Received: 18 May 2021 | Revised: 8 July 2021 | Accepted: 3 August 2021 | Published online: 27 September 2021

DOI: 10.1002/dad2.12243

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease

Kathryn V. Papp[1,2] | Aubryn Samaroo[2] | Hsiang-Chin Chou[2] | Rachel Buckley[1,2,3] | Olivia R. Schneider[2] | Stephanie Hsieh[2] | Daniel Soberanes[1] | Yakeel Quiroz[2] | Michael Properzi[2] | Aaron Schultz[2] | Iván García-Magariño[4,5] | Gad A. Marshall[1,2] | Jane G. Burke[2] | Raya Kumar[2] | Noah Snyder[2] | Keith Johnson[2,6] | Dorene M. Rentz[1,2] | Reisa A. Sperling[1,2] | Rebecca E. Amariglio[1,2]

[1] Center for Alzheimer Research and Treatment, Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

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

[3] Melbourne School of Psychological Science, University of Melbourne, Melbourne, Victoria, Australia

[4] Department of Software Engineering and Artificial Intelligence, Complutense University of Madrid, Madrid, Spain

[5] Instituto de Tecnología del Conocimiento, UCM, Madrid, Spain

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

Correspondence
Kathryn V. Papp, Center for Alzheimer's Research and Treatment, Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Hale Building, 60 Fenwood Rd., Boston, MA 02115, USA.
E-mail: kpapp@bwh.harvard.edu

Abstract

Introduction: Unsupervised digital cognitive testing is an appealing means to capture subtle cognitive decline in preclinical Alzheimer's disease (AD). Here, we describe development, feasibility, and validity of the Boston Remote Assessment for Neurocognitive Health (BRANCH) against in-person cognitive testing and amyloid/tau burden.

Methods: BRANCH is web-based, self-guided, and assesses memory processes vulnerable in AD. Clinically normal participants (n = 234; aged 50–89) completed BRANCH; a subset underwent in-person cognitive testing and positron emission tomography imaging. Mean accuracy across BRANCH tests (Categories, Face-Name-Occupation, Groceries, Signs) was calculated.

Results: BRANCH was feasible to complete on participants' own devices (primarily smartphones). Technical difficulties and invalid/unusable data were infrequent. BRANCH psychometric properties were sound, including good retest reliability. BRANCH was correlated with in-person cognitive testing ($r = 0.617$, $P < .001$). Lower BRANCH score was associated with greater amyloid ($r = -0.205$, $P = .007$) and entorhinal tau ($r = -0.178$, $P = .026$).

Discussion: BRANCH reliably captures meaningful cognitive information remotely, suggesting promise as a digital cognitive marker sensitive early in the AD trajectory.

KEYWORDS
digital biomarkers, mobile testing, preclinical Alzheimer's disease, unsupervised assessment

1 | INTRODUCTION

The continuum of Alzheimer's disease (AD) begins with pathological changes in amyloid and tau followed by subtle cognitive changes and subsequent clinical impairment.[1] Efforts to detect preclinical AD (i.e., biomarker evidence of elevated amyloid/tau without clinical impairment) have become increasingly urgent as interventions shift to this earliest stage of disease with the hope of delaying the onset of clinical impairment.[2] However, the large sample size and lengthy study durations needed to detect therapeutic benefit in secondary prevention trials poses a significant barrier to rapid progress.

Mobile cognitive assessment may expedite the screening and tracking of participants for secondary prevention.[3] By allowing participants to complete study assessments using their own electronic device, data collection exponentially increases, improving clinical trial efficiency. Interest in using digital and remote assessments is growing and multiple studies have demonstrated the feasibility and validity of computerized assessments in supervised[4–6] and unsupervised settings[7,8] using personal devices.[9,10] Several of these assessments are designed for specific symptomatic groups (e.g., for detection of mild cognitive impairment [MCI])[9] or broad populations.[6]

A smaller subset of mobile assessments have been specifically designed to capture subtle cognitive decline in preclinical AD.[11–14] Decrements in episodic memory, particularly paired associative memory (i.e., integrating contextual information such as linking a face with a name) as well as associative inference,[15] are observed in preclinical AD.[16–18] Regions critical for associative memory such as the entorhinal cortex and anterior hippocampus[19,20] are also early sites for phosphorylated tau neurofibrillary tangles.[21] Likewise, pattern separation (i.e., the ability to discriminate between previously learned items and those that are perceptually similar[22]) is an aspect of memory performance reliant on the dentate gyrus and connections to the entorhinal cortex and hippocampus.[23] Finally, measures which facilitate learning using semantic cues have been shown to be particularly useful in identifying decrements in memory encoding in preclinical AD.[24,25] Digital cognitive tasks that target cognitive processes that decline during preclinical AD, such as the Boston Remote Assessment for Neurocognitive Health (BRANCH), would be particularly relevant to prevention trials that are seeking cognitive outcomes that can track putative therapeutic benefit.

Here, we provide initial validation for BRANCH, a web-based cognitive battery designed for unsupervised completion on a personal electronic device (e.g., smartphone). Tasks were designed to challenge the aforementioned memory processes in clinically normal (CN) older adults. We designed an intuitive interface to facilitate unsupervised testing and used stimuli relevant to everyday life (faces, groceries, street signs). We deployed BRANCH to two groups of CN older adults: (1) a registry sample without in-person contact to assess BRANCH feasibility "in the wild" and an (2) observational sample to assess validity in relation to in-clinic traditional assessments and AD biomarkers. If truly a valid cognitive measure, we expected lower BRANCH performance to be associated with increasing age given well-known age-related cognitive decline, even in the absence of neurodegenerative disease.[26] Additionally, we expected BRANCH performance to be correlated with traditional in-clinic cognitive assessments. Furthermore, we expected those with higher AD pathologic burden (i.e., amyloid and tau measured with biomarkers) to be associated with lower BRANCH performance. In addition to testing the aforementioned hypotheses, BRANCH reliability was assessed using a re-test paradigm in a subset of registry participants. Finally, we examined the feasibility of BRANCH in MCI participants in a separate well-characterized cohort, to ensure BRANCH remained feasible for participants who progress to MCI over the course of a secondary prevention trial.

