

The Computerized Cognitive Composite (C3) in A4, an Alzheimer's Disease Secondary Prevention Trial

K.V. Papp^{1,2}, D.M. Rentz^{1,2}, P. Maruff^{3,4}, C.-K. Sun⁵, R. Raman⁵, M.C. Donohue⁵, A. Schembri⁴, C. Stark⁶, M.A. Yassa⁶, A.M. Wessels⁷, R. Yaari⁷, K.C. Holdridge⁷, P.S. Aisen⁵, R.A. Sperling^{1,2} on behalf of the A4 Study Team*

1. Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 2. Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 3. The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, Victoria, Australia; 4. Cogstate, Ltd, Melbourne, Victoria, Australia; 5. Alzheimer Therapeutic Research Institute, Keck School of Medicine, University of Southern California, San Diego, CA, USA; 6. Center for the Neurobiology of Learning and Memory and Department of Neurobiology and Behavior, University of California Irvine, Irvine, California, USA; 7. Eli Lilly and Company, Indianapolis, Indiana, USA; *Full listing of A4 Study team and site personnel available at A4STUDY.org

Corresponding Author: Kathryn V. Papp, Center for Alzheimer Research and Treatment; 60 Fenwood Road; Boston, MA 02115, Telephone: 617-643-5322; Fax: 857-5461, Email Address: kpapp@bwh.harvard.edu

J Prev Alz Dis 2021;1(8):59-67
Published online July 5, 2020, <http://dx.doi.org/10.14283/jpad.2020.38>

Abstract

BACKGROUND: Computerized cognitive assessments may improve Alzheimer's disease (AD) secondary prevention trial efficiency and accuracy. However, they require validation against standard outcomes and relevant biomarkers.

OBJECTIVE: To assess the feasibility and validity of the tablet-based Computerized Cognitive Composite (C3).

DESIGN: Cross-sectional analysis of cognitive screening data from the A4 study (Anti-Amyloid in Asymptomatic AD).

SETTING: Multi-center international study.

PARTICIPANTS: Clinically normal (CN) older adults (65-85; n=4486).

MEASUREMENTS: Participants underwent florbetapir-Positron Emission Tomography for A β +/- classification. They completed the C3 and standard paper and pencil measures included in the Preclinical Alzheimer's Cognitive Composite (PACC). The C3 combines memory measures sensitive to change over time (Cogstate Brief Battery-One Card Learning) and measures shown to be declining early in AD including pattern separation (Behavioral Pattern Separation Test- Object-Lure Discrimination Index) and associative memory (Face Name Associative Memory Exam- Face-Name Matching). C3 acceptability and completion rates were assessed using qualitative and quantitative methods. C3 performance was explored in relation to A β +/- groups (n=1323/3163) and PACC.

RESULTS: C3 was feasible for CN older adults to complete. Rates of incomplete or invalid administrations were extremely low, even in the bottom quartile of cognitive performers (PACC). C3 was moderately correlated with PACC (r=0.39). A β + performed worse on C3 compared with A β - [unadjusted Cohen's d=-0.22 (95%CI: -0.31,-0.13) p<0.001] and at a magnitude comparable to the PACC [d=-0.32 (95%CI: -0.41,-0.23) p<0.001]. Better C3 performance was observed in younger, more educated, and female participants.

CONCLUSIONS: These findings provide support for both the feasibility and validity of C3 and computerized cognitive outcomes more generally in AD secondary prevention trials.

Key words: Digital biomarkers, cognition, computerized testing, preclinical Alzheimer's disease, secondary prevention.

Introduction

Computerized cognitive assessments have the potential to significantly reduce data administration and scoring errors, site burden, and cost in Alzheimer's disease (AD) secondary prevention trials as cognitive screening tools and outcome measures. These assessments have yet to replace paper and pencil measures as primary outcomes given several remaining questions: How feasible are computerized assessments in normal older adults and older adults who progress to Mild Cognitive Impairment (MCI) over the course of a trial? How reliable is the data collected? And finally, how valid are computerized cognitive assessments, that is, are they related to gold-standard paper and pencil primary outcomes and AD pathology targeted in a given intervention?

The Anti-Amyloid in Asymptomatic Alzheimer's (A4) study (1, 2) offers a unique opportunity to address some of these questions by assessing the feasibility and validity of the Computerized Cognitive Composite (C3) in a very large multi-site AD secondary prevention study targeting clinically normal (CN) older adults with elevated cerebral amyloid (2). The C3 is derived using two well-validated memory paradigms from the cognitive neuroscience literature: the Face Name Associative Memory Exam (FNAME) and the Behavioral Pattern Separation Task-Object (BPS-O). It also includes measures from the Cogstate Brief Battery (CBB) which uses playing cards to assess visual memory in addition to reaction time (RT) and working memory and was designed to be sensitive to change over time with randomized alternate forms. The CBB has been studied in relationship to AD neuroimaging markers in several cohort studies of normal older adults (3, 4). Behavioral versions of the FNAME (5, 6) and a modified version of the BPS-O (7) were selected for inclusion in the C3 as they have been shown to elicit aberrant activity in the medial temporal

Table 1. Participant Characteristics by A β Status

	All n=4486	A β -n=3163	A β + n=1323	p-value
Age M(SD)	71.29 (4.67)	70.95 (4.53)	72.10 (4.89)	<0.0001
Sex (% female)	59%	60%	59%	0.641
Education M(SD)	16.58(2.84)	16.59 (2.85)	16.54(2.81)	0.564
APOE Genotype (% ϵ 4+)	35%	25%	58%	<0.001
Race (% Caucasian)	92%	91%	94%	<0.001

Note. Two-sample t-test with unequal variances were used for continuous variables and Fisher's Exact test for categorical variables. Values are Mean (Standard Deviation) unless otherwise indicated.

lobes during functional imaging studies in individuals at risk for AD based on biomarkers (8-10). More specifically, these individuals fail to habituate to repeated stimuli (FNAME) or during both correct rejections and false alarms (BPS-O), neural signatures consonant with successful memory formation. The C3 was identified a-priori to include one primary memory outcome from each component measure including: the BPS-O lure discrimination index, Face-Name Matching accuracy, and One-Card Learning accuracy.

