Computerized Cognitive Testing for Use in Clinical Trials: A Comparison of the NIH Toolbox and Cogstate C3 Batteries

R.F. Buckley[1,2,3,4], K.P. Sparks[1,5], K.V. Papp[1,2,5], M. Dekhtyar[1,5], C. Martin[6], S. Burnham[7], R.A. Sperling[1,2,5], D.M. Rentz[1,2,5]

1. Massachusetts General Hospital, Boston, Massachusetts, USA; 2. Harvard Medical School, Boston, Massachusetts, USA; 3. Florey Institutes of Neuroscience and Mental Health, Melbourne, Australia; 4. Melbourne School of Psychological Sciences, University of Melbourne, Australia; 5. Brigham and Women's Hospital, Boston, Massachusetts, USA; 6. Northeastern University, Boston, Massachusetts, USA; 7. Commonwealth Scientific and Industrial Research Organization, Perth, Australia

Corresponding Author: Dorene M. Rentz, Harvard Medical School, Department of Neurology, Brigham and Women's Hospital, 60 Fenwood Road, Boston, MA 02115, Phone 617-732-8235, Email: drentz@bwh.harvard.edu

Abstract

BACKGROUND: As prevention trials for Alzheimer's disease move into asymptomatic populations, identifying older individuals who manifest the earliest cognitive signs of Alzheimer's disease is critical. Computerized cognitive testing has the potential to replace current gold standard paper and pencil measures and may be a more efficient means of assessing cognition. However, more empirical evidence about the comparability of novel computerized batteries to paper and pencil measures is required.
OBJECTIVES: To determine whether two computerized IPad batteries, the NIH Toolbox Cognition Battery and Cogstate-C3, similarly predict subtle cognitive impairment identified using the Preclinical Alzheimer Cognitive Composite (PACC).
DESIGN, SETTING, PARTICIPANTS: A pilot sample of 50 clinically normal older adults (Mage=68.5 years±7.6, 45% non-Caucasian) completed the PACC assessment, and the NIH Toolbox and Cogstate-C3 at research centers of Massachusetts General and Brigham and Women's Hospitals. Participants made 3-4 in-clinic visits, receiving the PACC first, then the NIH Toolbox, and finally the Cogstate-C3.
MEASUREMENTS: Performance on the PACC was dichotomized by typical performance (>= 0.5SD), versus subtle cognitive impairment (<0.5SD). Composites for each computerized battery were created using principle components analysis, and compared with the PACC using non-parametric Spearman correlations. Logistic regression analyses were used to determine which composite was best able to classify subtle cognitive impairment from typical performance.
RESULTS: The NIH Toolbox formed one composite and exhibited the strongest within-battery alignment, while the Cogstate-C3 formed two distinct composites (Learning-Memory and Processing Speed-Attention). The NIH Toolbox and C3 Learning-Memory composites exhibited positive correlations with the PACC (ρ=0.49, p<0.001; ρ=0.58, p<0.001, respectively), but not the C3 Processing Speed-Attention composite, ρ=-0.18, p=0.22. The C3 Learning-Memory was the only composite that classified subtle cognitive impairment, and demonstrated the greatest sensitivity (62%) and specificity (81%) for that subtle cognitive impairment.
CONCLUSIONS: Preliminary findings suggest that the NIH Toolbox has the advantage of showing the strongest overall clustering and alignment with standardized paper-and-pencil tasks. By contrast, Learning-Memory tasks within the Cogstate-C3 battery have the greatest potential to identify cross-sectional, subtle cognitive impairment as defined by the PACC.

Key words: Cognition, neuropsychology, aging, computerized testing.

Introduction

Interest in using computerized cognitive testing as a potential outcome measure in clinical trials has steadily increased. Computerized testing has been proposed as a feasible and reliable way of testing older participants (1-4). Studies examining the validity of computerized cognitive composites in relation to performance on conventional neuropsychological instruments are accruing (5-8), and furthermore, computerized testing has already become a secondary outcome in a major clinical trial (9). Until recently, however, clinical trials have relied upon conventional paper and pencil neuropsychological tests, as they represent a gold-standard in clinical testing and diagnostic decision-making (for a discussion, see: 10). As technology advances, clinical trials are increasingly moving towards validated computerized testing for sensitively capturing cognitive performance in large-scale secondary prevention cohorts. Comparing computerized batteries against current measures used in large-scale clinical trials is critical as the field moves towards these large-scale, population-based cognitive screening and assessments (11, 12). The Alzheimer's Disease Cooperative Study Preclinical Alzheimer Cognitive Composite (PACC) (9, 13) is a composite of standard paper and pencil tests that are currently being used in a large-scale prevention trial (9). The PACC was originally designed as a multi-domain but memory-predominant cognitive composite that exhibited sensitivity to biomarker risk of AD in clinically-normal older adults (13).

