COGNITIVE & BEHAVIORAL ASSESSMENT

Associations between NIH Toolbox Cognition Battery and *in vivo* brain amyloid and tau pathology in non-demented older adults

Beth E. Snitz[1] | Dana L. Tudorascu[2] | Zheming Yu[3] | Elizabeth Campbell[3] | Brian J. Lopresti[3] | Charles M. Laymon[3,4] | Davneet S. Minhas[3] | Neelesh K. Nadkarni[2] | Howard J. Aizenstein[5,4] | William E. Klunk[5] | Sandra Weintraub[6] | Richard C. Gershon[7] | Ann D. Cohen[5]

[1]Department of Neurology, University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania

[2]Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

[3]Department of Radiology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

[4]Department of Bioengineering, University of Pittsburgh School of Engineering, Pittsburgh, Pennsylvania

[5]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania

[6]Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois

[7]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois

Correspondence
Ann D. Cohen, University of Pittsburgh, 3501 Forbes Avenue, Suite 830, Pittsburgh, PA. Email: cohenad@upmc.edu

Abstract

Introduction: The National Institutes of Health (NIH) Toolbox Cognition Battery (NIHTB-CB) was developed to be a common assessment metric across a broad array of research studies. We investigated associations between NIHTB-CB and brain amyloid and tau deposition in cognitively unimpaired older adults.

Methods: One hundred eighteen community-based volunteers completed magnetic resonance imaging (MRI), Pittsburgh compound B (PiB)-PET (positron emission tomography) and AV-1451-PET neuroimaging, a neuropsychological evaluation, NIHTB-CB, and the Clinical Dementia Rating (CDR) scale. Demographically adjusted regression models evaluated cognition–biomarker associations; standardized effect sizes allowed comparison of association strength across measures.

Results: No NIHTB-CB measures were associated with amyloid deposition. NIHTB-CB measures of fluid cognition, including Pattern Comparison Processing Speed, Dimensional Change Card Sort, and Fluid Cognition Composite, were associated with tau deposition in higher Braak regions. Pattern Comparison Processing Speed was the most robust association with sensitivity analyses.

Discussion: NIHTB-CB tasks of processing speed and executive functions may be sensitive to pathologic tau deposition on imaging in normal aging.

KEYWORDS
Alzheimer's disease, biomarkers, cognition, neuropsychology

The National Institutes of Health (NIH) Toolbox (NIHTB) was developed to assess neurologic and behavioral functions, in the domains of cognition, sensation, movement, and emotion, providing an available "common metric" for use across a broad array of research studies.[1] As part of the NIH Blueprint for Neuroscience Research initiative, its development was commissioned by 16 NIH Institutes to provide brief,

efficient, psychometrically sound, and accessible assessment measures for research use. These goals included the use of nonproprietary instruments, availability of both English and Spanish administration, and suitability for measuring constructs across the lifespan (ages 3 to 85).[1,2] Since its release in 2012, and migration to an iPad app–administered format in 2016, adoption into research communities has increased. There are currently 206 studies on ClinicalTrials.gov using NIHTB, more than 260 peer-reviewed publications, and active translation efforts in multiple languages. Of note: There is now an annual fee for the app subscription of approximately $500.00.

The National Institutes of Health (NIH) Toolbox Cognition Battery (NIHTB-CB) comprises seven primary tasks tapping sub-domains of attention/executive functions, language, processing speed, working memory, and episodic memory. Composite scores are computed for (1) crystallized abilities, reflecting tasks of semantic knowledge, such as vocabulary and word reading, thought to "hold" as relatively stable abilities in aging and disease; and (2) fluid abilities, reflecting tasks of novel problem solving, reasoning, processing speed, and memory, processes more vulnerable to decline in aging and disease. In addition, a total cognition composite is calculated. Previous work has shown good reliability for tasks[3] and composites scores,[4] with intraclass correlations for test-retest reliability ranging from .73 to .90. Construct validity for each task measure as well as composite scores has been evaluated and reported as metrics of convergent and divergent validity. Values for convergent validity range from .50 to .92, similar to construct validity for gold standard traditional neuropsychological (NP) measures.[2,5–8] The NIHTB-CB is used increasingly in a multitude of studies of behavioral and neurologic conditions across settings and ages.[9–11] This includes several studies of aging and in the Alzheimer's disease (AD) spectrum of risk,[12,13] although the number is limited. There are especially limited studies, to date, of utility and validity of the NIHTB-CB in pre-symptomatic and early symptomatic AD, or in association with AD biomarkers.[14] This is the case despite recognized need for novel sensitive cognitive measures in this disease stage[15,16] when AD pathologic change is present, including amyloid beta (Aβ) and possibly tau deposition.[17]

The goal of the present study was to begin to address the literature gap by investigating associations between current NIHTB-CB measures and AD neuroimaging biomarkers of Aβ and tau in older adults without clinically significant cognitive impairment. In particular, the need to identify cognitive measures with improved sensitivity to pre-clinical AD for use in early screening and prevention trials has been discussed widely in the literature.[18,19] In this study, we also compared the pattern and strength of AD biomarker associations with NIHTB-CB to those with traditional pencil-and-paper NP tests.

1 | METHODS

1.1 | Participants

Three ongoing neuroimaging studies contributed data to these analyses. The Monongahela-Youghiogheny Healthy Aging Team-

---

RESEARCH IN CONTEXT

1. Systematic review: The authors reviewed the literature using traditional (eg, PubMed, Google Scholar) sources. Although the National Institutes of Health (NIH) Toolbox Cognition Battery (NIHTB-CB) is used increasingly across different behavioral and neurologic conditions, there were no identified studies of associations between NIHTB and AD biomarkers amyloid beta (Aβ) and tau deposition in cognitively normal older adults. Several studies have investigated other biomarkers (hippocampal volume) and discrimination among clinical groups. These relevant studies are appropriately cited.

2. Interpretation of results: Our exploratory findings suggest that several measures of fluid cognition are associated with tau deposition in cortical regions in cognitively normal older adults. Pattern Comparison Processing Speed was the most robust association.

