






## Research Article

# Stricker Learning Span criterion validity: a remote self-administered multi-device compatible digital word list memory measure shows similar ability to differentiate amyloid and tau PET-defined biomarker groups as in-person Auditory Verbal Learning Test

Nikki H. Stricker<sup>1</sup> , John L. Stricker<sup>2</sup>, Ryan D. Frank<sup>3</sup>, Winnie Z. Fan<sup>3</sup>, Teresa J. Christianson<sup>3</sup>, Jay S. Patel<sup>1</sup>, Aimee J. Karstens<sup>1</sup> , Walter K. Kremers<sup>3</sup>, Mary M. Machulda<sup>1</sup>, Julie A. Fields<sup>1</sup> , Jonathan Graff-Radford<sup>4</sup>, Clifford R. Jack Jr.<sup>5</sup>, David S. Knopman<sup>4</sup>, Michelle M. Mielke<sup>6</sup> and Ronald C. Petersen<sup>4</sup>

<sup>1</sup>Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA, <sup>2</sup>Department of Information Technology, Mayo Clinic, Rochester, MN, USA, <sup>3</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA, <sup>4</sup>Department of Neurology, Mayo Clinic, Rochester, MN, USA, <sup>5</sup>Department of Radiology, Mayo Clinic, Rochester, MN, USA and <sup>6</sup>Department of Epidemiology and Prevention, Wake Forest University School of Medicine, Winston-Salem, NC, USA

### Abstract

**Objective:** The Stricker Learning Span (SLS) is a computer-adaptive digital word list memory test specifically designed for remote assessment and self-administration on a web-based multi-device platform (Mayo Test Drive). We aimed to establish criterion validity of the SLS by comparing its ability to differentiate biomarker-defined groups to the person-administered Rey's Auditory Verbal Learning Test (AVLT). **Method:** Participants ( $N = 353$ ; mean age = 71, SD = 11; 93% cognitively unimpaired [CU]) completed the AVLT during an in-person visit, the SLS remotely (within 3 months) and had brain amyloid and tau PET scans available (within 3 years). Overlapping groups were formed for 1) those on the Alzheimer's disease (AD) continuum (amyloid PET positive, A+,  $n = 125$ ) or not (A-,  $n = 228$ ), and those with biological AD (amyloid and tau PET positive, A+T+,  $n = 55$ ) vs no evidence of AD pathology (A-T-,  $n = 195$ ). Analyses were repeated among CU participants only. **Results:** The SLS and AVLT showed similar ability to differentiate biomarker-defined groups when comparing AUROCs ( $p's > .05$ ). In logistic regression models, SLS contributed significantly to predicting biomarker group beyond age, education, and sex, including when limited to CU participants. Medium (A- vs A+) to large (A-T- vs A+T+) unadjusted effect sizes were observed for both SLS and AVLT. Learning and delay variables were similar in terms of ability to separate biomarker groups. **Conclusions:** Remotely administered SLS performed similarly to in-person-administered AVLT in its ability to separate biomarker-defined groups, providing evidence of criterion validity. Results suggest the SLS may be sensitive to detecting subtle objective cognitive decline in preclinical AD.

**Keywords:** mobile health; Alzheimer's disease; neuropsychology; aging; cognition; telemedicine; smartphone; neuropsychological tests

(Received 14 November 2022; final revision 10 March 2023; accepted 26 April 2023)

### Introduction

The need for efficient and scalable approaches for identifying individuals at risk for preclinical and prodromal Alzheimer's disease (PAD) is paramount to ongoing clinical trial efforts, emerging decentralized trials, and for identifying individuals who will most benefit from currently available pharmacologic or behavioral treatments, or those on the horizon (Cummings et al.,

2021; Dorsey et al., 2020). Self-administered cognitive measures that can be completed "remotely" (i.e., outside of a typical clinical setting, including at home) are a critical component of an early PAD detection strategy since they require fewer resources to administer and provide easier access to cognitive screening compared to person-administered measures (Ashford et al., 2021; Papp et al., 2021; Sabbagh et al., 2020; Sabbagh et al.,

**Corresponding author:** Nikki H. Stricker; Email: [stricker.nikki@mayo.edu](mailto:stricker.nikki@mayo.edu)

Manuscript Tables, Figures and Supplementary Material used with permission of Mayo Foundation for Medical Education and Research; all rights reserved. A preprint of this manuscript is available: medRxiv 2022.10.26.22281534; doi: <https://doi.org/10.1101/2022.10.26.22281534>. A paper presentation of this work was presented at International Neuropsychological Society conference (Feb 2023). Early data from this work with a smaller sample size was also presented as a paper presentation at the July 2022 Alzheimer's Association International Conference.

**Cite this article:** Stricker N.H., Stricker J.L., Frank R.D., Fan W.Z., Christianson T.J., Patel J.S., Karstens A.J., Kremers W.K., Machulda M.M., Fields J.A., Graff-Radford J., Jack C.R., Jr., Knopman D.S., Mielke M.M., & Petersen R.C. Stricker Learning Span criterion validity: a remote self-administered multi-device compatible digital word list memory measure shows similar ability to differentiate amyloid and tau PET-defined biomarker groups as in-person Auditory Verbal Learning Test. *Journal of the International Neuropsychological Society*, 1–14, <https://doi.org/10.1017/S1355617723000322>

Copyright © INS. Published by Cambridge University Press 2023. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

2020); for a broader review of digital cognitive assessment for preclinical AD see Ohman *et al.* (2021). Frequently, tests originally designed for and validated within clinic settings are converted to remote use to increase access for those unable to readily visit research centers or due to necessity during the COVID-19 pandemic (Bauer & Bilder, 2023, *in press*; Mackin *et al.*, 2021; Marra *et al.*, 2020). The limitation of this “conversion” approach is that tests are not developed specifically with remote self-administration as a priority for test design decisions, which can contribute to mixed findings when performance is compared across settings (Cromer *et al.*, 2015; Mielke *et al.*, 2015; Stricker, Lundt, Alden, *et al.*, 2020). There is an urgent need for valid self-administered cognitive assessment tools designed specifically for remote use. Verbal memory measures are among the most sensitive to early changes in the Alzheimer’s disease (AD) process (Caselli *et al.*, 2020) but are also challenging to adapt to remote, self-administered methods (Bauer & Bilder, *in press*).

The Stricker Learning Span (SLS) is a digital computer-adaptive word list memory test specifically designed for remote assessment (Stricker *et al.*, 2022). The SLS is administered via a web-based multi-device platform designed for unsupervised self-administration of digital cognitive tests, Mayo Test Drive (MTD): Mayo Test Development through Rapid Iteration, Validation and Expansion (DRIVE) (Stricker *et al.*, 2022). Recent work has highlighted learning as a key deficit in PAD, conceptualized as a failure to benefit from repeated exposure (Lim *et al.*, 2020) or a lack of benefit from practice (Duff *et al.*, 2017; Machulda *et al.*, 2017). In line with this, the SLS was designed to emphasize learning. The SLS paradigm was influenced by cognitive science principles and neural network process simulations (Stricker *et al.*, 2022). The SLS stresses the contextual system during learning through use of high frequency word stimuli and variations in word item-level imagery to increase difficulty.

Preliminary support for the feasibility, reliability, and validity of the SLS was previously reported in an all-female older adult sample (Stricker *et al.*, 2022). Whereas that prior study used traditional approaches to test validation, the current study aimed to establish test validity using a novel approach to avoid the inherent issues with existing validation approaches. For example, one common approach is to correlate a new test with existing cognitive tests. However, existing tests, while well established, are themselves imperfect measures of hypothesized underlying constructs (Bilder & Reise, 2019). Another frequent approach is to establish validity by examining the ability of a new test to differentiate clinically defined groups. In the AD field, for example, it is common to establish the clinical validity of a new test by comparing individuals who are cognitively “normal” or unimpaired to individuals with mild cognitive impairment (MCI) or dementia; however, this introduces circularity because the use of cognitive tests is central to establishing those syndromal classifications. *In vivo* biomarkers offer an alternative ground truth for test validation studies that is completely independent of cognitive test performance. This is akin to validation with neuropathological diagnosis at autopsy given the correspondence between antemortem PET imaging and autopsy findings but has the notable benefit of being feasible during life (Chiotis *et al.*, 2017; Wolters *et al.*, 2021). A research framework is now available to use AD biomarkers to characterize participants using the amyloid (A), tau (T) and neurodegeneration (N), or the AT(N), system (Jack *et al.*, 2018). Imaging biomarkers of N are considered nonspecific to AD and will not be included in the current manuscript to limit the number of subgroups. Individuals with evidence of elevated amyloid (A+) are considered to show

Alzheimer’s pathologic change. An *in vivo* biological diagnosis of AD is defined by the presence of both A+ and elevated tau (T+).

The objective of this study was to determine the criterion validity of the SLS. Critically, this validation study was limited to unsupervised completion of the SLS in a remote environment outside of a typical clinical research setting. Our primary study hypothesis (Aim 1) was that remotely administered SLS and in-person-administered Rey’s Auditory Verbal Learning Test (AVLT) would differentiate AD biomarker-defined groups similarly. This hypothesis is tested in groups defined by biomarker status alone (A+ vs A– and A+T+ vs A–T–) to avoid circularity. That is, because the AVLT is considered for diagnostic decision-making as part of the consensus diagnosis process for study participants, it is important that the primary AVLT vs. SLS comparison is independent of diagnosis. Secondary hypotheses included that the SLS would be sensitive to preclinical AD in analyses limited to CU participants (Aim 2), SLS and AVLT would show significant correlations to support convergent validity (Aim 3), and that word list learning vs. delay indices would show similar sensitivity to biologically defined AD (A–T– vs A+T+; Aim 4).

## Methods

Most participants were recruited from the Mayo Clinic Study of Aging (MCSA), a longitudinal population-based study of aging among Olmsted County, Minnesota, residents. Participants are randomly sampled by age- and sex-stratified groups using the resources of the Rochester Epidemiology Project medical records-linkage system, which links the medical records from all county providers (St Sauver *et al.*, 2012). Participants with dementia are not eligible for MCSA enrollment. Participants complete study visits every 15 months that include a physician exam, study coordinator interview, and neuropsychological testing (Roberts *et al.*, 2008). The physician exam includes a medical history review, complete neurological exam, and the Short Test of Mental Status (STMS) (Kokmen *et al.*, 1991). The study coordinator interview with an informant includes the Clinical Dementia Rating® scale (Morris, 1993). Participants complete a multi-domain battery of nine neuropsychological tests administered by a psychometrist (Roberts *et al.*, 2008). The interviewing study coordinator, examining physician, and neuropsychologist initially each make an independent diagnostic determination. A final diagnosis of cognitively unimpaired, MCI (Petersen, 2004) or dementia (American Psychiatric Association, 1994) is then established through consensus agreement (Petersen, 2004; Roberts *et al.*, 2008). The diagnostic evaluation does not consider prior clinical information, prior diagnoses, SLS performance, or knowledge of biomarker status. Further details about the MCSA study protocol are available (Roberts *et al.*, 2008).

To enrich the sample for participants with cognitive impairment, additional participants were recruited from the Mayo Alzheimer’s Disease Research Center (ADRC).