It is unclear how computerized batteries perform in relation to one another against conventional paper-and-

pencil composites, such as the PACC. Secondly, it is not clearly understood how these batteries may compare in their ability to classify subtle cognitive impairment as defined by poor performance on paper-and-pencil composites. Two computerized batteries that are of particular relevance to these questions are the Cogstate Computerized Cognitive Composite (C3) battery (1), which is currently being used in the Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease (A4) secondary prevention trial (9), and the newly developed, non-proprietary iPad version of the National Institutes of Health (NIH) Toolbox Cognition Battery (NIHTB-CB) (for reference to the general computerized battery: 14, 15). The Cogstate C3 departs from the original Cogstate Brief Battery (7) as it includes the Face-Name Associative Memory Exam (FNAME), a challenging associative memory task found to be sensitive to neocortical amyloid burden in older adults (16), and the Behavioral Pattern Separation Task-Object (BPXT) (17), a pattern-separation memory task sensitive to treatment change in an MCI trial (17). The Cogstate Brief Battery is well-validated, and has been shown to capture AD-related cognitive changes in older adults (18), and those with MCI and AD (19). The desktop version of the NIHTB-CB has been validated against standard neuropsychological measures, and in a large and demographically diverse population ranging in age from 3 to 85 years (6, 14). The NIHTB-CB is intended to serve as a 'common currency' among longitudinal and epidemiological studies, however, it is yet to be tested in clinical trials or longitudinal observational studies of aging and dementia. Neither of these batteries are a direct replication or 'digitization' of conventional paper-and-pencil tests, but represent a novel approach to cognitive testing that that can be optimally translated to computerized technologies. As an example, Cogstate utilizes playing cards as a non-verbal assessment of working memory and processing speed that has wide cross-cultural applicability (18).

A critical component of early detection in preclinical Alzheimer's disease is the ability of neuropsychological tests to identify evidence of subtle cognitive impairment (20). Defined as Stage 3, after abnormal levels of both amyloidosis and neurodegeneration are apparent, the appearance of subtle changes in cognitive performance heralds the final phases of preclinical AD prior to a diagnosis of MCI. Targeting clinically-normal older adults at risk of AD-related cognitive decline over short term follow-up will require sophisticated cognitive batteries that are sensitive to subtle change, but will also need to meet the requirements of large-scale clinical trials in clinically-normal cohorts for being deployable across large populations. Before computerized batteries can be utilized in these environments, these batteries must demonstrate validity for identifying preclinical levels of subtle cognitive impairment (21, 22).

The aims of this pilot cross-sectional study were three-fold. First, we developed aggregate cognitive composites for both computerized batteries to measure overall cognitive performance in relation to the paper and pencil PACC. We also aimed to compare each of these computerized batteries against performance on the PACC in clinically normal older adults. Finally, using the PACC to define subtle cognitive impairment, our objective was to determine the ability of each of the computerized batteries to distinguish subtle cognitive impairment from typical cognitive performance. Evidence that these batteries similarly identify subtle cognitive impairment would support the validity of these instruments for large-scale screening and cognitive outcome protocols for clinical trials.

Materials and Methods

Participants

Fifty clinically normal, community-dwelling, older adults (age range: 54-97 years) were recruited from volunteers interested in research studies at the Center for Alzheimer Research and Treatment at Brigham and Women's Hospital and at the Massachusetts Alzheimer Disease Research Center at Massachusetts General Hospital. All subjects underwent informed consent procedures approved by the Partners Human Research Committee, the Institutional Review Board for Brigham and Women's Hospital and Massachusetts General Hospital. No prior computer or iPad knowledge was required. Subjects were excluded if they had a history of alcoholism, drug abuse, head trauma or current serious medical or psychiatric illnesses. All subjects met the age requirement (above 50 years old), and scored within age-specified norms on the Telephone Interview of Cognitive Status (TICS; 23). We set a minimum age of 50 years, as longitudinal research studies and clinical trials are beginning to include younger ages in their cohorts (i.e. the Australian Imaging Biomarker and Lifestyle (AIBL) study of ageing, the Harvard Aging Brain Study (HABS), and the ante-amyloid (A3) clinical trial (11)).

Procedures

In order to mimic a typical clinical trial setting, subjects participated in three-four clinic visits within a six-month time-frame, where they completed the PACC, the NIHTB-CB, and the Cogstate iPad C3 battery at the first, second and third visit, respectively. Each visit was separated from the next by approximately one week. The rationale for multiple clinic visits was to reduce cognitive fatigue when completing each neuropsychological battery. Both computerized batteries were performed from beginning to end in one visit. Participants made a fourth visit as part of a larger study that will not be covered in the current study. Participants were not extensively trained to use the iPad prior to testing, as the tests were overseen by

an examiner according to a standardized administration (CM, KPS, MD). Instructions were given if the participant was having trouble making selections (pressing too hard or too long).

Materials

The PACC includes Logical Memory–delayed recall (LM-DR), the Free and Cued Selective Reminding Test (FCSRT) total score, the Mini Mental State Exam (MMSE) total score, and Wechsler Adult Intelligence Scale-Revised Digit Symbol Coding Test (DSC) total score (13). This composite includes measures of general cognition (MMSE) and speeded executive function (DSC), but is 50% composed of episodic memory tests (13). All tests were z-transformed using the mean and standard deviation of performance by clinically normal older adults (n=256, age range: 61-90) years) participating in the Harvard Aging Brain Study (24, 25). This population served as an ideal normative sample by which to classify our current pilot sample as individuals were recruited from the same geographic area and recruited through the same centers. To form the PACC, all z-transformed variables were averaged together, with a higher score indicating better performance.