The aim of this study was to assess the feasibility and validity of the C3 in CN older adults participating in a secondary prevention trial. Specific goals included determining whether reliable C3 data was consistently captured using a touchscreen tablet and whether data reliability decreased in the lowest cognitive performers. To assess the validity of the C3, we investigated 1) whether the C3 was related to the primary study outcome: performance on traditional paper and pencil measures (i.e., the Preclinical Alzheimer's Cognitive Composite- PACC) 2) whether the C3 was related to cerebral amyloid (A β) and 3) whether the magnitude of this relationship was comparable to that observed between PACC and A β +/- . In addition to our main aims, we explored whether improved performance with C3 retesting using alternate forms differentiated between A β +/- individuals above and beyond cross-sectional performance. Finally, we explored performance on the constituent tests from the C3 and their relationships with A β status, demographic characteristics, and paper and pencil measures. The implications of these findings as they relate to the design and use of future computerized outcomes in secondary prevention trials are discussed.

Methods

Participants and Study Design

The A4 Study is a double-blind, placebo-controlled 240-week Phase 3 trial of an anti-A β monoclonal antibody in CN older adults with preclinical AD (2) occurring across 67 sites. Participants interested in enrolling in A4 were required to be aged 65 to 85 and were deemed clinically normal (CN) based on Mini Mental Status Exam (MMSE) ranging from 25-30 and Global Clinical Dementia (CDR) Rating Score of 0. During their initial screening visit, participants completed traditional

and computerized cognitive testing (detailed further below). Prior to enrollment, they underwent a florbetapir Positron Emission Tomography (PET) for classification of A β status (Table 1) at a second visit. On their third visit, all potential participants completed computerized testing and were subsequently provided with results of their AD biomarker imaging and informed about whether they were eligible (A β +) or ineligible (A β -) to enroll in the trial. The current study includes cognitive screening data at 2 timepoints for A β + and A β - individuals.

Cognitive Measures

The primary outcome for the A4 Study is performance on the PACC, a multi-domain composite of paper and pencil measures (11). Measures contributing to the C3 are administered on a touchscreen tablet using the Cogstate platform and serve as an exploratory outcome. All participants completed the PACC and C3 at the first screening visit (Visit 1) and an alternate C3 within 90 days (mean=55 days) at the study eligibility visit (Visit 3) prior to study eligibility disclosure.

Paper and Pencil Cognitive Testing: The PACC

The PACC, described in detail elsewhere (11), is calculated as the sum of mean performance across four measures normalized using a z-score including the MMSE (0–30), the WMS-R Logical Memory Delayed Recall (LMDR; 0–25), the Digit-Symbol Coding Test (DSC; 0–93), and the Free and Cued Selective Reminding Test-Free + Total Recall (FCSRT96; 0–96) (2).

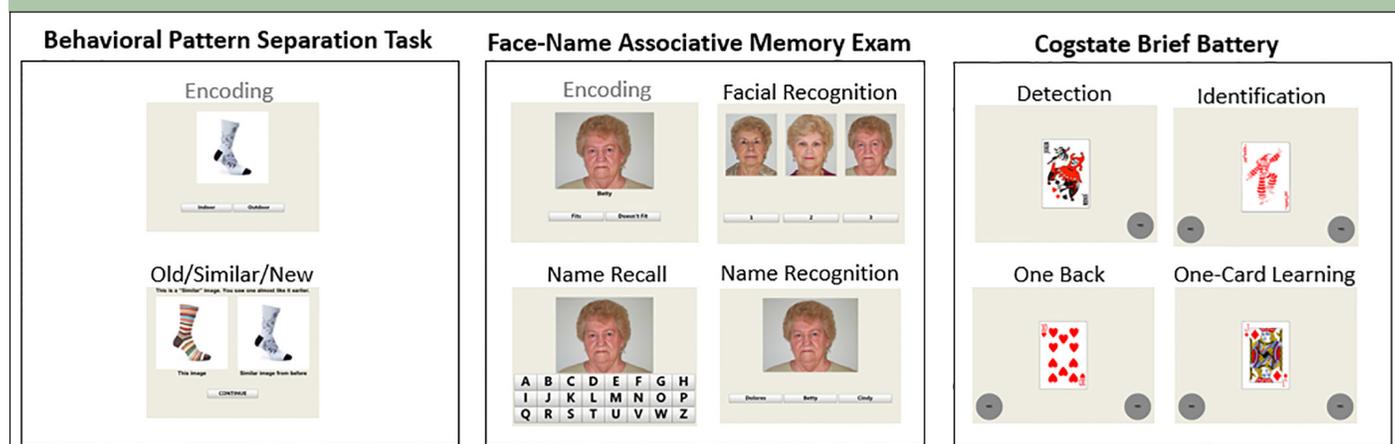
Computerized Testing: The C3

Figure 1 provides a schematic of C3 Components: BPS-O, FNAME and the CBB. An examiner is present in the testing room and initially guides administration, but the battery has the potential to be completed largely independently in the context of written on-screen instructions and automatic transitions between tasks (12).

Behavioral Pattern Separation- Object (BPS-O; more recently termed the Mnemonic Similarity Test)

Participants are presented with images of 40 everyday objects serially and are allotted 5 seconds to determine

Figure 1. C3 Task Schematic



Note. All tasks are completed on a tablet using a touchscreen. Stimuli in gray are not scored.

whether the item is for use “indoors” or “outdoors” to ensure adequate attentiveness to stimuli (7). Participants are subsequently shown 20 of the same items interspersed with both novel images and lure images. They are asked to categorize each image as: Old, Similar, or New within 5 seconds. Accuracy and RT measures are collected. Of interest is the rate at which participants can correctly identify lures as “Similar” rather than as “Old.” The lure discrimination index (LDI) is computed as the proportion of “Similar” responses given to lure items minus the ratio of “Similar” responses given to the foils (the latter is to correct for response bias). The LDI is the primary outcome from the BPS-O task. A higher LDI indicates better pattern separation performance.