RESEARCH IN CONTEXT

1. **Systematic review**: The authors searched the scientific literature for digital cognitive assessments in preclinical Alzheimer's disease (AD). Unsupervised digital cognitive testing on an individual's own device is an appealing means to capture AD-related memory decrements and a growing area of research.
2. **Interpretation**: Our findings indicate that Boston Remote Assessment for Neurocognitive Health (BRANCH) is a valid cognitive measure for preclinical AD; BRANCH exhibited moderate correlations with traditional cognitive tests and worse BRANCH performance was associated with greater global amyloid and entorhinal tau burden on imaging.
3. **Future directions**: BRANCH reliably captures meaningful cognitive information remotely, suggesting promise as a digital cognitive marker sensitive early in the AD trajectory. Future work will explore short-term learning curves by capturing BRANCH more frequently, developing BRANCH for use in different languages, and incorporating additional metrics (e.g., subjective report, environmental factors) to enrich and contextualize cognitive data.

2 | METHODS

2.1 | Participants

The registry sample was recruited from two online local registries; exclusion criteria included self-report of MCI/dementia and participants were presumed CN by self-report. The observational sample included CN participants from the Harvard Aging Brain Study (HABS; 2P01AG036694-11-Sperling, Johnson) and related studies (1R01AG058825-01A-Amariglio, R01AG053184-Marshall).[27] Observational participants underwent neuropsychological testing and neuroimaging described below. Study procedures were conducted in accordance with human subjects' protections and the study protocol was approved by the Mass General Brigham Institutional Review Board. All participants underwent informed consent. Participants from the observational sample were classified as CN by either study entry criteria[28] or via a multidisciplinary consensus meeting described elsewhere.[29] Briefly, study entry criteria included a Clinical Dementia Rating (CDR) global score = 0 and normal education-adjusted performance on the Mini-Mental State Examination (MMSE), and Logical Memory Delayed Recall (LMDR). BRANCH was also administered in-clinic to a separate group of MCI patients (described in Supp A in supporting information).

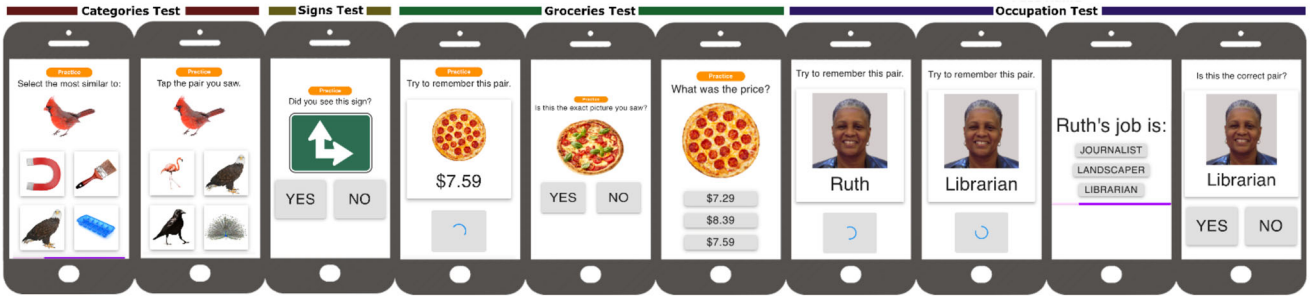

FIGURE 1 Schematic of Boston Remote Assessment for Neurocognitive Health (BRANCH) tasks.

2.2 | Platform

BRANCH initial piloting efforts are described in Supp A. BRANCH was developed using a web-based platform, which met hospital data privacy and security requirements. BRANCH was sent to participants via e-mail/text and can be completed on any web-enabled device (Figure 1). Accuracy serves as the primary outcome across tests; however, exploratory analyses were complete on reaction time outcomes.

2.3 | BRANCH tasks and task rationale

Categories Test: Modeled on the Similarities task from Weschler Adult Intelligence Scale and the Free and Cued Selective Reminding Test (FCSRT),[24,25] this measure uses a semantic-association paradigm to facilitate encoding of a pictured pair of items belonging to the same category (e.g., robin–eagle) among unrelated items (Figure 1). After a delay, participants must identify the previously learned pair among within-category distractors (e.g., peacock–crow), including category prototypes, requiring specific versus gist-like episodic memory.

Face-Name-Occupation Test: This is a modified version of the Face-Name Associative Memory Exam (FNAME)[17,30] incorporating both face–name and face–occupation pairs. Participants are asked to remember both a name and an occupation associated with faces. Each pairing is presented serially such that each target face is seen twice (once with a name; once with an occupation). After the learning phase, an associative inference component[31] assesses relational memory by requiring the participant to correctly match names with occupations without seeing faces, based on inferred associations from trained face–name/occupation pairs. After a delay, the participant must identify the correct face–name/occupation pair among counterbalanced distractors.

Groceries Test: This is an adapted paired associate learning test combining a visual and numerical element.[32] It differs from extant associative learning tasks that are limited to verbal-only[33] or artificial stimuli.[34] Participants are asked to remember a price (ranging from $1.09–$12.99, i.e., within approximately 15% of market value) paired with a pictured grocery item. After a delay, participants must recognize the correct price among counterbalanced incorrectly paired and partially novel price/grocery distractor pairs. Participants then complete an adapted pattern separation paradigm[35] whereby they are required to indicate whether a pictured grocery has been previously seen (equal number of targets, lures, same-category foils).

Signs Test: The Signs Test is a visual memory task using a continuous recognition paradigm.[36–38] Street sign stimuli were selected to (1) serve as a counter to the primarily verbal-based memory tests which confer a female advantage,[39] (2) mirror more clinically meaningful functions (i.e., driving), and (3) be potentially more valid across a range of educational backgrounds and minority groups.[40] Stimuli are learned based on serial and repeated exposures and performance is determined by the participant's ability to discriminate between previously learned versus novel street signs as they are provided with visual and auditory feedback for incorrect responses.