The NIHTB-CB included the Flanker Inhibitory Control and Attention Test (Flanker), the Picture Sequence Memory Test (PSMT), the Picture Vocabulary Test (PVT), the Pattern Comparison Processing Speed Test (PCPST) and the Dimensional Change Card Sort Test (DCCS) (14). Two other NIHTB-CB measures, the List Sorting Working Memory test and the Oral Reading Recognition test, were not included in the current study as they required the use of an additional keyboard. The Flanker is a measure of cognitive control, where the participant is asked to attend to a stimulus that is flanked by four identical stimuli that are either positioned congruently or incongruently to the target. The participant must select the direction in which the target stimulus is pointing. The PSMT is a measure of episodic memory in which participants are shown a series of images and asked to re-create the image order over two trials. The PVT is a measure of receptive vocabulary; participants are orally presented a word and are asked to select from one of four images that is closest to the meaning of the word. The outcome measure for PVT was age-scaled and standardized. The PCPST is a measure of processing speed, where participants are asked to match an object with response items by either color or shape. The DCCS is a measure of set shifting, where a participant matches a target visual stimulus to one or two choice stimuli according to shape or color (14). The PSMT, PCPST, DCCS and Flanker tasks were all computed scores provided by NIHTB-CB. These computed scores reflect a theta score, which reflects an individual's overall ability or performance, similar to a z-score.

The Cogstate C3 includes the FNAME and the Behavioral Pattern Separation-Object Task (BPXT), as well as the Detection Task (DET), the Identification Task (IDN), the One Card Learning Task (OCL) and the One-Back Task (ONB). The FNAME is an associative memory test that requires participants to associate (FNMT), and subsequently recall (FNLT), and recognize (FSBT) faces with corresponding names. The FNAME task measures frequency of correct responses. The BPXT assesses working and recognition memory; participants are iteratively presented with a series of repeated, novel and distractor images and are asked to categorize each into Old, Similar, or New. The outcome measure is frequency of correct responses. Additional tasks use playing cards as stimuli. The DET is a measure of reaction time and processing speed, where participants are asked to respond when a stimulus card is turned face up. The IDN is an attention paradigm in which a card is presented and the respondent must choose whether the card is red or is not red (black). The outcome measures for these two tasks were speed (sec:ms). The OCL task is a non-verbal memory task, which assesses short term recall of a set of repeated playing cards. OCL was measured using accuracy. The ONB task is a measure of working memory, where respondents are asked to serially match each card to the previous trial, and was also measured according to speed of response (18). These scores were not transformed, however, they were converted to z-scores.

Creating computerized battery composites

Our initial aim was to create cognitive composites from the computerized batteries in order to align with the PACC. Previous studies have created cognitive composites from the Cogstate Brief battery in older adults who were clinically-normal and patients with MCI and AD, so we investigated whether the Cogstate C3 could create similar composites. NIHTB-CB Crystallized Cognition Composites and Fluid Cognition Composites have been proposed in a previous study, however, these were created from a sample of children and required two extra tests that we did not include in our study. Global composite measures were created for each of the NIHTB-CB and Cogstate C3 batteries using principal component analysis (PCA), and Bartlett factor scores were extracted. These composites were created consistent with previous reports using the Cogstate Brief Battery (19) and the NIHTB-CB (15). PCA was used to reduce the NIHTB-CB and C3 into global composite scores. Using scree plots and eigenvalue cut-offs, we determined that the NIHTB-CB could be reduced to one composite, while the C3 exhibited a better fit with two composites. The NIHTB-CB composite accounted for 47% of the variance explained in the model, while the two C3 composites accounted for a total of 61% of the variance (with the first factor accounting for 32% variance). The first C3 factor, 'Learning-Memory', included the BPXT, FNMT, FNLT,

FSBT and OCL. The second factor, 'Processing Speed-Attention', included the ONB, IDN and DET. A clustering model using a two-dimensional PCA, which compared the similarities of the tasks in two-dimensional space was also used to explore how tasks clustered together. The results, displayed in Figure 1, suggested that IDN, DET and ONB created a distinct cluster, while the remainder of the NIHTB-CB and C3 tasks formed a second, tight cluster (see Figure 1). Using composite scores arising from this data reduction approach, allowed us to pursue our main hypotheses, i.e., using the composite scores to differentiate high and low performance in relation to the PACC.

Figure 1. Visualization of clusters using PCA of NIHTB-CB with C3 tasks. Arrows indicate the loading coefficients of each variable of interest
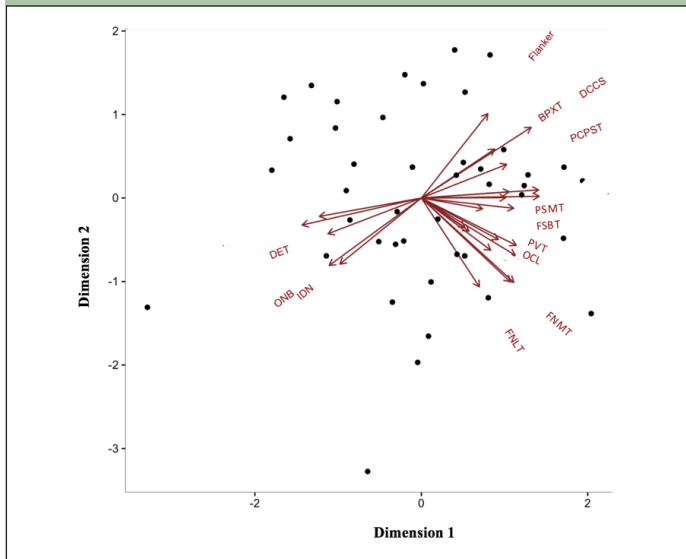
Statistical methodology