Face-Name Associative Memory Exam (FNAME)

Participants are shown 12 face-name pairs presented serially. For each face-name pair, the participant is asked whether the name “fits” or “doesn’t fit” the face to ensure adequate attentiveness to the stimuli. Participants are allowed 5 seconds to respond and are asked to try to remember the face-name pair. Following the learning phase, the CBB tests serve as a 12 to 15-minute delay. Subsequently, there are three measures of memory including face recognition (FSBT), first letter name recall (FNLT) and face-name matching (FNMT). In FSBT, participants are asked to identify the previously learned faces, presented alongside two distractor faces of matching age, race, and sex. The target face is subsequently presented with a touchscreen keyboard and the participant selects the first letter of the name paired with that face (FNLT). Finally, the target face is presented with three names (target name, a re-paired same-sex name, and an age and sex-matched foil name) and the participant must select the correct name (FNMT). Accuracy for each component is scored /12 with FNMT number of correct matches serving as the primary outcome of interest.

Cogstate Brief Battery (CBB)

The CBB (13, 14) uses playing cards as stimuli and includes a measure of attention (Detection-DET), reaction time RT (Identification-IDN), working memory (One-Back Test-ONB), and visual memory (One-Card Learning-OCL). Measures of RT and accuracy are recorded. To address skewness, a log₁₀ transformation is applied to RT measures and an arcsin sqrt transformation is applied to accuracy measures. In DET, participants are required to tap ‘Yes’ as quickly as possible in response to a stimulus card turning face-up. The task continues until 35 correct trials are recorded. The outcome is RT. In IDN, a participant must select whether the card is red or not red; thirty correct trials are required. RT is the primary outcome for IDN; IDN accuracy was also examined. In ONB, participants must indicate “yes” or “no” whether the current card is equivalent to the previously seen card. In OCL, participants must learn a series of playing cards by responding ‘yes’ or ‘no’ to whether the card has been previously seen in the task. For ONB and OCL, both RT and accuracy are computed. Here, we examined RT and Accuracy for both IDN and ONB. We examined only RT for DET and only Accuracy for OCL.

The C3

Constituents of the C3 were identified a-priori and include one primary memory outcome from each measure including the BPS-O LDI, FNMT, and OCL. The C3 is computed as the average of these z-scored outcomes derived from the study population at Visit 1.

Data Quality

Data from individual C3 measures were included in analyses if they met pre-specified task-specific completion checks (Supplementary Table 1). For example, OCL for a given participant is included in analyses if

the participant responds in $\geq 75\%$ of trials. Study rater comments were also reviewed to better determine C3 usability and acceptability.

Amyloid PET Imaging

Eligible participants completed a florbetapir PET scan at Visit 2. Scan acquisition occurred over 50-70 minutes following an injection of 10mCi of florbetapir-F18. $A\beta$ binding was assessed using mean standardized uptake value ratio (SUVr) with whole cerebellar gray as a reference region. Participants were deemed eligible ($A\beta+$) versus not eligible ($A\beta-$) using an algorithm combining both quantitative SUVr (>1.15) information and a centrally-determined visual read (2).

Statistical Analyses

Primary analyses were performed on the C3 at Visit 1. To assess C3 feasibility and data validity, test completion rates and performance checks were computed (Supplementary Table 1) and rates subsequently compared between $A\beta+/-$ groups using Chi-square tests. Rater comments were systematically reviewed and observations by raters were grouped into categories (e.g., technical issue, interruptions) and the frequency of observations made in each category were computed. To infer C3 feasibility and data validity in those who may develop impairment over the course of the A4 study, we compared test completion rates and performance checks between the lowest cognitive performers (bottom quartile on PACC) with typical cognitive performers using *chi* square tests.

Demographic differences between $A\beta+/-$ groups were assessed using Welch's two-sample t-tests for continuous variables and Fisher's Exact test for categorical variables (e.g., age, APOE). Linear models were fit to compare cognitive performance across males and females. Linear models were fit to compare cognitive performance across $A\beta+/-$ while adjusting for covariates: age, sex, and education. Effect size was computed as a Cohen's *d* (mean difference between $A\beta+$ and $A\beta-$ groups divided by the pooled standard deviation) with 0.01 representing a "very small" effect, 0.20 representing a "small" effect, and 0.5 representing a "medium" effect (15). Comparable linear models were performed and effect sizes calculated for individual C3 components to examine $A\beta+/-$ group differences on individual C3 measures (e.g., OCL, ONB, BPS-O). No adjustments were made for multiple comparisons; however, results are reported as point estimates and 95% confidence intervals.

Differences in performance between Visit 1 and Visit 3 were examined using linear models of difference scores with $A\beta$ status, age, sex, and education as covariates.

Pearson correlation coefficients were computed to assess the relationships between C3 and demographic characteristics as well as C3 and the PACC. Pearson

correlation coefficients were similarly used to assess the relationships among C3 components and PACC components to assess the convergent and discriminant validity between memory versus non-memory tasks on C3 versus PACC.

Linear models were also fit to compare cognitive performance between $\epsilon 4+/-$ while adjusting for covariates: age, sex, and education.

All analyses were conducted using R version 3.6.1 (R-project.org).