BRANCH composite: The BRANCH composite included mean accuracy across: Categories, Face–Name–Occupation Inference, Face–Name–Occupation Recognition, Groceries Price Recall, Groceries Pattern Separation, and Signs.

2.4 | Post-BRANCH survey and criteria for data validity

After BRANCH, participants were surveyed regarding technical difficulties. Participants rated instruction clarity (yes/no), task difficulty (very easy–very difficult; five options), and task engagement (not engaging–highly engaging; five options) by task.

To determine the usability of BRANCH data, we implemented two cutoffs, both of which were required for data to be usable. First, tasks were considered complete if > 90% of items were completed. Second, participants were required to exhibit > 66% accuracy on the learning portion of the Categories test.

2.5 | Standard neuropsychological testing

Observational sample participants completed the Preclinical Alzheimer Cognitive Composite (PACC-5),[41,42] which includes two memory measures, LMDR and the FCSRT(/96); a measure of global cognition, MMSE; a measure of processing speed, Digit Symbol Substitution Test (DSST); and a language measure, category fluency (CAT). Participants also completed measures of processing speed (Trail Making Test A [TMTA]) and mental flexibility (TMTB).

TABLE 1 Participant characteristics by sample

	Registry	Observational	Group differences between registry and observational
n	79	155	
Race			$\chi^2 = 5.13, P = .162$
Black	21.5%	11.6%	
White	70.9%	84.5%	
Asian	2.5%	2.6%	
Native American	0%	0.6%	
MMSE	–	29.15 (1.06) Range:25–30	
CDR- Global (0/0.5)		146/3	
Age	67.2 (10.00) Range: 50–83	73.7 (10.63) Range: 53–90	t = 6.06, $P < .001$
Sex (% female)	54.4%	62.6%	$\chi^2 = 1.26, P = .262$
Education (y)	16.2 (4.0)	16.7 (2.0)	t = 1.34, $P = .183$

Abbreviations: CDR, Clinical Dementia Rating; MMSE , Mini-Mental State Examination.

2.6 | AD biomarkers of amyloid and tau: positron emission tomography data acquisition and analysis

Observational sample participants underwent positron emission tomography (PET)[7] with ^{11C}Pittsburg Compound-B (PiB; n = 144) and ^{F18}Flortaucipir (FTP; n = 129) using previously published procedures[43] within 3 and 2 years of BRANCH administration, respectively. FTP images were acquired from 75 to 105 minutes and PiB images were acquired using a 60-minute dynamic acquisition on a Siemens ECAT HR+ PET scanner. PET images were co-registered to corresponding T1 images using FreeSurfer-based(v6) structural regions of interest (ROIs) mapped into native PET space using SPM12. FTP was expressed as a standardized uptake volume ratio (SUVR) and PiB as the distribution volume ratio (DVR). The reference region was cerebellar gray using an magnetic resonance imaging (MRI)-based method; FTP-PET data were corrected for partial volume effects. For PiB, a global cortical aggregate was calculated. For FTP, entorhinal cortex was used because of its importance to memory and status as an early site of tau deposition.

2.7 | Statistics

Statistical analyses were completed using R (v4.0.3). To determine the feasibility of BRANCH, we examined rates of administration errors, self-reported technical issues and task difficulty, and engagement in the registry sample.

To characterize BRANCH psychometric properties, we computed descriptive statistics for each BRANCH test accuracy score across both samples. We expected accuracy measures to be approximately 70% and that minimal floor or ceiling effects would be observed. One exception is the less challenging Categories test, for which we expected ceiling effects.

Correlations were used to examine whether BRANCH captured age-related memory decrements[26] across registry and observational samples. BRANCH retest reliability was assessed on re-testing among a subset of participants. Pearson correlation coefficient with an *r* of 0.8 to 0.9 was considered "good" reliability.[44]

Correlations were used to examine BRANCH convergent/discriminant validity in relation to traditional cognitive tests in the observational sample. We expected composites (i.e., BRANCH, PACC-5) to be more strongly correlated compared to individual tests. Similarly, we expected traditional memory measures (LMDR, FCSRT) to be more strongly correlated with memory-oriented BRANCH tasks (convergent validity), whereas traditional processing speed measures (TMTA) would be less correlated with BRANCH memory measures (discriminant validity). Given the large number of comparisons for BRANCH measures versus traditional measures, statistical significance was set at $P < .001$.

Regarding AD-biomarker validity, partial correlations controlling for age were used to examine the relationship between BRANCH composite performance and amyloid and entorhinal tau. Exploratory analyses examined individual BRANCH measures.

Finally, we examined the feasibility of BRANCH in MCI patients by quantifying the number of discontinuations and the presence of floor effects (Supp A).

3 | RESULTS

3.1 | Participant characteristics

Demographics for registry (n = 79) participants and observational (n = 155) participants are shown in Table 1. Registry participants were slightly younger. Participant characteristics for the in-clinic sample (n = 22 MCI) are provided in supporting information.

TABLE 2 BRANCH feasibility: registry sample

Self-report of difficulty completing			**2.97%**		
Self-report of difficulty understanding instructions			**2.97%**		
Self-report of technical difficulties			**15.84%**		
Self-reported task difficulty					
	Very difficult	**Somewhat difficult**	**Average**	**Somewhat easy**	**Very easy**
Categories	4%	40%	34%	16%	7%
Groceries	23%	45%	22%	7%	4%
Face–Name–Occupation	26%	50%	18%	4%	3%
Signs	6%	36%	41%	16%	2%
Self-reported task engagement					
	Not engaging	**A little engaging**	**Average**	**Somewhat engaging**	**Highly engaging**
Categories	2%	3%	14%	24%	33%
Groceries	3%	11%	21%	37%	29%
Face–Name–Occupation	2%	7%	18%	41%	33%
Signs	3%	5%	24%	33%	36%

Note: percentages may not equal 100% because of rounding; registry sample (n = 78).
Abbreviation: BRANCH, Boston Remote Assessment for Neurocognitive Health.