Analyses were conducted using IBM SPSS version 22.0, and R (version 3.3.0). Due to a smaller sample size, a series of non-parametric Spearman correlations were conducted to ascertain relationships between computerized battery composites and PACC performance. Performance on the PACC was dichotomized into normal and subtle cognitive impairment according to a cut-off of 0.5SD below the normative group mean, which was derived according to the Harvard Aging Brain Study cohort (an entirely separate cohort from the one used in the current study). These participants were not considered to meet criteria for a diagnostic classification of mild cognitive impairment, but demonstrated a very subtle cognitive decrement in comparison with their peers. This classification was chosen to align with Stage 3 Preclinical AD criteria, which states: "Evidence of subtle cognitive decline, that does not meet criteria for MCI or dementia" (20). Choosing a 0.5SD cut-off allowed us to define subtly poorer performance, while maintaining samples

with enough power for analytical purposes. This sits in contrast to diagnostic classifications for MCI which typically note performance below 1-1.5SD age-adjusted norms (26). Three logistic regression analyses were performed to determine how well the NIHTB-CB and C3 composites could detect the subtle cognitive impairment group. Although age and education levels were not found to be significantly different between typical PACC performers and those with subtle cognitive impairment, we ran analyses with these covariates included in order to portray our results within the context of age and education-adjustment. Receiver Operating Characteristic (ROC) curves were computed to determine sensitivity, specificity and area under the curve (AUC) parameters of each composite to classify normal and subtle cognitive impairment. Post-hoc analyses were run to ascertain which tests within the best performing composites were driving better classification outcomes.

Results

Participant Characteristics

No demographic differences were found between those who were classified as subtle cognitive impairment or exhibiting typical performance according to the PACC (see Table 1). However, a marginally greater number of non-Caucasian individuals (n = 9) were found to be classified as subtly impaired on the PACC in comparison with Caucasians (n = 4), $\chi 2(1) = 3.91$, p = .10, however this difference was not statistically significant. The non-Caucasian group was not found to be significantly different from the Caucasian group on any demographics, although, there was a trend for lower education levels (ranging from 12-20 years), $\chi 2(4) = 8.13$, p = .09.

Associations between computerized batteries and PACC performance

The NIHTB-CB and C3 Learning-Memory were both associated with the PACC (ρNIHTB-CB(47) = 0.49, p < .001 and ρC3 Learning-Memory(47) = 0.58, p <.001). There was no significant relationship found between the PACC and C3 Processing Speed-Attention, $\rho(47) = -0.18$, p =.22.

Ability of computerized tasks to distinguish subtle cognitive impairment according to the PACC

Logistic regression analyses showed that the NIHTB-CB and Cogstate C3 Learning-Memory models were significantly able to distinguish subtle cognitive impairment from typical PACC performance, and

Table 1. Participant characteristics and cognitive performance

	Full group n = 50	Subtle cognitive impairment n = 13	Typical performance n = 37	p
Sex (% F)	62	46	67	.30
Age (yrs)	68.5 (7.5)	70.2 (9.1)	68.0 (7.0)	.46
Education (yrs)	15.6 (3.1)	14.8 (3.1)	15.9 (2.7)	.28
Race (% Black)	46	69	38	.10
Ethnicity (% Hispanic)	4	15	0	-*
PACC	-0.24 (0.5)	-0.96 (0.4)	0.02 (0.3)	<.001
MMSE	29 (1.2)	27.6 (1.6)	29.2 (0.7)	.002
Logical memory delayed recall	11.89 (4.1)	7.54 (3.5)	13.41 (3.2)	<.001
FCSRT total recall	47.86 (0.4)	47.62 (0.7)	47.95 (0.2)	.09
Digit Symbol Coding	47.50 (11.4)	36.61 (9.4)	51.32 (9.6)	<.001
Cogstate C3 Learning-Memory	0.04 (1.0)	-0.94 (0.8)	0.39 (0.8)	<.001
BPXT	-0.01 (1.0)	-0.46 (1.2)	0.15 (0.9)	.13
FNMT	0.03 (1.0)	-0.75 (0.9)	0.31 (0.9)	.004
FNLT	0.02 (1.0)	-0.40 (0.2)	0.16 (1.1)	.006
FSBT	0.04 (0.9)	-0.73 (1.1)	0.31 (0.7)	.008
OCL	0.001 (1.0)	-0.58 (1.1)	0.20 (0.9)	.04
Cogstate C3 Speed-Attention	0.02 (0.9)	0.15 (1.3)	-0.03 (0.9)	.65
IDN	0.04 (1.0)	0.04 (1.3)	0.04 (0.9)	.99
DET	-0.02 (1.0)	0.20 (1.3)	-0.10 (0.9)	.45
ONB	0.02 (1.0)	0.20 (1.3)	-0.04 (0.9)	.49
NIHTB-CB	-0.01 (1.0)	-0.53 (0.9)	0.16 (0.9)	.04
PVT	0.00 (1.0)	-0.63 (1.0)	0.22 (1.0)	.02
PSMT	0.03 (1.0)	-0.51 (0.7)	0.23 (1.0)	.01
DCCS	-0.03 (1.0)	-0.32 (1.0)	0.07 (0.9)	.25
Flanker	-0.01 (1.0)	-0.10 (1.1)	0.02 (1.0)	.74