3. Future directions: This initial report sets a benchmark for further confirmatory studies with other cohorts. In particular, larger cohorts may have increased power to detect very small effects with Aβ. Furthermore, investigating the utility of longitudinal change in NIHTB-CB measures relative to change in biomarkers and to clinical progression is an important future direction.

Neuroimaging (MYHAT-NI) study enrolls participants from the population-based parent MYHAT cohort study of risk factors for mild cognitive impairment in small towns southeast of Pittsburgh, PA.[20] The other two neuroimaging studies recruit from the Heart Strategies Concentrating on Risk Evaluation (Heart SCORE) parent study,[21] a longitudinal study of cardiovascular disease risk and prevention (neuroimaging studies Heart SCORE-A and Heart SCORE-B). Inclusion criteria for all three studies were the same: age $\geq$65 and current enrollment in the parent study. The same exclusion criteria for all three studies included contraindication for magnetic resonance imaging (MRI) and diagnosis of dementia. The MYHAT-NI cognitive impairment criterion was stricter, excluding participants with Clinical Dementia Rating (CDR)[22] sum-of-box score >1.0. Heart SCORE-A further excluded participants with current substance-use disorder.

Participants were selected for the present analyses if (1) their behavioral and imaging data were collected and processed as of October 15, 2018; and (2) they were classified as cognitively unimpaired by CDR global score = 0 (see below).

1.2 | Diagnostic methods

All three neuroimaging studies included a comprehensive NP evaluation and the interviewer-based CDR scale, capturing cognitively

driven daily function in the home and community. Both Heart SCORE-A and Heart SCORE-B studies used an identical NP test battery, whereas MYHAT-NI used a similar battery assessing the same cognitive domains (memory, language, visuospatial abilities, executive function, and attention) with some overlapping and some distinct tests (see subsequent text). Both Heart SCORE-A and Heart SCORE-B studies adjudicated cognitive status via multi-disciplinary consensus conference using published diagnostic criteria.[23,24] However, MYHAT-NI operationalized cognitive status with the CDR. Therefore, we applied the CDR as the common methodologic definition of cognitive status across all three studies, defining CDR global score of 0 as cognitively unimpaired (CU). Furthermore, using the CDR to operationalize cognitive status allows for less-restricted test score range and potential selection bias than using NP tests that are also dependent variables of interest. Other studies have operationalized cognitive status with the CDR.[25,26]

NIHTB-CB tests were not a component of the cognitive diagnostic process in all three neuroimaging studies.

1.3 | Cognitive assessments

NIH Toolbox Cognition Battery (version 1.17) was administered via iPad to all participants by trained interviewers within 3 months of PET imaging. Measures include Picture Vocabulary and Oral Reading Recognition Tests, measuring language functions; Dimensional Change Card Sort (DCCS) and Flanker Inhibitory Control and Attention Tests, measuring executive functions and attention; List Sorting Working Memory Test, measuring working memory; Picture Sequence Memory Test, measuring episodic memory; and Pattern Comparison Processing Speed Test, measuring processing speed. Composite scores include Crystallized Cognition (averaging Picture Vocabulary and Oral Reading Recognition Tests), Fluid Cognition (averaging DCCS, Flanker, List Sorting, Picture Sequence Memory, and Pattern Comparison Processing Speed Tests), and Total Cognition (averaging all seven primary tests of Cognition) Composites.[27] The supplemental measures, immediate recall of the Rey Auditory Verbal Learning Test and Oral Symbol-Digit Test, were not administered because of time and participant burden constraints. Present analyses used uncorrected standard scores for individual tests and composite scores, since the age range of the pooled sample exceeded age 85, which is the upper limit of the demographically corrected standardized scores available from the NIHTB. The uncorrected standard scores are standardized to the NIHTB-CB nationally representative normative sample and are scaled to a mean of 100 and standard deviation (SD) of 15 for interpretability and comparability.[28]

Traditional NP tests in common among the three neuroimaging studies include the Mini Mental State Exam (MMSE), Wechsler Memory Scale Revised (WMS-R) Logical Memory Story A immediate and delayed recall, semantic fluency (animals), Trail Making Tests A & B, and clock drawing.[29,30] Traditional NP test data were analyzed as raw scores, with the exception of two standardized composite scores, a processing speed composite (comprising Trail Making Tests and semantic

fluency), and a global composite (comprising all seven traditional NP variables), following published methods.[31]

1.4 | Neuroimaging

Before the PET-imaging session, and for the purpose of anatomical region of interest (ROI) definition, a T1-weighted magnetization prepared rapid gradient echo (MPRAGE) MRI scan was obtained for each participant using a 3T Siemens PRISMA scanner. [^{18}F]AV-1451 PET scanning was performed on a four-ring Siemens Biograph mCT PET/CT scanner (22.1 cm field-of-view, reconstructed image resolution ≈5 mm). [^{11}C]PiB PET scans were performed on either a Siemens ECAT Exact HR+ scanner (15.2 cm field-of-view, reconstructed image resolution ≈6 mm) or the Siemens mCT PET/CT scanner.

[^{11}C]PiB (15 mCi nominal) or [^{18}F]AV-1451 (7 to 10 mCi) were administered as slow bolus injections via the antecubital vein. [^{11}C]PiB PET image data were collected over 50 to 70 minutes post-injection, and [^{18}F]AV-1451 PET imaging over 75 to 105 minutes. PET emission data for both tracers were binned into sets of 5-minute time frames spanning the acquisition duration. Scan sessions included acquisition of a low-dose CT scan (mCT) or a 511 keV transmission scan (HR+, using rotating ^{68}Ge/^{68}Ga rod sources) for attenuation and scatter correction. All scans were acquired in three-dimensional mode and reconstructed via analytic methods (filtered backprojection [mCT] or direct Fourier transform [HR+]) using the manufacturer's software.