Results

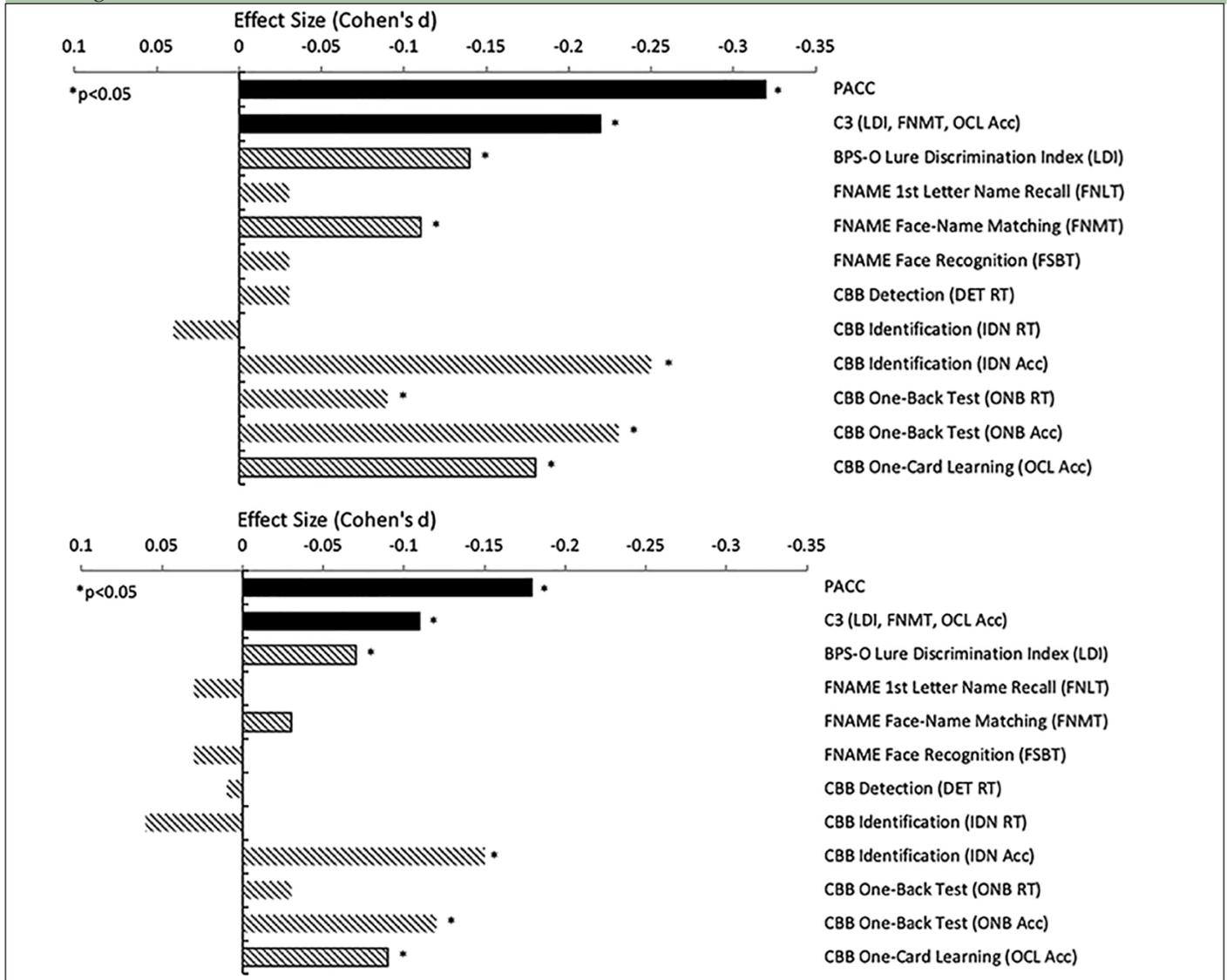
Feasibility of the C3

Completion and performance checks were met in $>98\%$ of individual test administrations within the C3 (Supplementary Table 1) and equivalent by $A\beta+/-$. Raters reported issues in approximately 4% of C3 administrations. The most commonly reported problem (reflecting 0.7% of administrations) was that the tablet was insufficiently responsive to a participant's finger taps and/or the participant was mis-tapping by either hovering their fingers too closely to the screen or by tapping too quickly. The second most commonly reported issue (0.5% of administrations) was overly deliberative responding on BPS-O and FNAME causing items to time-out. This was followed by non-specific technical issues (e.g., frozen program, interruptions from low battery signal or software update, glitches such as stimulus not loading or items auto-proceeding). Report of confusion with task instructions was very low (reported in 0.3% of administrations). Participants most commonly had difficulty understanding instructions for ONB and OCL; additionally, some reported confusion regarding the goal of the judgment component of BPS-O and FNAME learning components (i.e., indoor vs. outdoor, fits vs. doesn't fit). Despite this, few participants ($<3\%$) failed to make an "indoor/outdoor" or a "fits" judgment on more than 3 items. Participants refused to continue C3 testing in $<0.002\%$ of administrations with the most common reasons including frustration and fatigue.

Predictions for the Feasibility of the C3 Longitudinally

To preliminarily estimate whether the C3 (to be completed at 6-month intervals for the A4 study duration) will remain feasible in participants experiencing cognitive decline, we examined C3 performance in the lowest cognitive performers on PACC. The magnitude of the C3 $A\beta$ group difference increased by a factor of 5.2 when restricting the $A\beta+$ group to the bottom quartile of PACC [adjusted Cohen's $d=-0.57$ (95% CI: -0.68, -0.45) $p<0.001$], however, no significant changes in rates of performance completion and performance checks were observed.

Figure 2. Covariate-Unadjusted and Adjusted Group Differences (Effect Sizes: Cohen's d) Between Aβ+ / Aβ- Groups at Screening Visit 1



Note. Smaller effect size (Cohen's d) is associated with worse performance in Aβ+ (n=1323) relative to Aβ- (n=3163). Top (unadjusted) and bottom (covariate-adjusted). PACC=Preclinical Alzheimer's Cognitive Composite; C3= Computerized Cognitive Composite; FNAME=Face-Name Associative Memory Exam; CBB=Cogstate Brief Battery; RT=reaction time; Acc=Accuracy

Demographic and Clinical Characteristics

Aβ+ were older compared with Aβ- (Table 1). There were no group differences for sex or education level. Aβ+ exhibited a higher rate of ε4 positivity and higher proportion of Caucasians compared with Aβ-.

C3 Performance

Aβ+ performed worse on the C3 compared with Aβ- (unadjusted d=-0.22, adjusted d=-0.11), mirroring the Aβ+/- performance difference on the PACC (unadjusted d=-0.32, adjusted d=-0.18) (Figure 2; Table 2). Importantly, the majority of participants were performing in the normal range, with performance in Aβ+ on average only -0.08 standard deviations below the

mean. In addition to Aβ positivity (Beta=-0.07 p=0.002), older age (Beta= -0.04 p<0.0001), less education (Beta= 0.03 p<0.0001), and male sex (Beta=-0.10 p<0.0001) contributed to overall worse C3 performance. Models adjusted for demographic features generally resulted in smaller Aβ+/- effect sizes compared with unadjusted models (Figure 2). For example, there was 66% decrease in effect size between the unadjusted (d=-0.22) and adjusted C3 (d=-0.11). C3 and PACC were moderately correlated (r=0.39, p<0.001). However, both contributed unique explanatory variance about Aβ+/- status when modeled together (Supplementary Table 2 Model A).

Improved performance at re-testing was observed for C3 with an average increase of 0.25 standard deviations between visits (Beta=0.25, p<0.0001). However, there was no relationship between Aβ status and differential improvement on C3 re-testing (Beta= 0.00, p=0.961).