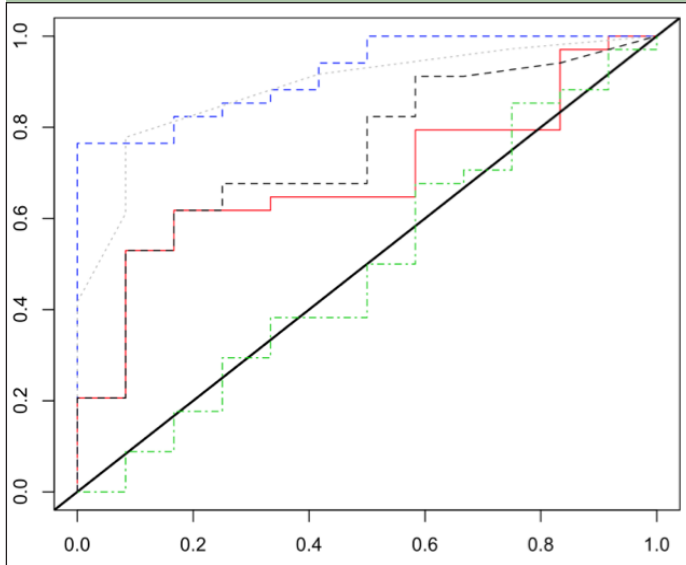Note: Subtle cognitive impairment is PACC performance below 0.5SD *Cell sizes are too small to count

Table 2. Regression and ROC analyses with each computerized composite to predict subtle cognitive impairment or typical performance on the PACC

	Regression			ROC analysis		
	AIC	Model χ2 (p)	OR (p) [CI95%]	Sensitivity	Specificity	AUC
NIHTB-CB	56.2	48.2 (.04)	2.2 (.05) [1.1, 5.4]	0.55	0.64	0.69
NIH PSMT	54.2	46.2 (.04)	3.3 (.04) [1.3, 12.5]	0.50	0.74	.74
C3 Learning-Memory	34.9	26.9 (<.001)	40.1 (.003) [5.8, 844.5]	0.61	0.80	0.92
C3 Proc Speed-Att	58.3	50.3 (.65)	0.9 (.66) [0.4, 1.8]	0.50	0.50	0.49
C3 FNLT	39.4	31.4 (.004)	19989.0 (.003) [85, 9.1e+07]	0.45	0.81	88.7

Note: The large confidence intervals in this analysis are driven largely by the sample size, and so the OR should be interpreted with caution

explained 9% and 49% of variance in their respective models ($\chi 2$NIHTB-CB(42) = 48.22, p = .04 and $\chi 2$COGSTATE-C3(42) = 23.61, p < .001; see Table 2 for all three model fits and estimates). Greater NIHTB-CB performance related to better classification of those with subtle cognitive impairment, however, this finding did not survive multiple comparisons (B (SE) = 0.79 (0.4), p = .05). Better Learning-Memory performance significantly increased the chance of being classified with typical (i.e., better) PACC performance (B (SE) = 3.71 (1.2), p = .003). Our findings showed the same pattern of results with or without age and education included in the models (see Table 2).

Figure 2. ROCs for the NIHTB-CB and Cogstate C3 composites, and the C3 FNLT task alone to distinguish between high and low PACC performance (Blue = C3 Learning-Memory, Red = NIHTB-CB, Green = C3 Processing Speed-Attention, Grey = C3 FNLT, Black-dash = NIH PSMT)


ROC curves showed that performance on the C3 Learning-Memory composite accounted for the largest AUC (92%), and exhibited the greatest sensitivity (61%) and specificity (80%) indices for classifying subtle cognitive impairment (see Table 2 for all sensitivity and specificity parameters). Figure 3 depicts a scatterplot between the NIHTB-CB and Cogstate C3 (by averaging performance on both C3 composites) according to the PACC groups. Scores sitting in the top right-hand quadrant depict high performance on both computerized batteries; all but one of these scores included individuals with typical PACC performance, illustrating high specificity.

As the C3 Learning-Memory composite exhibited the highest odds ratio and ROC parameters, we ran a post-hoc logistic regression to determine which measures within the Learning-Memory composite (FNMT, FNLT, FSBT, BPXT and OCL) were driving these results. Better performance on the Face Name Letter Task, a measure

of delayed free recall, was the only measure within the Learning-Memory composite found to significantly increase the likelihood of typical PACC performance (B (SE) = 5.6 (3.1), p = .05). As a comparison, we also conducted a post-hoc analysis with the NIHTB PSMT task, a free recall memory task, and found that better PSMT performance significantly predicted typical PACC performance, (OR = 3.3, p = .04, CI95%: 1.3-12.5). Neither the FNLT task nor the NIHTB PSMT task were better able to classify subtle cognitive impairment in comparison with the full composite measures, with AUC, sensitivity and specificity parameters comparable to their counterparts (see Table 2 and Figure 2).

Figure 3. Scatterplot of association between NIHTB-CB and Cogstate C3 battery, with slopes estimating group effect of high and low PACC performance
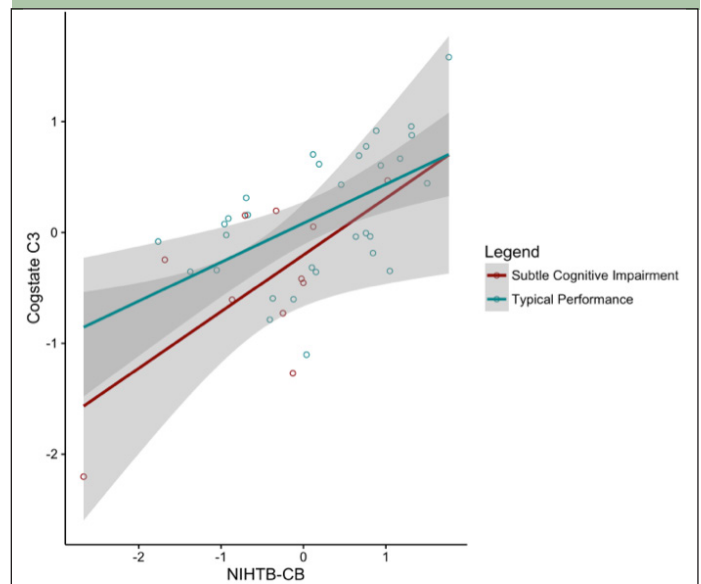

