


Identifying Sensitive Measures of Cognitive Decline at Different Clinical Stages of Alzheimer's Disease

Roos J. Jutten^{1,*} , Sietske A.M. Sikkes^{1,2}, Rebecca E. Amariglio^{2,3}, Rachel F. Buckley^{2,3,4,5}, Michael J. Properzi², Gad A. Marshall^{2,3}, Dorene M. Rentz^{2,3}, Keith A. Johnson^{2,6}, Charlotte E. Teunissen⁷, Bart N.M. Van Berckel⁸, Wiesje M. Van der Flier¹, Philip Scheltens¹, Reisa A. Sperling^{2,3}, Kathryn V. Papp^{2,3} and for the Alzheimer Disease Neuroimaging Initiative** National Alzheimer's Coordinating Center, the Harvard Aging Brain Study, and the Alzheimer Dementia Cohort

¹Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

²Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

³Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

⁴Florey Institute of Neuroscience and Mental Health, Melbourne, VIC, Australia

⁵Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, VIC, Australia

⁶Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

⁷Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

⁸Department of Radiology and Nuclear Medicine, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

(RECEIVED February 27, 2020; FINAL REVISION August 3, 2020; ACCEPTED August 4, 2020; FIRST PUBLISHED ONLINE October 13, 2020)

Abstract

Objective: Alzheimer's disease (AD) studies are increasingly targeting earlier (pre)clinical populations, in which the expected degree of observable cognitive decline over a certain time interval is reduced as compared to the dementia stage. Consequently, endpoints to capture early cognitive changes require refinement. We aimed to determine the sensitivity to decline of widely applied neuropsychological tests at different clinical stages of AD as outlined in the National Institute on Aging – Alzheimer's Association (NIA-AA) research framework. **Method:** Amyloid-positive individuals (as determined by positron emission tomography or cerebrospinal fluid) with longitudinal neuropsychological assessments available were included from four well-defined study cohorts and subsequently classified among the NIA-AA stages. For each stage, we investigated the sensitivity to decline of 17 individual neuropsychological tests using linear mixed models. **Results:** 1103 participants (age = 70.54 ± 8.7, 47% female) were included: $n = 120$ Stage 1, $n = 206$ Stage 2, $n = 467$ Stage 3 and $n = 309$ Stage 4. Neuropsychological tests were differentially sensitive to decline across stages. For example, Category Fluency captured significant 1-year decline as early as Stage 1 ($\beta = -.58, p < .001$). Word List Delayed Recall ($\beta = -.22, p < .05$) and Trail Making Test ($\beta = 6.2, p < .05$) became sensitive to 1-year decline in Stage 2, whereas the Mini-Mental State Examination did not capture 1-year decline until Stage 3 ($\beta = -1.13, p < .001$) and 4 ($\beta = -2.23, p < .001$). **Conclusions:** We demonstrated that commonly used neuropsychological tests differ in their ability to capture decline depending on clinical stage within the AD continuum (preclinical to dementia). This implies that stage-specific cognitive endpoints are needed to accurately assess disease progression and increase the chance of successful treatment evaluation in AD.

Keywords: Alzheimer's disease, Cognitive decline, Outcome measures, Disease progression

*Correspondence and reprint requests to: Roos J. Jutten, PhD, Alzheimer Center Amsterdam, Amsterdam UMC, Vrije Universiteit Amsterdam, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands. E-mail: r.jutten@amsterdamumc.nl

**Data used in preparation of this article were obtained from the Alzheimer Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

INTRODUCTION

Accurately capturing cognitive changes along the Alzheimer's disease (AD) continuum is essential for monitoring clinical progression and evaluating the efficacy of treatments to slow or halt cognitive decline (Evans et al., 2018; Food and Drug Administration, 2018). Future AD research and clinical trials will likely target a more narrow spectrum of participants at earlier disease stages, in which AD

pathology is present but clinical symptoms remain subtle or absent (Bateman et al., 2012; Dubois et al., 2016). In these (pre)clinical stages, the expected degree of observable cognitive decline over a certain time interval is likely reduced as compared to later clinical stages in which cognitive impairment is (more) evident and decline more rapid (Sperling et al., 2011). As most existing cognitive tests were primarily designed to track decline in the mild cognitive impairment (MCI) and dementia stages, they are probably suboptimal to track subtle cognitive decline observed in preclinical stages of AD (Evans et al., 2018; Mortamais et al., 2017; Rentz et al., 2013). In addition, several studies have shown that tests addressing specific cognitive domains such as episodic memory and semantic memory are more prone to capture decline in early clinical stages of AD, as compared to, for example, tests addressing global cognition (Lim et al., 2016; Mortamais et al., 2017; Papp, Rentz, Orlovsky, Sperling, & Mormino, 2017). Consequently, optimal cognitive endpoints to assess cognitive changes over time are likely to differ across clinical stages on the AD continuum.

In the recently updated National Institute of Aging – Alzheimer's Association (NIA-AA) research framework, Jack et al. proposed a novel clinical staging scheme to classify individuals along the continuum of AD, based on the presence of AD pathophysiology and severity of clinical symptoms (Jack et al., 2018). In this 4+ clinical staging scheme, Stage 1 is described as a preclinical stage in which overt clinical symptoms are absent. Stage 2 is defined by subjective concerns regarding previous level of functioning and/or subtle abnormalities detectable on (longitudinal) sensitive cognitive testing, without the presence of any functional impairment. In Stage 3, abnormalities on cognitive tests are more apparent and mild functional impairment may be detectable. Finally, Stages 4, 5, and 6 are described as overt dementia, corresponding to mild, moderate, and severe dementia, respectively. Grouping individuals into these refined clinical stages may be beneficial in optimizing the selection and assessment of participants for a given treatment target and may be similarly beneficial in refining the cognitive assessment procedures optimal for a given stage (Jack et al., 2019). However, specific procedures to operationalize these stages have yet to be delineated. Moreover, it is yet unknown to what extent currently used cognitive outcomes vary in their ability to capture decline at the different clinical stages defined in the NIA-AA research framework.

A more refined understanding of the differential sensitivity of existing neuropsychological tests at the different NIA-AA-defined clinical stages is needed to provide guidance on the selection of outcome measures when this framework is applied to define the treatment population. Therefore, the current study aimed to identify sensitive measures to detect cognitive change for each of the four NIA-AA defined clinical stages (Jack et al., 2018). To achieve this, we operationalized the NIA-AA clinical staging schema into measurable criteria and applied these criteria across AD biomarker-positive individuals obtained from four large cohorts. Subsequently, we investigated the sensitivity to decline of commonly used

neuropsychological tests by clinical stage, with a focus on the pre-dementia stages 1 to 3. We hypothesized that with increasing clinical stage, more tests would show greater sensitivity to decline.

METHODS

Study Participants

Data were obtained from the Harvard Aging Brain Study (HABS), the Alzheimer's Disease Neuroimaging Initiative (ADNI), and National Alzheimer's Coordinating Center (NACC) databases, and the Amsterdam Dementia Cohort (ADC). For each cohort, specific recruitment criteria and data collection have been described in detail elsewhere (Aisen et al., 2010; Beekly et al., 2007; Dagley et al., 2017; van der Flier, 2018). Briefly, the HABS cohort is a community-based sample of individuals aged ≥ 63 years, who are considered clinically normal at baseline by (1) a global Clinical Dementia Rating score of 0; and (2) performance above education-adjusted cut-offs on Logical Memory Story A Delayed Recall and the Mini-Mental State Examination (MMSE) (Dagley et al., 2017). ADNI is a multicenter longitudinal cohort study with the primary goal of testing whether serial neuroimaging and other biological, clinical, and neuropsychological markers can be combined to measure clinical progression on the AD spectrum (<http://adni.loni.usc.edu/wp-content/uploads/2008/07/adni2-procedures-manual.pdf>). The NACC has developed and maintains a large database of standardized clinical and neuropathological research data, obtained from NIA-funded Alzheimer's Disease Centers across the United States (Beekly et al., 2007; Morris et al., 2006). The NACC database contains mostly memory clinic referred participants with some additional community-based recruitment. The ADC is a Dutch memory clinic-based cohort of individuals visiting the Alzheimer Center Amsterdam (van der Flier, 2018). All participants in the ADC had undergone a standard diagnostic work-up, including medical history, neurological examination, laboratory screening tests, and neuropsychological evaluation. All studies were approved by an ethical review board, and all participants provided written informed consent to use their clinical data for research purposes. All data included in this manuscript were obtained in compliance with the Helsinki Declaration. All data were collected between June 2002 and April 2018.