3.2 | Task feasibility: registry sample

Seventy-one percent of participants used smartphones/tablets. Remaining participants used a desktop/laptop. Rates of unusable data (as defined above) were low at 3% suggesting good feasibility (few task discontinuations). Rates of self-reported difficulty completing the task (2.97%) and having difficulty understanding task instructions (2.97%) were low (Table 2) further supporting feasibility. Self-reported technical difficulties were higher (15.84%) with the most common concern being finger tap response latency followed by slow-loading task images.

Participants reported that the Face–Name–Occupation and Groceries tasks were the most challenging tasks while Categories was least challenging (Table 2). Most participants found the tasks engaging (Table 2) suggesting good acceptability. Comparable feasibility data for the observational group was observed (Supp B in supporting information).

3.3 | BRANCH descriptive statistics: registry and observational samples

Psychometric properties of BRANCH are shown in Table 3. BRANCH required an average of 22±4.97 minutes. Mean accuracy on the BRANCH composite was 0.74±0.08 (range: 0.50–.92). The registry and observational groups performed comparably on the BRANCH composite (t = 1.67, $P = .096$), although there was a nonsignificant trend toward observational participants performing better compared to registry participants, which may be attributable to their enhanced familiarity with cognitive and digital test-taking.

Four of the six accuracy outcomes exhibited mean performance within 70% to 80%, consistent with our original goals. An exception, by design, was the Categories recall task, which was much less challenging (mean = 93%). Groceries recall was more challenging (mean = 53%) than initially intended. However, incorporating both an easier and more challenging task is useful for populations that may include both CN and MCI. There were no floor effects. Ceiling effects were observed for the Categories test (65% at ceiling) but otherwise minimal across other measures. Apart from the categories test, all BRANCH outcomes were normally distributed (defined by skewness ranging from −0.5 to 0.5; Table 3).

3.4 | BRANCH validity in relation to age: registry and observational samples

Older age was associated with worse BRANCH performance ($r = -0.190$, $P = .004$; Figure 2). The magnitude of the age effect on cognition was similar across samples (observational: $r = -0.262$, $P = .001$; registry: $r = -0.217$, $P = .055$). Across both cohorts, slower reaction time on BRANCH outcomes was associated with older age (Supp C in supporting information).

3.5 | BRANCH re-test reliability: registry sample

Registry participants re-taking BRANCH were of comparable age, sex, and education levels compared to the overall registry sample (Supp D in supporting information). Mean time between administrations was 3.6±.73 months. Re-test reliability was good (Figure 3; $r = 0.81$, $P < .001$).

TABLE 3 Psychometric properties of BRANCH (A), convergent/discriminant validity against paper and pencil cognitive measures (B), associations with amyloid and tau (C)

A: Psychometric properties (n = 234)							
	BRANCH composite	Categories Recall	Face–Name–Occupation Inference	Face–Name–Occupation Recognition	Groceries: Price Recognition	Groceries: Pattern Separation	Signs Test
% ceiling	0%	65%	6.40%	1.23%	2.13%	0%	0%
% floor	0%	0%	0%	0%	0%	0%	0%
Mean (% accuracy)	74	93	71	76	53	78	77
SD	8	13	18	13	19	9	7
Range	43	67	80	60	80	53	50
Skewness	−0.441	−2.69	−0.45	−0.23	0.16	−0.43	−0.49
B: Correlations (Pearson *r*) with Paper and Pencil Cognitive Measures (n = 160)							
	BRANCH composite	Categories Recall	Face–Name–Occupation Inference	Face–Name–Occupation Recognition	Groceries: Price Recognition	Groceries: Pattern Separation	Signs Test
PACC-5	0.617**	0.354**	0.441**	0.524**	0.278**	0.325**	0.255
LMDR	0.302**	0.119	0.241	0.305**	0.095	0.225	0.121
DSST	0.415**	0.298**	0.262**	0.382**	0.188	0.220	0.238
FCSRT	0.472**	0.303**	0.328**	0.351**	0.285**	0.148	0.104
MMSE	0.348**	0.234	0.322**	0.276**	0.043	0.169	0.207
CAT	0.575**	0.258	0.381**	0.484**	0.329**	0.361**	0.205
TMTA	−0.287	−0.169	−0.090	−0.246	−0.250	−0.137	−0.260
TMTB	−0.379**	−0.382**	−0.112	−.315**	−0.142	−0.266	−0.321**
C: Correlations (Pearson r) with biomarkers of amyloid and tau (n = 144 for PiB and n = 129 for FTP)							
	BRANCH composite	Categories Recall	Face–Name–Occupation Inference	Face–Name–Occupation Recognition	Groceries: Price Recognition	Groceries: Pattern Separation	Signs Test
Cortical amyloid (PiB)	−0.205, *P* = .007*	−0.215, *P* = .006*	−0.122, *P* = .078	−0.073, *P* = .199	−0.177, *P* = .019*	−0.142, *P* = .049*	0.050, *P* = .281
Entorhinal tau (FTP)	−0.178, *P* = .026*	−0.020, *P* = .414	−0.306, *P* = .000*	−0.163, *P* = .037*	−0.133, *P* = .073	0.117, *P* = .100	.219, *P* = .008

Notes: For B, Pearson *r*, two-tailed, **multiple-comparison corrected; $P < .001$; on all cognitive measures, higher score reflects better performance with the exception of TMTA and TMTB for which lower score reflects better performance. *For C*, analyses controlled for age. Results not multiple-comparison corrected; *$P < .05$.