This study was completed in accordance with the Helsinki Declaration. The study protocols were approved by the Mayo Clinic and Olmsted Medical Center Institutional Review Boards. All participants provided written informed consent for the primary study protocols (MCSA, ADRC); oral consent (which includes consent provided after reading informed consent elements sent in an email or described verbally) was obtained for the ancillary Mayo Test Drive study protocol approved by Mayo Clinic that covered collection of remote cognitive assessment data. No compensation was provided for participation in the ancillary study.

### ***In vivo neuroimaging markers of amyloid and tau***

The most recent imaging available  $\pm 3$  years of baseline SLS was used. Amyloid and tau positivity is determined using Pittsburgh Compound B PET (PiB-PET) and tau PET (flortaucipir) (Jack et al., 2008; Jack et al., 2017; Vemuri et al., 2017). PET images are acquired using a GE Discovery RX or DXT PET/CT scanner. A global cortical PiB PET standard uptake value ratio (SUVR) is computed by calculating the median uptake over voxels in the prefrontal, orbitofrontal, parietal, temporal, anterior cingulate, and posterior cingulate/precuneus regions of interest (ROIs) for each participant and dividing this by the median uptake over voxels in the cerebellar crus gray matter. For tau PET, we utilize median uptake over the voxels in the meta regions consisting of entorhinal, amygdala, parahippocampal, fusiform, inferior temporal, and middle temporal ROIs normalized to the cerebellar crus gray matter (Jack et al., 2017). Cutoffs to determine amyloid and tau positivity are  $SUVR \geq 1.48$  (centiloid 22) (Klunk et al., 2015) and  $\geq 1.25$  (Jack et al., 2017) respectively, to maintain consistency with our past Cogstate-focused work (Alden et al., 2021; Pudumjee et al., 2021; Stricker, Lundt, Albertson, et al., 2020).

### ***Person-administered AVLT completed in clinic***

The psychometrist reads a list of 15 words (List A) aloud and asks the participant to repeat back as many words as they can recall. This is repeated 5 times (learning trials 1–5 total). A distractor list (List B) is presented, followed by short delay (Trial 6) of list A words. Recall of List A is again tested after 30 minutes (30-minute delay), followed by written recognition (Ferman et al., 2005; Stricker, Christianson, et al., 2020). The primary variable for this study is AVLT sum of trials (trials 1–5 total + trial 6 + 30-min recall; range 0–105), which is sensitive to early changes in memory (Jack et al., 2015). Additional variables include correct words on trials 1–5 total (thought to reflect learning), as well as 30-minute delay. Long-term percentage retention (AVLT 30-minute delay / trial 5), thought to reflect storage/savings, is also reported.

### ***Self-administered SLS completed remotely (not in clinic)***

All participants completed the SLS remotely and without supervision or assistance. Participants followed a link provided in an email to complete the test session. The SLS is administered via the Mayo Test Drive (MTD) platform (Stricker et al., 2022).

The SLS is a 5-trial adaptive list learning task (Figure 1). Single words are visually presented sequentially during learning trials. After each list presentation, memory for the word list is tested with 4-choice recognition. Following a computer adaptive testing approach, the SLS starts with eight items and then the number of words either stays the same, increases by five or decreases by two according to pre-specified rules based on percentage of correct responses to extend the floor and ceiling relative to traditional word list memory tests (range 2–23 words; Figure 2). Short delay follows the Symbols Test (Nicosia et al., 2023; Stricker et al., 2022); all items presented on any learning trial are tested during delay (range 8–23). Select screenshots have been previously published (Stricker et al., 2022).

The SLS uses common, high frequency words that are easier to recall, but harder to recognize (Lohnas & Kahana, 2013), as previously described (Stricker et al., 2022). There are 23 item bins with 4 words each, and words within a bin have similar imagery ratings (Clark & Paivio, 2004). Each successive item bin has lower

imagery ratings, thus increasing the difficulty of subsequent items. Most (90%) of the 92 total words used are on the Dolch sight words reading list (preschool through Grade 3), with half at the preschool level (Dolch, 1936). Each test session randomly selects 1 word from each item bin as the “target,” the 3 others serve as the foils and a 23-item target word list is generated, even if not all items are presented due to the adaptive procedure. To reduce recency effects, the order of item presentation is randomized for each trial and the last item presented is never the first tested. The primary variable is SLS sum of trials (total correct across learning trials 1–5 plus delay, range 0–108). Secondary variables include maximum (max) learning span across any learning trial (range 0–23), total correct across learning trials (1–5 total, range 0–85), and total correct short delay (range 0–23). Percent retention (max span/delay) is also reported to allow within-test comparisons, but this measure is not meant to be compared to AVLT percent retention as differences are expected based on differences in test design.

### ***Inclusion criteria***

To be included in this study, participants had to have both SLS sum of trials and AVLT sum of trials available and an amyloid PET scan within 3 years. All but two participants also had a tau PET scan available within 3 years. Participants who completed the SLS as of 7/7/22 were included in this study. Parent study data available as of 8/22/22 were included.

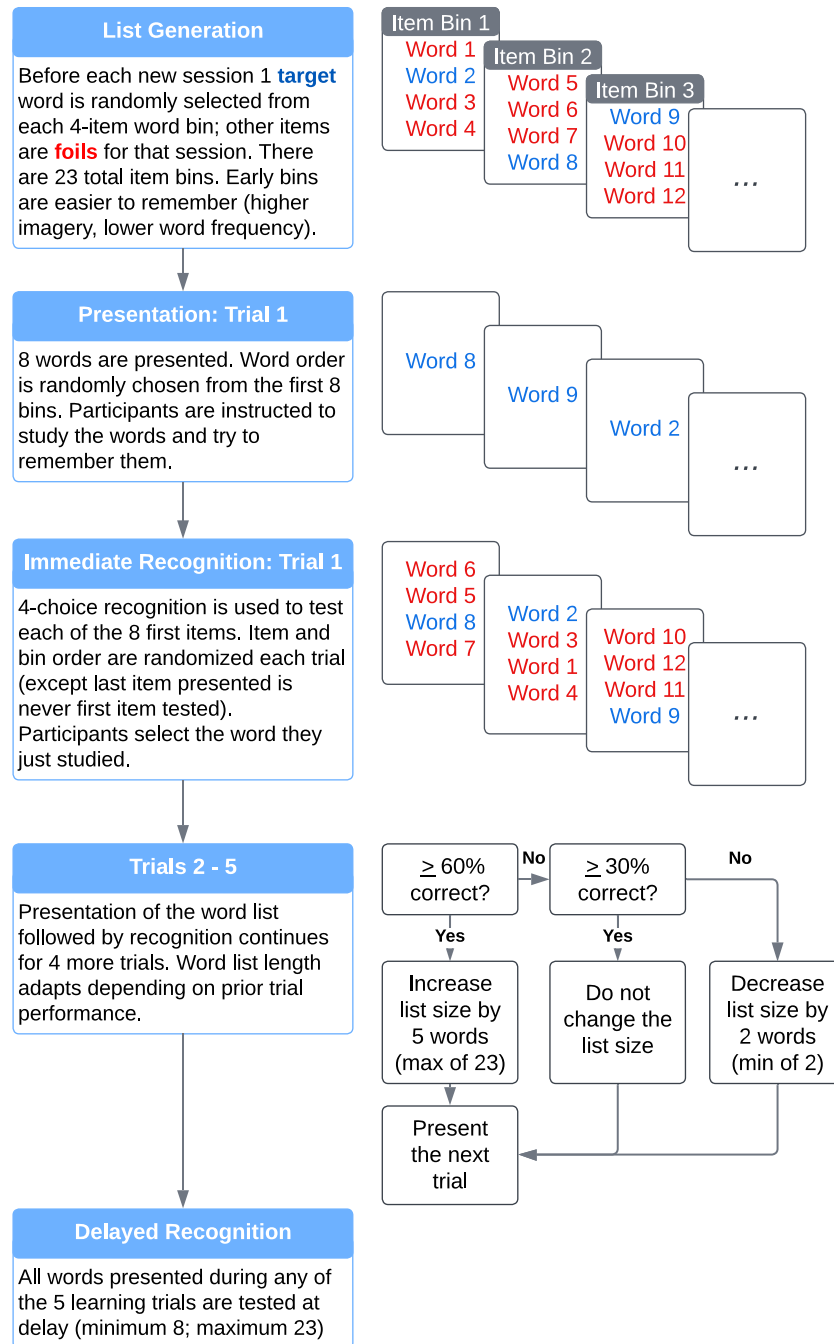
### ***Statistical methods***

Demographics and clinical characteristics were descriptively summarized using counts and percentages for categorical data and means and standard deviations for continuous data. Data distributions across groups (A– vs A+ and A–T– vs A+T+) were compared using chi-square tests for categorical variables and linear model ANOVA tests for continuous variables. Pearson correlation coefficients were used to characterize the linear relationship between AVLT and SLS variables. Unadjusted and adjusted Hedge’s G with weighted and pooled standard deviation was used to assess effect size for group comparisons. Unadjusted and adjusted logistic regression models were used to determine the predictive accuracy of AVLT and SLS sum of trials in predicting abnormal amyloid PET (A+ vs A–) and abnormal amyloid and tau PET (A+T+ vs A–T–). To formally compare the ability of both tests to differentiate biomarker-defined groups, the AUROC from models with AVLT were directly compared to models with SLS (Therneau, 2021). For models that adjusted for demographic variables, age, sex, and education were the adjustment terms for both Hedge’s G and logistic regression models. A two-sided  $p$ -value  $< 0.05$  was considered statistically significant. All analyses were performed using R version 4.1.2.

## **Results**

### ***Participant characteristics***

The mean age of the 353 participants was 71.8 (SD = 10.8) years, mean education was 15.7 (SD = 2.4), 53.5% were male, 98.0% were White and 92.6% were cognitively unimpaired (Table 1). MTD remote testing was completed within half a month (mean) of the in-person visit. Participant characteristics by biomarker subgroups are reported in Table 2. As expected, based on known increases in A+ and T+ rates with increased age (Jack et al., 2008; Jack et al., 2017; Vemuri et al., 2017), biomarker positive groups were older



**Figure 1.** Stricker Learning Span methods. Figure used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

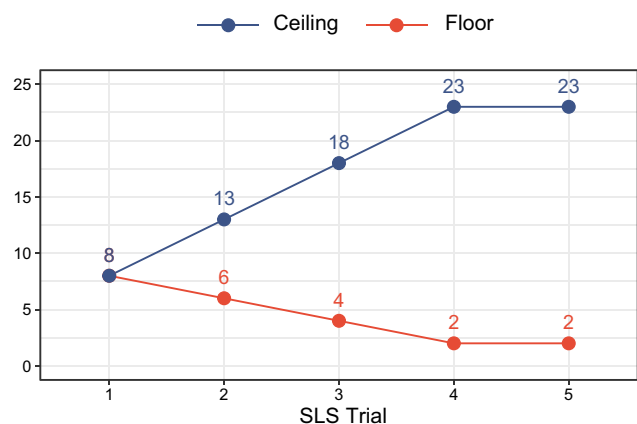
than biomarker negative groups ( $p$ 's < .05). Biomarker positive and negative groups showed similar years of education and sex distribution ( $p$ 's > .05).