Selection criteria for the current study included (1) amyloid positivity as determined by at least one abnormal marker of amyloid accumulation in positron emission tomography (PET) imaging or amyloid- β levels in cerebrospinal fluid (CSF) using previously published cohort-specific summary measures and cut-offs [19–21] (detailed amyloid assessment methods for each cohort are specified below); (2) at least one follow-up visit with neuropsychological testing available; and (3) MMSE > 10 or Montreal Cognitive Assessment > 2 at baseline (Folstein, Folstein, & McHugh, 1975; Nasreddine

et al., 2005). Neuropsychological baseline performance was anchored to time of first amyloid assessment.

AD Biomarker Classification

The HABS and ADNI cohorts used PET imaging data only, while both PET and CSF measures were used in NACC and ADC. Amyloid binding in HABS was measured using Pittsburgh compound B PET scanning, and in ADNI using Florbetapir AV-45 PET scanning (summary data were obtained from the ADNI Laboratory of Neuroimaging database: <http://www.loni.ucla.edu/ADNI/>). In HABS, amyloid positivity was based on GMM of the distribution volume ratio of mean uptake in frontal, lateral parietal and temporal, and retrosplenial regions (cut-off value ≥ 1.20 (Mormino et al., 2014)). In ADNI, amyloid positivity was based on standard uptake value ratios of mean uptake in four cortical regions (frontal, cingulate, parietal, and temporal cortices) normalized to the whole cerebellum uptake (cut-off value ≥ 1.11 (Clark et al., 2011)). In the NACC cohort, amyloid positivity in PET or CSF was based on each center's local standard for positivity (Besser et al., 2018). In the ADC, amyloid positivity on Florbetaben, Flutemetamol, PIB-PET, or Florbetapir AV-45 PET scans was based on whole-brain visual assessment performed by an experienced nuclear medicine physician (B.N.M.B.) who was blinded to clinical information. Amyloid positivity in CSF in the ADC was based on drift-corrected amyloid- β 1–42 values, with a cut-off of 813 pg/ml (Tijms et al., 2018).

Clinical Stages

To operationalize the NIA-AA clinical staging scheme, we used baseline measures of subjective cognitive decline (SCD), cognitive impairment, and functional impairment available across cohorts (Table 1).

Subjective cognitive decline

SCD (yes/no) was quantified as either a memory clinic visit or by an SCD screening score for the community-based cohorts. That is, by definition, all individuals from the ADC memory clinic cohort were classified as having SCD. For the NACC cohort, all individuals that were referred by an ADRC ($n = 154$, 75%) were classified as having SCD as well. For the remaining NACC individuals, as well as for all individuals from the ADNI and HABS cohorts, an SCD screening score was calculated. This SCD screening score was extracted from single items from available self-report questionnaires (i.e. the Everyday Cognition scale (ECog) (Farias et al., 2008), the Geriatric Depression Scale – 12 item version (GDS-12) (Yesavage et al., 1983), and Memory Questionnaire (MemQ), addressing (1) recent change in memory functioning (MemQ item 1); (2) consistent change over the last few months (GDS-12 item 6); and (3) concern associated with this change (ECog item 1). Answers were scored as yes = 1 and no = 0,

leading to a total score range from 0 to 3 with higher scores reflecting higher level of SCD. A score of 2 or higher was labeled as having SCD.

Cognitive impairment

Level of cognitive impairment was quantified using (1) the MMSE (or a MOCA transformed score if MMSE was unavailable) (Folstein et al., 1975; Nasreddine et al., 2005; Trzepacz, Hochstetler, Wang, Walker, & Saykin, 2015); and (2) a memory retention score reflecting the proportion of items recalled from either story or word list on immediate and delayed recall (Schmidt, 1996; Wechsler, 1987).

Functional impairment

Severity of functional impairment was determined using the CDR[®] Dementia Staging Instrument global score or the CDR sum of boxes (CDR-SB) (Hughes, Berg, Danziger, Coben, & Martin, 1982; M. M. Williams, Storandt, Roe, & Morris, 2013).

Classification criteria

For all measures, we created stage-specific cut-off scores based on previously published data or the highest or lowest quintiles in our sample (Table 1). To summarize, Stage 1 was quantified as (1) no visit to a memory clinic or a SCD screening score < 2 ; (2) an MMSE score of ≥ 26 (Chapman et al., 2016), a proportion of $\geq 52\%$ items learned on a story or word list (highest quintile in our sample), and a delayed recall of > 11 items on the logical memory tests (Aisen et al., 2010) and (3) a CDR-SB score $\leq .05$ and global CDR of 0 (M. M. Williams et al., 2013). Stage 2 was quantified as (1) an MMSE score of ≥ 26 (Chapman et al., 2016), a proportion of $\geq 52\%$ items learned on a story or word list and a delayed recall; and (2) a CDR-SB score ≤ 1 and global CDR of $< .5$ [31] only if caused by endorsement of memory box of the CDR-SB of 0.5. Stage 3 was quantified as (1) a MMSE score ≥ 24 (Chapman et al., 2016) and a proportion items learned $\geq 20\%$ (lowest quintile in our sample); (2) a CDR-SB score between 1.5 and 4 and global CDR of .5 (M. M. Williams et al., 2013). Stage 4 was quantified as (1) a MMSE score < 26 [31] and a proportion items learned $< 20\%$ (lowest quintile in our sample); (2) a CDR-SB score ≥ 4.5 and global CDR of ≥ 1 (M. M. Williams et al., 2013).

We initially applied a strict approach in that at least one variable from all clinical features (SCD, cognition and function) was required for categorization, in order to create distinct categories in line with clinical trial screening procedures. Participants who remained unclassified due to incongruencies among clinical features were classified in the stage in which the majority of their classification measures fit best. For example, an individual with an MMSE of 23, a proportion items learned of .45, a CDR-SB of 2, and a global CDR score of 0.5 was classified as Stage 3.

Table 1. Operationalization of the NIA-AA clinical staging scheme

		Stage 1	Stage 2	Stage 3	Stages 4–6
		No evidence of clinical impact	Subjective report of decline or subtle cognitive abnormalities	Apparent cognitive abnormalities with mild functional impairment	Overt dementia
Pathophysiologic changes	Amyloid pathology	Abnormal	Abnormal	Abnormal	Abnormal
Subjective complaints	Visit memory clinic	No	Not required	Not required	Not required
	OR SCD screening score*	<2	Not required	Not required	Not required
Cognitive abnormalities	MMSE AND	26–30	26–30	24–30	<26
	Story or list learning	Proportion items learned $\geq .52^*$ AND Delayed word recall >11 items	Proportion items learned $\geq .52$	Proportion items learned $\geq .20^*$	Proportion items learned <.20
Functional impact	CDR-SB	$\leq .5$	$\leq 1.0^{**}$	1.5–4.5	≥ 4.5
	Global CDR	0	0–0.5**	0.5–1	≥ 1

*Based on the highest and lowest quintile in our study sample.

