initiatives to improve cognitive measurement in the MCI and mild dementia stages, as its composition has been based on both theoretical constructs (i.e. the combination of cognition and IADL measure) and empirical research [18, 23].

There are some limitations to consider. First, not all MCI and AD participants had AD biomarkers available, and therefore it is unknown whether they had AD pathology. However, in these circumstances we relied on an extensive clinical assessment and excluded participants with other conditions that could have caused or contributed to the cognitive or functional symptoms. Second, some heterogeneity in our sample may have been caused by minor differences in the recruitment strategies employed across the centers as well as from including SCD and DLB participants. Consequently, our sample may not perfectly mirror the composition of an ideal clinical trial sample. It should also be noted that the SCD and DLB samples were relatively small, and that the CFC investigations in these groups were of explorative nature. This limits the interpretation of the CFC results in those groups. Additionally, the sample-size of participants with an MRI scan available was relatively small, as this was not required for participation in our study. Therefore, we should be careful with interpreting these findings, particularly in the SCD group. A further limitation is that we have only investigated a single weighting method to create the CFC score, whereas it is likely that the optimal scoring method involves different weights for the different components. For example, our data showed that the memory component seems to be relatively easy compared to the EF component, which might be something that we need to account for when tracking changes over time, in particular for the MCI group. Lastly, the fact that the CFC in its current composition focuses less on other domains than episodic memory and EF may limit its usefulness for measuring progression at more severe stages of dementia.

Strengths of this study include our construct validation approach, which is a unique aspect of the Catch-Cog study. Given the lack of a gold standard for ‘disease severity’, we used different reference measures of disease severity to compare the CFC with, which led to converging evidence for the clinical relevance of the CFC. Additionally, we were able to perform a direct head-to-head
comparison between the CFC and traditional tests, which, to our knowledge, has never been done in previous studies. The advantage of this is to reveal both the strengths and weaknesses of different clinical measurements. Furthermore, our investigation of the CFC in an independent, prospective cohort is an essential aspect of this study. Although the CFC consists of tests that have been validated as part of other test batteries and across several study cohorts, it cannot be assumed that all measures perform similarly in a novel composition. For example, tests may perform differently when assessed in a different test order. An independent validation of a novel composite measure such as the CFC measure is thus needed to enhance future implementation. We are currently assessing the CFC longitudinally in our Catch-Cog prospective study-cohort, and we will investigate its sensitivity to change after 3, 6 and 12 months, and compare its sensitivity with that of the traditional tests. This longitudinal data will also enables us to explore whether different weights for the subtests can improve the sensitivity to change over time, as well as whether different weights are useful when tracking change in different diagnostic groups. For example, putting more weight on the cognitive parts and activities of daily living that decline relatively late in the disease course, may enhance the use of the CFC to track progression in later disease stages.

5. Conclusions

We demonstrated that the CFC has good construct validity and captures clinically relevant aspects of disease severity. We also showed its improved suitability for the target population as compared to traditional tests, as reflected by fewer range restrictions in scoring. These findings illustrate that the CFC has good potential to be a sensitive, clinically meaningful outcome measure. It is therefore better indicated for use to evaluate cognition and function as compared to traditional tests, as in line with the recent FDA recommendations. Ultimately, the CFC can yield more accurate and useful measurement of clinically relevant changes, which will aid the monitoring of disease progression and evaluation of novel treatments.
Declarations

Ethics approval and consent to participate
The Medical-Ethical Committee of the VU University Medical Center approved the study for all Dutch centers. The South East Scotland Research Ethic Committee approved the study for the Scottish site. All participants and study partners provided written and oral informed consent for the use of their data for research purposes.

Consent for publication
Not applicable.