**Aim 1: SLS shows similar ability to differentiate PET-defined biomarker groups compared to the AVLT (all participants)**

**AUROC comparisons.** Total AUROC values for SLS sum of trials vs. AVLT sum of trials were similar for differentiating biomarker groups ( $p$ 's > .05; Table 3, Figure 3). This similarity was seen for all pairwise AUROC comparisons (adjusted and unadjusted models; A- vs A+ and A-T- vs A+T+). These four direct AUROC comparisons

support our hypothesis and represent the primary test of Aim 1 given that inclusion of all available participants limits concerns about circularity present when the sample is restricted to CU individuals only.

**Amyloid groups.** Unadjusted models using only the primary cognitive variable as the predictor show that both the SLS and AVLT significantly differentiate A- vs A+ (AUROCs of 0.63 and 0.64, respectively). Adjusted models that include demographic variables increase the overall AUROC values of the full model (both AUROCs = 0.76), and both the SLS and AVLT significantly improve biomarker group prediction over and above the demographic variables.



**Figure 2.** Stricker Learning Span (SLS) computer adaptive testing approach provides an expanded ceiling and floor of items presented relative to traditional word list memory tests. High performers are exposed to an increasing number of words across trials as shown in blue. Low performers are shown a decreasing number of words across each trial as shown in orange. Figure used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

**Amyloid and tau groups.** Unadjusted models using only the primary cognitive variable as the predictor show that both the SLS and AVLT significantly differentiate individuals without AD biomarkers (A–T–) from those with biological AD (A+T+; AUROCs of 0.72–0.73). Adjusted models that include demographic variables increase the overall AUROC values of the full model (AUROCs of 0.83–0.84), and both the SLS and AVLT significantly improve biomarker group prediction over and above the demographic variables.

**Descriptive effect sizes from group difference analyses.** We report effect sizes from mean group comparisons to additionally characterize the magnitude of these effects for both unadjusted and adjusted models (see Figure 4 and Supplemental Table 2). The pattern of results for these parametric analyses is similar to that of the non-parametric AUROC analyses.

### Aim 2: SLS shows sensitivity to preclinical AD (CU participants only)

Findings show that the SLS is sensitive to preclinical AD, consistent with our Aim 2 hypothesis.

**AUROC comparisons.** Total AUROC values for SLS sum of trials vs. AVLT sum of trials were similar for differentiating biomarker groups in CU participants ( $p$ 's > .05 for each pairwise comparison; Table 3). Note that direct comparisons of SLS and AVLT should be viewed cautiously when results are limited to CU participants given that AVLT has some circularity with diagnosis (is considered by the neuropsychologist in conjunction with eight other in-person neuropsychological tests, and then discussed in consensus meeting) whereas the SLS is independent of diagnosis (results are not available to consensus team members). Thus, SLS results for analyses limited to CU participants can be readily interpreted. AVLT results are presented for reference, but circularity may impact findings.

**Amyloid groups.** Unadjusted models using only the primary cognitive variable as the predictor show that both the SLS and AVLT significantly differentiate CU A– vs CU A+ (both AUROCs = 0.63). Adjusted models that include demographic variables increase the overall AUROC values of the full model

**Table 1.** Participant demographic, clinical and neuroimaging characteristics and devices used for Mayo Test Drive (mean, SD except where otherwise noted) for all participants ( $N = 353$ )

	Mean (SD)	Range
Age at MCSA or ADRC visit <sup>1</sup>	70.84 (10.82)	35–94
In-person visit to MTD, months	0.55 (0.42)	–0.30 – 2.99
In-person visit to imaging, months	–0.78 (1.03)	–2.83 – 0.32
Sex, N (%) Male	189 (53.5%)	–
Education, years	15.72 (2.36)	11–20
Race (% White) <sup>2</sup>	346 (98.0%)	–
Ethnicity (% Non-Hispanic) <sup>3</sup>	351 (99.7%)	–
Kokmen Short Test of Mental Status <sup>4</sup>	35.6 (2.4)	22–38
N (%) first MCSA or ADRC visit (visit # range)	60 (17.0%)	1–14
Diagnosis		–
Cognitively Unimpaired (N, %)	326 (92.6%)	–
MCI (N, %)	21 (6.0%)	–
Dementia (N, %)	5 (1.4%)	–
Unavailable (N, %) <sup>5</sup>	1 (0.3%)	–
CDR Global score		–
0 (N, %)	332 (94.1%)	–
0.5 (N, %)	20 (5.7%)	–
1 (N, %)	1 (0.3%)	–
Subjective memory concerns (N, % Yes) <sup>6</sup>	201 (58.3%)	–
Amyloid meta-ROI	1.54 (0.35)	1.16 – 3.31
Tau meta-ROI <sup>7</sup>	1.21 (0.13)	0.90 – 2.31
Entorhinal Tau <sup>7</sup>	1.13 (0.16)	0.78 – 2.30
SLS sum of trials	74.06 (18.27)	16–108
AVLT sum of trials	64.38 (18.77)	17–105
SLS Max Span	16.43 (4.18)	3–23
AVLT Trial 5	11.06 (2.75)	0–15
SLS 1-5 Total	59.45 (14.07)	13–85
AVLT 1-5 Total	45.99 (12.11)	15–75
SLS Delay	14.60 (4.61)	3–23
AVLT 30-min delay	9.06 (3.98)	0–15
SLS Retention	88.28 (15.11)	36.4–133.3
AVLT Retention <sup>8</sup>	79.03 (27.80)	0–214.3
AVLT Recognition Percent Correct	91.31 (9.34)	53–100
Mayo Test Drive device type used		–
Desktop computer or laptop (N, %)	220 (62.3%)	–
Smartphone (N, %)	95 (26.9%)	–
Tablet (N, %)	37 (10.5%)	–
Other / Not Sure (N, %)	1 (0.3%)	–

Note. ADRC = Mayo Alzheimer's Disease Research Center; AVLT = Auditory Verbal Learning Test; AVLT Sum of Trials = AVLT 1–5 total + Trial 6 + 30-minute delay; AVLT Recognition Percent Correct =  $\{[\text{recognition hits} + (15 - \text{recognition false positive errors})] / 30\} \times 100$ ; AVLT Retention = AVLT 30-minute delay / Trial 5; CDR = Clinical Dementia Rating Scale; MCSA = Mayo Clinic Study of Aging; MTD = Mayo Test Drive; SLS = Stricker Learning Span; SLS Max Span = maximum number of words recognized across any learning trial; SLS 1–5 Total = sum of words correctly recognized across trials 1–5; SLS Retention = SLS Delay / SLS Max Span; SLS Sum of Trials = SLS 1–5 total + delay. Table used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

<sup>1</sup> $n = 8$  Mayo Alzheimer's Disease Research Center (ADRC);  $n = 345$  Mayo Clinic Study of Aging (MCSA)

<sup>2</sup> $n = 2$  Asian,  $n = 3$  Black,  $n = 2$  More than one

<sup>3</sup> $n = 1$  missing

<sup>4</sup> $n = 1$  missing

<sup>5</sup>consensus diagnosis not yet available for this participant; CDR = 0; amyloid and tau PET negative status.

<sup>6</sup> $n = 8$  missing; subjective memory concern is "yes" when subjective memory complaints are on the Blessed Memory Test questions 1–4 is marked as worse or question 5 indicates any other problems with thinking or memory.

<sup>7</sup> $n = 2$  missing

<sup>8</sup> $n = 1$  missing due to AVLT Trial 5 score of 0.

(AUROCs = 0.76–0.77), and both the SLS and AVLT significantly improve biomarker group prediction over and above the demographic variables.

**Amyloid and tau groups.** Unadjusted models using only the primary cognitive variable as the predictor show that both the SLS and AVLT significantly differentiate CU individuals without AD biomarkers (A–T–) from CU participants with biological AD (A+T+; AUROCs = 0.67–0.69). Adjusted models that include

**Table 2.** Demographic and clinical characteristics of biomarker subgroups and means (SDs) for cognitive variables

	A- vs A+				A-T- vs A+T+			
	All participants		CU participants only		All participants		CU participants only	
	A- (N = 228)	A+ (N = 125)	CU A- (N = 215)	CU A+ (N = 111)	A-T- (n = 195)	A+T+ (n = 55)	CU A-T- (n = 185)	CU A+T+ (n = 42)
Age at MCSA visit	67.6 (10.6)	76.8 (8.4)	67.3 (10.6)	76.7 (8.4)	66.3 (10.3)	77.1 (8.7)	66.1 (10.3)	77.3 (8.8)
Sex, N (%) Male	119 (52.2%)	70 (56.0%)	111 (51.6%)	63 (56.8%)	102 (52.3%)	32 (58.2%)	96 (51.9%)	26 (61.9%)
Education, years	15.8 (2.4)	15.7 (2.5)	15.9 (2.3)	15.6 (2.3)	15.9 (2.3)	15.8 (2.5)	16.0 (2.3)	15.7 (2.0)
Visit to imaging, months	-0.8 (1.0)	-0.7 (1.0)	-0.9 (1.0)	-0.8 (1.1)	-0.8 (1.1)	-0.5 (0.9)	-0.9 (1.1)	-0.6 (1.0)
STMS	35.8 (2.2)	35.2 (2.6)	36.07 (1.79)	35.72 (1.94)	35.3 (2.1)	34.6 (3.0)	36.06 (1.75)	35.67 (1.78)
CDR 0, N (%)	222 (97.4%)	112 (86.8)	215 (100.0%)	106 (95.5%)	190 (97.4%)	43 (78.2%)	185 (100.0%)	40 (95.2%)
CDR 0.5, N (%)	6 (2.6%)	16 (12.4%)	0 (0.0%)	5 (4.5%)	5 (2.6%)	11 (20.0%)	0 (0.0%)	2 (4.8%)
CDR 1, N (%)	0 (0%)	1 (0.8%)	-	-	0 (0.0%)	1 (1.8%)	-	-
CU, N (%)	215 (94.7%)	111 (88.8%)	-	-	185 (95.4%)	42 (76.4%)	-	-
MCI, N (%)	12 (5.3%)	9 (7.2%)	-	-	9 (4.6%)	8 (14.5%)	-	-
Dementia, N (%)	0 (0.0%)	5 (4.0%)	-	-	0 (0.0%)	5 (9.1%)	-	-
Unavailable, N (%)	1 (0.4%)	0 (0.0%)	-	-	1 (0.5%)	0 (0.0%)	-	-
<b>Primary Variables</b>								
SLS sum of trials	77.25 (16.64)	68.23 (19.67)	78.77 (15.45)	71.06 (17.69)	77.92 (16.62)	62.42 (20.50)	79.17 (15.67)	67.31 (17.78)
AVLT sum of trials	67.92 (17.89)	57.92 (18.68)	69.50 (17.03)	61.30 (16.81)	68.65 (17.66)	53.13 (17.62)	70.01 (16.94)	59.81 (13.93)
<b>Secondary Variables</b>								
SLS 1-5 Total	61.88 (12.88)	55.03 (15.08)	63.07 (11.91)	57.11 (13.62)	62.29 (12.83)	50.67 (15.83)	63.28 (12.05)	54.26 (13.76)
AVLT 1-5 Total	48.24 (11.67)	41.89 (11.88)	49.24 (11.12)	43.88 (10.87)	48.59 (11.62)	39.11 (10.99)	49.44 (11.19)	43.00 (9.00)
SLS Delay	15.37 (4.19)	13.20 (5.00)	15.71 (3.98)	13.95 (4.49)	15.63 (4.20)	11.75 (5.07)	15.90 (4.05)	13.05 (4.38)
AVLT 30-min delay	9.72 (3.70)	7.84 (4.19)	10.06 (3.50)	8.64 (3.71)	9.94 (3.58)	6.75 (4.08)	10.23 (3.41)	8.36 (3.14)