PET image data sets were inspected for interframe motion. If required, framewise registration was performed using the image registration tool (PFUS) in PMOD version 3.709 software (PMOD Technologies, Zurich, Switzerland). For each tracer, a single frame image was produced by summing frames over 50 to 70 minutes post-injection for [^{11}C]PiB[32] and 80 to 100 minutes for AV-1451.[33-35] For each subject, the single-frame [^{11}C]PiB and [^{18}F]AV-1451 images were registered to the corresponding T1 MR image using the normalized mutual information algorithm implemented in PMOD software.

Each participant's MR image was parcellated into a set of ROIs using the default FreeSurfer 5.3 pipeline, with the exception of striatal subregions. To produce a more finely detailed parcellation of the striatum, components from the Imperial College London Clinical Imaging Centre (CIC) atlas[36] were substituted for the striatum of the (FreeSurfer default) Desikan-Killiany atlas[37] as described previously.[38] All FreeSurfer ROIs were visually inspected and manually edited where appropriate.

The ROIs generated were used to sample radioactivity concentrations in the summed PET images. Nine composite regional values were generated for [^{11}C]PiB PET scans (anterior cingulate, posterior cingulate, insula, superior frontal cortex, orbitofrontal cortex, lateral temporal cortex, parietal cortex, precuneus, and ventral striatum) by volume-weighted averaging of standard FreeSurfer and CIC ROIs. These nine regions were selected to most closely align with the standardized amyloid imaging Centiloid ROI.[39] A global [^{11}C]PiB retention index was computed by volume-weighted averaging of all nine composite PiB regions. Three composite Braak regional values were generated for

[18F]AV-1451 PET scans (Braak 1/2, Braak 3/4 , and Braak 5/6) from a volume-weighted average of standard FreeSurfer ROIs, as described by Schöll et al.,[40] except striatal subregions (accumbens, caudate, putamen, and pallidum) were not included in the Braak 5/6 region. Composite regional values were converted to standardized uptake value ratios (SUVRs) by normalizing to FreeSurfer cerebellar gray matter activity.

[11C]PiB PET scans were classified as regionally PiB-negative or PiB-positive using regional SUVR cutoffs determined by a previously described sparse k-means clustering and resampling method applied to 62 cognitively normal controls.[41] The regional [11C]PiB SUVR cutoffs used were 1.47 for anterior cingulate, 1.50 for posterior cingulate, 1.30 for insula, 1.33 for superior frontal cortex, 1.39 for orbitofrontal cortex, 1.28 for lateral temporal cortex, 1.34 for parietal cortex, 1.51 for precuneus, and 1.37 for ventral striatum. Participants were classified as regionally PiB-positive if any one composite region SUVR exceeded the corresponding regional cutoff.

1.5 | Analysis

Descriptive statistics were computed for cognitive and demographic variables and for PET predictors of interest ([11C]PiB and [18F]AV-1451 SUVR values). Multiple linear regression models were used to assess the association between each cognitive outcome of interest and predictors, adjusting for age, sex, race, and education; model assumptions were evaluated using residuals. Due to a high number of models, analyses were considered exploratory. Outliers of undue influence were evaluated by regression diagnostic indexes, including Cook's D > 1.0. Models with identified outliers were run excluding outliers in sensitivity analyses (see Results). The standardized β coefficients from each linear multiple regression model were interpreted as effect sizes reflecting strength of association, with 95% confidence intervals (CIs) providing information about statistical reliability.

2 | RESULTS

In the present analyses, n = 63 participants from the MYHAT-NI study, n = 36 participants from Heart SCORE-A and n = 19 from Heart SCORE-B studies contributed data, yielding a total combined sample of n = 118. Table 1 shows demographic characteristics and mean cognitive measures of the combined sample. Of note, about 25% of the sample had only a high school diploma or fewer years of education. About 18% were non-white, mostly African American. Supplemental Table 1 shows demographic measures by each study. Heart SCORE-A and Heart SCORE-B studies had a higher proportion of African American participants than MYHAT-NI, which had a higher mean age than the Heart SCORE cohorts. The MYHAT-NI study had fewer participants with education level beyond college. These differences reflect the demographic characteristics of the two parent studies.[20,21] Of note, despite differences in distributions, all demographic variables were largely overlapping among the three neuroimaging studies (eg, see age ranges, Supplemental Table 1).

TABLE 1 Demographic characteristics (mean, SD/n, %) and mean (SD) cognitive measures for n = 118 cognitively unimpaired older adults

Age (years)	76.29 (5.68), range 65 to 91
Sex	
Male	50 (42.4%)
Female	68 (57.63%)
Race	
White/Caucasian	97 (82.2%)
African American	19 (16.1%)
Other	2 (1.6%)
Education	
High school or L = less	30 (25.6%)
Some college	29 (24.8%)
4-year college	22 (18.8%)
Greater than college	36 (30.8%)
Traditional neuropsychological tests (raw scores)	
MMSE (30 points)	28.3 (1.5)
Logical Memory IR (Story A, 25 points)	14.16 (3.77)
Logical Memory DR (Story A, 25 points)	12.77 (3.82)
Semantic fluency (animals), # words per minute	19.62 (5.08)
Trail Making A, connections per second	0.76 (0.24)
Trail Making B, connections per second	0.32 (0.12)
Clock drawing (15 points)	14.23 (0.88)
Processing speed composite score[a] (0 to 100)	56.11 (16.75)
Global composite score[b] (0 to 100)	66.69 (9.75)
NIH Toolbox Cognition, uncorrected standard scores	
Total cognition composite	97.68 (9.38)
Fluid cognition composite	86.09 (10.26)
Crystallized cognition composite	111.28 (7.68)
Picture Vocabulary Test	112.7 (8.8)
Flanker Inhibitory Control and Attention Test	89.08 (9.22)
List Sorting Working Memory Test	95.19 (10.88)
Dimensional Change Card Sort Test	96.04 (7.37)
Pattern Comparison Processing Speed Test	80.63 (14.62)
Picture Sequence Memory Test	93.32 (11.94)
Oral Reading Recognition Test	109.08 (6.73)

Notes: NIH Toolbox Cognition Battery uncorrected standard scores have a mean of 100 and SD of 15.

CDR = Clinical Dementia Rating scale; DR = delayed recall; IR = immediate recall; MMSE = Mini Mental State Exam.