Table 2. Group Differences Between A β + versus A β - on C3 at Screening Visit 1

		n=3163	n=1323	Unadjusted		Covariate-Adjusted	
		A β - M (SD)	A β + M (SD)	Cohen's d [95% C.I.]	p-value	Cohen's d [95% C.I.]	p-value
	PACC	0.18 (2.45)	-0.43 (2.68)	-0.32 [-0.41, -0.23]	<0.001	-0.18 [-0.25, -0.12]	<0.001
	C3 (LDI, FNMT, OCL)	0.04 (0.65)	-0.07 (0.68)	-0.22 [-0.31, -0.13]	<0.001	-0.11 [-0.17, -0.04]	<0.001
BPS-O	LDI	0.41 (0.20)	0.39 (0.21)	-0.14 [-0.23, -0.05]	0.002	-0.07 [-0.14, -0.01]	0.033
FNAME	FNLT	3.76 (2.24)	3.71 (2.27)	-0.03[-0.12, 0.06]	0.526	0.03 [-0.04, 0.09]	0.402
	FNMT	8.17 (1.92)	8.01 (2.04)	-0.11[-0.20, -0.02]	0.017	-0.03 [-0.10, 0.03]	0.332
	FSBT	10.48 (1.71)	10.44 (1.75)	-0.03 [-0.12, 0.06]	0.526	0.03 [-0.03, 0.1]	0.330
CBB	DET RT	2.60 (0.10)	2.60 (0.11)	-0.03 [0.06, -0.12]	0.570	0.01 [0.08, -0.05]	0.686
	IDN RT	2.78(0.08)	2.77(0.08)	-0.04 [0.13, -0.04]	0.332	-0.06 [0.13, -0.00]	0.055
	IDN Acc	1.43(0.15)	1.40(0.16)	-0.25 [-0.34, -0.16]	<0.0001	-0.14[-0.21, -0.08]	<0.001
	ONB RT	2.96(0.09)	2.96 (0.10)	-0.09[-0.01, -0.18]	0.037	-0.03[-0.04, -0.09]	0.384
	ONB Acc	1.38 (0.16)	1.35(0.17)	-0.23 [-0.32, -0.15]	<0.0001	-0.13[-0.19, -0.06]	<0.001
	OCL Acc	0.97(0.12)	0.95 (0.12)	-0.18[-0.26, -0.09]	<0.001	-0.09[-0.16, -0.03]	0.005

Note. M=mean, SD=standard deviation; PACC=Preclinical Alzheimer's Cognitive Composite; C3= Computerized Cognitive Composite; BPS-O= Behavioral Pattern Separation Task-Object; LDI=Lure Discrimination Index; FNAME=Face-Name Associative Memory Exam; FNLT=1st letter Name Recall; FNMT=Face-Name Matching; FSBT=Facial Recognition; CBB=Cogstate Brief Battery; RT=reaction time; Acc=Accuracy; DET=Detection; IDN=Identification; ONB=One-Back Test; OCL=One-Card Learning.

Table 3. Pearson correlation coefficients (r) Among C3 Components and Demographics

		Age	Education	PACC	MMSE	FCSRT	Logical Memory	DSST
	C3 (LDI, FNMT, OCL)	-0.29	0.12	0.39	0.20	0.27	0.27	0.25
BPS-O	LDI	-0.15	0.05	0.21	0.08	0.14	0.17	0.14
FNAME	FNLT	-0.20	0.02	0.34	0.17	0.29	0.23	0.17
	FNMT	-0.22	0.06	0.28	0.15	0.20	0.2	0.16
	FSBT	-0.22	0.06	0.27	0.13	0.20	0.13	0.22
CBB	DET RT	-0.13	0.07	0.19	0.08	0.08	0.05	0.26
	IDN RT	-0.11	0.04	0.23	0.09	0.13	0.05	0.31
	IDN Acc	-0.17	0.01	0.12	0.07	0.09	0.06	0.09
	ONB RT	-0.16	-0.04	0.29	0.11	0.16	0.09	0.37
	ONB Acc	-0.19	0.06	0.21	0.11	0.14	0.12	0.17
	OCL Acc	-0.16	0.13	0.25	0.14	0.18	0.17	0.15

Note. Higher value represents better performance. PACC=Preclinical Alzheimer Cognitive Composite; C3= Computerized Cognitive Composite; BPS-O= Behavioral Pattern Separation Task-Object; LDI=Lure Discrimination Index; FNAME=Face-Name Associative Memory Exam; FNLT=1st letter Name Recall; FNMT=Face-Name Matching; FSBT=Facial Recognition; CBB=Cogstate Brief Battery; RT=reaction time; Acc=Accuracy; DET=Detection; IDN=Identification; ONB=One-Back Test; OCL=One-Card Learning; FCSRT=Free and Cued Selective Reminding Test; DSST=Digit Symbol Substitution Test

Importantly, A β + continued to perform worse on the C3 compared with A β - and this group difference was at a comparable magnitude as compared with initial testing (re-testing cohen's d=-0.21, p<0.0001).

Individual C3 Components

Individual C3 components which showed statistically significant differences between groups were BPS-O LDI, FNAME FNMT, CBB IDN accuracy, ONB accuracy and RT, and OCL accuracy. When adjusting for demographics, FNAME FNMT and ONB RT were no longer significant. Interestingly, for IDN RT, A β + exhibited a statistical trend towards unexpectedly faster RT compared with A β - (adjusted d=-0.06, p=0.055). Despite a trend towards being slightly faster, A β + were less accurate for IDN compared with A β - (unadjusted d=-0.25, adjusted d=-0.14). IDN Accuracy was correlated

with IDN RT (r= -0.30, p<0.001) such that generally faster RT for correct responses was associated with reduced overall accuracy. However, when both IDN Accuracy and IDN RT were incorporated into the sample model to predict A β status, only reduced IDN Accuracy was a significant predictor (Supplementary Table 2 Model B).

Correlations Among C3 Components, Demographics, PACC

Age

Greater age was associated with worse performance across all C3 outcomes (Table 3). This association was strongest for the overall C3 Composite (r=-0.29, p<0.001). Age was least associated with RT tasks including DET (r=-0.13, p<0.001) and IDN (r=-0.11, p<0.001).