** Only if caused by endorsement of memory box of the CDR-SB with 0.5.

Neuropsychological Tests

We selected all neuropsychological tests available in at least two cohorts. This resulted in a total of 12 different tests providing 17 individual measures. Supplementary Table 1 provides an overview of which tests were available by cohort. Global cognition measures included the MMSE and MOCA (Folstein et al., 1975; Nasreddine et al., 2005). Episodic memory measures included the Wechsler memory subscale Logical Memory (LM) immediate recall and delayed recall (Wechsler, 1987), and the Rey Auditory Verbal Learning Test (RAVLT) (Schmidt, 1996) with version 1 used in ADC, and version 1 and 2 alternated in ADNI. Additionally, we standardized immediate recall, delayed recall, and recognition measures obtained from list learning tests that were available across cohorts (Selective Reminding Test in HABS, the ADAS-Cog Word Lists in ADNI and RAVLT in ADC), by calculating *z*-scores by cohort with total baseline mean and standard deviation as reference values. Subsequently, these *z*-scores were combined into an overall immediate recall score, an overall delayed recall score, and overall recognition score. Semantic memory measures included the Category Fluency Test (CFT) Animals and Vegetables (Lezak, 2004) and the Boston Naming Test (BNT) (B. W. Williams, Mack, & Henderson, 1989). EF measures included the controlled oral word association test (COWAT) (Ruff, Light, Parker, & Levin, 1996) and Trail Making Test (TMT) part B (Tombaugh, 2004). Attention and working memory measures included the Wechsler Adult Intelligence Scale (WAIS-IV) subscales Digit Span Forward and Backward (Wechsler, 2008). Measures of processing speed included the TMT part A (Tombaugh, 2004) and WAIS-IV Symbol Substitution (Wechsler, 2008).

Statistical Analyses

Statistical analyses were performed using R version 3.5.3 (R Core Team, 2018). Statistical significance was set at $p < .05$. Demographic and clinical differences between stages were investigated using Chi-square tests and one-way analyses of variance followed by Tukey's HSD test for *post hoc* comparisons.

To investigate the sensitivity to change over time of each neuropsychological test, a series of linear mixed models (LMM) were performed with a random intercept and slope for each subject. Separate models were run for each test score (dependent), with time (measured on a continuous level), age (centered at overall mean), sex, education (centered at overall mean), stage (as categorical variable), and the interaction between time and stage as independent variables. To examine whether tests were differentially sensitive to change over time across stages, we focused on the time and time \times stage estimates. Subsequently, LMM were repeated for each neuropsychological test separately for each stage, to identify which tests were sensitive within 12 months of follow-up within each stage. For tests that were identified as sensitive to 12-months decline, mean to standard deviation ratios (MSDRs) of change over baseline to 12, 24, and 36 months follow-up were calculated as a measure of effect-size.

RESULTS

A total of 1103 participants were included (HABS $n = 74$, ADNI $n = 506$, NACC = 204, ADC $n = 319$). The majority of the US individuals were Caucasian (93% in ADNI, 92% in NACC, and 78% in HABS). Overall, years of follow-up

Table 2. Demographic and clinical characteristics for each clinical stage

	Stage 1	Stage 2	Stage 3	Stage 4	<i>p</i> -Value	<i>Post hoc</i> comparisons
<i>N</i>	120	206	467	309	N.a.	N.a.
Years of FU (M ± SD)	2.6 ± 1.6	2.8 ± 1.8	2.3 ± 1.6	1.6 ± 1.1	N.a.	N.a.
Age (M ± SD)	74.5 ± 6.4	70.9 ± 9	70.8 ± 8.2	68.4 ± 9.2	<.001	Stage 1 > Stage 3 > Stage 4, Stage 1 > Stage 2 > Stage 4
Female, <i>n</i> (%)	68 (57)	107 (52)	204 (44)	135 (44)	.02	N.a.
Education (M ± SD)	16.2 ± 2.9	15.1 ± 3.3	15.1 ± 3.3	13.6 ± 3.6	<.001	Stage 1 > Stage 3 > Stage 4, Stage 1 > Stage 2 > Stage 4
MMSE (M ± SD)	29.1 ± 1	28.4 ± 1.2	26.9 ± 2	21.6 ± 3.1	<.001	Stage 1 > Stage 3 > Stage 4, Stage 2 > Stage 3
CDR-SB (M ± SD)	0.04 ± 0.13	0.29 ± 0.38	2.05 ± 1.3	5.5 ± 2.1	<.001	Stage 1 > Stage 3 > Stage 4, Stage 2 > Stage 3
HABS, <i>n</i> (%)	42 (35)	32 (16)	0 (0)	0 (0)	N.a.	N.a.
ADC, <i>n</i> (%)	0 (0)	79 (38)	94 (20)	145 (47)	N.a.	N.a.
ADNI, <i>n</i> (%)	49 (41)	72 (35)	288 (62)	97 (31)	N.a.	N.a.
NACC, <i>n</i> (%)	29 (24)	23 (11)	85 (18)	67 (22)	N.a.	N.a.
CN	120 (100)	91 (44)	11 (2)	0	N.a.	N.a.
SCD	0	44 (22)	11 (2)	2 (1)	N.a.	N.a.
MCI	0	71 (34)	317 (68)	4 (2)	N.a.	N.a.
Dementia	0	0	128 (28)	300 (97)	N.a.	N.a.

HABS, Harvard Aging Brain Study; ADC, Amsterdam Dementia Cohort; ADNI, Alzheimer's Disease Neuroimaging Initiative; NACC, National Alzheimer's Coordinating Center.

ranged from 1 to 12.9, with a mean of 2.3 ± 1.6 years, and number of follow-up visits ranged from 1 to 8, with a mean of 2.3 ± 1.4 visits. After applying our defined staging criteria, $n = 1005$ (91%) participants were initially classified in one of the four stages, while $n = 99$ (9%) remained unclassified due to incongruencies among classification measures and were then classified in the closest stages. Ultimately, $n = 120$ were classified as Stage 1, $n = 206$ as Stage 2, $n = 467$ as Stage 3, and $n = 309$ as Stage 4+.

Table 2 presents the demographic and clinical characteristics for the different stages. Groups differed regarding age, sex, and education, with *post hoc* comparisons revealing that Stage 1 participants were older than all other participants and Stage 2 participants were older than Stage 4 participants. *Post hoc* comparisons also showed that Stage 1 participants had higher education levels compared with all other participants and Stage 2 had higher education than Stage 4 participants. By definition, we found worse MMSE and CDR-SB scores with increasing clinical stage. All HABS participants fell into Stages 1 and 2, all ADC participants were classified into Stages 2 to 4, and ADNI and NACC participants were distributed among all stages. Overall, the four stages were in correspondence with clinical diagnoses as established in the original cohorts, given that Stage 1 only included cognitively normal participants, and that most SCD, MCI, and dementia participants fell into Stages 2, 3, and 4, respectively (Table 2).

Sensitivity to Decline Over Time by Clinical Stage

Overall, LMM results revealed that neuropsychological tests were differentially sensitive to change over time at different

clinical stages, as indicated by the Time and Time \times Stage estimates in Table 3 (and Supplementary Table 2). For example, CFT Animals detected annual decline at all stages, showing a significant decline of .58 points (i.e. animals named in 60 seconds) in Stage 1 and Stage 2 ($p < .001$), and marginally significant steeper decline in Stage 3 ($-.28$ points, $p = .06$), and significantly steeper decline in Stage 4 (-1.27 points, $p < .001$). In contrast, the MMSE did not capture decline in Stages 1 and 2, but showed steeper decline in Stage 3 (-1.13 points, $p < .001$) and Stage 4 (-2.23 points, $p < .001$). Figure 1 illustrates the difference in sensitivity to decline over time for these two measures.

Second, LMM results indicated that Time \times Stage estimates escalated by increasing clinical stage (Table 3), with most tests exhibiting a steeper decline in Stage 3 and 4 as compared to Stage 1.