[a]Component measures = Trail Making Tests A and B, semantic fluency.

[b]Component measures = all traditional NP tests.

Table 1 also shows mean cognitive measures in the combined sample. NIHTB-CB uncorrected standard scores were relatively lower for fluid cognitive measures (eg, Fluid Cognition Composite, Pattern Comparisons, and Flanker Tests) compared to crystallized cognitive measures (eg, Crystallized Cognition Composite, Picture Vocabulary, and

TABLE 2 Summary of PiB-PET and AV-1451-PET biomarkers, mean (SD), by regionally defined PiB status

	PiB-negative (n = 84)	PiB-positive (n = 34)	P	Total sample (n = 118)
Global PiB SUVR	1.12 (0.05)	1.58 (0.33)	<0.0001	1.25 (0.27)
AV-1451 SUVR Braak 1/2	1.15 (0.11)	1.25 (0.14)	0.001	1.18 (0.13)
AV-1451 SUVR Braak 3/4	1.11 (0.07)	1.18 (0.07)	<0.0001	1.13 (0.08)
AV-1451 SUVR Braak 5/6	1.02 (0.07)	1.07 (0.07)	0.0002	1.03 (0.07)

Note: Participants were classified as regionally PiB-positive if at least one of nine composite regions exceeded the corresponding regional cutoff, including anterior cingulate, posterior cingulate, insula, superior frontal cortex, orbitofrontal cortex, lateral temporal cortex, parietal cortex, precuneus, and ventral striatum. (See Methods for details and SUVR cutoff values.)

TABLE 3 Standardized regression estimates reflecting association effect size between NIHTB-CB tests and PET biomarkers

	Global PiB SUVR	AV-1451 SUVR Braak 1/2	AV-1451 SUVR Braak 3/4	AV-1451 SUVR Braak 5/6
Total cognition composite	0.07 [−0.08, 0.24]	−0.09 [−0.24, 0.06]	−0.09 [−0.24, 0.07]	−0.13 [−0.28, 0.02]
Fluid cognition composite	0.01 [−0.16, 0.19]	−0.10 [−0.26, 0.07]	−0.14 [−0.31, 0.03]	**−0.20 [−0.37, −0.04]**
Crystallized cognition composite	0.21 [−0.02, 0.28]	−0.07 [−0.21, 0.08]	0.01 [−0.14, 0.15]	0.01 [−0.14, 0.15]
Picture Vocabulary Test	0.12 [−0.02, 0.26]	−0.06 [−0.21, 0.09]	0.02 [−0.13, 0.17]	0.03 [−0.12, 0.17]
Flanker Inhibitory Control and Attention Test	0.10 [−0.08, 0.27]	−0.04 [−0.21, 0.14]	−0.05 [−0.23, 0.13]	−0.06 [−0.25, 0.12]
List Sorting Working Memory Test	0.07 [−0.09, 0.24]	0.05 [−0.12, 0.22]	0.01 [−0.16, 0.18]	−0.06 [−0.23, 0.11]
Dimensional Change Card Sort Test	−0.11 [−0.29, 0.06]	−0.03 [−0.21, 0.15]	**−0.19 [−0.37, −0.02]**	**−0.20 [−0.38, −0.03]**
Pattern Comparison Processing Speed Test	−0.10 [−0.28, 0.08]	−0.12 [−0.30, 0.07]	**−0.25 [−0.42, −0.07]**	**−0.28 [−0.46, −0.11]**
Picture Sequence Memory Test	−0.07 [−0.27, 0.12]	−0.14 [−0.32, 0.04]	0.01 [−0.17, 0.19]	−0.02 [−0.20, 0.17]
Oral Reading Recognition Test	0.08 [−0.07, 0.26]	−0.06 [−0.22, 0.09]	−0.01 [−0.17, 0.15]	−0.01 [−0.16, 0.15]

Note: Standardized estimates are adjusted for age, sex, race, and education; standardized partial coefficients can be interpreted as the number of standard deviations the outcome increases (cognitive measure) for every standard deviation increase in the predictor (PET measure), holding all other predictors constant. Bolded estimates are those whose 95% CI does not contain zero, consistent with $P < .05$.

Oral Reading Tests), an expected pattern of normal cognitive aging reflected in the non–age-adjusted standard scores.[42,43]

Table 2 shows mean [11C]PiB and [18F]AV-1451 SUVR measures. There was approximately 29% PiB (Aβ) -positivity. [18F]AV-1451 retention, reflecting tau pathology, was higher in the PiB-positive participants compared to PiB-negative in all combined Braak-defined regions, reflecting a significant association between Aβ and tau deposition.

Table 3 presents associations between NIHTB-CB measures and PET biomarkers, adjusted for age, sex, race, and education. The standardized regression estimates are effect sizes of association between biomarkers and cognitive tests, holding covariates constant. The 95% CIs are provided and estimates can be interpreted as statistically significant at the $P < .05$ level if the 95% CI does not include zero. Results show that many of the association effects across cognitive measures and PET biomarkers were close to zero. Exceptions to this include Pattern Comparison Processing Speed Test, which showed small but reliable associations with tau deposition in Braak regions 3/4 and Braak regions 5/6, such that higher tau predicted lower performance (also see the Figure 1). Dimensional Change Card Sort also showed small associations with tau in Braak regions 3 /4 and 5/6 reaching significance, as did Fluid Cognition Composite with Braak region 5/6 only. None of the associations with global Aβ deposition

were significant. Picture Sequence Memory Test showed the largest association with tau in Braak 1/2 (entorhinal cortex/hippocampus) compared to its association with Aβ or with tau in other brain regions, although the effect was not statistically reliable.

Table 4 shows associations between neuropsychological paper and pencil measures and PET biomarkers, adjusted for age, sex, race, and education. Effect sizes can be compared to those presented in Table 3 with NIHTB-CB measures. Association effects were close to zero across most tests and PET biomarkers and none were statistically reliable.