Discussion

This pilot study in normal older adults sought to directly compare performance on computerized batteries, the NIHTB-CB and Cogstate C3 batteries, to the PACC, a clinical trial outcome measure composed of conventional paper and pencil cognitive tasks. The Learning-Memory composite from the Cogstate C3 battery was able to distinguish between normal PACC performance and subtle cognitive impairment (see Fig 4 for a diagrammatic representation of findings). The composite also showed particularly high specificity and AUCs for correctly classifying normal individuals. These findings were found to be primarily driven by the delayed free recall index from the Face-Name task that was featured within the composite. By contrast, the NIHTB-CB yielded a moderate level of specificity, with a sensitivity at chance level, while the C3 Processing Speed-Attention composite was poor on both parameters. We did find, however, that the NIHTB-CB showed a comparable level of correlation

with the PACC as was found with C3 Learning-Memory. By contrast, the C3 Processing Speed-Attention composite did not show any affinity with the PACC. This supports other findings suggesting that processing speed and attention domains are less sensitive to AD-related change very early in the trajectory (18), and perhaps are more sensitive to age-related etiologies (27). These results most likely reflect the nuanced differences in 'intended purpose' for the NIHTB-CB and Cogstate (C3) batteries. The NIHTB-CB has been proposed as a well-validated measure that can be utilized in a broad range of age-groups and education levels (14), while the Cogstate C3 is a battery primarily intended for clinical trials, and which has been shown to be sensitive to AD-related cognitive change (28).

Figure 4. Diagrammatic representation of each composite arising from the Cogstate C3 and NIHTH-CB computerized batteries, and their corresponding tests. Each composite is also attached to an odds ratio (OR) which represents the ability of each composite to distinguish between typical and subtly impaired PACC performance. The pink boxes denote the tasks that were most contributory to the variance explained in the logistic regression model
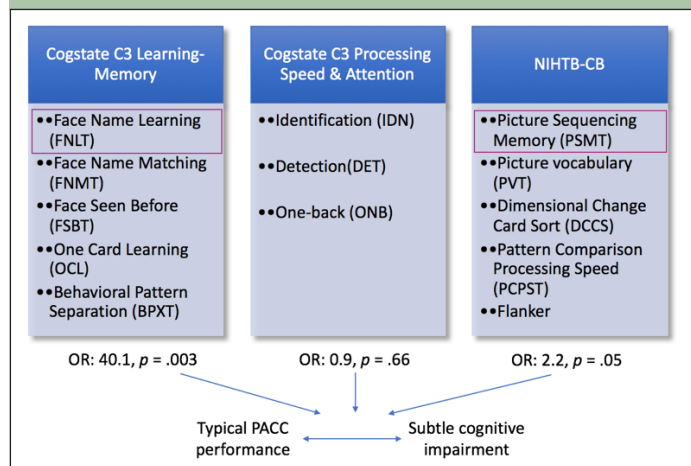
One strength of the NIHTB-CB in this study was that it formed a clear singular composite, and displayed largely unified within-battery alignment as suggested by clustering methods. The NIHTB-CB has shown strong convergent validity with other standard neuropsychological paper-and-pencil tests along the broad developmental trajectory (14), and was originally designed to complement measures used in research studies of cognition or to serve as a brief adjunct measure in longitudinal and epidemiologic studies (14, 29). It was not, however, specifically developed as an early diagnostic tool for AD-related cognitive impairment or as a target for disease outcomes. The NIHTB composite was able to identify subtle cognitive impairment, particularly using the NIHTB memory task. This supports the notion that the NIHTB-CB is a suitable measure of cognitive performance in clinically-normal older adults. Sensitivity for classification of subtle cognitive impairment was not as high in comparison with the Cogstate C3 Learning-Memory composite. An additional advantage of the NIHTB-CB battery is that it includes a measure of IQ, which is not covered by the C3. As such, this battery has the unique potential to efficiently measure cognitive reserve outcomes, and may well have the ability to inform an individuals' likely compensatory duration for increasing pathology over time. Our findings highlight the different possible utilities of these computerized batteries within the context of secondary prevention clinical trials. It is possible that the NIHTB-CB will be more sensitive to early longitudinal cognitive decline, however, the current pilot study is unable to investigate this question.