Sensitivity to Decline by Follow-Up Duration

Annual decline on the individual neuropsychological tests summarized by cognitive domain is shown in Figure 2, with corresponding regression coefficients presented in Supplementary Table 3. Only three measures captured significant decline after 12 months in Stage 1 (CFT Animals, CFT Vegetables, LM Immediate Recall), and only four measures captured decline after 12 months in Stage 2 (CFT Animals, CFT Vegetables, Word List Delayed Recall and TMT-B). Six additional measures detected decline in Stage 3, (Digit Span Backward, Digit Span Forward, Word List Recognition, MMSE, MOCA, and TMT-A). In Stage 4, all measures detected decline after 12 months, except for

Table 3. Regression coefficients from linear mixed models investigating change over time in cognitive test scores

	Global cognition						Semantic memory								
	MMSE			MOCA			CFT animals			CFT vegetables			BNT total 30		
	Beta	95% CI	p-Value	Beta	95% CI	p-Value	Beta	95% CI	p-Value	Beta	95% CI	p-Value	Beta	95% CI	p-Value
Intercept	28.97	28.50 to 29.44	<0.001	25.66	24.29 to 27.03	<0.001	20.37	19.47 to 21.28	<0.001	12.84	11.89 to 13.79	<0.001	28.97	28.03 to 29.91	<0.001
Time (Stage 1)	-0.13	-0.43 to 0.17	0.409	-0.52	-1.33 to 0.29	0.215	-0.58	-0.84 to -0.32	<0.001	-0.63	-0.85 to -0.41	<0.001	-0.16	-0.41 to 0.09	0.198
Stage 2	-0.45	-0.99 to 0.09	0.101	-1.90	-3.75 to -0.05	0.048	-0.97	-2.01 to 0.07	0.068	-0.62	-1.80 to 0.57	0.312	-1.22	-2.18 to -0.25	0.014
Stage 3	-1.86	-2.34 to -1.38	<0.001	-4.71	-6.20 to -3.22	<0.001	-4.00	-4.92 to -3.07	<0.001	-3.87	-5.00 to -2.74	<0.001	-2.29	-3.16 to -1.42	<0.001
Stage 4	-6.68	-7.22 to -6.15	<0.001	-11.67	-13.26 to -10.09	<0.001	-7.76	-8.77 to -6.75	<0.001	-7.30	-8.52 to -6.08	<0.001	-6.18	-7.22 to -5.13	<0.001
Time × Stage2	-0.35	-0.74 to 0.04	0.076	0.68	-0.52 to 1.88	0.274	-0.00	-0.35 to 0.34	0.988	-0.02	-0.34 to 0.31	0.927	-0.11	-0.47 to 0.26	0.565
Time × Stage3	-1.13	-1.47 to -0.80	<0.001	-0.48	-1.43 to 0.46	0.318	-0.28	-0.58 to 0.02	0.064	-0.38	-0.92 to 0.17	0.175	-0.50	-0.78 to -0.21	0.001
Time × Stage4	-2.23	-2.64 to -1.82	<0.001	-1.63	-2.66 to -0.60	0.002	-1.27	-1.67 to -0.87	<0.001	-0.53	-1.16 to 0.10	0.102	-2.02	-2.47 to -1.57	<0.001
Observations	3551			387			3890			837			2420		
	Episodic memory (immediate recall)						Episodic memory (delayed recall and recognition)								
	Logical memory immediate recall			Word list immediate recall			Logical memory delayed recall			Word list delayed recall			Word list recognition		
	Beta	95% CI	p-Value	Beta	95% CI	p-Value	Beta	95% CI	p-Value	Beta	95% CI	p-Value	Beta	95% CI	p-Value
Intercept	16.37	15.70 to 17.04	<0.001	0.52	0.34 to 0.71	<0.001	12.45	9.97 to 14.92	<0.001	0.78	0.59 to 0.98	<0.001	0.65	0.45 to 0.84	<0.001
Time (Stage 1)	0.04	-0.17 to 0.25	0.734	-0.03	-0.07 to 0.02	0.212	2.80	0.48 to 5.13	0.018	-0.12	-0.16 to -0.08	<0.001	-0.05	-0.10 to -0.00	0.040
Stage 2	-3.71	-4.49 to -2.92	<0.001	-0.08	-0.29 to 0.12	0.420	-3.38	-6.25 to -0.51	0.021	-0.22	-0.44 to -0.00	0.046	-0.24	-0.46 to -0.02	0.036
Stage 3	-7.34	-8.00 to -6.68	<0.001	-0.58	-0.77 to -0.39	<0.001	-5.52	-8.24 to -2.80	<0.001	-0.85	-1.05 to -0.66	<0.001	-0.57	-0.77 to -0.37	<0.001
Stage 4	-12.41	-13.20 to -11.63	<0.001	-1.23	-1.43 to -1.02	<0.001	-10.53	-13.64 to -7.41	<0.001	-1.36	-1.58 to -1.15	<0.001	-1.20	-1.42 to -0.99	<0.001
Time × Stage2	-0.04	-0.34 to 0.27	0.816	-0.12	-0.18 to -0.06	<0.001	0.77	-2.03 to 3.57	0.591	0.01	-0.05 to 0.07	0.716	-0.04	-0.10 to 0.03	0.294
Time × Stage3	-0.45	-0.69 to -0.20	<0.001	-0.14	-0.19 to -0.09	<0.001	-3.80	-6.40 to -1.21	0.004	0.03	-0.01 to 0.08	0.161	-0.09	-0.15 to -0.03	0.003
Time × Stage4	-0.71	-1.15 to -0.28	0.001	-0.28	-0.35 to -0.20	<0.001	-3.76	-6.76 to -0.76	0.014	-0.02	-0.09 to 0.05	0.652	-0.05	-0.14 to 0.03	0.219
Observations	2688			3524			2675			3424			3419		
	Attention and working memory						Executive functioning								
	Digit span forward			Digit span backward			TMT B			COWAT total			COWAT 1		
	Beta	95% CI	p-Value	Beta	95% CI	p-Value	Beta	95% CI	p-Value	Beta	95% CI	p-Value	Beta	95% CI	p-Value
Intercept	6.68	6.41 to 6.95	<0.001	4.86	4.62 to 5.11	<0.001	86.61	73.16 to 100.05	<0.001	42.30	38.61 to 45.99	<0.001	14.41	13.56 to 15.25	<0.001
Time (Stage 1)	-0.05	-0.10 to 0.01	0.080	-0.01	-0.07 to 0.05	0.745	4.02	0.34 to 7.70	0.033	-0.59	-1.27 to 0.10	0.095	-0.07	-0.29 to 0.15	0.514
Stage 2	-0.28	-0.60 to 0.03	0.079	0.11	-0.18 to 0.40	0.477	20.40	5.25 to 35.55	0.008	-2.38	-6.58 to 1.81	0.267	-1.05	-2.01 to -0.08	0.034
Stage 3	-0.67	-0.97 to -0.36	<0.001	-0.56	-0.84 to -0.28	<0.001	48.56	34.90 to 62.22	<0.001	-7.20	-11.71 to -2.68	0.002	-1.84	-2.70 to -0.97	<0.001
Stage 4	-1.10	-1.41 to -0.80	<0.001	-1.13	-1.41 to -0.84	<0.001	108.29	92.84 to 123.73	<0.001	-10.63	-15.11 to -6.15	<0.001	-3.84	-4.79 to -2.89	<0.001
Time × Stage2	-0.02	-0.09 to 0.05	0.551	-0.07	-0.14 to 0.01	0.076	2.17	-2.73 to 7.07	0.385	0.22	-0.64 to 1.09	0.611	-0.07	-0.36 to 0.22	0.630
Time × Stage3	-0.09	-0.16 to -0.02	0.015	-0.15	-0.23 to -0.07	<0.001	9.32	5.07 to 13.56	<0.001	-0.80	-1.74 to 0.13	0.095	-0.29	-0.55 to -0.04	0.023
Time × Stage4	-0.15	-0.24 to -0.06	0.001	-0.15	-0.25 to -0.05	0.003	27.46	20.18 to 34.73	<0.001	-1.98	-3.10 to -0.87	0.001	-1.02	-1.39 to -0.65	<0.001
Observations	1726			1722			3415			1232			3744		

