



Published in final edited form as:

J Alzheimers Dis. 2020 ; 76(1): 261–274. doi:10.3233/JAD-200087.

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

Nikki H. Stricker¹, Emily S. Lundt², Sabrina M. Albertson², Mary M. Machulda¹, Shehroo B. Pudumjee¹, Walter K. Kremers², Clifford R. Jack Jr.³, David S. Knopman⁴, Ronald C. Petersen⁴, Michelle M. Mielke^{4,5}

¹Division of Neurocognitive Disorders, Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota, USA

²Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

³Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA

⁴Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA

⁵Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

Abstract

Objective: We aimed to determine whether cross-sectional performance on the Cogstate Brief Battery (CBB) and Auditory Verbal Learning Test (AVLT) could identify (1) cognitively unimpaired (CU) participants with preclinical Alzheimer's disease defined by neuroimaging biomarkers of amyloid and tau, and (2) incident mild cognitive impairment (MCI)/dementia.

Method: CU participants age 50+ were eligible if they (1) had amyloid (A) and tau (T) imaging within two years of their baseline CBB or (2) had at least one follow-up visit. AUROC analyses assessed the ability of measures to differentiate groups. We also explored the frequency of cross-sectional subtle objective cognitive impairment (sOBJ) defined as performance -1 SD on CBB Learning/Working Memory Composite (Lrn/WM) or AVLT delayed recall using age-corrected normative data.

Results: A+T+ (n=33, mean age 79.5) and A+T- (n=61, mean age 77.8) participants were older than A-T- participants (n=146, mean age 66.3), and comparable on sex and education. Lrn/WM did not differentiate A+T+ or A+T- from A-T- participants. AVLT differentiated both A+T+ and A+T- from A-T- participants; 45% of A+T+ and 25% of A+T- participants met sOBJ criteria. The follow-up cohort included 150 CU individuals who converted to MCI/dementia; 450 age, sex, and

Corresponding Author: Nikki H. Stricker, Ph.D., Mayo Clinic, 200 First Street SW, Rochester, MN 55905; 507-284-2649 (phone), 507-284-4158 (fax), Stricker.Nikki@mayo.edu.

Conflict of Interest/Disclosure Statement

The authors have no conflicts of interest to report.

education matched controls were identified. Lrn/WM and AVLT both differentiated between stable and converter CU participants.

Conclusions: Among CU participants, AVLT helps differentiate A+T+ and A+T- from A-T- participants. The CBB did not differentiate biomarker subgroups, but showed potential for predicting incident MCI/dementia. Results will help inform future definitions of sOBJ.

Keywords

Sensitivity and Specificity; Neuropsychology; Memory; Subtle Cognitive Decline; Amyloid; Tau; Cognigram; One Card Learning; One Back; Conversion

INTRODUCTION

There is a genuine need to develop an efficient way to screen the population for elevated brain amyloid and tau in order to 1) identify appropriate individuals with preclinical Alzheimer's disease (AD) for anti-amyloid and anti-tau clinical trials, and 2) identify those with AD pathology and thus at highest risk of developing AD dementia once anti-amyloid or anti-tau therapies are identified. Simple cognitive assessment methods show promise in this regard. First, there are detectable cognitive differences in amyloid positive versus negative cognitively unimpaired (CU) individuals, which are most robust for memory measures [1, 2]. Effect sizes are small in amyloid positive CU individuals ($d = -.17$), but are larger in individuals with both amyloid and additional disease burden in the form of tau or neurodegeneration ($d = -.47$) [1]. Second, subtle objective cognitive impairment (sOBJ), also referred to as subtle cognitive decline, shows promise for contributing to detection of preclinical AD. sOBJ based on a single baseline assessment confers risk of converting from CU to MCI/AD at follow-up [3]. For example, Edmonds and colleagues showed that 46% of individuals with sOBJ alone, and 84% of individuals with sOBJ, neurodegeneration and amyloidosis converted to mild cognitive impairment (MCI)/AD dementia [3]. Further, individuals with cross-sectional sOBJ show faster accumulation of amyloid per PET neuroimaging, faster entorhinal cortical thinning, and greater risk of conversion to MCI/AD dementia over follow-up relative to CU participants without sOBJ [4].