2.1 | Sensitivity analyses

Three models with one potential outlier were identified for Braak 5/6 tau and Fluid Cognition Composite, Dimensional Change Card Sort, and Pattern Comparison Processing Speed. Removing these observations in sensitivity analyses resulted in no change for Pattern Comparison Processing Speed, but smaller effects in the other two outcomes which were no longer significant. Sensitivity analyses results are presented in Supplemental Table 2; the influence of the outlier on the Braak 5/6 tau–Pattern Comparison Processing Speed association is illustrated in Supplemental Figure 2.
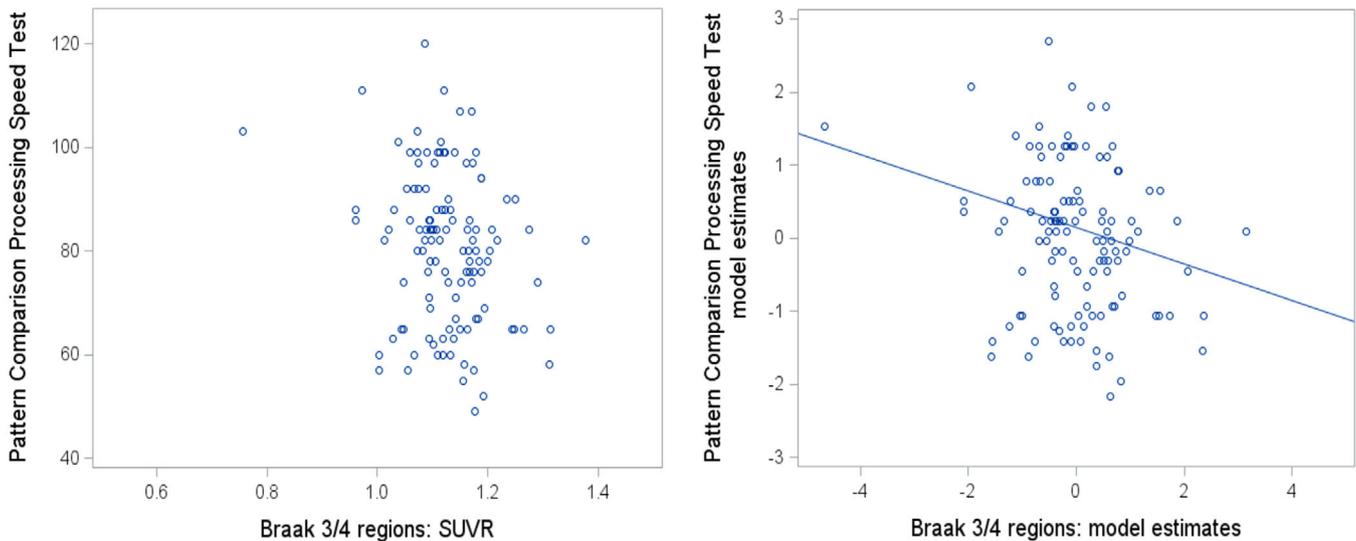
FIGURE 1 Association between Braak region 3/4 tau (AV-1451 retention) and Pattern Comparison Processing Speed Test performance. Scatterplot on the left shows raw data points. Scatterplot on the right shows standardized model estimates with age, sex, education, and race as covariates

TABLE 4 Standardized regression estimates reflecting association effect size between traditional neuropsychological tests and PET biomarkers

	Global PiB SUVR	AV-1451 SUVR Braak 1/2	AV-1451 SUVR Braak 3/4	AV-1451 SUVR Braak 5/6
MMSE	−0.03 [−0.2, 0.14]	−0.19 [−0.36, −0.02]	−0.15 [−0.32, 0.02]	−0.14 [−0.31, 0.04]
Logical Memory IR	−0.06 [−0.23, 0.12]	−0.07 [−0.26, 0.11]	−0.01 [−0.19, 0.17]	0.02 [−0.16, 0.20]
Logical Memory DR	0.02 [−0.16, 0.20]	−0.03 [−0.23, 0.16]	0.01 [−0.17, 0.20]	0.07 [−0.12, 0.25]
Semantic fluency	−0.07 [−0.25, 0.11]	−0.1 [−0.27, 0.09]	−0.08 [−0.25, 0.1]	−0.01 [−0.19, 0.17]
Trail Making Test A	0.04 [−0.15, 0.22]	0.07 [−0.12, 0.26]	0.06 [−0.13, 0.24]	−0.01 [−0.19, 0.18]
Trail Making Test B	−0.005 [−0.18, 0.17]	−0.06 [−0.23, 0.12]	−0.07 [−0.24, 0.1]	0.01 [−0.17, 0.18]
Clock drawing	0.08 [−0.11, 0.27]	−0.09 [−0.28, 0.1]	−0.01 [−0.19, 0.18]	0.05 [−0.14, 0.24]
Processing speed composite score[a]	−0.02 [−0.19, 0.16]	−0.04 [−0.22, 0.13]	−0.05 [−0.22, 0.12]	−0.03 [−0.20, 0.15]
Global composite score[b]	−0.04 [−0.22, 0.13]	−0.09 [−0.28, 0.09]	−0.06 [−0.24, 0.12]	−0.03 [−0.21, 0.15]

Note: Standardized estimates are adjusted for age, sex, race, and education; standardized partial coefficients can be interpreted as the number of SDs the outcome increases (cognitive measure) for every SD increase in the predictor (PET measure), holding all other predictors constant. Bolded estimates are those whose 95% CI does not contain zero, consistent with $P < .05$.
DR = delayed recall; MMSE = Mini Mental State Exam.
[a]Component measures = Trail Making Tests A and B, semantic fluency.
[b]Component measures = all NP tests listed in Table.