Note. A = amyloid; ADRC = Mayo Alzheimer's Disease Research Center; AVLT = Auditory Verbal Learning Test; AVLT 1-5 Total = sum of words correctly recalled across trials 1-5; AVLT Sum of Trials = AVLT 1-5 total + Trial 6 + 30-minute delay; CDR = Clinical Dementia Rating Scale; CU = Cognitively Unimpaired; MCI = Mild Cognitive Impairment; MCSA = Mayo Clinic Study of Aging; MTD = Mayo Test Drive; SLS = Stricker Learning Span; SLS 1-5 Total = sum of words correctly recognized across trials 1-5; SLS Sum of Trials = SLS 1-5 total + delay; STMS = Kokmen Short Test of Mental Status; T = tau. See Supplemental Table 2 for group difference comparison results for additional SLS and AVLT variables. Table used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

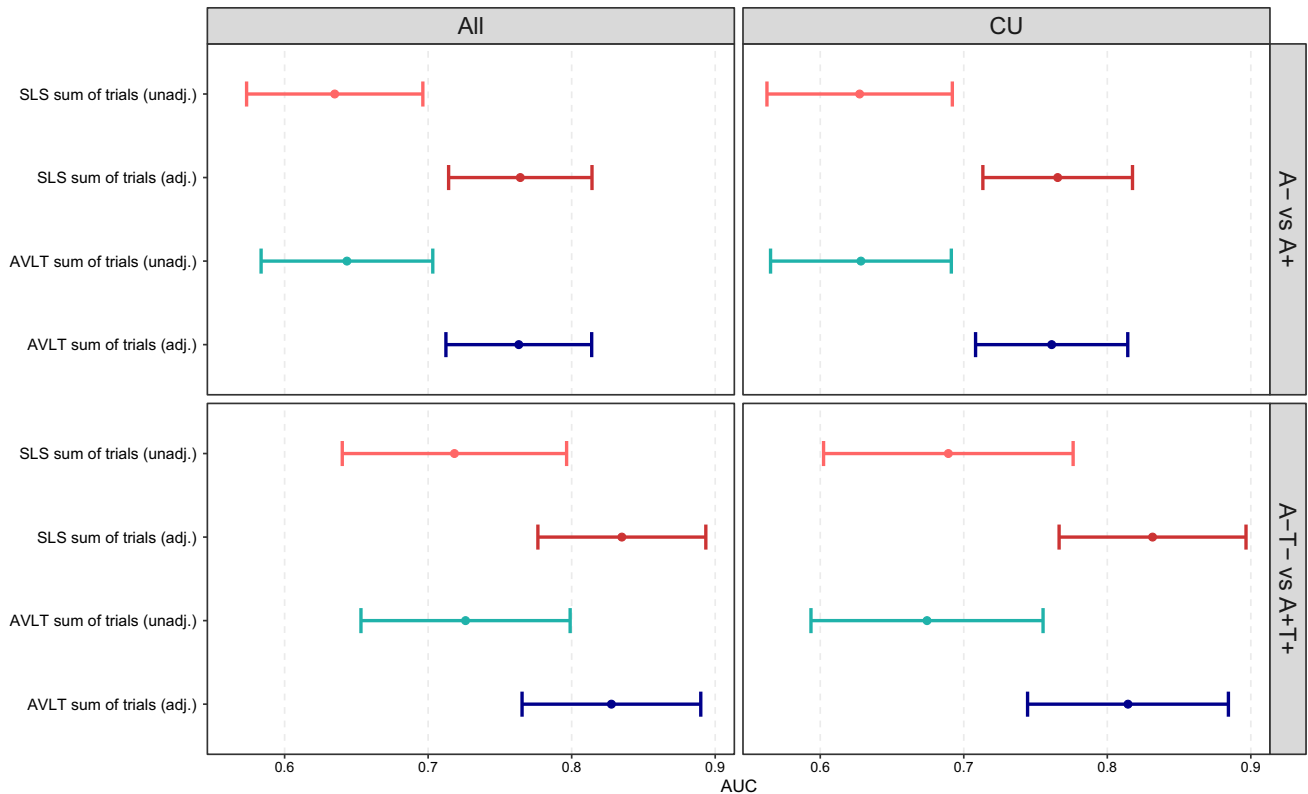
**Table 3.** Logistic regression analysis predicting a) A+ vs. A- and b) A+ T+ vs A-T- for both unadjusted models (cognition as the only predictor, on left) and models that also include age, sex, education, in addition to cognition (on right). The *p*-value comparing the AUROCs of the remotely administered Stricker Learning Span and in-person administered Auditory Verbal Learning Test provides a direct comparison of test performance for differentiating biomarker-defined groups

A. Accuracy of SLS and AVLT in Predicting Abnormal Amyloid PET (A+ vs A-)								
Characteristic	Unadjusted				Age+Sex+Educ Adjusted			
	Odds Ratio <sup>1</sup> (95% CI)	<i>p</i> -value	AUROC <sup>2</sup> (95% CI)	<i>p</i> -value comparing AUROCs	Odds Ratio <sup>1</sup> (95% CI)	<i>p</i> -value	AUROC <sup>2</sup> (95% CI)	<i>p</i> -value comparing AUROCs
<b>All participants (N = 353)</b>								
SLS sum of trials	0.760 (0.669, 0.860)	<.001	0.63 (0.57, 0.70)	0.76	0.966 (0.940, 0.992)	.013	0.76 (0.71, 0.81)	0.90
AVLT sum of trials	0.741 (0.653, 0.841)	<.001	0.64 (0.58, 0.70)		0.965 (0.939, 0.992)	.012	0.76 (0.71, 0.81)	
<b>CU only (N = 326)</b>								
SLS sum of trials	0.753 (0.652, 0.869)	<.001	0.63 (0.56, 0.69)	0.98	0.969 (0.940, 1.000)	.013	0.77 (0.71, 0.82)	0.57
AVLT sum of trials	0.752 (0.653, 0.866)	<.001	0.63 (0.57, 0.69)		0.974 (0.945, 1.004)	.012	0.76 (0.71, 0.81)	
B. Accuracy of SLS and AVLT in Predicting Abnormal Amyloid and Tau PET (A+T+ vs A-T-)								
Characteristic	Unadjusted				Age+Sex+Educ Adjusted			
	Odds Ratio <sup>1</sup> (95% CI)	<i>p</i> -value	AUROC <sup>2</sup> (95% CI)	<i>p</i> -value comparing AUROCs	Odds Ratio <sup>1</sup> (95% CI)	<i>p</i> -value	AUROC <sup>2</sup> (95% CI)	<i>p</i> -value comparing AUROCs
<b>All participants (N = 250)</b>								
SLS sum of trials	0.638 (0.536, 0.759)	<.001	0.72 (0.64, 0.80)	0.81	0.944 (0.920, 0.970)	<.001	0.84 (0.78, 0.90)	0.67
AVLT sum of trials	0.610 (0.503, 0.738)	<.001	0.73 (0.65, 0.80)		0.946 (0.921, 0.972)	<.001	0.83 (0.77, 0.90)	
<b>CU only (N = 227)</b>								
SLS sum of trials	0.659 (0.537, 0.810)	<.001	0.69 (0.60, 0.78)	0.68	0.960 (0.931, 0.988)	.007	0.83 (0.77, 0.90)	0.26
AVLT sum of trials	0.679 (0.545, 0.847)	<.001	0.67 (0.59, 0.76)		0.971 (0.942, 1.001)	.059	0.81 (0.74, 0.88)	

Note. The biomarker negative group is the reference group (e.g., A- vs A+ for part A; A-T- vs A+T+ for part B). AUROC = area under the receiver operating characteristic curve; AVLT = Rey's Auditory Verbal Learning Test; AVLT sum of trials = trials 1-5 total correct + trial 6 short-delay correct + 30-minute delay correct, in raw score units. SLS = Stricker Learning Span; SLS sum of trials = 1-5 correct + delay correct, in raw score units. Table used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

<sup>1</sup>Note that SLS and AVLT odds ratios cannot be compared directly since these two scores are on a different scale (raw scores are used). AUROC data are the focus of test comparisons. Lower test performance is associated with higher odds of being in the biomarker positive group. For example, for the A-T- vs A+T+ comparison (all participants, unadjusted model OR = 0.638), each 10-point decrease in SLS sum of trials is associated with 57% increased odds of being A+T+ (95% CI 32% - 87%) calculated as  $e^{-1 \cdot \ln(OR)}$ .

<sup>2</sup>Note, for yes/no responses the AUROC is equivalent to concordance, the probability that a randomly selected participant with a yes outcome (positive biomarker group) will have a larger predicted probability than a randomly selected participant with a no outcome (negative biomarker group).



**Figure 3.** Area Under the Receiving Operating Curve (AUROC, 95% CI) values for remotely-administered SLS sum of trials and in-person administered AVLT sum of trials. *Note.* All models significantly differentiate biomarker groups better than chance (no AUROC confidence intervals include 0.5). Stricker Learning Span (SLS) sum of trials = 1–5 total correct + delay. Auditory Verbal Learning Test (AVLT) sum of trials = 1–5 total + Trial 6 + 30-minute delay recall. Unadj. = unadjusted models. Adj. = model adjusts for age, education and sex. Figure used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

**Table 4.** Convergent validity: Pearson correlations ( $r$ ) between remotely administered Stricker Learning Span measures and in-person administered Auditory Verbal Learning Test measures for all participants.

In-Person AVLT	Remotely administered Stricker Learning Span (SLS)			
	SLS Sum of Trials	SLS 1-5 Total	SLS Max Span	SLS Delay
AVLT Sum of Trials	<b>0.62</b>	0.61	0.60	0.61
AVLT 1-5 Total	0.60	<b>0.60</b>	0.58	0.58
AVLT Trial 5	0.56	0.55	<b>0.53</b>	0.55
AVLT 30-min Delay	0.58	0.57	0.56	<b>0.59</b>

*Note.* All correlations are significant ( $p$ 's < 0.001). Correlations in bold show the relationship between the most similar AVLT and SLS measures. AVLT = Auditory Verbal Learning Test; AVLT Sum of Trials = AVLT 1–5 total + Trial 6 + 30-minute delay; SLS = Stricker Learning Span; SLS Max Span = maximum number of words recognized across any of the five learning trials; SLS 1–5 Total = sum of words correctly recognized across trials 1–5; SLS Sum of Trials = SLS 1–5 total + delay. See Supplemental Figure 1 for a full correlation matrix with additional measures. Table used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

demographic variables increase the overall AUROC values of the full model (AUROCs = 0.81–0.83). The SLS significantly improved biomarker group prediction over and above the demographic variables; the AVLT approached significance ( $p = 0.06$ ).

*Descriptive effect sizes.* Effect sizes from mean group comparisons are also reported (see Figure 4 and Supplemental Table 2).