Various definitions of sOBJ have been suggested based on traditional neuropsychological tests [3, 5–7], but there is no clear agreement on the best definition or measures to use. Cross-sectional definitions have included a $<10^{\text{th}}$ percentile cut-off score on composite measures of memory [7] or global cognition [6]. Other approaches have combined cognitive and functional measures, requiring two out of six scores <1 SD below published normative data in two separate cognitive domains or a subtle decline in functional abilities based on a Functional Assessment Questionnaire score of 6–8 [3]. Intra-individual definitions of objective longitudinal cognitive decline (sOBJ) may confer greater sensitivity relative to cross-sectional approaches and represent an important method for defining transitional cognitive decline for individuals with preclinical AD [8]. However, whether intra-individual decline is superior compared to a cross-sectional approach remains a hypothesis to be tested and no clear a priori cutoffs have been recommended for how to define longitudinal decline at this time. Despite this ambiguity and prior evidence of the utility of cross-sectional sOBJ [3, 4], there is no cross-sectional option for defining Stage 2 transitional cognitive decline

based on cognitive measures as recently proposed by the National Institute on Aging-Alzheimer's Association (NIA-AA) Working Group [8]. Therefore, additional information about the frequency of cross-sectional sOBJ in preclinical AD will help address whether cross-sectional definitions of sOBJ should be considered as another way to define transitional cognitive decline and will inform development of a generalizable, widely accepted definition of sOBJ.

Computerized cognitive measures, such as the Cogstate Brief Battery (CBB), may be a more efficient method for identifying sOBJ relative to traditional, standardized person-administered measures [9, 10]. Further, computerized measures can be incorporated into study designs independent of diagnosis, which helps avoid the frequent circularity between measures used to define sOBJ and those used for current and future diagnosis. If shown to be reliable and valid, computerized measures that provide the option for unsupervised administration may be better suited for large-scale cognitive screening and monitoring than traditional neuropsychological measures or other computerized measures that require in-person administration. The CBB is one such tool that can be administered with or without supervision and consists of 4 cognitive tasks that measure psychomotor function, attention, working memory, and visual memory [11]. The CBB is already being used for at-home enrichment of clinical trial enrollment through the Brain Health Registry and has been implemented in the next phase of the Alzheimer's Disease Neuroimaging Initiative (ADNI-3) [12–14]. Early Cogstate Ltd.-affiliated studies suggest that the CBB may show promise for detecting early preclinical cognitive change; individuals identified as memory decliners based on longitudinal performance on the One Card Learning subtest from the CBB had a higher frequency of amyloid positivity in CU individuals relative to individuals with stable performance [15, 16]. Cogstate has FDA approval to market the CBB as a medical device under the name Cognigram™ for use as a digital cognitive assessment tool that can be completed in clinic or at home through prescription use in individuals 6–99 years of age. Although the CBB was originally developed to detect change over time, Cognigram is marketed for use on a single occasion or to determine cognitive change over time. One study showed that a single supervised administration of the Learning/Working Memory (Lrn/Wm) Composite from the CBB showed good diagnostic accuracy for Alzheimer's dementia and MCI [17], but this has not yet been replicated. It remains unclear whether a single CBB administration can help predict elevated brain amyloid or conversion to MCI/dementia in CU individuals.

The primary aim of this study was to help inform evolving definitions of sOBJ by determining whether baseline performance on a computerized measure, the CBB, can help identify CU participants with (1) preclinical AD defined by neuroimaging biomarkers of amyloid and tau, and (2) incident MCI/dementia. We predicted that the CBB Lrn/Wm Composite and a traditional neuropsychological measure, AVLT delayed recall, would show comparable 1) diagnostic accuracy and 2) prognostic accuracy based on AUROC analyses. We also applied a conventional cut-off of -1 SD below age-adjusted normative scores to the CBB and AVLT to investigate the frequency of sOBJ and to facilitate clinical translation of study results. We predicted that CU individuals meeting criteria for a biological diagnosis of Alzheimer's disease (CU A+T+), but not those with Alzheimer pathologic change alone (CU A+T-), would show a higher frequency of sOBJ relative to CU A-T- individuals. We

also predicted that individuals with incident MCI/dementia defined as CU individuals who later convert to MCI or dementia (CU converters) would show a higher frequency of sOBJ relative to CU stable participants at baseline.