3 | DISCUSSION

The aim of this study was to investigate associations between NIH Toolbox Cognition Battery and AD biomarkers Aβ and tau PET in older adults without cognitive symptoms. The importance of identifying sensitive cognitive measures that can capture subtle deficits or decline has been discussed within the context of preclinical AD staging: for example, consistent with clinical Stage 2 in the National Institute on Aging - Alzheimer's Association (NIA-AA) research framework,[44] and in the context of AD prevention trials.[15,45,46]

Present findings indicate small but reliable associations between several fluid cognition measures of NIHTB-CB and tau deposition in the brain, but not Aβ deposition, in this cognitively unimpaired sample of older adults. The most robust findings observed were associations between Pattern Comparison Processing Speed and tau deposition in higher Braak (ie, extra-medial temporal lobe) regions. These were observed in the primary analysis and in sensitivity analyses excluding a potential outlier. Dimensional Change Card Sort was also associated with tau in higher Braak regions, as was the Fluid Cognition Composite score; however, these associations were influenced by a few extreme observations of tau in the Braak 5/6 region, so they must be interpreted with caution. By comparison with associations between traditional paper-and-pencil NP tests and AD biomarkers, reliable NIHTB-CB effect sizes were larger.

Generally, the associations of an AD biomarker with fluid cognitive tasks is consistent with the cognitive aging literature, showing that fluid cognition declines with age, and likely latent aging-associated pathophysiology, whereas crystallized cognition remains stable across the lifespan.[47,48] To date, there have been relatively few focused investigations of NIHTB-CB in the AD spectrum, including normal aging. The battery was not designed for the goal of clinical diagnosis of AD or MCI. In particular, only one task assesses episodic memory (Picture Sequence Memory Test), which has no delayed recall component, a key cognitive measure sensitive to and predictive of early AD.[49,50] Supplemental NIHTB-CB tests include immediate recall with the Rey Auditory Verbal Learning Test. Including this supplemental measure, and with an additional delayed recall condition, Hackett et al.[13] reported that the NIHTB-CB showed group discrimination among older adults with normal cognition, subjective cognitive decline, mild cognitive impairment (MCI), and AD. Buckley et al.[12] found that the NIHTB-CB discriminated among subtle cognitive impairment and normal cognition in a small sample of clinically normal older adults, although not with the sensitivity of another computerized battery, the Cogstate C3 Learning and Memory composite, in a head-to-head comparison. To our knowledge, there have been even fewer investigations of NIHTB and AD biomarkers. A study examining hippocampal volume in 93 self-reported unimpaired older volunteers found associations with the Fluid Cognition composite.[14] Of interest, among individual tests, the strongest predictor of hippocampal volume was the Pattern Comparison Processing Speed. Present results are also strongest for this test in its association with tau deposition, albeit in brain regions outside the hippocampus.

Although most studies of novel computerized tasks sensitive to preclinical AD focus on memory and associative learning paradigms,[16,18] there is a smaller but longstanding literature on computerized measures of reaction time, processing speed, and attentional control.[51–53] More recently, Kochan et al.[54] reported that both simple and choice reaction time predicted time-to-dementia, AD being the predominant etiology, over 4 years in a large population-based study. Regarding Aβ and tau imaging, Mishra et al.[55] found that computerized tasks of attentional control were moderately correlated with a summary measure of [^{18}F]AV-1451 retention in cognitively normal older adults, with comparable association effect size as episodic memory. More broadly, meta-analytic studies of Aβ-cognition relationships in unimpaired older adults indicate very small-to-small associations with processing speed[56] and executive function domains.[56,57] These biomarker meta-analyses are consistent with the literature showing that, in addition to episodic memory impairment, preclinical deficits in executive functioning and perceptual speed predict future progression to AD.[58]

With regard to differential associations with cognition between Aβ and tau pathologies, the latter is more proximally related to neurodegeneration and cognitive impairment in the theoretical AD biomarker model[59] across the disease spectrum, and this is empirically supported by postmortem studies.[60] Among cognitively unimpaired individuals, however, the extant literature of cross-sectional studies is mixed. In a large multi-cohort data set of 907 participants ages 40 and older, most traditional neuropsychological tests were associated with Aβ status

and none with tau, regardless of biomarker method (cerebrospinal fluid [CSF] or PET).[61] In contrast, an earlier smaller CSF study showed that both total-tau and phosphorylated-tau were associated with episodic memory, whereas Aβ was not associated with any cognitive domain performance.[62] Rentz et al.[63] recently reported that both inferotemporal tau and global Aβ imaging predicted MMSE scores in models controlling for verbal IQ (a proxy for cognitive reserve). It is clear from the past 10 to 15 years of amyloid imaging research that cross-sectional Aβ-cognition association effect sizes in cognitively unimpaired cohorts are small and require power for detection, relative to larger effects of association with longitudinal cognitive change.[56]

The present analyses were limited in that data were combined from three separate studies, with some minor differences in exclusion criteria, as well as in traditional NP test batteries. As a result, there was no episodic list-learning memory test in common as a traditional comparison measure, which may have been more sensitive to Aβ or tau biomarkers in this cohort. A strength of this study is that the combined study samples have a somewhat higher proportion of under-represented minorities than do typical AD biomarker studies, as well as a more broadly representative distribution of education.[64,65] Because this study was exploratory in nature, replication and confirmation of findings is needed. Future research might also include traditional NP tests with closer methodologic correspondence to NIHTB-CB tests (eg, Ruff 2 & 7 Test[66] as an analog to the Flanker Test; Peabody Picture Vocabulary Test[67] as an analog for Picture Vocabulary), to more meaningfully compare tests of the same construct against AD biomarkers.

In summary, in this initial report characterizing NIHTB-CB relationships to AD PET biomarkers, several measures of fluid cognition were associated with brain tau deposition in older adults without dementia. One measure, Pattern Comparison Processing Speed, was robust to outlier sensitivity analyses and was associated with cortical tau deposition in cognitively normal older adults. Future efforts, including the ongoing multi-site ARMADA (Advancing Reliable Measurement in Alzheimer's Disease and Cognitive Aging) study will address broader questions of the utility of NIHTB measures in aging and AD, including longitudinal change across the disease spectrum, increasing the normative age range above age 85 in English and Spanish versions, and validating new tasks. This large-scale effort is anticipated to be completed in 2022.

REFERENCES

1. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH toolbox for assessment of neurological and behavioral function. *Neurology*. 2013;80(11 supplement 3):S2-S6.

2. Weintraub S, Dikmen SS, Heaton RK, et al. Cognition assessment using the NIH Toolbox. *Neurology*. 2013;80(11 Supplement 3):S54-S64.