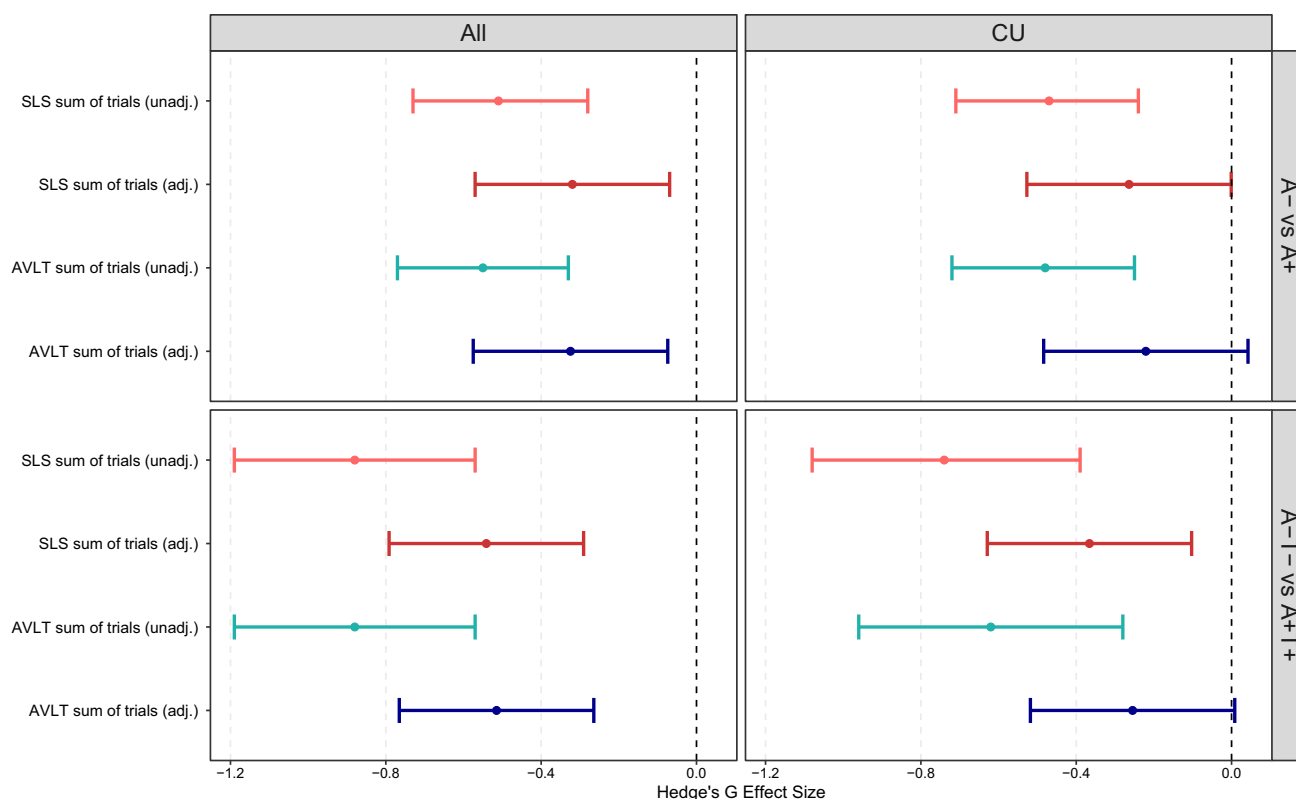
### Aim 3: Convergent validity

SLS sum of trials and AVLT sum of trials were strongly correlated ( $r = 0.62$ ,  $p < .001$ ). Additional correlations are reported in Table 4 and Supplemental Figure 1.

### Aim 4: Learning measures show similar ability as delay memory measures to differentiate A–T– and A+T+ groups for both the SLS and the AVLT (all participants)

All secondary SLS and AVLT variables show results in the expected direction with lower performance in the A+T+ compared to A–T– group ( $p$ 's < .05 for both unadjusted and adjusted analyses), with generally similar effect sizes across comparable SLS and AVLT variable pairs (Figure 5). Within-test descriptive comparisons show that learning variables (1–5 total) show effect sizes that are similar in magnitude as delay variables. For example, SLS 1–5 ( $g = -.86$ ) and delay ( $g = -.88$ ) both show large unadjusted effect





**Figure 4.** Hedge's *g* effect sizes with 95% confidence intervals to show the magnitude of group differences for remotely administered Stricker Learning Span (SLS, red shades) and in-person Auditory Verbal Learning Test (AVLT) measures (blue shades) across biomarker groups (A- vs A+ top; A-T- vs A+T+ bottom) in all participants (left) and Cognitively Unimpaired participants (right). Note. Groups do not significantly differ when the CI includes 0 (dashed line). Unadj. = unadjusted models. Adj. = model adjusts for age, education and sex. Remote SLS Sum of Trials = SLS trials 1-5 correct + delay correct. In-person AVLT Sum of Trials = AVLT 1-5 correct + Trial 6 (short-delay) correct + 30-minute recall correct). See Tables 4 and 5 for direction of effect sizes and confidence intervals. See Supplemental Table 2 for numeric results. Figure used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

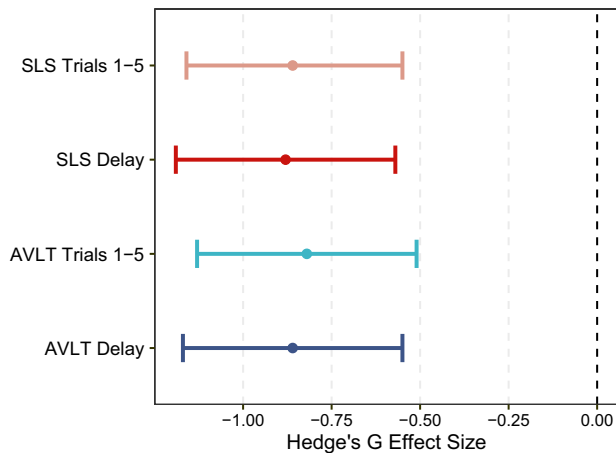
sizes, and AVLT 1-5 ( $g = -.82$ ) and delay ( $g = -.86$ ) also show large unadjusted effect sizes (Figure 4). Trials 1-5 total ( $g = -.86$  SLS and  $-.82$  AVLT) may be a slightly more advantageous learning measure for group discrimination relative to SLS max span ( $-.81$ ) or AVLT Trial 5 ( $-.74$ ) (Supplemental Table 2). Similarly, comparison of two different types of delayed memory measures suggests a slight advantage for delay total correct relative to retention ( $-.88$  vs  $-.55$  for SLS, respectively and  $-.86$  vs  $-.79$  for AVLT, respectively). See Figure 6 for a visualization of trial-by-trial data for both the SLS and AVLT. See Supplemental Tables 1 and 2 for results of adjusted analyses, other subgroup comparisons, and results of other memory tests.

## Discussion

This study followed a novel approach to test validation and established the criterion validity of an unsupervised computer adaptive word list memory test (SLS) completed outside of a clinic setting. The SLS differentiates AD biomarker-defined groups as well as a traditional word list recall test administered by trained psychometrists in a clinic setting (AVLT). Specifically, our Aim 1 hypothesis was supported by AUROC comparisons that showed remotely administered SLS sum of trials and in-person-administered AVLT sum of trials have comparable ability to differentiate individuals on the Alzheimer's continuum (A+) or not (A-) and individuals meeting a research framework for a biological

diagnosis of AD (A+T+) or not (A-T-) in a predominantly cognitively unimpaired sample.

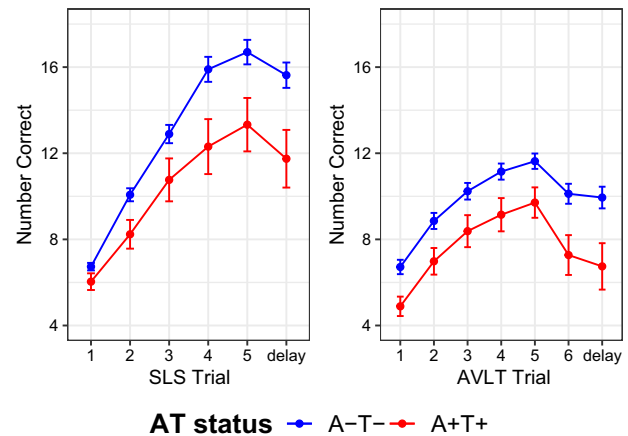
In line with our prior results showing that the AVLT has the potential to be useful for detecting subtle objective cognitive decline in preclinical AD (Stricker, Lundt, Albertson, et al., 2020), our current results extend this prior AVLT finding and suggest that the SLS also has promise in this regard (Aim 2). Specifically, when limiting the sample to CU participants, our AUROC results show that the SLS by itself could help predict, better than chance, which individuals had elevated brain amyloid, vs did not and which had elevated brain amyloid and tau vs did not. In contrast, our prior work examining the utility of the Learning/Working Memory index comprised of visual recognition and working memory tasks from the Cogstate Brief Battery administered in clinic did not significantly differentiate biomarker groups better than chance (CU A-T- vs CU A+T+ or CU A-T- vs CU A+T-) and showed that the AVLT was significantly better than the Learning/Working Memory Index for differentiating CU A-T- vs CU A+T- when comparing total AUROCs. While the predictive ability of the SLS by itself is relatively modest, predictive ability improves when demographic variables are added to the model, and the SLS continues to show an independent effect over and above demographic variables. For example, a model with age, sex, education, and SLS sum of trials together had an AUROC of 0.83 for predicting A+T+ status in CU participants. Thus, our current results suggest that the SLS could be a scalable, easily accessible



**Figure 5.** Hedge's *g* unadjusted effect sizes comparing ability of learning and delay trials to differentiate individuals without AD biomarkers from those with biological AD (A-T- vs A+T+). Note. Stricker Learning Span (SLS) Trials 1-5 = trials 1-5 total correct; Auditory Verbal Learning Test (AVLT) Trials 1-5 = 1-5 total correct; AVLT Delay = 30-minute delayed recall. See Supplemental Table 2 for numeric results. Figure used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

addition to a multivariable model approach that improves overall prediction of AD risk. For example, the addition of a word list recall measure has previously shown added utility for predicting elevated brain amyloid in individuals without dementia, over and above age alone or age combined with APOE  $\epsilon 4$  carrier status (Maserejian et al., 2019). Given its capacity for remote self-administration, the SLS would be a good candidate screening measure in nonspecialty care settings to use in combination with such predictive models to inform the need for further work-up. Plasma biomarkers will also likely be a critical component of future predictive models, particularly for preclinical disease stages (Brand et al., 2022). Establishing evidence of some independent utility for cognitive measures of interest is an important first step prior to inclusion in research to develop such models in the future. It will also be critical for any such future investigations to include adequate representation of individuals from under-represented groups to ensure broader applicability of results (Ashford et al., 2021).