	Processing speed			
	TMT A		Symbol substitution	
	Beta	95% CI	Beta	95% CI
Intercept	36.48	29.57 to 43.39	45.77	42.21 to 49.34
Time (Stage 1)	0.44	-2.74 to 3.61	-1.15	-1.78 to -0.52
Stage 2	4.58	-3.24 to 12.41	-2.12	-6.15 to 1.91
Stage 3	10.02	2.96 to 17.07	-11.08	-15.52 to -6.63
Stage 4	33.86	26.16 to 41.55	-17.71	-22.11 to -13.31
Time × Stage2	2.27	-1.90 to 6.43	-0.21	-1.06 to 0.64
Time × Stage3	6.64	3.08 to 10.20	-0.94	-2.02 to 0.15
Time × Stage4	25.49	21.04 to 29.94	-0.90	-2.37 to 0.56
Observations		3817		867
				p-Value
				<0.001
				<0.001
				0.304
				<0.001
				<0.001
				0.634
				0.092
				0.227

Word Recognition. Table 4 presents the tests that detected decline after 12 months separately for Stages 1, 2, and 3, including their MSDRs at 12, 24, and 36 months follow-up, showing that, overall, MSDR's tended to increase over time.

DISCUSSION

We operationalized the NIA-AA clinical scheme into measurable criteria, which enabled classification of individuals with biomarker evidence of AD into the four different stages of clinical severity. Subsequently, we demonstrated that neuropsychological tests differed in their sensitivity to decline at these different clinical stages. We found that more tests were sensitive to decline by increasing clinical stage. Moreover, in general, longer time intervals were needed to capture greater magnitude of change. Sensitive tests for Stage 1 focused primarily on semantic memory, whereas sensitive tests in succeeding stages also covered episodic memory, executive functioning (Stage 2), working memory, processing speed, and global cognition (Stages 3 and 4).

We found that most tests were sensitive to change over time in Stages 3 and 4 and that the majority of those were capable of detecting decline after 1 year. This is unsurprising, as most of those measures were originally designed to detect frank cognitive decline which is assumed to become apparent from Stage 3 on (Jack et al., 2018). However, we also identified a few measures that were sensitive to decline at Stages 1 and 2. Sensitive tests for Stage 1 included the CFT Animals and Vegetables (semantic memory) and LM immediate recall (episodic memory), whereas sensitive tests in Stage 2 included both CFT measures as well, plus Delayed Recall (episodic memory) and TMT-B (executive functioning). These results are largely in line with other studies showing that tests addressing the cognitive domains semantic memory, episodic memory, and executive functioning are among the first to decline in early AD (Mortamais et al., 2017; Papp et al., 2017; Ritchie et al., 2017).

Notably, not all measures addressing a particular cognitive domain were similarly sensitive to change across stages. For instance, a discrepancy between two semantic memory measures was observed, as the CFT detected decline in Stages 1 and 2, whereas the BNT did not. Possible explanations for this include differences regarding the difficulty level of both tests, as well as differences in measurement characteristics; for example, the BNT is untimed and has a fixed maximum score. As such, the BNT might be more susceptible to ceiling effects as compared to the CFT, thereby limiting its ability to detect decline in individuals with only subtle cognitive impairment. However, this is not to suggest that measurement properties alone influence the sensitivity of tests, as illustrated by the differences in sensitivity to change between the CFT and COWAT measures, which have similar measurement properties but rely on different cognitive strategies (Henry, Crawford, & Phillips, 2004).

Multi-domain cognitive composite measures have been designed in attempt to improve the measurement of cognitive

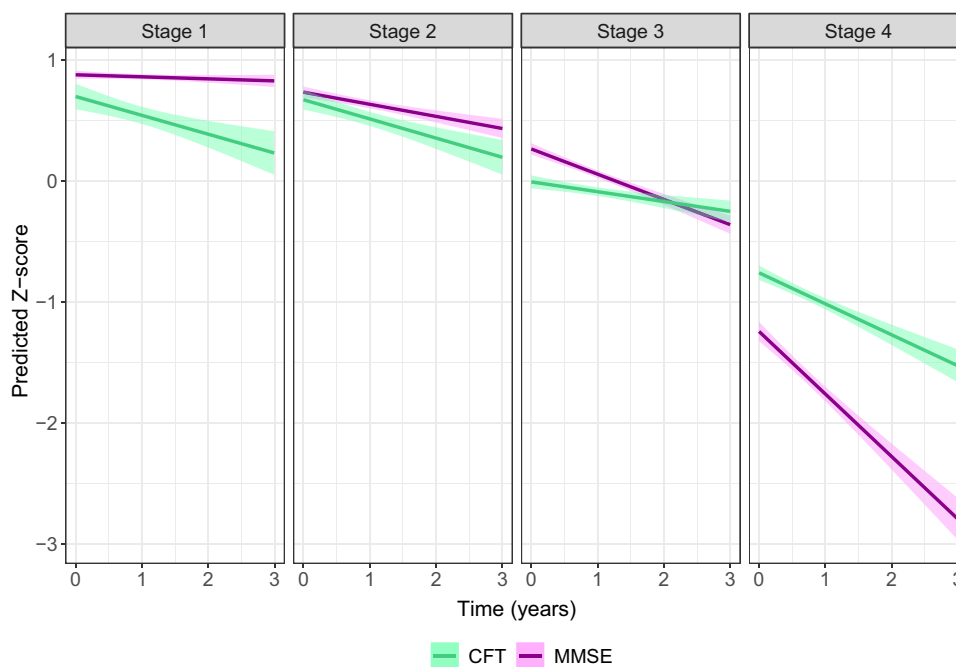


Fig. 1. Comparison between Category Fluency Test (CFT), Word List Delayed Recall, and Mini-Mental State Examination (MMSE), based on LMM adjusting for age, sex and education. Scores were z-transformed using overall group mean and SD, to facilitate comparisons between tests.

change in early stages of AD. Examples include the Preclinical Alzheimer's Cognitive Composite (PACC) (Donohue et al., 2014) and the Alzheimer's Prevention Initiative (API) composite for preclinical AD (Langbaum et al., 2014), and the Alzheimer's disease Composite Score (Wang et al., 2016) and the Cognitive-Functional Composite for prodromal stages of AD (Jutten et al., 2017). While composite scores resulting from these measures seem capable of tracking decline in preclinical and prodromal stages of AD (Donohue et al., 2017; Lim et al., 2016), few studies have focused on the individual components that contribute to these composites. In the current study, we identified sensitive measures that largely correspond with tests that have been included in the aforementioned composite measures (Donohue et al., 2014; Jutten et al., 2017; Langbaum et al., 2014). However, our results also indicated that some of the measures included in composites for preclinical AD may actually be less useful for detecting decline in this stage. For instance, the PACC includes the MMSE, and the API includes both the MMSE and BNT, whereas in our data set, these measures were not found to be sensitive to decline before Stage 3. This suggests that existing composites could potentially be further optimized to track short-term disease progression in Stages 1 and 2. Our findings could be used to select most sensitive measures in order to optimize composites or create novel composites, which could then be validated in an independent sample.