Education

Higher education was associated with better performance on all individual C3 outcomes, with the largest impact on OCL accuracy ($r=0.13$, $p<0.001$) followed by the overall C3 ($r=0.12$, $p<0.001$). The only exception was ONB RT where faster performance was associated with lower education.

Sex

Women outperformed men on all components of FNAME including FNLT ($d=-0.46$, $p<0.0001$), FNMT ($d=-0.36$, $p<0.0001$), and FSBT ($d=-0.39$, $p<0.0001$). Women also outperformed men on IDN Accuracy ($d=-0.16$, $p<0.0001$) and ONB Accuracy ($d=-0.08$, $p=0.019$). Interestingly, however, men outperformed women on DET ($d=-0.23$, $p<0.0001$) and ONB RT ($d=-0.12$, $p<0.001$). Performance between the sexes was comparable for BPS-O, IDN RT, and OCL Accuracy.

On OCL, $A\beta+$ females did not perform differently compared with $A\beta-$ females [Estimate= -0.00 (0.01), $p=0.468$]. However, $A\beta-$ males performed worse compared with $A\beta+$ males [Estimate= -0.02 (0.01), $p=0.0006$]. This suggests that OCL captures subtle decrements in memory between $A\beta+/-$ men but not women. A non-significant statistical trend toward the same pattern was observed in BPS-O.

PACC and C3

Correlations among components of the 2 composites tended to be more strongly-related in a domain-specific manner providing support for convergent and discriminant validity (Table 3). For example, DET and IDN were correlated with DSST at $r=0.26$ and 0.31 , respectively while not being significantly related to memory components of the PACC (FCSRT, Story Memory) or MMSE.

The C3 and APOE Status

There was no difference in performance between $APOE\epsilon4$ carriers vs. non-carriers on the C3 [adjusted $d=-0.03$ (95% CI: -0.09 , 0.03), $p=0.379$] or on individual C3 outcomes (not shown). The model for carrier vs. non-carrier group differences did not improve with the removal of demographic covariates in contrast with models for $A\beta+/-$ [unadjusted $d=0.03$ (95% CI: -0.05 , 0.10), $p=0.470$]. Finally, we did not observe an interaction between $E4$ and $A\beta$ status on the C3.

Discussion

Among a large sample of CN older adults screening for an AD secondary prevention trial, assessment of cognition using a tablet-based measure (C3) was feasible. Diminished C3 performance was associated with worse PACC performance and elevated $A\beta$. Although the magnitude of the $A\beta+/-$ group difference was statistically

small ($d=-0.11$, once adjusted for covariates) it was comparable to that observed on well-established and clinically meaningful paper and pencil measures included in the primary outcome, i.e., the PACC ($d=-0.18$). Performance on the C3 was also reliable, with an equal $A\beta+/-$ group effect on the C3 at retesting within 90 days. More broadly, these findings suggest that computerized testing has the potential to replace traditional paper and pencil primary outcomes in future trials- representing a potential shift in clinical trial cognitive assessment methodology. Additionally, these results further confirm the small but consistent association between $A\beta$ burden and cognition cross-sectionally within a CN population.

Usability/Acceptability of the C3

The very low rates of incomplete and/or invalid administrations for the C3 battery indicate that in the older adults assessed, even those with little computer literacy, the supervised tablet-based cognitive testing has high acceptability. Rates of completion and performance check failures remained low in a subset of low PACC performers, providing early evidence for C3 feasibility longitudinally as some participants show progressive cognitive decrements over the course of the study. Study procedures required a rater to supervise C3 testing, however, raters noted that many participants did not require significant assistance after completing the first few measures. This was further evidence by improved performance on re-testing as participants gained familiarity with the device and tasks. Future trials may consider further optimizing computerized tasks to be self-guided to reduce rater training and time. Potential barriers to tablet-based testing were infrequent, largely addressable, and unlikely to systematically affect performance on the C3. These included inexperience with tablets leading both to mis-tapping and difficulty registering finger taps. Many older adults emphasized accuracy over speed during learning trials, resulting in time-outs. Several of these issues can be addressed with modifications to instructions and design (e.g., including a timer indicator) while others will diminish over time with secular trends toward increased familiarity with digital technology.

The C3 Composite and Individual C3 Measures by $A\beta+/-$

Components of C3 tests which differed between $A\beta+/-$ groups were primarily in memory (BPS-O; OCL) but also included working memory (ONB). The difference in pattern separation memory performance between $A\beta+/-$ participants extends previous fMRI works showing an association between AD biomarkers (including $A\beta$ -PET) and aberrant fMRI activity during learning on a pattern separation task in normal older adults (9) to a difference in frank performance. The BPS-O (10) was designed in part to capture a weakened

“novelty signal”, that is, a reduced ability to correctly discriminate between stimuli that are similar but not identical to previously encountered targets. This tendency to misidentify similar lures as targets has been conceptualized as an error in pattern separation (16). A β group differences were also observed on face-name memory but this effect was significantly attenuated when controlling for demographic features. In contrast with other C3 memory measures (OCL Accuracy and BPS-O) there was a significant sex effect whereby women generally performed better on all aspects of FNAME compared with men. This may be attributable to a general female advantage in verbal memory (17), however, it may be related to the nature of the information. Previous work with FNAME indicates a diminishment of the sex effect when requiring memory for occupation-face versus name-face pairs (5, 18). Our findings from the CBB measures were consistent with previous results examining this battery in relationship to AD neuroimaging markers in normal older adults. Poorer performance on OCL has been associated with higher levels of CSF phosphorylated-tau/A β 42 in late middle-aged participants in the Wisconsin Registry for Alzheimer’s Prevention (4). Similarly, we found that OCL was sensitive. However, we also found that working memory (ONB) was also relatively strongly associated with elevated A β . While C3 constituents were selected theoretically and a-priori, ONB may be considered for inclusion in future optimized and/or data-driven C3 versions. Interestingly, the A β + group made more errors on a Cogstate RT task (IDN) but paradoxically also performed the task more quickly compared with the A β - group. These findings suggest that faster RT may, in fact, be a sign of subtle decrements. One explanation for this finding is an age-associated decrease in inhibition of pre-potent responses (19) may be more pronounced in preclinical AD. More broadly, it confirms that early cognitive changes in preclinical AD extend beyond memory (20, 21).