Abbreviations: BRANCH, Boston Remote Assessment for Neurocognitive Health; CAT, category fluency ; DSST, Digital Symbol Substitution Test; FCSRT, free recall on the Free and Cued Selective Reminding Test Free Recall; FTP, flortaucipir; LMDR, Logical Memory Delayed Recall; MMSE, Mini-Mental State Examination; PACC-5, Preclinical Alzheimer Cognitive Composite; PiB, Pittsburgh compound B; TMT, Trail Making Test.

3.6 | BRANCH versus traditional cognitive measures: convergent and discriminant validity—observational sample

BRANCH composite performance was moderately correlated ($r = 0.617$, $P < .001$) with PACC-5 (Table 3, Figure 2). BRANCH exhibited convergent validity with memory measures whereby the composite was correlated with memory measures (LMDR: $r = 0.302$, $P < .001$; FCSRT: $r = 0.472$, $P < .001$). Additionally, each individual BRANCH test accuracy metric (apart from Signs) was correlated with the PACC-5 with the magnitude of the correlation ranging from small (Groceries Price Recognition; $r = 0.278$, $P < .001$) to medium (Face–Name–Occupation Recognition; $r = 0.524$, $P < .001$). Face–Name–Occupation Inference and Recognition were most broadly correlated with PACC-5 measures (positive correlations were observed for each PACC component), whereas Signs was least related to PACC-5 measures. No individual BRANCH measures were correlated with TMTA, providing evidence for discriminant validity. However, worse performance on the BRANCH composite and some subtests were associated with worse performance on a task of mental

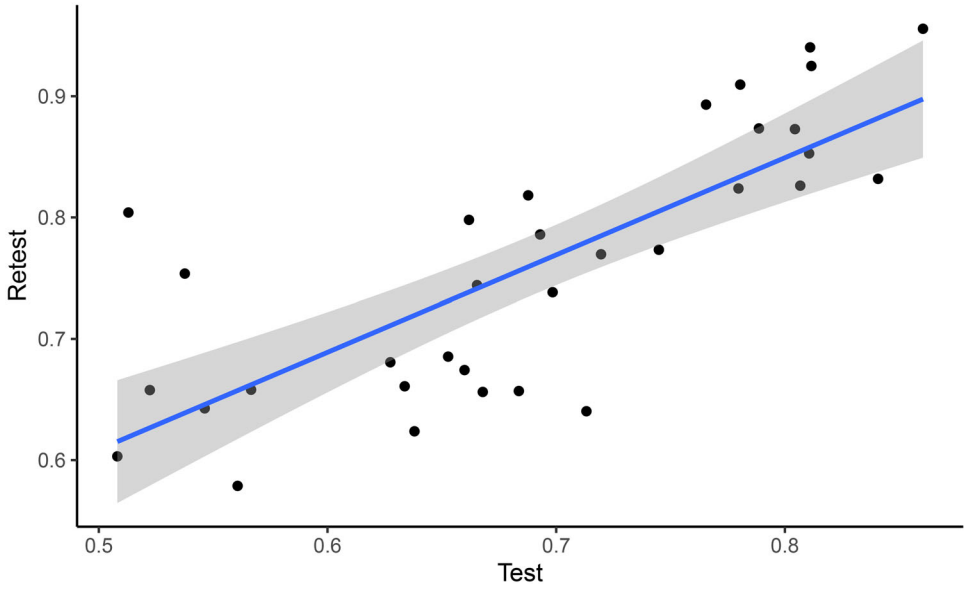

FIGURE 2 Boston Remote Assessment for Neurocognitive Health (BRANCH) Test-Retest Reliability. Note: The graph shows the correlation ($r = 0.81$, $P < .001$) between BRANCH composite performance at first (Test) and second (Retest) administration among 31 registry participants indicating good retest reliability

flexibility, TMTB. Reaction time outcomes for BRANCH tasks were generally correlated in the expected directions with DSST CAT, TMTA, and TMTB (Supp C).

3.7 | BRANCH versus AD biomarkers: observational sample

Among the observational sample with neuroimaging (Supp E in supporting information), lower BRANCH composite performance was associated higher cortical amyloid ($r = -0.205$, $P = .007$) and entorhinal tau ($r = -0.178$, $P = .026$; Figure 2; Table 3). As a point of comparison, the association between PACC-5 and cortical amyloid was $r = -0.191$, $P = .033$ and $r = -0.159$, $P = .076$ for entorhinal tau. For individual BRANCH tests, amyloid was negatively correlated with both Groceries test outcomes and Categories Recall. Greater entorhinal tau was associated with lower performance on both Face–Name–Occupation tasks. Signs was not related to amyloid, but was related to tau but in an unexpected direction. Examining BRANCH reaction time data, slower response speed across a few outcomes (Categories, Groceries Pattern Separation, Signs) was associated with higher amyloid (Supp E). No significant associations between reaction times and BRANCH tasks were observed for entorhinal tau.

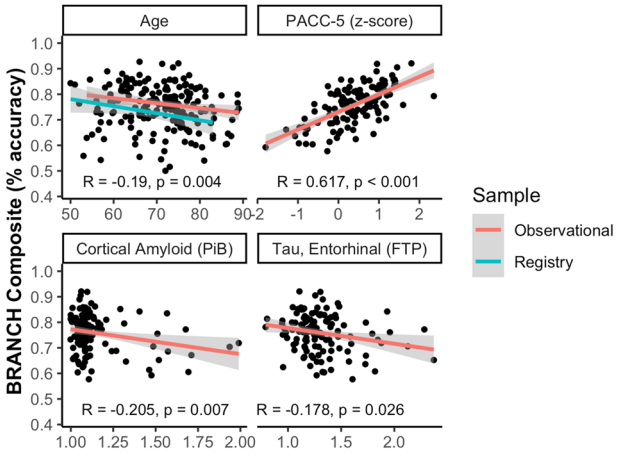

FIGURE 3 Associations between Boston Remote Assessment for Neurocognitive Health (BRANCH) composite performance and age, positron emission tomography amyloid and tau, and Preclinical Alzheimer Cognitive Composite (PACC-5) score. For age sample, n = 234; for PACC-5 sample, n = 160; for Pittsburgh Compound B (PiB) sample, n = 144; for flortaucipir (FTP) sample, n = 129; *r* values for PiB and FTP are controlled for age

3.8 | Feasibility of BRANCH: in-clinic MCI sample

Twenty-two MCI participants (mean age = 78.67; mean CDR Sum of Boxes = 1.41, range 0.5–4) were able to complete BRANCH in-clinic on an iPad with initial guidance by a rater (Supp A). Three participants discontinued because the task was too confusing/difficult. No floor effects were observed for those who completed BRANCH. MCI participants performed worse than an in-clinic CN sample ($t = 3.43$, $P < .01$, Cohen's $d = 0.45$).