There are some limitations that should be taken into account. First, our combination of community-based and memory clinic-based cohorts may have led to some heterogeneity in our sample, especially in Stages 1 and 2. With

regard to the memory clinic-based cohorts, a potential limitation includes the fact that individuals who were labeled as having SCD may include individuals who actually did not have a memory concern themselves, but went to the memory clinic because a family member or other person expressed such a concern. Additionally, our results might have been biased by other differences between study cohorts, for example regarding amyloid assessment methods, follow-up timeframes, and cross-cultural differences. However, we explored cohort effects on the selected neuropsychological tests in an initial stage of our analyses, which we found did not affect our results. Another potential limitation includes the fact that we only looked at measures that were available across cohorts, thereby excluding potentially sensitive measures, such as the Free and Cued Selective Reminding Test (Grober, Veroff, & Lipton, 2018) which was only available in the HABS cohort. However, we chose this approach to avoid that results on a specific tests would be driven by a single cohort. Last, it should be noted that the majority of our sample is Caucasian, which limits the generalizability of our findings across populations. On the other hand, our study represents a more global sample than other studies that rely on single cohorts, by combining four data sets originating from Europe and the US, and also regional diversity within the US. Another major strength of this study is our large sample of AD biomarker positive individuals covering the entire AD continuum. Furthermore, our classification of clinical severity as defined by the NIA-AA framework might have provided a more objective method to classify individuals instead of relying on clinical syndrome diagnoses that could have been biased by cohort (Jack et al., 2019).

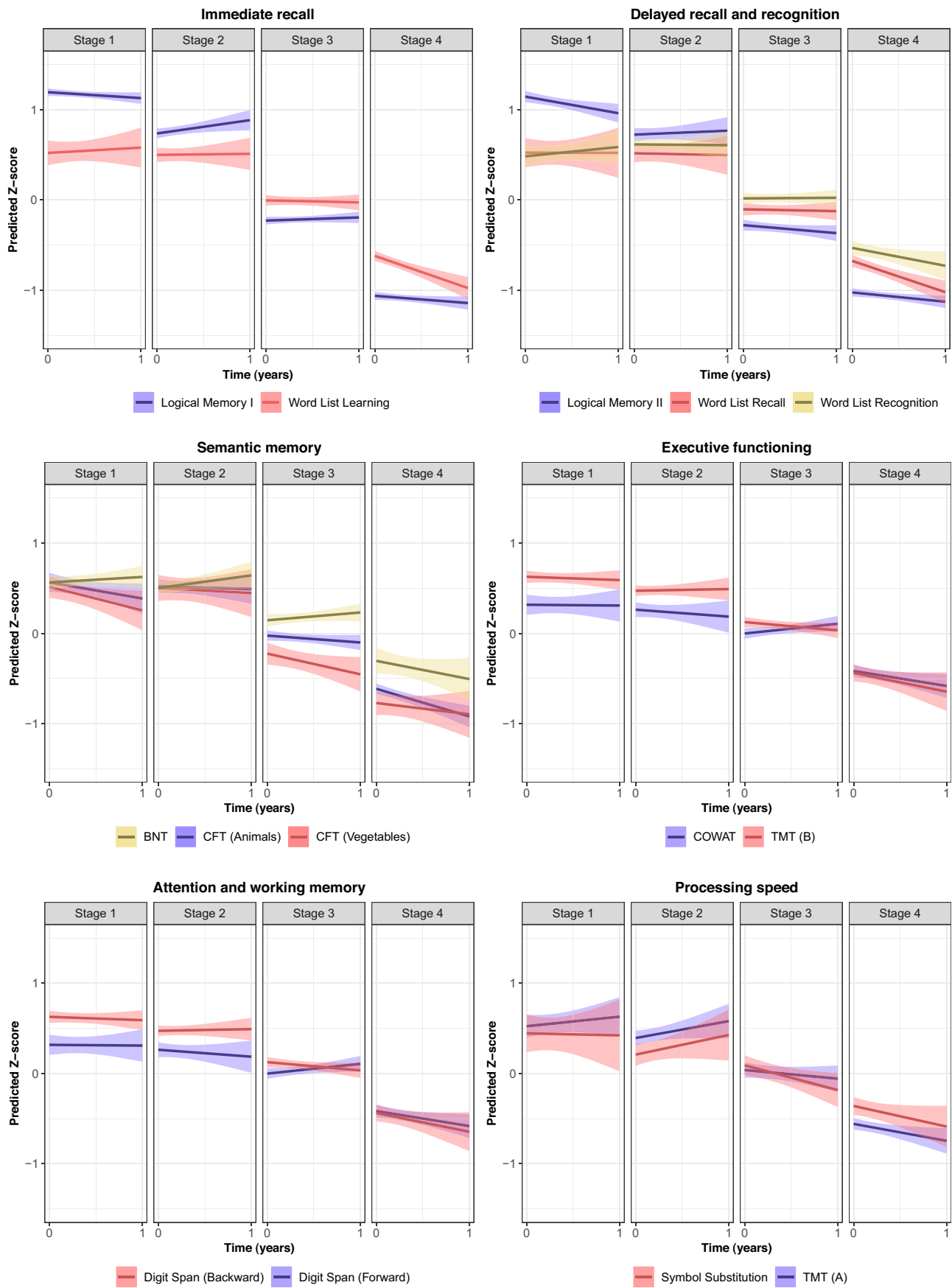


Fig. 2. Annual change for individual tests based on LMM correcting for, age, sex and education, stratified per stage. For each test, scores were z-transformed using baseline overall group mean and SD, to facilitate comparisons between tests.

Table 4. Cognitive tests identified as sensitive to decline after 12 months, including mean to standard deviation ratios (MSDRs) of change from baseline over succeeding follow-up time points

	Baseline – 12 months	Baseline – 24 months	Baseline – 36 months
Stage 1			
CFT (animals)	0.30 (<i>n</i> = 107)	0.29 (<i>n</i> = 90)	0.43 (<i>n</i> = 44)
CFT (vegetables)	0.38 (<i>n</i> = 68)	0.41 (<i>n</i> = 44)	0.38 (<i>n</i> = 39)
LM immediate recall	0.19 (<i>n</i> = 104)	0.13 (<i>n</i> = 89)	0.14 (<i>n</i> = 43)
Stage 2			
CFT (animals)	0.12 (<i>n</i> = 172)	0.26 (<i>n</i> = 133)	0.57 (<i>n</i> = 82)
CFT (vegetables)	0.39 (<i>n</i> = 54)	0.59 (<i>n</i> = 39)	0.74 (<i>n</i> = 24)
Word list delayed recall	0.22 (<i>n</i> = 121)	0.19 (<i>n</i> = 93)	0.42 (<i>n</i> = 57)
TMT part B	0.15 (<i>n</i> = 171)	0.24 (<i>n</i> = 130)	0.47 (<i>n</i> = 79)
Stage 3			
CFT (animals)	0.22 (<i>n</i> = 434)	0.34 (<i>n</i> = 274)	0.45 (<i>n</i> = 191)
CFT (vegetables)	0.49 (<i>n</i> = 78)	0.55 (<i>n</i> = 18)	
Digit span backward	0.12 (<i>n</i> = 141)	0.35 (<i>n</i> = 63)	0.41 (<i>n</i> = 29)
Digit span forward	0.20 (<i>n</i> = 141)	0.34 (<i>n</i> = 63)	0.39 (<i>n</i> = 29)
Word list delayed recall	0.43 (<i>n</i> = 348)	0.46 (<i>n</i> = 255)	0.52 (<i>n</i> = 185)
Word list recognition	0.21 (<i>n</i> = 346)	0.24 (<i>n</i> = 252)	0.34 (<i>n</i> = 184)
MMSE	0.24 (<i>n</i> = 363)	0.53 (<i>n</i> = 261)	0.54 (<i>n</i> = 195)
MOCA	0.37 (<i>n</i> = 69)	0.57 (<i>n</i> = 17)	N.a.
TMT part A	0.16 (<i>n</i> = 428)	0.24 (<i>n</i> = 275)	0.30 (<i>n</i> = 188)
TMT part B	0.17 (<i>n</i> = 397)	0.31 (<i>n</i> = 248)	0.39 (<i>n</i> = 172)

CFT, Category Fluency Test; LM, Logical Memory; TMT, Trail Making Test; MMSE, Mini-Mental State Examination; MOCA, Montréal Cognitive Assessment.