The highly overlapping results for the ability of the SLS and AVLT to discriminate AD biomarker groups are particularly interesting given that although the SLS was designed to mimic the sensitivity of the AVLT, it was not designed to be a one-to-one adaptation of the AVLT. Results of correlation analyses align with this intent and support our hypothesis that the SLS and AVLT would show a significant correlation ( $r = 0.62$ ), and therefore further support convergent validity (Aim 3). Our initial pilot study similarly supported the convergent validity of the SLS, but with slightly lower correlation coefficients likely due to homogeneity of that sample (e.g., all female, restricted age range, excluded individuals with dementia), and potentially due to a longer duration between in-person and remote testing in that study (average 10 months) (Stricker et al., 2022). Our results are particularly notable given that a previous study using an exact computerized replication of the AVLT facilitated by audio recording and speech recognition to allow self-administration on an iPad showed only slightly higher correlations ( $r = 0.63$ – $0.70$ ) with the AVLT as typically administered in a well-controlled cross-over design completed in a clinic setting (Morrison et al., 2018).



**Figure 6.** Stricker Learning Span (SLS, left panel) and Auditory Verbal Learning Test (AVLT, right panel) learning slopes and delayed memory performances across all participants without AD biomarkers (A-T-, blue) and all participants with biological AD (A+T+, red). Note. Sum of trials is the primary variable for each test; these figures show the data comprising sum of trials for each measure (sum of trials = total correct items for all trials displayed in each respective figure). SLS items presented differ by trial based on computer adaptive testing rules, thus the highest possible number of presented across trials vary (see Figure 1 for ceiling and floor values for Trials 1-5; the delay trial can range from 8-23). The SLS uses 4-choice recognition to test item memory. In contrast, 15 words are presented each time for AVLT trials (constant range of 0-15) and item memory is tested with free recall. See Supplemental Table 2 for numeric results. AVLT Trial 6 = short delay. AVLT delay = 30-minute delay. Figure used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

We also qualitatively compared commonly derived supraspan wordlist indices within each test given that learning indices may have equivalent utility as delayed memory indices for the early detection of AD (Belleville et al., 2017; Weissberger et al., 2017). Results supported our hypothesis that word list learning measures would show similar sensitivity as word list delay memory measures to biologically defined AD (A-T- vs A+T+; Aim 4). The effect sizes of learning trials and delayed memory are similar (Figure 5). We chose to focus on delay items correct instead of percent retention (i.e., savings) for the comparison to learning indices. It is important to note that retention is largely dependent on specific test design characteristics including the influence of serial position effects (Atkinson & Shiffrin, 1968; Gavett & Horwitz, 2012; Greene et al., 1996). The SLS randomizes word order to minimize the recency effects often observed in individuals with Alzheimer's dementia that leads to an over-estimate of true forgetting (Cunha et al., 2012). For example, learning to criterion studies demonstrate that individuals in the early stages of AD take a longer time to reach criterion, which is reflective of lower learning ability; however, once learning is equated by reaching criterion rates of forgetting are similar to healthy control participants (Greene et al., 1996; Grober & Kawas, 1997; Stamate et al., 2020). Accordingly, SLS learning indices (1-5, max span) demonstrate a larger effect size than SLS retention in biologically defined AD versus those without AD biomarkers, whereas AVLT retention effect size is more similar to AVLT learning indices (see Supplemental Table 2).

This study has several strengths. Most studies examining the ability of cognitive measures to differentiate biomarker groups have focused on differentiation of A+ versus A- individuals (Baker et al., 2017; Duke Han et al., 2017). Our inclusion of tau status defined by PET imaging, in addition to amyloid, is a strength. Our population-based sample helps to increase generalizability to clinical settings where comorbidities are common. Our approach of reporting both

unadjusted and adjusted effect sizes illustrates the robust biomarker-group difference effect sizes observed in unadjusted analyses. For example, the SLS showed a large group difference effect size across A-T- and A+T+ groups in all participants (-0.88), that decreased to a medium effect size when controlling for demographics (-0.53) and was further attenuated when limited to CU only (-0.74 unadjusted, -0.37 adjusted). There is growing evidence that cognitive decline is not a normal part of the aging process, but rather is reflective of previously undetected neuropathologies that increase in prevalence with increasing age (Bos et al., 2018; Boyle et al., 2021; Harrington et al., 2018). To maximize the utility of cognitive measures for informing risk of AD biomarker positivity, we recommend the use of raw scores and argue that the effect of age should not be routinely "adjusted" away as it decreases the predictive power of cognition.

Several limitations should also be noted. First, because this is a population-based study, the racial and ethnic characteristics reflect that of Olmsted County from which participants are randomly sampled, resulting in a predominantly White, Non-Hispanic sample. Second, the predominantly CU composition of this sample may have decreased AUROC values and magnitude of effect sizes for differentiating biomarker positive and negative groups, as suggested by the highly similar results seen when limiting analyses only to CU participants. A more balanced sample design, with more inclusion of individuals with mild to moderate dementia, could support greater utility of these memory measures for identifying individuals at risk for biomarker positivity; the current results are more relevant for preclinical detection given that 93% of our sample is CU. In addition, the population-based nature of the MCSA sample may produce lower AUROC values for predicting elevated brain biomarkers than studies that use a convenience sample that frequently includes a higher number of individuals at risk of Alzheimer's disease based on family history of subjective concerns and studies that have strict inclusion criteria to limit potential comorbidities (Maserejian et al., 2019). MCSA participants have higher rates of comorbid conditions given that exclusionary criteria are limited to terminal illness or hospice care (Roberts et al., 2008). Third, a majority of the sample (83%) had prior exposure to the AVLT given the longitudinal nature of the MCSA and ADRC studies, thus practice effects could have impacted the ability of the AVLT to discriminate biomarker groups. Because biomarker negative participants benefit more from practice effects than biomarker positive participants, it is possible this could have amplified group difference effects for the AVLT (Alden et al., 2022; Machulda et al., 2017). Future work is needed to replicate these results in a setting where both the SLS and AVLT are baseline administrations. Similarly, the entirely unsupervised and remote approach for the SLS could dampen the sensitivity of the SLS as the results presented in this study include all available remote data.

We capture participant-reported information about test interference, noise in the test environment, and participant comments that can provide additional information about test interruptions or environmental considerations. However, because our goal was to establish the robust criterion validity of the SLS "in the wild," we did not apply any exclusionary criterion based on this information in the present study. Another reason we did not apply such exclusionary criteria is that individuals who are less able to follow instructions provided for the recommended test environment may be more likely to have cognitive impairment. Thus, increased likelihood of lower test performance in an uncontrolled

environment, worsened by environmental distractions, could also be related to risk of cognitive decline. If cognitive screening/risk for cognitive decline is the goal, worse performance in remote settings may help identify risk in a way not captured by controlled clinical settings, adding an element of ecological validity or ability to adapt to a new task without assistance. Future work will examine whether and to what degree these factors may influence test performance, as increased distractions in the home environment can negatively impact performance (Madero et al., 2021). Similarly, individuals with low technological literacy may perform more poorly on the SLS because of a lack of familiarity or comfort with mobile devices or computers. Our approach of allowing individuals choose to use the device they are most comfortable with helps address this to some degree. Even though most adults in the U.S. have access to some device, individuals from disadvantaged backgrounds may not have access to cellular service, wifi or broadband internet at home, although only 7% of Americans report they do not use the internet across any of these access methods and this has dramatically declined since 2000 (Perrin & Atske, 2021). We also cannot rule out the possibility that some individuals may have written down words to benefit their performance; however, given that this is a research study there would be no apparent incentive to artificially increase performance. In addition, there are elements of test design that help to deter this to some extent. First, word order is randomized for each trial; since the words are not presented in the same order, this makes it more difficult to write them down each time. Second, by the 4<sup>th</sup> learning trial, for high performers there are 23 words. Thus, it would be quite burdensome to write all the words down. In addition, if this was occurring, this would greatly increase the time to complete the measure. We review the data for outliers with regards to time to completion overall and for each test, and we did not have specific concerns this occurred in the current sample. Finally, while use of a strictly biomarker-defined ground truth is a novel aspect of this study, in vivo PET biomarkers also have some limitations such as high but imperfect reliability (manifested by "noise" in the trajectories of imaging results over time in some individuals), and the fact that PET measures of amyloid and tau pathology have a sensitivity floor and medically significant pathology can exist that lies beneath this detection threshold (Lee et al., 2022). Also, we adopted a liberal window for inclusion of available biomarker data to allow for some missed scanning opportunities during the COVID-19 pandemic, to maximize the sample size and because of the generally good stability of amyloid and tau classifications (Jack et al., 2019), but this also decreases study precision relative to a narrower time window.

In summary, SLS test design prioritized remote assessment needs and a computer-adaptive approach (Stricker et al., 2022). Even though the SLS is not a direct adaptation of the AVLT, our results show highly similar ability of the remotely-administered SLS and in-person-administered AVLT to differentiate AD biomarker-defined groups. These results challenge preconceived notions about memory assessment by showing that creative use of a recognition memory paradigm that emphasizes learning in an all-remote unsupervised sample differentiates AD biomarker-defined groups as effectively as a traditional word list memory measure based on free recall responses.

**Acknowledgements.** Research reported in this publication was supported by the Kevin Merszei Career Development Award in Neurodegenerative Diseases Research IHO Janet Vittone, MD, the Rochester Epidemiology Project (R01 AG034676), the National Institute on Aging of the National Institutes of Health (grant numbers R21 AG073967, P30 AG062677, P50 AG016574, U01 AG006786,

RF1 AG55151, R01 AG041851, R37 AG011378), the Robert Wood Johnson Foundation, The Elsie and Marvin Dekelboum Family Foundation, GHR Foundation, Alzheimer's Association, and the Mayo Foundation for Education and Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. A Mayo Clinic invention disclosure has been submitted for the Stricker Learning Span and the Mayo Test Drive platform (NHS, JLS). We have no other conflicts of interest to disclose related to this work. The authors wish to thank the participants and staff at the Mayo Clinic Study of Aging and Mayo Alzheimer's Disease Research Center.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S1355617723000322>.