4 | DISCUSSION

Here, we showed the feasibility of BRANCH to be deployed remotely in an unsupervised setting, both among participants well-versed in cognitive testing in an observational sample and a purely remote registry

sample. Few participants had invalid/unusable data (3%) or technical difficulties. BRANCH was feasible to complete on personal devices, with the majority using smartphones. Few discontinued and most participants found the tasks to be engaging.

Our results support the validity of BRANCH as an unsupervised web-based cognitive measure as evidenced by its moderate correlation ($r = 0.617$, $P < .001$) with traditional neuropsychological measures (PACC-5). BRANCH exhibited convergent validity with memory measures (i.e., LMDR, FCSRT) and discriminant validity with processing speed measures (i.e., TMTA). Correlations between BRANCH versus traditional memory tests ranged from $r = 0.302$ to 0.472, which is comparable, if not higher, compared to correlations observed in other studies with digital assessments.[45,46] For example, correlations between Cambridge Neuropsychological Test Automated Battery (CANTAB) memory measures and traditional memory measures ranged from 0.14 to 0.39[47] whereas a digital version of a list-learning test showed somewhat stronger correlations with the traditional task, exhibiting correlations ranging from .37 to .62.[48] However, it appears that the Signs Test, in its current configuration as a continuous learning paradigm, was not as strongly associated with gold standard cognitive measures; as such, we are considering modifying the paradigm in the future.

Most BRANCH tasks exhibited sound psychometric properties. Retest reliability was good ($r = 0.81$, $P < .001$). Apart from the Categories Test, all outcomes were normally distributed. Ceiling effects were observed for the less challenging Categories Test. Otherwise, average task difficulty fell within our goal of 70% accuracy (mean BRANCH composite accuracy score = 74%) and there was good range between the most challenging (Groceries Price Recognition; 53%) and least challenging tasks (Categories Recall; 93%). Having this range in difficulty for tasks is desirable for an instrument useful across those who are CN to early MCI. Further, as a basic measure of test validity, we found worse performance on BRANCH was associated with increasing age. Given well-known associations between older age and lower cognitive test performance,[26,49] this observed relationship provides further converging evidence that BRANCH is a valid cognitive measure.

Finally, worse BRANCH performance was associated with greater global amyloid ($r = -0.205$, $P = .007$) and entorhinal tau ($r = -0.178$, $P = .026$). These findings suggest that BRANCH captures memory performance that corresponds with biomarker burden in CN individuals who may be at risk for future disease progression. Small correlations are on par with other studies examining single timepoint performance on computerized tests and PET markers among CN.[4,50,51] To compare, a meta-analysis of studies of CN older adults found an association of 0.12 between traditional memory measures and PET amyloid.[52] Interestingly, associations between AD biomarkers were numerically stronger for the BRANCH composite versus the PACC-5, which may be attributable to the emphasis on sensitive memory measures in BRANCH.

We did not initially set out to examine reaction time data because web-based cognitive testing is confounded by internet speed/connectivity and inter-device variability. However, we were encouraged to see that slower reaction time on BRANCH tasks was associated with older age and slower performance on convergent processing speed measures (e.g., DSST, TMTB). These data suggest that reaction time data from a web-based program may be more promising than initially imagined and, in the future, may supplement accuracy measures.

4.1 | Limitations and future directions

A potential limitation of any unsupervised cognitive testing is whether the person assigned to the task is the person completing the task. To address this, we now ask participants to "attest" that they are the designated participant and are completing BRANCH unassisted. While there are possibilities to further guarantee fidelity with technology (e.g., recording the participant taking the test via web video[53]) the benefits of these strategies are outweighed at present by data privacy concerns. Additionally, a limitation of unsupervised cognitive testing is that contextual information (e.g., anxiety, boredom, task approach) is not collected by a rater/clinician. In future work, we plan to incorporate additional metrics (e.g., subjective report, environmental factors) to enrich and contextualize cognitive data.[54]

Our studies recruited individuals with their own digital devices who were comfortable with this technology, which is unlikely representative of the larger US older adult population. Another limitation of our study was the high (college) education level. However, we were able to achieve a balanced representation by sex (54% female) and a relatively racially diverse group (28% non-White registry sample).[55] Further work will be needed to determine feasibility of BRANCH in samples with lower educational attainment.

Finally, while we demonstrated the feasibility of BRANCH in MCI in-clinic, future work must confirm BRANCH feasibility among MCI in remote/unsupervised settings. Initial remote testing of BRANCH among MCI patients ($n < 5$) has been promising and we are continuing to collect this data. Additionally, we plan to examine to extent which BRANCH may track cognitive decline over time, whether BRANCH change predicts MCI progression, and the diagnostic utility of BRANCH in differentiating between CN and MCI. Future work will explore short-term learning curves by capturing BRANCH more frequently and developing BRANCH for use in different languages.

4.2 | Summary and conclusions

In this proof-of-concept study, we demonstrated that BRANCH provides reliable and meaningful cognitive data among CN individuals and shows small but persistent associations with AD biomarkers despite being deployed remotely/unsupervised. Digital capture of cognition, unlike traditional measures, has multiple benefits including increased accessibility, automated data capture and storage, scalability, and cost-effectiveness. There is a need in AD secondary prevention for digital cognitive tests which target fundamental aspects of early memory

decline and which can be completed on a personal device. Here, we provided evidence for the feasibility, reliability, and validity of BRANCH as a cognitive measure for use in preclinical AD.