Availability of data and materials
All data used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests
R.J.J., P.R.L, R.V., R.A.J.D, F.J.J, E.M.O., C.W.R. report no relevant disclosures. In the past two years, J.E.H. has received honoraria and paid consultancy from 23andMe, Abbvie, A2Q, AlzCure, Amgen, Anavex, Aptinyx, Astellas, AstraZeneca, Avraham, Axon, Axovant, Biogen, Boehringer Ingelheim, Bracket, Catenion, Cognition Therapeutics, CRF Health, Curasen, DeNDRoN, Enzymotec, Eisai, Eli Lilly, GfHEu, Heptares, Johnson & Johnson, Kaasa Health, Lysosome Therapeutics, Lundbeck, Merck, MyCognition, Mind Agilis, Neurocog, Neurim, Neuroscios, Neurotrack, Novartis, Nutricia, Pfizer, PriceSpective, Probiodrug, Regeneron, Rodin Therapeutics, Roche, Sanofi, Servier, Shire, Takeda & vTv Therapeutics. S.I. received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking EPAD grant agreement n° 115736. A.A. has received speaker fees from Lundbeck. P.S. has acquired grant support (for the institution) from GE Healthcare and Piramal. In the past 2 years, he has received consultancy/speaker fees (paid to the institution) from Novartis, Probiodrug, Biogen, Roche, and EIP Pharma, LLC. B.M.T. receives grant support from ZonMW. S.A.M.S. is supported by grants from JPND and Zon-MW, and has provided consultancy services in the past 2 years for Lundbeck. All funds were paid to her institution.

Funding
The present study is supported by a grant from Memorabel (grant n° 733050205), which is the research program of the Dutch Deltaplan for Dementia.
Author's contributions
R.J.J., J.E.H., and S.A.M.S. designed the study, performed the statistical analyses, interpreted the results, and drafted the manuscript. P.R.L, S.I., R.V., R.A.J.D, F.J.J, E.M.O., A.A., C.W.R and P.S. helped interpret the results and edit the manuscript. R.J.J. wrote the final version of the manuscript. All authors read and approved the final manuscript.

Acknowledgements
The authors would like to thank Mandy Ter Haar, Larissa Masselink, Judith Meurs, Anne Brunner, Mieke Geertsma, Nina Schimmel, Ilya de Groot, Judy van Hemmen, Kate Forsyth, Sarah Gregory, Neil Fullerton, Clare Dolan and Matthew Hunter for their help with the data collection. Additionally, we would like to acknowledge Stichting Buytentwist for their support.

Research of the Alzheimer Center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. The Alzheimer Center Amsterdam is supported by Alzheimer Nederland and Stichting VUmc Fonds.
References


16. McKeith, I.G., et al., Diagnosis and


30. Albert, M.S., et al., The diagnosis


33. Lezak, M.D., Neuropsychological assessment. 2004: Oxford University Press, USA.


## Supplementary data

**Supplementary File 1. MRI settings per scanner type**

<table>
<thead>
<tr>
<th>MRI sequence</th>
<th>Settings</th>
<th>Philips 3T</th>
<th>GE 3T</th>
<th>Siemens 3T</th>
</tr>
</thead>
<tbody>
<tr>
<td>3DT1</td>
<td>Sagittal 3D turbo field echo sequence (1.00 mm isotropic voxels, TR = 7.9 ms, TE = 4.5 ms, FA = 8 degrees)</td>
<td>Sagittal 3D FSPGR sequence (1.00 mm isotropic voxels, TR = 8.2 ms, TE = 3.2 ms / TI = 450 ms, FA = 12 degrees)</td>
<td>3D MPRAGE (1.00 mm x 1.00 mm x 1.00 mm voxels, TR = 2300 ms, TE = 3 ms)</td>
<td></td>
</tr>
<tr>
<td>3D Fluid-attenuated inversion recovery</td>
<td>3D sagittal VISTA sequence (1.12 mm isotropic voxels, TR = 4800 ms, TE = 279 ms, inversion time = 1650 ms)</td>
<td>3D sagittal CUBE sequence (1.2 mm isotropic voxels, TR = 8000 ms, TE = 130 ms, inversion time = 2340 ms)</td>
<td>3D sagittal SPACE sequence (1.2 mm isotropic voxels, TR = 5000 ms, TE = 387 ms, inversion time = 2340 ms)</td>
<td></td>
</tr>
<tr>
<td>2D PD-T2:</td>
<td>Oblique axial orientation, 45-48 slices of 3mm thickness (TR = 3925 ms, TE = 19 and 100 ms)</td>
<td>Oblique axial orientation, 45-48 slices of 3mm thickness (TR = 8600 ms, TE = 20 and 112 ms)</td>
<td>Oblique axial orientation, 45-48 slices of 3mm thickness (TR = 4730 ms, TE = 20 ms)</td>
<td></td>
</tr>
</tbody>
</table>

TR: repetition time; TE: echo time; FA: flip angle.