## References

- Alden, E. C., Lundt, E. S., Twohy, E. L., Christianson, T. J., Kremers, W. K., Machulda, M. M., & Jack, C. R. Knopman, D. S. Mielke, M. M. Petersen, R. C. Stricker, N. H. (2022). Mayo normative studies: A conditional normative model for longitudinal change on the Auditory Verbal Learning Test and preliminary validation in preclinical Alzheimer's disease. *Alzheimers Dement (Amst)*, 14(1), e12325. <https://doi.org/10.1002/dad2.12325>
- Alden, E. C., Pudumjee, S. B., Lundt, E. S., Albertson, S. M., Machulda, M. M., Kremers, W. K., Jack, C. R., Knopman, D. S., Petersen, R. C., Mielke, M. M., & Stricker, N. H. (2021). Diagnostic accuracy of the Cogstate Brief Battery for prevalent MCI and prodromal AD (MCI A(+) T(+)) in a population-based sample. *Alzheimers & Dementia*, 17(4), 584–594. <https://doi.org/10.1002/alz.12219>
- American Psychiatric Association 1994, *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*, 4th edn. Washington, D.C: American Psychiatric Association.
- Ashford, M. T., Veitch, D. P., Neuhaus, J., Nosheny, R. L., Tosun, D., & Weiner, M. W. (2021). The search for a convenient procedure to detect one of the earliest signs of Alzheimer's disease: A systematic review of the prediction of brain amyloid status. *Alzheimers & Dementia*, 17(5), 866–887. <https://doi.org/10.1002/alz.12253>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Spence, K. W. S. J. T. (Eds.), *The psychology of learning and motivation: II*. Cambridge, MA: Academic Press.
- Baker, J. E., Lim, Y. Y., Pietrzak, R. H., Hassenstab, J., Snyder, P. J., Masters, C. L., & Maruff, P. (2017). Cognitive impairment and decline in cognitively normal older adults with high amyloid-beta: A meta-analysis. *Alzheimers Dement (Amst)*, 6, 108–121. <https://doi.org/10.1016/j.dadm.2016.09.002>
- Bauer, R. M., Bilder, R. M. 2023, Automated neuropsychological testing. In Brown, G. G., King, T. Z., Haaland, K. Y., & Crosson, B. (Eds.), *APA Handbook of Neuropsychology (Vol. 2: Neuroscience and Neuromethods)*. Washington, DC: American Psychological Association.
- Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H. T. V., & Croteau, J. Consortium for the Early Identification of Alzheimer's, d-Q (2017). Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: A systematic review and meta-analysis. *Neuropsychology Review*, 27(4), 328–353. <https://doi.org/10.1007/s11065-017-9361-5>
- Bilder, R. M., & Reise, S. P. (2019). Neuropsychological tests of the future: How do we get there from here? *Clinical Neuropsychologist*, 33(2), 220–245. <https://doi.org/10.1080/13854046.2018.1521993>
- Bos, I., Vos, S. J. B., Jansen, W. J., Vandenbergh, R., Gabel, S., Estanga, A., Ecay-Torres, M., Tomassen, J., den Braber, A., Lleó, A., Sala, I., Wallin, A., Kettunen, P., Molinuevo, J. L., Rami, L., Chetelat, G., de la Sayette, V., Tsolaki, M., Freund-Levi, Y., . . . , & Visser, P. J. (2018). Amyloid-beta, tau, and cognition in cognitively normal older individuals: Examining the necessity to adjust for biomarker status in normative data. *Frontiers in Aging Neuroscience*, 10, 193. <https://doi.org/10.3389/fnagi.2018.00193>
- Boyle, P. A., Wang, T., Yu, L., Wilson, R. S., Dawe, R., Arfanakis, K., Schneider, J. A., & Bennett, D. A. (2021). To what degree is late life cognitive decline driven by age-related neuropathologies? *Brain*, 144(7), 2166–2175. <https://doi.org/10.1093/brain/awab092>
- Brand, A. L., Lawler, P. E., Bollinger, J. G., Li, Y., Schindler, S. E., Li, M., Lopez, S., Ovod, V., Nakamura, A., Shaw, L. M., Zetterberg, H., Hansson, O., Bateman, R. J., & Bateman, R. J. (2022). The performance of plasma amyloid beta measurements in identifying amyloid plaques in Alzheimer's disease: A literature review. *Alzheimer's Research & Therapy*, 14(1), 195. <https://doi.org/10.1186/s13195-022-01117-1>
- Caselli, R. J., Langlais, B. T., Dueck, A. C., Chen, Y., Su, Y., Locke, D. E. C., Woodruff, B. K., & Reiman, E. M., (2020). Neuropsychological decline up to 20 years before incident mild cognitive impairment. *Alzheimers & Dementia*, 16(3), 512–523. <https://doi.org/10.1016/j.jalz.2019.09.085>
- Chiotis, K., Saint-Aubert, L., Boccardi, M., Gietl, A., Picco, A., Varrone, A., Garibotto, V., Herholz, K., Nobili, F., Nordberg, A., & Geneva Task Force for the Roadmap of Alzheimer's, B. (2017). Clinical validity of increased cortical uptake of amyloid ligands on PET as a biomarker for Alzheimer's disease in the context of a structured 5-phase development framework. *Neurobiology of Aging*, 52, 214–227. <https://doi.org/10.1016/j.neurobiolaging.2016.07.012>
- Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan, 1968 norms. *Behavior Research Methods, Instruments, & Computers*, 36(3), 371–383. <https://doi.org/10.3758/BF03195584>
- Cromer, J. A., Harel, B. T., Yu, K., Valadka, J. S., Brunwin, J. W., Crawford, C. D., Mayes, L. C., & Maruff, P. (2015). Comparison of cognitive performance on the cogstate brief battery when taken in-clinic, in-group, and unsupervised. *Clinical Neuropsychologist*, 29(4), 542–558. <https://doi.org/10.1080/13854046.2015.1054437>
- Cummings, J., Lee, G., Zhong, K., Fonseca, J., & Taghva, K. (2021). Alzheimer's disease drug development pipeline: 2021. *Alzheimers Dement (N Y)*, 7(1), e12179. <https://doi.org/10.1002/trc2.12179>
- Cunha, C., Guerreiro, M., de Mendonca, A., Oliveira, P. E., & Santana, I. (2012). Serial position effects in Alzheimer's disease, mild cognitive impairment, and normal aging: predictive value for conversion to dementia. *Journal of Clinical and Experimental Neuropsychology*, 34(8), 841–852. <https://doi.org/10.1080/13803395.2012.689814>
- Dolch, E. W. (1936). A basic sight vocabulary. *Elementary School Journal*, 36, 456–460. <https://doi.org/10.1086/457353>
- Dorsey, E. R., Kluger, B., & Lipset, C. H. (2020). The new normal in clinical trials: Decentralized studies. *Annals of Neurology*, 88(5), 863–866. <https://doi.org/10.1002/ana.25892>
- Duff, K., Hammers, D. B., Dalley, B. C. A., Suhrie, K. R., Atkinson, T. J., Rasmussen, K. M., Horn, K. P., Beardmore, B. E., Burrell, L. D., Foster, N. L., Hoffman, J. M. (2017). Short-term practice effects and amyloid deposition: Providing information above and beyond baseline cognition. *The Journal of Prevention of Alzheimer's Disease*, 4(2), 87–92. <https://doi.org/10.14283/jpad.2017.9>
- Duke Han, S., Nguyen, C. P., Stricker, N. H., & Nation, D. A. (2017). Detectable neuropsychological differences in early preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology Review*, 27(4), 305–325. <https://doi.org/10.1007/s11065-017-9345-5>
- Ferman, T. J., Lucas, J. A., Ivnik, R. J., Smith, G. E., Willis, F. B., Petersen, R. C., & Graff-Radford, N. R. (2005). Mayo's older African American normative studies: Auditory Verbal Learning Test norms for African American elders. *Clinical Neuropsychologist*, 19(2), 214–228. <https://doi.org/10.1080/13854040590945300>
- Gavett, B. E., & Horwitz, J. E. (2012). Immediate list recall as a measure of short-term episodic memory: Insights from the serial position effect and item response theory. *Archives of Clinical Neuropsychology*, 27(2), 125–135. <https://doi.org/10.1093/arclin/acr104>
- Greene, J. D., Baddeley, A. D., & Hodges, J. R. (1996). Analysis of the episodic memory deficit in early Alzheimer's disease: evidence from the doors and people test. *Neuropsychologia*, 34(6), 537–551. [https://doi.org/10.1016/0028-3932\(95\)00151-4](https://doi.org/10.1016/0028-3932(95)00151-4)
- Grober, E., & Kawas, C. (1997). Learning and retention in preclinical and early Alzheimer's disease. *Psychology and Aging*, 12(1), 183–188. <https://doi.org/10.1037//0882-7974.12.1.183>
- Harrington, K. D., Schembri, A., Lim, Y. Y., Dang, C., Ames, D., Hassenstab, J., Laws, S. M., Rainey-Smith, S., Robertson, J., Rowe, C. C., Sohrabi, H. R., Salvado, O., Weinborn, M., Villemagne, V. L., Masters, C. L., Maruff, P.