N.a. = Data not available for this time point.

To conclude, we demonstrated that commonly used neuropsychological tests differ in their ability to capture short-term cognitive decline and therefore disease progression depending on clinical stage (preclinical to symptomatic) within the AD continuum. This implies that stage-specific cognitive endpoints are needed to accurately assess change over time at different clinical stages of AD. The current study results can provide guidance on the selection of cognitive endpoints when the NIA-AA 2018 framework is applied to define the treatment population. Since only few existing tests were identified as sensitive in Stages 1 and 2, future directions include the development and validation of optimized composite measures, to capture the subtle cognitive decline observed in those preclinical stages. Furthermore, novel test paradigms that rely on digital assessment and scoring software are being developed, which could potentially aid in detecting and tracking of a greater magnitude of decline in preclinical stages of AD. This will ultimately lead to more accurate and fine-grained measurement and thereby increase the chance of successful evaluation of potentially preventive treatments for AD.

FUNDING

HABS: The Harvard Aging Brain Study is funded by the National Institute on Aging (P01AG036694; Principal Investigators Reisa Sperling, Keith Johnson) with additional support from several philanthropic organizations. KVP (1K23AG053422-01) is supported by a K23 award from NIA and an award from the Alzheimer's Association.

ADC: The Alzheimer Center Amsterdam is supported by Alzheimer Nederland and Stichting VUMC funds. Research of the Alzheimer Center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. The clinical database structure was developed with funding from Stichting Dioraphte. The SCIENCE project is supported by a research grant from Gieskes-Strijbis fonds. The present study is supported by a grant from Memorabel (grant no. 733050205), which is the research program of the Dutch Deltaplan for Dementia.

ADNI: Data collection and sharing for this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol, Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann, La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis

Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

NACC: The NACC database is funded by NIA/NIH grant U01 AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P30 AG062428-01 (PI James Leverenz, MD) P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P30 AG062421-01 (PI Bradley Hyman, MD, PhD), P30 AG062422-01 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI Robert Vassar, PhD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P30 AG062429-01 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P30 AG062715-01 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), and P50 AG047270 (PI Stephen Strittmatter, MD, PhD).

ACKNOWLEDGEMENTS

The authors are very thankful to all patients and participants in the studies included in the paper, as well as to everyone involved in the data collection and data sharing.

CONFLICTS OF INTEREST

RJJ, REA, RFB, MJP, DMR, KAJ, BNMB, PS and RS report no disclosures relevant to this manuscript. SAMS is supported by grants from JPND and Zon-MW and has provided consultancy services in the past 2 years for Nutricia and Takeda. All funds were paid to her institution. GAM has received research salary support from Eisai Inc., Eli Lilly and Company, Janssen Alzheimer Immunotherapy,

Novartis, and Genentech. Additionally, GAM has served as a consultant for Grifols Shared Services North America, Inc., Eisai Inc., and Pfizer. WMF holds the Pasman chair. CET received grants from the European Commission, the Dutch Research Council (ZonMW), Association of Frontotemporal Dementia/Alzheimer's Drug Discovery Foundation, The Weston Brain Institute, Alzheimer Nederland. CET has functioned in advisory boards of Roche, received non-financial support in the form of research consumables from ADxNeurosciences and Euroimmun, performed contract research or received grants from Probiobio, Biogen, Eisai, Toyama, Janssen prevention center, Boehringer, AxonNeurosciences, EIP farma, PeopleBio, Roche. KVP has served as a paid consultant for Biogen.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1355617720000934>

REFERENCES

- Aisen, P.S., Petersen, R.C., Donohue, M.C., Gamst, A., Raman, R., Thomas, R.G., ... Alzheimer's Disease Neuroimaging Initiative. (2010). Clinical core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 6(3), 239–246. <https://doi.org/10.1016/j.jalz.2010.03.006>
- Bateman, R.J., Xiong, C., Benzinger, T.L.S., Fagan, A.M., Goate, A., Fox, N.C., ... for the Dominantly Inherited Alzheimer Network. (2012). Clinical and biomarker changes in Dominantly Inherited Alzheimer's Disease. *The New England Journal of Medicine*, 367(9), 795–804. <https://doi.org/10.1056/NEJMoa1202753>
- Beekly, D.L., Ramos, E.M., Lee, W.W., Deitrich, W.D., Jacka, M.E., Wu, J., ... Raskind, M. (2007). The National Alzheimer's Coordinating Center (NACC) database: The uniform data set. *Alzheimer Disease and Associated Disorders*, 21(3), 249–258. <https://doi.org/10.1080/00380768.1987.10557563>
- Besser, L., Kukull, W., Knopman, D.S., Chui, H., Galasko, D., Weintraub, S., ... Morris, J.C. (2018). Version 3 of the National Alzheimer's Coordinating Center's uniform data set. *Alzheimer Disease and Associated Disorders*, 32(4), 351–358. <https://doi.org/10.1097/WAD.0000000000000279>
- Chapman, K.R., Bing-Canar, H., Alosco, M.L., Steinberg, E.G., Martin, B., Chaisson, C., ... Stern, R.A. (2016). Mini Mental State Examination and Logical Memory scores for entry into Alzheimer's disease trials. *Alzheimer's Research & Therapy*, 8, 9. <https://doi.org/10.1186/s13195-016-0176-z>
- Clark, C.M., Schneider, J.A., Bedell, B.J., Beach, T.G., Bilker, W.B., Mintun, M.A., ... for the AV45-A07 Study Group. (2011). Use of Florbetapir-PET for imaging β -amyloid pathology. *JAMA*, 305(3), 275–283. <https://doi.org/10.1001/jama.2010.2008>
- Dagley, A., LaPoint, M., Huijbers, W., Hedden, T., McLaren, D.G., Chatwal, J.P., ... Schultz, A.P. (2017). Harvard aging brain study: Dataset and accessibility. *NeuroImage*, 144(Pt B), 255–258. <https://doi.org/10.1016/j.neuroimage.2015.03.069>
- Donohue, M.C., Sperling, R.A., Salmon, D.P., Rentz, D.M., Raman, R., Thomas, R.G., ... Alzheimer's Disease Cooperative Study