3. Weintraub S, Dikmen SS, Heaton RK, et al. The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: validation in an adult sample. *J Int Neuropsychol Soc*. 2014;20(6):567-578.

4. Heaton RK, Akshoomoff N, Tulsky D, et al. Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *J Int Neuropsychol Soc*. 2014;20(6):588-598.

5. Zelazo PD, Anderson JE, Richler J, et al. NIH toolbox cognition battery (CB): validation of executive function measures in adults. *J Int Neuropsychol Soc*. 2014;20(6):620-629.

6. Carlozzi NE, Tulsky DS, Chiaravalloti ND, et al. NIH toolbox cognitive battery (NIHTB-CB): the NIHTB pattern comparison processing speed test. *J Int Neuropsychol Soc*. 2014;20(6):630-641.

7. Dikmen SS, Bauer PJ, Weintraub S, et al. Measuring episodic memory across the lifespan: NIH toolbox picture sequence memory test. *J Int Neuropsychol Soc*. 2014;20(6):611-619.

8. Gershon RC, Cook KF, Mungas D, et al. Language measures of the NIH toolbox cognition battery. *J Int Neuropsychol Soc*. 2014;20(6):642-651.

9. Tulsky DS, Carlozzi NE, Holdnack J, et al. Using the NIH toolbox cognition battery (NIHTB-CB) in individuals with traumatic brain injury. *Rehabil Psychol*. 2017;62(4):413.

10. Hessl D, Sansone SM, Berry-Kravis E, et al. The NIH toolbox cognitive battery for intellectual disabilities: three preliminary studies and future directions. *J Neurodev Disord*. 2016;8(1):35.

11. Carlozzi N, Goodnight S, Casaletto K, et al. Validation of the NIH toolbox in individuals with neurologic disorders. *Arch Clin Neuropsychol*. 2017;32(5):555-573.

12. 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. *J Prev Alzheimers Dis*. 2017;4(1):3.

13. Hackett K, Krikorian R, Giovannetti T, et al. Utility of the NIH toolbox for assessment of prodromal Alzheimer's disease and dementia. *Alzheimer's Dement*. 2018;10:764-72.

14. O'Shea A, Cohen R, Porges EC, Nissim NR, Woods AJ. Cognitive aging and the hippocampus in older adults. *Front Aging Neurosci*. 2016;8:298.

15. Weintraub S, Carrillo MC, Farias ST, et al. Measuring cognition and function in the preclinical stage of Alzheimer's disease. *Alzheimers Dement*. 2018;4:64-75.

16. Loewenstein DA, Curiel RE, Duara R, Buschke H. Novel cognitive paradigms for the detection of memory impairment in preclinical Alzheimer's disease. *Assessment*. 2018;25(3):348-359.

17. Jack CR, Bennett DA, Blennow K, et al. A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*. 2016;87(5):539-547.

18. Rentz DM, Parra Rodriguez MA, Amariglio R, Stern Y, Sperling R, Ferris S. Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. *Alzheimers Res Ther*. 2013;5:58.

19. Mortamais M, Ash JA, Harrison J, et al. Detecting cognitive changes in preclinical Alzheimer's disease: a review of its feasibility. *Alzheimers Dement*. 2017;13(4):468-492.

20. Ganguli M, Chang C-CH, Snitz BE, Saxton JA, Vanderbilt J, Lee C-W. Prevalence of mild cognitive impairment by multiple classifications: the Monongahela-Youghiogheny Healthy Aging Team (MYHAT) project. *Am J Geriatr Psychiatry*. 2010;18(8):674-683.

21. Bambs C, Kip KE, Dinga A, Mulukutla SR, Aiyer AN, Reis SE. Low prevalence of "ideal cardiovascular health" in a community-based population: the heart strategies concentrating on risk evaluation (Heart SCORE) study. *Circulation*. 2011;123(8):850-857.

22. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 1993;43(11):2412-2414.

23. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):270-279.

24. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):263-269.

25. Schindler SE, Jasielec MS, Weng H, et al. Neuropsychological measures that detect early impairment and decline in preclinical Alzheimer disease. *Neurobiol Aging*. 2017;56:25-32.

26. Ganguli M, Jia Y, Hughes TF, et al. Mild cognitive impairment that does not progress to Dementia: a population-based study. *J Am Geriatr Soc*. 2019;67(2):232-238.

27. Bauer PJ, Zelazo PD. Ix. Nih toolbox cognition battery (cb): summary, conclusions, and implications for cognitive development. *Monogr Soc Res Child Dev*. 2013;78(4):133-146.

28. Slotkin J, Nowinski C, Hays R, et al. NIH Toolbox Scoring and Interpretation Guide. Washington (DC): National Institutes of Health; 2012:6-7.

29. Lezak MD, Howieson DB, Bigler ED, Tranel D. Neuropsychological Assessment: OUP USA; 2012.

30. Ganguli M, Snitz B, Vander Bilt J, et al. How much do depressive symptoms affect cognition at the population level? The Monongahela-Youghiogheny Healthy Aging Team (MYHAT) study. *Int J Geriatr Psychiatry*. 2009;24(11):1277-1284.

31. Malek-Ahmadi M, Chen K, Perez SE, He A, Mufson EJ. Cognitive composite score association with Alzheimer's disease plaque and tangle pathology. *Alzheimers Res Ther*. 2018;10(1):90.

32. McNamee RL, Yee S-H, Price JC, et al. Consideration of optimal time window for Pittsburgh compound B PET summed uptake measurements. *J Nucl Med*. 2009;50(3):348-355.

33. Baker SL, Lockhart SN, Price JC, et al. Reference tissue-based kinetic evaluation of F-18-AV-1451 for tau imaging. *J Nucl Med*. 2017;58(2):332-338.

34. Hahn A, Schain M, Erlandsson M, et al. Modeling strategies for quantification of in vivo F-18-AV-1451 binding in patients with tau pathology. *J Nucl Med*. 2017;58(4):623-631.