Within the Cogstate C3 battery, two distinct composites were extracted, similar to previous studies (18, 19, 30), supporting the notion that the Cogstate Battery was intended to measure distinct cognitive domains. The C3 Learning-Memory composite, however, showed an association with PACC performance, and an ability to classify subtle cognitive impairment. The Cogstate Brief battery has been shown to reliably highlight increasing magnitude of impairment in MCI and AD diagnostic groups, and that computerized performance tracks well with performance on conventional tests (18, 28). Our findings suggest that the FNAME component of the Cogstate C3 battery may be of particular interest for clinical trials of preclinical AD. Although evidence of subtle cognitive impairment was defined in our study, it is not solely an indication of stage 3 preclinical AD as we do not have indications of AD biomarker status. Furthermore, exhibiting subtle cognitive impairment does not by itself indicate progressive cognitive decline. As such, sensitivity to the classification of subtle cognitive impairment will need to be more fully determined by larger, longitudinal investigations. In addition, validation studies will be required in comparison populations of MCI and AD dementia. It may be that the ADAS-Cog and screening tools such as the MMSE are sufficient for clinical populations, but that more challenging neuropsychological tasks included in computerized batteries are more relevant for large-scale clinical trials of clinically-normal individuals. Our findings further suggest that not all C3 tasks have the ability to identify subtle cognitive decline, and as such, may not be necessary for inclusion in large-scale screening procedures for preclinical AD trials.

We found that the driving predictor of sensitivity to subtle cognitive impairment in the current study was the delayed free recall index from the C3 FNAME task. The Cogstate C3 departs from the Cogstate Brief Battery in that it includes the FNAME (1), which has been shown to be sensitive to amyloid-ß deposition (12). The addition of the FNAME measures in the C3 battery may have

increased the ability of the C3 to capture variation in PACC performance, which is is the current standard for clinical trials (1). As the PACC is a composite that is more heavily weighted towards memory (by including two memory measures), and is honed to detect amyloid-related change (13), it is not surprising that memory components of the C3 battery are able to classify subtle impairment on the conventional composite. In the current study, delayed free recall from the FNLT was found to drive the group classification, which provides support for the recommendation that the FNAME be included in the Cogstate Brief Battery for longitudinal studies of memory in preclinical AD. Although it was a significant component of the composite to classify group performance, neither the C3 FNLT task nor the NIHTB PSMT task performed significantly better than their composite counterparts. While parsimonious neuropsychological batteries are advantageous, we currently recommend that full Cogstate Learning-Memory or NIHTB-CB batteries are performed.

The current study is a pilot study of clinically normal older adults, and as such we were limited to studying the classification of subtle cognitive impairment as defined by the PACC. Although, the sample size is small, the strength of this study is that it covers a broad range of older ages and maximizes the racial diversity of subjects. As no major demographic differences were present in typical and subtle cognitive impairment PACC performers, we did not covary for race in our analyses, although we acknowledge that more sophisticated examinations of diversity-related cognitive profiles should be conducted in larger samples (31). In addition, we did not acquire AD biomarkers, and cannot conclude on the extent to which these tests measure biological markers of interest. In the future, we plan to include the NIHTB-CB and C3 in a larger cohort of clinically normal older adults who have undergone AD biomarkers and intend to follow the performance of these individuals over time. In addition, it will be important to counterbalance for battery administration, and assess in-home compared to in-clinic testing performance. The trend is moving towards large-scale online cognitive testing, as evidenced by registries that include online testing such as the Brain Health Registry (32) and the UKBioBank (33). Determining test-retest reliability between at-home and in-clinic testing will be vital. Large secondary prevention trials that require access to trial-ready cohorts who are identified based on cognitive performance are needed. Computerized on-line testing, that is well validated, will make this feasible. We believe that both iPad batteries presented in this study, show promise as valid cognitive assessments in the clinical trial setting. However, more work will be needed before they can be effectively utilized as on-line cognitive tests for large-scale prevention trials.

References

1. Rentz DM, Dekhtyar M, Sherman J, Burnham S, Blacker D, Aghjayan SL, Papp KV, Amariglio RE, Schembri A, Chenhall T. The Feasibility of At-Home iPad Cognitive Testing For Use in Clinical Trials. JPAD 2016;3, 8-12.
2. Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: a systematic review. Alzheimer's & Dementia 2008;4, 428-437.
3. Sano M, Egelko S, Ferris S, Kaye J, Hayes TL, Mundt JC, Donohue M, Walter S, Sun S, Sauceda-Cerda L. Pilot Study to Show the Feasibility of a Multicenter Trial of Home-based Assessment of People Over 75 Years Old. Alzheimer disease and associated disorders 2010;24, 256-263.
4. Fredrickson J, Maruff P, Woodward M, Moore L, Fredrickson A, Sach J, Darby D. Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. Neuroepidemiology 2009;34, 65-75.
5. Steinberg SI, Negash S, Sammel MD, Bogner H, Harel BT, Livney MG, McCoubrey H, Wolk DA, Kling MA, Arnold SE. Subjective Memory Complaints, Cognitive Performance, and Psychological Factors in Healthy Older Adults. American Journal of Alzheimer's Disease & Other Dementias 2013;28, 776-783.
6. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Slotkin J, Carlozzi NE, Bauer PJ, Wallner-Allen K, Fox N. The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: Validation in an adult sample. JINS 2014;20, 567-578.
7. Maruff P, Thomas E, Cysique L, Brew B, Collie A, Snyder P, Pietrzak RH. Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. Archives of Clinical Neuropsychology 2009;24, 165-178.
8. Hammers D, Spurgeon E, Ryan K, Persad C, Barbas N, Heidebrink J, Darby D, Giordani B. Validity of a brief computerized cognitive screening test in dementia. Journal of geriatric psychiatry and neurology 2012;25, 89-99.
9. Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, Aisen P. The A4 Study: Stopping AD Before Symptoms Begin? Science Translational Medicine 2014;6, 228fs213.
10. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized Neuropsychological Assessment Devices: Joint Position Paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology(). Archives of Clinical Neuropsychology 2012;27, 362-373.
11. Sperling RA, Mormino E, Johnson K. The Evolution of Preclinical Alzheimer's Disease: Implications for Prevention Trials. Neuron 2014;84, 608-622.
12. Rentz DM, Parra Rodriguez MA, Amariglio R, Stern Y, Sperling RA, Ferris S. Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. Alzheimer's Research & Therapy 2013;5, 58.
13. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical alzheimer cognitive composite: Measuring amyloid-related decline. JAMA Neurology 2014;71, 961-970.
14. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, Carlozzi NE, Slotkin J, Blitz D, Wallner-Allen K. Cognition assessment using the NIH Toolbox. Neurology 2013;80, S54-S64.
15. Akshoomoff N, Beaumont JL, Bauer PJ, Dikmen SS, Gershon RC, Mungas D, Slotkin J, Tulsky D, Weintraub S, Zelazo PD, Heaton RK. NIH Toolbox cognition battery (CB): Composite scores of crystallized, fluid, and overall cognition. Monographs of the Society for Research in Child Development 2013;78, 119-132.
16. Rentz DM, Amariglio RE, Becker JA, Frey M, Olson LE, Frishe K, Carmasin J, Maye JE, Johnson KA, Sperling RA. Face-name associative memory performance is related to amyloid burden in normal elderly. Neuropsychologia 2011;49, 2776-2783.
17. Stark SM, Yassa MA, Lacy JW, Stark CE. A task to assess behavioral pattern

separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. Neuropsychologia 2013;51, 2442-2449.

18. Lim YY, Ellis KA, Harrington K, Ames D, Martins RN, Masters CL, Rowe C, Savage G, Szoeke C, Darby D. Use of the CogState Brief Battery in the assessment of Alzheimer's disease related cognitive impairment in the Australian Imaging, Biomarkers and Lifestyle (AIBL) study. Journal of Clinical and Experimental Neuropsychology 2012;34, 345-358.

19. Maruff P, Lim YY, Darby D, Ellis KA, Pietrzak RH, Snyder PJ, Bush AI, Szoeke C, Schembri A, Ames D. Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. BMC Psychology 2013;1, 30.

20. Sperling RA, Beckett L, Bennett D, Craft S, Fagan A, Kaye J, Montine T, Park D, Reiman E, Siemers E, Stern Y, Yaffe K. Criteria for preclinical Alzheimer's disease. Alzheimer's Association, 2010.

21. Carrillo MC, Vellas B. New and different approaches needed for the design and execution of Alzheimer's clinical trials. Alzheimer's & dementia: the journal of the Alzheimer's Association 2013;9, 436-437.

22. Reiman EM, Langbaum J, Fleisher AS, Caselli RJ, Chen K, Ayutyanont N, Quiroz YT, Kosik KS, Lopera F, Tariot PN. Alzheimer's Prevention Initiative: a plan to accelerate the evaluation of presymptomatic treatments. Journal of Alzheimer's Disease 2011;26, 321-329.

23. Knopman DS, Roberts RO, Geda YE, Pankratz VS, Christianson TJ, Petersen RC, Rocca WA. Validation of the telephone interview for cognitive status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. Neuroepidemiology 2010;34, 34-42.

24. Dagley A, LaPoint M, Huijbers W, Hedden T, McLaren DG, Chatwal JP, Papp KV, Amariglio RE, Blacker D, Rentz DM. Harvard aging brain study: Dataset and accessibility. In press, 2015;S1053-8119.

25. Hedden T, Mormino EC, Amariglio RE, Younger AP, Schultz AP, Becker JA, Buckner RL, Johnson KA, Sperling RA, Rentz DM. Cognitive profile of amyloid burden and white matter hyperintensities in cognitively normal older adults. The Journal of Neuroscience 2012;32, 16233-16242.

26. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia 2011;7, 270-279.

27. Hedden T, Gabrieli JDE. Insights into the ageing mind: a view from cognitive neuroscience. Nat Rev Neurosci 2004;5, 87-96.

28. Lim YY, Ellis KA, Harrington K, Pietrzak RH, Gale J, Ames D, Bush AI, Darby D, Martins RN, Masters CL, Rowe CC, Savage G, Szoeke C, Villemagne VL, Maruff P. Cognitive Decline in Adults with Amnestic Mild Cognitive Impairment and High Amyloid-β: Prodromal Alzheimer's Disease? J Alz Dis 2013;33, 1167-1176.

29. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH toolbox for assessment of neurological and behavioral function. Neurol 2013;80, S2-S6.

30. Lim YY, Ellis KA, Ames D, Darby D, Harrington K, Martins RN, Masters CL, Rowe CC, Savage G, Szoeke C, Villemagne VL, Maruff P. Aβ amyloid, cognition, and APOE genotype in healthy older adults. Alzheimer Dem 2013;9, 538-545.

31. Manly JJ, Schupf N, Tang M-X, Stern Y. Cognitive decline and literacy among ethnically diverse elders. Journal of Geriatric Psychiatry and Neurology 2005;18, 213-217.

32. Nosheny RL, Flennkiken D, Insel PS, Finley S, Mackin S, Camacho M, Truran-Sacrey D, Maruff P, Weiner MW (2015) Internet-based recruitment of subjects for prodromal and secondary prevention Alzheimer's disease trials using the brain health registry. Alzheimer's & Dementia: The Journal of the Alzheimer's Association 2015;11, P156.

33. Matthews PM, Sudlow C. The UK Biobank. Brain 2015;138, 3463-3465.