- Group. (2014). The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA Neurology*, 71(8), 961–970. <https://doi.org/10.1001/jamaneurol.2014.803>
- Donohue, M.C., Sun, C.K., Raman, R., Insel, P.S., Aisen, P.S., the North American Alzheimer's Disease Neuroimaging Initiative, Australian Imaging Biomarkers and Lifestyle, the Japanese Alzheimer's Disease Neuroimaging Initiative. (2017). Cross-validation of optimized composites for preclinical Alzheimer's disease. *Alzheimer's & Dementia*, 3(1), 123–129. <https://doi.org/10.1016/j.trci.2016.12.001>
- Dubois, B., Hampel, H., Feldman, H.H., Scheltens, P., Aisen, P., Andrieu, S., ... Jack, C.R. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 12(3), 292–323. <https://doi.org/10.1016/J.JALZ.2016.02.002>
- Evans, S., McRae-McKee, K., Wong, M.M., Hadjichrysanthou, C., De Wolf, F., & Anderson, R. (2018). The importance of endpoint selection: How effective does a drug need to be for success in a clinical trial of a possible Alzheimer's disease treatment? *European Journal of Epidemiology*, 33(7), 635–644. <https://doi.org/10.1007/s10654-018-0381-0>
- Farias, S.T., Mungas, D., Reed, B.R., Cahn-Weiner, D., Jagust, W., Baynes, K., & DeCarli, C. (2008). The measurement of everyday cognition (ECog): Scale development and psychometric properties. *Neuropsychology*, 22(4), 531.
- Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189–198.
- Food and Drug Administration. (2018). *Early Alzheimer's Disease: Developing Drugs for Treatment. Guidance for Industry*.
- Grober, E., Veroff, A.E., & Lipton, R.B. (2018). Temporal unfolding of declining episodic memory on the Free and Cued Selective Reminding Test in the prodementia phase of Alzheimer's disease: Implications for clinical trials. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, 161–171.
- Henry, J.D., Crawford, J.R., & Phillips, L.H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia*, 42(9), 1212–1222. <https://doi.org/10.1016/j.neuropsychologia.2004.02.001>
- Hughes, C.P., Berg, L., Danziger, W.L., Coben, L.A., & Martin, R.L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140, 566–572. Retrieved from <http://bjpr.rcpsych.org/content/bjprpsych/140/6/566.full.pdf>
- Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., ... Contributors. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4), 535–562. <https://doi.org/10.1016/j.jalz.2018.02.018>
- Jack, C.R., Thorneau, T.M., Weigand, S.D., Wiste, H.J., Knopman, D.S., Vemuri, P., ... Petersen, R.C. (2019). Prevalence of biologically vs clinically defined Alzheimer spectrum entities using the National Institute on Aging–Alzheimer's Association Research Framework. *JAMA Neurology*, 76(10), 1174–1183. <https://doi.org/10.1001/jamaneurol.2019.1971>
- Jutten, R.J., Harrison, J., de Jong, F.J., Aleman, A., Ritchie, C.W., Scheltens, P., & Sikkes, S.A.M. (2017). 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*, 3(1), 130–138. Retrieved from http://ac.els-cdn.com/S2352873717300045/1-s2.0-S2352873717300045-main.pdf?_tid=f0bd2a46-2bf4-11e7-9482-00000a0b0f6b&acdnat=1493371846_0ae2f35af7861b14f06d410655ae2cb2
- Langbaum, J.B., Hendrix, S.B., Ayutyanont, N., Chen, K., Fleisher, A.S., Shah, R.C., ... Reiman, E.M. (2014). An empirically derived composite cognitive test score with improved power to track and evaluate treatments for preclinical Alzheimer's disease. *Alzheimer's & Dementia*, 10(6), 666–674. <https://doi.org/10.1016/j.jalz.2014.02.002>
- Lezak, M.D. (2004). *Neuropsychological Assessment*. New York: Oxford University Press.
- Lim, Y.Y., Snyder, P.J., Pietrzak, R.H., Ukiqi, A., Villemagne, V.L., Ames, D., ... Maruff, P. (2016). Sensitivity of composite scores to amyloid burden in preclinical Alzheimer's disease: Introducing the Z-scores of attention, verbal fluency, and episodic memory for nondemented older adults composite score. *Alzheimer's & Dementia*, 2, 19–26. <https://doi.org/10.1016/j.dadm.2015.11.003>
- Mormino, E.C., Betensky, R.A., Hedden, T., Schultz, A.P., Ward, A., Huijbers, W., ... Sperling, R.A. (2014). Amyloid and APOE $\epsilon 4$ interact to influence short-term decline in preclinical Alzheimer disease. *Neurology*, 82(20), 1760–1767. <https://doi.org/10.1212/WNL.0000000000000431>
- Morris, J.C., Weintraub, S., Chui, H.C., Cummings, J., DeCarli, C., Ferris, S., ... Ramos, E.M. (2006). The Uniform Data Set (UDS): Clinical and cognitive variables and descriptive data from Alzheimer disease centers. *Alzheimer Disease and Associated Disorders*, 20(4), 210–216. <https://doi.org/http://dx.doi.org/10.1097/01.wad.0000213865.09806.92>
- Mortamais, M., Ash, J.A., Harrison, J., Kaye, J., Kramer, J., Randolph, C., ... Ritchie, K. (2017). Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. *Alzheimer's & Dementia*, 13(4), 468–492. <https://doi.org/10.1016/J.JALZ.2016.06.2365>
- Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695–699.
- Papp, K.V., Rentz, D.M., Orlovsky, I., Sperling, R.A., & Mormino, E.C. (2017). Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: The PACC5. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3(4), 668–677. <https://doi.org/10.1016/J.TRCL.2017.10.004>
- Rentz, D.M., Parra Rodriguez, M.A., Amariglio, R., Stern, Y., Sperling, R., & Ferris, S. (2013). Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. *Alzheimer's Research & Therapy*, 5(6), 58. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3978443/pdf/alzrt222.pdf>
- Ritchie, K., Ropacki, M., Alcala, B., Harrison, J., Kaye, J., Kramer, J., ... Ritchie, C.W. (2017). Recommended cognitive outcomes in preclinical Alzheimer's disease: Consensus statement from the European prevention of Alzheimer's dementia project. *Alzheimer's & Dementia*, 13. <https://doi.org/10.1016/j.jalz.2016.07.154>
- Ruff, R.M., Light, R.H., Parker, S.B., & Levin, H.S. (1996). Benton controlled oral word association test: Reliability and updated norms. *Archives of Clinical Neuropsychology*, 11(4), 329–338.
- Schmidt, M. (1996). *Rey Auditory Verbal Learning Test: A Handbook*. Los Angeles, CA: Western Psychological Services.
- Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., ... Phelps, C.H. (2011). Toward defining the

- preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3), 280–292. <https://doi.org/10.1016/j.jalz.2011.03.003>
- Tijms, B.M., Willemse, E.A.J., Zwan, M.D., Mulder, S.D., Visser, P.J., van Berckel, B.N.M., ... Teunissen, C.E. (2018). Unbiased approach to counteract upward drift in cerebrospinal fluid amyloid-beta 1–42 analysis results. *Clinical Chemistry*, 64(3), 576–585. <https://doi.org/10.1373/clinchem.2017.281055>
- Tombaugh, T.N. (2004). Trail making test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19(2), 203–214. [https://doi.org/https://doi.org/10.1016/S0887-6177\(03\)00039-8](https://doi.org/https://doi.org/10.1016/S0887-6177(03)00039-8)
- Trzepacz, P.T., Hochstetler, H., Wang, S., Walker, B., & Saykin, A.J. (2015). Relationship between the Montreal cognitive assessment and mini-mental state examination for assessment of mild cognitive impairment in older adults. *BMC Geriatrics*, 15(1), 1–9. <https://doi.org/10.1186/s12877-015-0103-3>
- van der Flier, W.M. (2018). Amsterdam dementia cohort: Performing research to optimize care. *Journal of Alzheimer's Disease: JAD*, 62(3), 1091–1111.
- Wang, J., Logovinsky, V., Hendrix, S.B., Stanworth, S.H., Perdomo, C., Xu, L., ... Satlin, A. (2016). ADCOMS: A composite clinical outcome for prodromal Alzheimer's disease trials. *J Neurol Neurosurg Psychiatry*, 87(9), 993–999. <https://doi.org/10.1136/jnnp-2015-312383>
- Wechsler, D. (1987). *WMS-R: Wechsler Memory Scale-Revised*. New York: Psychological Corporation.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) (Vol. 22, p. 498)*. San Antonio, TX: NCS Pearson.
- Williams, B.W., Mack, W., & Henderson, V.W. (1989). Boston naming test in Alzheimer's disease. *Neuropsychologia*, 27(8), 1073–1079. [https://doi.org/https://doi.org/10.1016/0028-3932\(89\)90186-3](https://doi.org/https://doi.org/10.1016/0028-3932(89)90186-3)
- Williams, M.M., Storandt, M., Roe, C.M., & Morris, J.C. (2013). Progression of Alzheimer's disease as measured by Clinical Dementia Rating Sum of Boxes scores. *Alzheimers Dement*, 9(1 Suppl), S39–S44. <https://doi.org/10.1016/j.jalz.2012.01.005>
- Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M., & Leirer, V.O. (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research*, 17(1), 37–49.