ACKNOWLEDGMENTS

This work was funded through awards by the NIH/NIA and Alzheimer's Association including 5K23AG053422-03 (Papp), 1R01AG058825-01A (Amariglio), 2P01AG036694-11 (Sperling, Johnson), R01AG067021-01 (Marshall), a Massachusetts ADRC Pilot Award Grant, and the Davis Family Alzheimer's Prevention Project. We would like to thank Emma Weizenbaum, Paul Maruff, and Jason Hassenstab for discussions surrounding cognitive test development and Kevin Grinberg and Logan Benson for programming help.

CONFLICTS OF INTEREST

A. Samaroo, H.C. Chou, O.R. Schneider, D. Soberanes, M. Properzi, J. Burke, R. Kumar, S. Hsieh, N. Snyder, A. Schultz, Iván García-Magariño, R. Buckley, Y. Quiroz, and R. Amariglio report no disclosures relevant to this manuscript. K. Papp has served as a paid consultant for Biogen and Digital Cognition Technologies. G. Marshall has served as a paid consultant for Grifols Shared Services North America, Inc. and Eisai Inc. He has received honoraria/payments from Miller Medical, South Shore Hospital, and Metrowest Medical Center. D. Rentz has served as a paid consultant for Biogen, Digital Cognition Technologies, Eli Lilly, and Janssen. R. Sperling has received honoraria from Shionogi, Genentech, Oligomerix, Inc., Cytox, Prothena, Acumen, JOMDD, Renew, Alnylam Pharmaceuticals, Neuraly, Janssen, Neurocentria, AC Immune, Biogen, Eisai, Roche, and Takeda Pharmaceuticals. K. Johnson has received consulting fees from Novartis and Cerveau.

REFERENCES

1. Jack CR, 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.
2. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):280-292.
3. Doraiswamy PM, Narayan VA, Manji HK. Mobile and pervasive computing technologies and the future of Alzheimer's clinical trials. *NPJ Digit Med*. 2018;1(1):1-4.
4. Snitz BE, Tudorascu DL, Yu Z, et al. Associations between NIH Toolbox Cognition Battery and in vivo brain amyloid and tau pathology in non-demented older adults. *Alzheimers Dement (Amst)*. 2020;12(1):e12018-e12018.
5. Buckley RF, 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.
6. 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):65-75.
7. Rentz DM, 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):8-12.
8. Samaroo A, Amariglio R, Burnham S, et al. Diminished learning over repeated exposures (LORE) in preclinical Alzheimer's disease. *Alzheimers Dement (Amst)*. 2021;12(1):e12132.
9. Mackin RS, Insel PS, Truran D, et al. Unsupervised online neuropsychological test performance for individuals with mild cognitive impairment and dementia: results from the Brain Health Registry. *Alzheimers Dement (Amst)*. 2018;10:573-582.
10. Zygouris S, Tsolaki M. Computerized cognitive testing for older adults: a review. *Am J Alzheimers Dis Other Demen*. 2015;30(1):13-28.
11. Lancaster C, Koychev I, Blane J, et al. Gallery Game: smartphone-based assessment of long-term memory in adults at risk of Alzheimer's disease. *J Clin Exp Neuropsychol*. 2020(4):329-343.
12. Güsten J, Ziegler G, Düzel E, Berron D. Age impairs mnemonic discrimination of objects more than scenes: a web-based, large-scale approach across the lifespan. *cortex*. 2021;137:138-148.
13. Öhman F, Hassenstab J, Berron D, Schöll M, Papp KV. Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimers Dement (Amst)*. in press.
14. Hassenstab J, Aschenbrenner AJ, Balota DA, et al. Remote cognitive assessment approaches in the Dominantly Inherited Alzheimer Network (DIAN) Using digital technology to drive clinical innovation in brain-behavior relationships: a new era in neuropsychology. *Alzheimers Dement*. 2020;16:e038144.
15. Schlichting ML, Preston AR. Memory integration: neural mechanisms and implications for behavior. *Curr Opin Behav Sci*. 2015;1:1-8.
16. de Rover M, Pironti VA, McCabe JA, et al. Hippocampal dysfunction in patients with mild cognitive impairment: a functional neuroimaging study of a visuospatial paired associates learning task. *Neuropsychologia*. 2011(7):2060-2070.
17. Sperling R, Chua E, Cocchiarella A, et al. Putting names to faces: successful encoding of associative memories activates the anterior hippocampal formation. *Neuroimage*. 2003;20(2):1400-1410.
18. 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.
19. Carr VA, Bernstein JD, Favila SE, Rutt BK, Kerchner GA, Wagner AD. Individual differences in associative memory among older adults explained by hippocampal subfield structure and function. *Proc Natl Acad Sci*. 2017;114(45):12075-12080.
20. Van Strien N, Cappaert N, Witter M. The anatomy of memory: an interactive overview of the parahippocampal–hippocampal network. *Nat Rev Neurosci*. 2009;10(4):272-282.
21. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol*. 1991;82(4):239-259.
22. 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(12):2442-2449.
23. Berron D, Schütze H, Maass A, et al. Strong evidence for pattern separation in human dentate gyrus. *J Neurosci*. 2016;36(29):7569-7579.
24. Grober E, Lipton RB, Hall C, Crystal H. Memory impairment on free and cued selective reminding predicts dementia. *Neurology*. 2000;54(4):827-832.
25. Papp KV, Rentz DM, Mormino EC, et al. Cued memory decline in biomarker-defined preclinical Alzheimer disease. *Neurology*. 2017;88(15):1431-1438.
26. Boyle P, Yu L, Wilson R, Schneider J, Bennett DA. Relation of neuropathology with cognitive decline among older persons without dementia. *Front Aging Neurosci*. 2013;5:50.
27. Dagley A, LaPoint M, Huijbers W, et al. Harvard Aging Brain Study: dataset and accessibility. *Neuroimage*. 2015.
28. Dagley A, LaPoint M, Huijbers W, et al. Harvard Aging Brain Study: dataset and accessibility. *Neuroimage*. 2017;144(Pt B):255-258.
29. Papp KV, Buckley R, Mormino E, et al. Clinical meaningfulness of subtle cognitive decline on longitudinal testing in preclinical AD. *Alzheimers Dement*. 2020;16(3):552-560.
30. Rentz DM, Locascio JJ, Becker JA, et al. Cognition, reserve, and amyloid deposition in normal aging. *Ann Neurol*. 2010;67(3):353-364.