Part of the impetus for combining outcomes from the BPS-O, FNAME, and CBB into a C3, is aligned with the rationale for cognitive composites as primary endpoints (22) to maximize signal to noise ratio in a population expected to exhibit subtle cognitive decrements. This was confirmed in our data whereby the combination of FNMT, BPS-O, and OCL into the C3 resulted in a numerically larger effect size compared with any single one of these measures alone. However, there are multiple means of constructing composites including data-driven approaches; for example, selecting measures most associated with A β cross-sectionally or measures most sensitive to change. The current C3 was theoretically derived on the basis of previous literature and longitudinal data is needed to confirm its sensitivity over time. Importantly, different memory measures provided related but partially unique information about A β status. For example, both BPS-O and OCL were significant

predictors of A β status when included in the same model (Supplementary Table 2 Model C). More recent work examining the heterogeneity of cognitive decline in early AD suggests that different atrophy patterns are associated with different cognitive trajectories (23). A cognitive composite would thus benefit from being sufficiently broad to avoid under/overestimating decline in a given subgroup.

Our finding that OCL differentiated A β + vs A β - men but not women highlights the issue of heterogeneity in a different light. Males and females performed equivalently for visual memory of playing cards (OCL) but females outperformed males on face-name memory. We hypothesize that visual card-based tasks may be both more engaging and an area of relative strength for males versus females in contrast with name memory (17). Regardless, these findings highlight the rationale for composite scores and the opportunity to use C3 to better understand demographic and individual differences in performance and cognitive trajectories.

C3 Performance and $\epsilon 4$ Status

The lack of a group difference in C3 performance between $\epsilon 4$ carriers vs. non-carriers is not unexpected given the specific recruitment of CN older adults and the current cross-sectional analysis. This is evidenced by the further diminishment of group differences between $\epsilon 4+$ vs. $\epsilon 4-$ participants when including age as a covariate. In contrast, removal of age as a covariate systematically increased the A β + vs. A β - group differences.

C3 and Re-testing

Consistent with the literature, participants performed slightly better on re-testing which is consistent with increased familiarity with the tablet and task demands (3). Diminished practice effects have been shown to predict incident MCI and/or dementia (24, 25) and have been suggested as a screening tool (26). However, we did not observe differential improvement in performance by A β group status. Future adjustments to the FNAME paradigm emphasizing item versus task familiarity may increase the relevance of a diminished practice effect. More specifically, using repeated versus alternate stimuli may capture more AD-specific learning over repeated exposures to the same material (27). C3 practice effects are likely to diminish significantly after the second administration (24). Likewise, item familiarity practice effects are unlikely to contribute to C3 trajectories over time given that all remaining versions are unique.

Conclusions

Within the context of AD secondary prevention trials, our results indicate that computerized (tablet-based) cognitive testing is feasible in older adults in a secondary prevention trial setting and we provide support for the validity of such testing as the C3 was

1) correlated with the primary outcome of paper and pencil composite performance (PACC), 2) related to AD pathological burden ($A\beta+/-$) and 3) related to $A\beta+/-$ at a similar magnitude as the PACC. Positive relationships with AD biomarkers and PACC suggest that the C3 is capturing meaningful cognitive decrements and, has the potential to serve as a proxy for paper and pencil measures in future trials. In addition to reducing staff time and allowing the possibility for remote assessment, computerized testing has the potential to capture a greater quantity and more nuanced quality of data for each measure. Future work will determine the sensitivity of the C3 to change over time in the context of an anti-amyloid treatment trial.

Acknowledgments and funding: The A4 Study is a secondary prevention trial in preclinical Alzheimer's disease, aiming to slow cognitive decline associated with brain amyloid accumulation in clinically normal older individuals. The A4 Study is funded by a public-private-philanthropic partnership, including funding from the National Institutes of Health-National Institute on Aging (U19AG010483; R01AG063689), Eli Lilly and Company, Alzheimer's Association, Accelerating Medicines Partnership, GHR Foundation, an anonymous foundation and additional private donors, with in-kind support from Avid, Cogstate, Albert Einstein College of Medicine, US Against Alzheimer's disease, and Foundation for Neurologic Diseases. The companion observational Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN) Study is funded by the Alzheimer's Association and GHR Foundation. The A4 and LEARN Studies are led by Dr. Reisa Sperling at Brigham and Women's Hospital, Harvard Medical School and Dr. Paul Aisen at the Alzheimer's Therapeutic Research Institute (ATRI), University of Southern California. The A4 and LEARN Studies are coordinated by ATRI at the University of Southern California, and the data are made available through the Laboratory for Neuro Imaging at the University of Southern California. The participants screening for the A4 Study provided permission to share their de-identified data in order to advance the quest to find a successful treatment for Alzheimer's disease. We would like to acknowledge the dedication of all the participants, the site personnel, and all of the partnership team members who continue to make the A4 and LEARN Studies possible. The complete A4 Study Team list is available on: a4study.org/a4-study-team.

Conflicts of interest: K Papp has served as a consultant for Biogen Idec and Digital Cognition Technologies. D Rentz has served as a consultant for Eli Lilly, Biogen Idec, Lundbeck Pharmaceuticals, and serves as a member of the Scientific Advisory Board for Neurotrack. P Maruff is a full-time employee of Cogstate Ltd. C-K. Sun has no disclosures to report. R. Raman has no disclosures to report. M. Donohue has served on scientific advisory boards for Biogen, Eli Lilly, and Neurotrack; and has consulted for Roche. His spouse is a full-time employee of Janssen. A. Schembri is a full-time employee of Cogstate Ltd. C. Stark has no disclosures to report. M Yassa has served as a consultant for Pfizer, Eli Lilly, Lundbeck and Dart Neuroscience and is chief scientific officer of Signa Therapeutics, LLC. A. Wessels is a full-time employee of Eli Lilly and Company. R. Yaari is a full-time employee of Eli Lilly and Company. K. Holdridge is a full-time employee of Eli Lilly and Company. P. Aisen has received research funding from NIA, FNII, the Alzheimer's Association, Janssen, Lilly and Eisai, and personal fees from Merck, Roche, Biogen, ImmunoBrain Checkpoint and Samus. R.A. Sperling has received research funding from NIH, Alzheimer's Association and Eli Lilly for this research. She has served as a consultant for AC Immune, Biogen, Eisai, Janssen, Neurocentria and Roche. Her spouse has served as a consultant to Biogen, Janssen, and Novartis.