- (2018). Estimates of age-related memory decline are inflated by unrecognized Alzheimer's disease. *Neurobiology of Aging*, 70, 170–179. <https://doi.org/10.1016/j.neurobiolaging.2018.06.005>
- Jack, C. R., Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., Sperling, R., & Sperling, R. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers & Dementia*, 14(4), 535–562. <https://doi.org/10.1016/j.jalz.2018.02.018>
- Jack, C. R., Jr., Wiste, H. J., Weigand, S. D., Therneau, T. M., Lowe, V. J., Knopman, D. S., Gunter, J. L., Senjem, M. L., Jones, D. T., Kantarci, K., Machulda, M. M., Mielke, M. M., Roberts, R. O., Vemuri, P., Reyes, D. A., & Petersen, R. C. (2017). Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers & Dementia*, 13(3), 205–216. <https://doi.org/10.1016/j.jalz.2016.08.005>
- Jack, C. R. Jr., Lowe, V. J., Senjem, M. L., Weigand, S. D., Kemp, B. J., Shiung, M. M., Knopman, D. S., Boeve, B. F., Klunk, W. E., Mathis, C. A., Petersen, R. C. (2008). 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnesic mild cognitive impairment. *Brain*, 131(Pt 3), 665–680. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18263627>
- Jack, C. R. Jr., Wiste, H. J., Therneau, T. M., Weigand, S. D., Knopman, D. S., Mielke, M. M., Lowe, V. J., Vemuri, P., Machulda, M. M., Schwarz, C. G., Gunter, J. L., Senjem, M. L., Graff-Radford, J., Jones, D. T., Roberts, R. O., Rocca, W. A., Petersen, R. C. (2019). Associations of amyloid, tau, and neurodegeneration biomarker profiles with rates of memory decline among individuals without dementia. *JAMA*, 321(23), 2316–2325. <https://doi.org/10.1001/jama.2019.7437>
- Jack, C. R. Jr., Wiste, H. J., Weigand, S. D., Knopman, D. S., Vemuri, P., Mielke, M. M., Lowe, V. J., Senjem, M. L., Gunter, J. L., Machulda, M. M., Gregg, B. E., Pankratz, V. S., Rocca, W. A., Petersen, R. C. (2015). Age, sex, and APOE epsilon4 effects on memory, brain structure, and beta-amyloid across the adult life span. *Jama Neurology*, 72(5), 511–519. <https://doi.org/10.1001/jamaneurol.2014.4821>
- Klunk, W. E., Koeppe, R. A., Price, J. C., Benzinger, T. L., Devous, M. D., Sr., Jagust, W. J., Johnson, K. A., Mathis, C. A., Minhas, D., Pontecorvo, M. J., Rowe, C. C., Skovronsky, D. M., Mintun, M. A., & Mintun, M. A. (2015). The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. *Alzheimers & Dementia*, 11(1), 1–15 e11–14. <https://doi.org/10.1016/j.jalz.2014.07.003>
- Kokmen, E., Smith, G. E., Petersen, R. C., Tangalos, E., & Ivnik, R. C. (1991). The short test of mental status: Correlations with standardized psychometric testing. *Archives of Neurology*, 48(7), 725–728. <https://doi.org/10.1001/archneur.1991.00530190071018>
- Lee, J., Burkett, B. J., Min, H.-K., Lundt, E. S., Albertson, S. M., Botha, H., Senjem, M. L., Gunter, J. L., Schwarz, C. G., Jones, D. T., Knopman, D. S., Jack, C. R. Jr., Petersen, R. C., Lowe, V. J. (2022). The Overlap Index as a means of evaluating early tau-PET signal reliability. *Journal of Nuclear Medicine*, 63(11), 1748–1753. <https://doi.org/10.2967/jnumed.121.263136>
- Lim, Y. Y., Baker, J. E., Bruns, L., Mills, A., Fowler, C., Fripp, J., Rainey-Smith, S. R., Ames, D., Masters, C. L., & Maruff, P. (2020). Association of deficits in short-term learning and A $\beta$  and hippocampal volume in cognitively normal adults. *Neurology*, <https://doi.org/10.1212/WNL.000000000010728>.doi
- Lohns, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency in memory for mixed frequency lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(6), 1943–1946. <https://doi.org/10.1037/a0033669>
- Machulda, M. M., Hagen, C. E., Wiste, H. J., Mielke, M. M., Knopman, D. S., Roberts, R. O., Vemuri, P., Lowe, V. J., Jack, C. R. Jr., Petersen, R. C. (2017). Practice effects and longitudinal cognitive change in clinically normal older adults differ by Alzheimer imaging biomarker status. *Clinical Neuropsychologist*, 31(1), 99–117. <https://doi.org/10.1080/13854046.2016.1241303>
- Mackin, R. S., Rhodes, E., Insel, P. S., Nosheny, R., Finley, S., Ashford, M., Camacho, M. R., Truran, D., Mosca, K., Seabrook, G., Morrison, R., Narayan, V. A., & Weiner, M. (2021). Reliability and validity of a home-based self-administered computerized test of learning and memory using speech recognition. *Neuropsychology, Development, and Cognition. Section B: Aging, Neuropsychology and Cognition*, 1–15. <https://doi.org/10.1080/13825585.2021.1927961>
- Madero, E. N., Anderson, J., Bott, N. T., Hall, A., Newton, D., Fuseya, N., Harrison, J. E., Myers, J. R., & Glenn, J. M. (2021). Environmental distractions during unsupervised remote digital cognitive assessment. *The Journal of Prevention of Alzheimer's Disease*, 1–4. <https://doi.org/10.14283/jpad.2021.9>
- Marra, D. E., Hamlet, K. M., Bauer, R. M., & Bowers, D. (2020). Validity of teleneuropsychology for older adults in response to COVID-19: A systematic and critical review. *Clinical Neuropsychologist*, 1–42. <https://doi.org/10.1080/13854046.2020.1769192>
- Maserejian, N., Bian, S., Wang, W., Jaeger, J., Syrjanen, J. A., Aakre, J., Jack, C. R., Mielke, M. M., Gao, F., & Alzheimer's Disease Neuroimaging Initiative and the AIBL research team (2019). Practical algorithms for amyloid beta probability in subjective or mild cognitive impairment. *Alzheimers Dement (Amst)*, 11, 180. <https://doi.org/10.1016/j.dadm.2019.09.001>
- Mielke, M. M., Machulda, M. M., Hagen, C. E., Edwards, K. K., Roberts, R. O., Pankratz, V. S., Knopman, D. S., Jack, C. R., & Petersen, R. C. (2015). Performance of the CogState computerized battery in the Mayo Clinic Study on Aging. *Alzheimer's & Dementia*, 11(11), 1367–1376. <https://doi.org/10.1016/j.jalz.2015.01.008>
- Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, 43(11), 2412–2414. <https://doi.org/10.1212/WNL.43.11.2412-a>
- Morrison, R. L., Pei, H., Novak, G., Kaufer, D. I., Welsh-Bohmer, K. A., Ruhmel, S., & Narayan, V. A. (2018). 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)*, 10, 647–656. <https://doi.org/10.1016/j.dadm.2018.08.010>
- Nicosia, J., Aschenbrenner, A. J., Balota, D., Sliwinski, M., Tahan, M., Adams, S., Stout, S. H., Wilks, H., Gordon, B., Benzinger, T., Fagan, A., Xiong, C., Bateman, R., Morris, J.H. (2023). Unsupervised high-frequency smartphone-based cognitive assessments are reliable, valid, and feasible in older adults at risk for Alzheimer disease. *Journal of the International Neuropsychological Society*, 29(5), 459–471. <https://doi.org/10.1017/S135561772200042X>
- Ohman, F., Hassenstab, J., Berron, D., Scholl, M., & Papp, K. V. (2021). Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimers Dement (Amst)*, 13(1), e12217. <https://doi.org/10.1002/dad2.12217>
- Papp, K. V., Samaroo, A., Chou, H.-C., Buckley, R., Schneider, O. R., Hsieh, S., Soberanes, D., Quiroz, Y., Properzi, M., Schultz, A., García-Magariño, Ián, Marshall, G. A., Burke, J. G., Kumar, R., Benzyner, N., Johnson, K., Rentz, D. M., Sperling, R. A., Amariglio, R. E. (2021). Unsupervised mobile cognitive testing for use in preclinical Alzheimer's disease. *Alzheimers Dement (Amst)*, 13(1), e12243. <https://doi.org/10.1002/dad2.12243>
- Perrin, A., & Atske, S. (2021). 7% of Americans don't use the internet. *Who are they?*. Retrieved from <https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/>
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3), 183–194. <https://doi.org/10.1111/j.1365-2796.2004.01388.x>
- Pudumjee, S. B., Lundt, E. S., Albertson, S. M., Machulda, M. M., Kremers, W. K., Jack, C. R., Knopman, D. S., Petersen, R. C., Mielke, M. M., & Stricker, N. H. (2021). A comparison of cross-sectional and longitudinal methods of defining objective subtle cognitive decline in preclinical Alzheimer's disease based on cogstate one card learning accuracy performance. *Journal of Alzheimer's Disease*, 83(2), 861–877. <https://doi.org/10.3233/JAD-210251>
- Roberts, R. O., Geda, Y. E., Knopman, D. S., Cha, R. H., Pankratz, V. S., Boeve, B. F., Ivnik, R. J., Tangalos, E. G., Petersen, R. C., & Rocca, W. A. (2008). The Mayo Clinic Study of Aging: Design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*, 30(1), 58–69. <https://doi.org/10.1159/000115751>
- Sabbagh, M., Boada, M., Borson, P., Doraiswamy, P. M., Dubois, B., Ingram, J., Iwata, A., Porsteinsson, A. P., Possin, K. L., Rabinovici, G. D., Vellas, B., Chao, S., Vergallo, A., & Hampel, H. (2020). Early detection of mild cognitive impairment MCI in an at home setting. *The Journal of Prevention of Alzheimer's Disease*, <https://doi.org/10.14283/jpad.2020.21>

- Sabbagh, M. N., Boada, M., Borson, S., Chilukuri, M., Doraiswamy, P. M., Dubois, B., Ingram, J., Iwata, A., Porsteinsson, A. P., Possin, K. L., Rabinovici, G. D., Vellas, B., Chao, S., Vergallo, A., Hampel, H. (2020). Rationale for early diagnosis of Mild Cognitive Impairment (MCI) supported by emerging digital technologies. *The Journal of Prevention of Alzheimer's Disease*, 7(3), 158–164. <https://doi.org/10.14283/jpad.2020.19>
- St Sauver, J. L., Grossardt, B. R., Yawn, B. P., Melton, L. J., Pankratz, J. J., & Brue, S. M. (2012). Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system. *International Journal of Epidemiology*, 41(6), 1614–1624. <https://doi.org/10.1093/ije/dys195>
- Stamate, A., Logie, R. H., Baddeley, A. D., & Della Sala, S. (2020). Forgetting in Alzheimer's disease: Is it fast? Is it affected by repeated retrieval? *Neuropsychologia*, 138, 107351. <https://doi.org/10.1016/j.neuropsychologia.2020.107351>
- Stricker, J. L., Corriveau-Lecavalier, N., Wiepert, D. A., Botha, H., Jones, D. T., & Stricker, N. H. (2022). Neural network process simulations support a distributed memory system and aid design of a novel computer adaptive digital memory test for preclinical and prodromal Alzheimer's disease. *Neuropsychology*, <https://doi.org/10.1037/neu0000847>
- Stricker, N. H., Christianson, T. J., Lundt, E. S., Alden, E. C., Machulda, M. M., Fields, J. A., Kremers, W. K., Jack, C. R., Knopman, D. S., Mielke, M. M., & Petersen, R. C. (2020). Mayo normative studies: Regression-based normative data for the auditory verbal learning test for ages 30-91 years and the importance of adjusting for sex. *Journal of the International Neuropsychological Society*, 1-16, <https://doi.org/10.1017/S1355617720000752>
- Stricker, N. H., Lundt, E. S., Albertson, S. M., Machulda, M. M., Pudumjee, S. B., Kremers, W. K., Jack, C. R., Knopman, D. S., Petersen, R. C., Mielke, M. M. (2020). Diagnostic and prognostic accuracy of the cogstate brief battery and auditory verbal learning test in preclinical Alzheimer's disease and incident mild cognitive impairment: Implications for defining subtle objective cognitive impairment. *Journal of Alzheimer's Disease*, 76(1), 261–274. <https://doi.org/10.3233/JAD-200087>
- Stricker, N. H., Lundt, E. S., Alden, E. C., Albertson, S. M., Machulda, M. M., Kremers, W. K., Knopman, D. S., Petersen, R. C., & Mielke, M. M. (2020). Longitudinal comparison of in clinic and at home administration of the cogstate brief battery and demonstrated practice effects in the Mayo Clinic study of aging. *The Journal of Prevention of Alzheimer's Disease*, 7(1), 21–28. <https://doi.org/10.14283/jpad.2019.35>
- Stricker, N. H., Stricker, J. L., Karstens, A. J., Geske, J. R., Fields, J. A., Hassenstab, J., Schwarz, C. G., Tosakulwong, N., Wiste, H. J., Jack, C. R. Jr., Kantarci, K., Mielke, M. M. (2022). A novel computer adaptive word list memory test optimized for remote assessment: Psychometric properties and associations with neurodegenerative biomarkers in older women without dementia. *Alzheimer's Dementia (Amst)*, 14(1), e12299. <https://doi.org/10.1002/dad2.12299>
- Therneau, T. M. 'A package for survival analysis in R 2021, =<https://CRAN.R-project.org/package=survival>.
- Vemuri, P., Lowe, V. J., Knopman, D. S., Senjem, M. L., Kemp, B. J., Schwarz, C. G., Przybelski, S. A., Machulda, M. M., Petersen, R. C., Jack, C. R. (2017). Tau-PET uptake: Regional variation in average SUVR and impact of amyloid deposition. *Alzheimer's Dementia (Amst)*, 6, 21–30. <https://doi.org/10.1016/j.dadm.2016.12.010>
- Weissberger, G. H., Strong, J. V., Stefanidis, K. B., Summers, M. J., Bondi, M. W., & Stricker, N. H. (2017). Diagnostic accuracy of memory measures in Alzheimer's dementia and mild cognitive impairment: A systematic review and meta-analysis. *Neuropsychology Review*, 27(4), 354–388. <https://doi.org/10.1007/s11065-017-9360-6>
- Wolters, E. E., Dodich, A., Boccardi, M., Corre, J., Drzezga, A., Hansson, O., Nordberg, A., Frisoni, G. B., Garibotto, V., Ossenkoppele, R. (2021). Clinical validity of increased cortical uptake of [(18)F]flortaucipir on PET as a biomarker for Alzheimer's disease in the context of a structured 5-phase biomarker development framework. *European Journal of Nuclear Medicine and Molecular Imaging*, 48(7), 2097–2109. <https://doi.org/10.1007/s00259-020-05118-w>