35. Shcherbinin S, Schwarz AJ, Joshi A, et al. Kinetics of the tau PET tracer F-18-AV-1451 (T807) in subjects with normal cognitive function, mild cognitive impairment, and Alzheimer disease. *J Nucl Med*. 2016;57(10):1535-1542.

36. Tziortzi AC, Searle GE, Tzimopoulou S, et al. Imaging dopamine receptors in humans with [11C]-(+)-PHNO: Dissection of D3 signal and anatomy. *NeuroImage*. 2011;54(1):264-277.

37. Desikan RS, Segonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31(3):968-980.

38. Tudorascu DL, Minhas DS, Lao PJ, et al. The use of Centiloids for applying 11C PiB classification cutoffs across region-of-interest delineation methods. *Alzheimers dement*. 2018;10:332-339.

39. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers dement*. 2015;11(1):1-15.e4.

40. Schöll M, Lockhart SN, Schonhaut DR, et al. PET imaging of tau deposition in the aging human brain. *Neuron*. 2016;89(5):971-982.

41. Cohen AD, Mowrey W, Weissfeld LA, et al. Classification of amyloid-positivity in controls: comparison of visual read and quantitative approaches. *Neuroimage*. 2013;71:207-215.

42. Li S-C, Lindenberger U, Hommel B, Aschersleben G, Prinz W, Baltes PB. Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychol Sci*. 2004;15(3):155-163.

43. Slotkin J, Kallen M, Griffith J, et al. *NIH Toolbox Technical Manual*. Bethesda, MD: National Institutes of Health; 2012.

44. Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14(4):535-562.

45. Sperling R, Mormino E, Johnson K. The evolution of preclinical Alzheimer's disease: implications for prevention trials. *Neuron*. 2014;84(3):608-622.

46. Snyder PJ, Kahle-Wrobleski K, Brannan S, et al. Assessing cognition and function in Alzheimer's disease clinical trials: do we have the right tools? *Alzheimers Dement*. 2014;10(6):853-860.

47. Craik F. *Aging and Cognitive Processes*, Springer Science & Business Media; 2012.

48. McDonough IM, Bischof GN, Kennedy KM, Rodrigue KM, Farrell ME, Park DC. Discrepancies between fluid and crystallized ability in healthy adults: a behavioral marker of preclinical Alzheimer's disease. *Neurobiol Aging*. 2016;46:68-75.

49. Knopman DS, Ryberg S. A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Arch Neurol*. 1989;46(2):141-145.

50. Laakso MP, Hallikainen M, Hänninen T, Partanen K, Soininen H. Diagnosis of Alzheimer's disease: MRI of the hippocampus vs delayed recall. *Neuropsychologia*. 2000;38(5):579-584.

51. Nestor PG, Parasuraman R, Haxby JV. Speed of information processing and attention in early Alzheimer's dementia. *Dev Neuropsychol*. 1991;7(2):243-256.

52. Parasuraman R, Haxby JV. Attention and brain function in Alzheimer's disease: a review. *Neuropsychology*. 1993;7(3):242.

53. Tse C-S, Balota DA, Yap MJ, Duchek JM, McCabe DP. Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology*. 2010;24(3):300.

54. Kochan NA, Bunce D, Pont S, Crawford JD, Brodaty H, Sachdev PS. Reaction time measures predict incident dementia in community-living older adults: the Sydney Memory and Ageing Study. *Am J Geriatr Psychiatry*. 2016;24(3):221-231.

55. Mishra S, Gordon BA, Su Y, et al. AV-1451 PET imaging of tau pathology in preclinical Alzheimer disease: defining a summary measure. *Neuroimage*. 2017;161:171-178.

56. Baker JE, Lim YY, Pietrzak RH, et al. Cognitive impairment and decline in cognitively normal older adults with high amyloid-β: a meta-analysis. *Alzheimers Dement*. 2017;6:108-121.

57. Hedden T, Oh H, Younger AP, Patel TA. Meta-analysis of amyloid-cognition relations in cognitively normal older adults. *Neurology*. 2013;80(14):1341-1348.

58. Backman L, Jones S, Berger A-K, Laukka EJ, Small BJ. Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology*. 2005;19(4):520-531.

59. Jack CR Jr, Knopman DS, Jagust WJ, et al. Update on hypothetical model of Alzheimer's disease biomarkers. *Lancet Neurol*. 2013;12(2):207.

60. Nelson PT, Alafuzoff I, Bigio EH, et al. Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *J Neuropathol Exp Neurol*. 2012;71(5):362-381.

61. Bos I, Vos SJ, Jansen WJ, et al. Amyloid-β, tau, and cognition in cognitively normal older individuals: examining the necessity to adjust for biomarker status in normative data. *Front Aging Neurosci*. 2018; 10:193.

62. Pettigrew C, Soldan A, Moghekar A, et al. Relationship between cerebrospinal fluid biomarkers of Alzheimer's disease and cognition in cognitively normal older adults. *Neuropsychologia*. 2015;78:63-72.

63. Rentz DM, Mormino EC, Papp KV, Cognitive resilience in clinical and preclinical Alzheimer's disease: the Association of Amyloid and Tau Burden on cognitive performance. *Brain Imaging Behav*. 2017;11(2):383-390.

64. Ganguli M, Albanese E, Seshadri S, et al. Population neuroscience: dementia epidemiology serving precision medicine and population health. *Alzheimer Dis Assoc Disord*. 2018;32(1):1-9.

65. Honningsvåg L-M, Linde M, Håberg A, Stovner LJ, Hagen K. Does health differ between participants and non-participants in the MRI-HUNT study, a population based neuroimaging study? The Nord-Trøndelag health studies 1984-2009. *BMC Med Imaging*. 2012;12(1):23.

66. Knight RG, McMahon J, Skeaff CM, Green TJ. Reliable change indices for the ruff 2 and 7 selective attention test in older adults. *Appl Neuropsychol*. 2010;17(4):239-245.

67. Snitz BE, Bieliauskas LA, Crossland A, Basso MR, Roper B. PPVT-R as an estimate of premorbid intelligence in older adults. *Clin Neuropsychol*. 2000;14(2):181-186.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.