Ethical Standards: Study procedures were conducted in accordance with consensus ethics principles derived from international ethics guidelines, including the Declaration of Helsinki and Council for International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines.

Open Access: This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

References

- Sperling, R.A., Rentz, D.M., Johnson, K.A., et al., The A4 study: stopping AD before symptoms begin? *Sci Transl Med*, 2014. 6(228): p. 228f13.
- Sperling, R.A., Donohue, M., Raman, R., et al., Factors associated with elevated amyloid burden in clinically normal older individuals in the A4 Study screening cohort. *JAMA Neurology*, in press.
- Mielke, M.M., Weigand, S.D., Wiste, H.J., et al., Independent comparison of CogState computerized testing and a standard cognitive battery with neuroimaging. *Alzheimers Dement*, 2014. 10(6): p. 779-89.
- Racine, A.M., Clark, L.R., Berman, S.E., et al., Associations between Performance on an Abbreviated CogState Battery, Other Measures of Cognitive Function, and Biomarkers in People at Risk for Alzheimer's Disease. *J Alzheimers Dis*, 2016. 54(4): p. 1395-1408.
- Papp, K.V., Amariglio, R.E., Dekhtyar, M., et al., Development of a psychometrically equivalent short form of the Face-Name Associative Memory Exam for use along the early Alzheimer's disease trajectory. *Clin Neuropsychol*, 2014. 28(5): p. 771-85.
- Rentz, D.M., Locascio, J.J., Becker, J.A., et al., Cognition, reserve, and amyloid deposition in normal aging. *Ann Neurol*, 2010. 67(3): p. 353-64.
- Stark, S.M., Yassa, M.A., Lacy, J.W., and Stark, C.E., A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, 2013. 51(12): p. 2442-9.
- Vannini, P., Hedden, T., Becker, J.A., et al., Age and amyloid-related alterations in default network habituation to stimulus repetition. *Neurobiol Aging*, 2012. 33(7): p. 1237-52.
- Marks, S.M., Lockhart, S.N., Baker, S.L., and Jagust, W.J., Tau and beta-Amyloid Are Associated with Medial Temporal Lobe Structure, Function, and Memory Encoding in Normal Aging. *J Neurosci*, 2017. 37(12): p. 3192-3201.
- Kirwan, C.B. and Stark, C.E., Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. *Learn Mem*, 2007. 14(9): p. 625-33.
- Donohue, M.C., Sperling, R.A., Salmon, D.P., et al., The Preclinical Alzheimer Cognitive Composite: Measuring Amyloid-Related Decline. *JAMA Neurol*, 2014. 71(8): p. 961-970.
- Rentz, D., Dekhtyar, M., Sherman, J., et al., The Feasibility of At-Home iPad Cognitive Testing For Use in Clinical Trials. *J Prev Alzheimers Dis*, 2016. 3(1): p. 8-12.
- Fredrickson, J., Maruff, P., Woodward, M., et al., Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. *Neuroepidemiology*, 2010. 34(2): p. 65-75.
- Maruff, P., Lim, Y.Y., Darby, D., et al., Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC Psychol*, 2013. 1(1): p. 30.
- Sawilowsky, S.S., New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods*, 2009. 8(2): p. 26.
- Yassa, M.A., Lacy, J.W., Stark, S.M., et al., Pattern separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus*, 2011. 21(9): p. 968-79.
- Sundermann, E.E., Biegon, A., Rubin, L.H., et al., Does the Female Advantage in Verbal Memory Contribute to Underestimating Alzheimer's Disease Pathology in Women versus Men? *J Alzheimers Dis*, 2017. 56(3): p. 947-957.
- Buckley, R., Sparks, K., Papp, K., et al., Computerized cognitive testing for use in clinical trials: a comparison of the NIH Toolbox and Cogstate C3 batteries. *The journal of prevention of Alzheimer's disease*, 2017. 4(1): p. 3.
- Butler, K.M. and Zacks, R.T., Age deficits in the control of prepotent responses: evidence for an inhibitory decline. *Psychol Aging*, 2006. 21(3): p. 638-43.
- Petersen, R.C., Department of Neurology, M.C.a.F., Rochester, Minnesota, Department of Health Sciences Research, M.C.a.F., Rochester, Minnesota, et al., Association of Elevated Amyloid Levels With Cognition and Biomarkers in Cognitively Normal People From the Community. *JAMA Neurology*, 2016. 73(1): p. 85-92.
- Baker, J.E., Lim, Y.Y., Pietrzak, R.H., et al., Cognitive impairment and decline in cognitively normal older adults with high amyloid-beta: A meta-analysis. *Alzheimers Dement (Amst)*, 2017. 6: p. 108-121.
- Kozauer, N. and Katz, R., Regulatory innovation and drug development for early-stage Alzheimer's disease. *N Engl J Med*, 2013. 368(13): p. 1169-71.
- Zhang, X., Mormino, E.C., Sun, N., et al., Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proc Natl Acad Sci U S A*, 2016. 113(42): p. E6535-e6544.
- Machulda, M.M., Pankratz, V.S., Christianson, T.J., et al., Practice effects and longitudinal cognitive change in normal aging vs. incident mild cognitive impairment and dementia in the Mayo Clinic Study of Aging. *Clin Neuropsychol*, 2013. 27(8): p. 1247-64.
- Hassenstab, J., Ruvolo, D., Jasielc, M., et al., Absence of practice effects in preclinical Alzheimer's disease. *Neuropsychology*, 2015. 29(6): p. 940-8.
- Duff, K., Beglinger, L.J., Schultz, S.K., et al., Practice effects in the prediction of long-term cognitive outcome in three patient samples: a novel prognostic index. *Arch Clin Neuropsychol*, 2007. 22(1): p. 15-24.
- Pihlajamaki, M., O'Keefe, K., O'Brien, J., Blacker, D., and Sperling, R.A., Failure of repetition suppression and memory encoding in aging and Alzheimer's disease. *Brain Imaging Behav*, 2011. 5(1): p. 36-44.