31. Armstrong K, Williams LE, Heckers S. Revised associative inference paradigm confirms relational memory impairment in schizophrenia. *Neuropsychology*. 2012;26(4):451-458.
32. Castel AD. Memory for grocery prices in younger and older adults: the role of schematic support. *Psychol Aging*. 2005;20(4):718-721.
33. Gallo DA, Sullivan AL, Daffner KR, Schacter DL, Budson AE. Associative recognition in Alzheimer's disease: evidence for impaired recall-to-reject. *Neuropsychology*. 2004;18(3):556-563.
34. Gauthier I, Tarr MJ. Becoming a "Greeble" expert: exploring mechanisms for face recognition. *Vision Res*. 1997;37(12):1673-1682.
35. Azab M, Stark SM, Stark CE. Contributions of human hippocampal subfields to spatial and temporal pattern separation. *Hippocampus*. 2014;24(3):293-302.
36. Maruff P, Collie A, Darby D, Weaver-Cargin J, Masters C, Currie J. Subtle memory decline over 12 months in mild cognitive impairment. *Dement Geriatr Cogn Disord*. 2004;18(3-4):342-348.
37. Salthouse TA. Aging associations: influence of speed on adult age differences in associative learning. *J Exp Psychol Learn Mem Cogn*. 1994;20(6):1486-1503.
38. Ashford JW, Tarpin-Bernard F, Ashford CB, Ashford MT. A computerized continuous-recognition task for measurement of episodic memory. *J Alzheimers Dis*. 2019;69(2):385-399.
39. Sundermann EE, Biegon A, Rubin LH, Lipton RB, Landau S, Maki PM. Does the Female Advantage in Verbal Memory Contribute to Underestimating Alzheimer's Disease Pathology in Women versus Men?. *J Alzheimers Dis*. 2017;56(3):947-957.
40. Parker C, Philp I. Screening for cognitive impairment among older people in black and minority ethnic groups. *Age Ageing*. 2004;33(5):447-452.
41. Papp KV, Rentz DM, Orlovsky I, Sperling RA, Mormino EC. Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: the PACC5. *Alzheimers Dement (N Y)*. 2017;3(4):668-677.
42. Mormino EC, Papp KV, Rentz DM, et al. Early and late change on the preclinical Alzheimer's cognitive composite in clinically normal older individuals with elevated amyloid beta. *Alzheimers Dement*. 2017;13(9):1004-1012.
43. Johnson KA, Schultz A, Betensky RA, et al. Tau positron emission tomographic imaging in aging and early A lzheimer disease. *Ann Neurol*. 2016;79(1):110-119.
44. Guttman L. A basis for analyzing test-retest reliability. *Psychometrika*. 1945;10(4):255-282.
45. Gills JL, Glenn JM, Madero EN, Bott NT, Gray M. Validation of a digitally delivered visual paired comparison task: reliability and convergent validity with established cognitive tests. *GeroScience*. 2019;41(4):441-454.
46. Dorociak KE, Mattek N, Lee J, et al. The Survey for Memory, Attention, and Reaction Time (SMART): development and Validation of a Brief Web-Based Measure of Cognition for Older Adults. *Gerontology*. 2021:1-13.
47. Smith PJ, Need AC, Cirulli ET, Chiba-Falek O, Attix DK. A comparison of the Cambridge Automated Neuropsychological Test Battery (CANTAB) with "traditional" neuropsychological testing instruments. *J Clin Exp Neuropsychol*. 2013;35(3):319-328.
48. Morrison RL, Pei H, Novak G, et al. A computerized, self-administered test of verbal episodic memory in elderly patients with mild cognitive impairment and healthy participants: a randomized, crossover, validation study. *Alzheimers Dement (Amst)*. 2018:647-656.
49. Salthouse TA. When does age-related cognitive decline begin?. *Neurobiol Aging*. 2009;30(4):507-514.
50. Papp KV, Rentz DM, Maruff P, et al. The computerized cognitive composite (c3) in a4, an alzheimer's disease secondary prevention trial. *J Prev Alzheimers Dis*. 2021;8(1):59-67.
51. Bischof GN, Rodrigue KM, Kennedy KM, Devous MD, Park DC. Amyloid deposition in younger adults is linked to episodic memory performance. *Neurology*. 2016(24):2562-2566.
52. 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.
53. Lee J, Kim RJ, Park SY, Henning MA. Using technologies to prevent cheating in remote assessments during the COVID-19 pandemic. *J Dent Educ*. 2020.
54. Weizenbaum E, Torous J, Fulford D. Cognition in context: understanding the everyday predictors of cognitive performance in a new era of measurement. *JMIR Mhealth Uhealth*. 2020;8(7):e14328.
55. Langbaum JB, High N, Nichols J, Kettenhoven C, Reiman EM, Tariot PN. The Alzheimer's Prevention Registry: a large internet-based participant recruitment registry to accelerate referrals to Alzheimer's-focused studies. *J Prev Alzheimers Dis*. 2020;7:242-250.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Papp KV, Samaroo A, Chou H-C, et al. Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. *Alzheimer's Dement*. 2021;13:e12243. https://doi.org/10.1002/dad2.12243