METHOD

The Mayo Clinic Study of Aging (MCSA) is a prospective population-based study of cognitive aging among Olmsted County, MN, residents following an age- and sex-stratified random sampling design [18]. Individuals randomly chosen for recruitment were invited to participate in the MCSA and those without a medical contraindication were invited to participate in imaging studies. PiB-PET began in 2008 and tau-PET imaging began in 2016. The MCSA began administering Cogstate measures in the clinic in 2012 to the newly enrolled 50–69 year olds during clinic visits. Cogstate administration for those aged 70 and older began the following year.

MCSA study visits included a neurologic evaluation by a physician, an interview by a study coordinator, and neuropsychological testing by a psychometrist [18]. The physician examination included a medical history review, complete neurological examination, and administration of the Short Test of Mental Status (STMS) [19]. The study coordinator interview included demographic information and medical history, and questions about memory to both the participant and informant using the Clinical Dementia Rating scale [20]. Details about the neuropsychological battery have been previously reported [18].

For each participant, performance in a cognitive domain was compared with age-adjusted scores of cognitively unimpaired (CU) individuals using Mayo's Older Americans Normative Studies [21]. Participants with scores of ≥ 1.0 SD below the age-specific mean in the general population were considered for possible cognitive impairment, taking into account education, prior occupation, visual or hearing deficits, and other information. A diagnostic determination of CU, MCI, or dementia is made by a consensus agreement between the study coordinator, examining physician, and neuropsychologist using published criteria [18, 22]. The diagnosis at each study visit is conducted blind to prior clinical information, diagnosis, or knowledge of biomarkers. Performance on the neuropsychological battery is considered by the neuropsychologist for diagnostic recommendation, but not by the physician or study coordinator. Performance on Cogstate was not considered for diagnosis.

The study protocols were approved by the Mayo Clinic and Olmsted Medical Center Institutional Review Boards. All participants provided written informed consent.

Inclusion Criteria

Two partially overlapping study samples from the MCSA were derived to address study hypotheses: 1) participants with biomarker data, and 2) all available participants, grouped by future diagnosis (see Table 1, which includes information about sample overlap).

Participants with biomarker data—Individuals were eligible if they were CU, aged 50 or older at the time of their baseline CBB session, and had amyloid PET (using Pittsburgh

Compound B [PiB-PET]) and tau PET (AV1451) scans within two years of their baseline Cogstate evaluation (n = 240). After application of the biomarker cut-offs described below to eligible participants, biomarker subgroups included CU individuals with a biological definition of Alzheimer's disease (n = 33 A+T+), CU individuals with AD pathologic change (n = 61 A+T-), and CU individuals with normal AD biomarkers (n = 146 A-T-).

Participants grouped by future diagnosis—For this analysis, MCSA participants aged 50+ who were CU at baseline CBB and had at least one MCSA follow-up visit were eligible for inclusion (N = 2,328). A cohort of 150 individuals with a diagnosis of MCI/dementia at any follow-up visit were identified and labeled as CU converters; the majority of these participants converted to MCI (n = 141). The remaining participants (n = 2,178) represented a pool of participants who remained CU at all available follow-up visits and were eligible for matching (CU stable). Each converter was matched to three stable CU participants on age (± 5 years), sex and education (≤ 12 years, 13–14, 15–16, 17+).

Biomarker methods

PiB-PET and tau-PET are acquired with a PET/CT operating in three-dimensional mode. The details of the acquisition, processing, and cut-off derivation have been previously published [23–25]. Individuals are considered amyloid positive if they have an SUVR > 1.48 in a previously defined meta-ROI [24] and tau positive if they have an SUVR > 1.25 in a previously defined meta-ROI including temporal lobe structures [24]. To simplify the study design and maintain a focus on individuals on the Alzheimer's disease continuum, we grouped participants according to amyloid and tau (AT) imaging status, instead of using the full AT(N) classification scheme [26]. Unspecified or unavailable biomarker data is denoted *; we did not consider N status and thus use N* to denote this. We limited the current study to individuals meeting the biological definition of Alzheimer's disease (A+T+N*) or with AD pathologic change (A+T-N*) per the recently proposed research framework for a biological diagnosis of Alzheimer's disease [8]. We refer to individuals who are A-T- as biomarker negative, but in doing so likely include some individuals with non-AD pathologic change (A-T-N+). Because of our focus on individuals within the Alzheimer's continuum, we did not include A-T+ individuals.

Cognitive Measures

Cogstate Brief Battery (CBB)—Only baseline Cogstate performance is included in the current study. Cogstate version 7 (v7) was used, which provides personal computer (PC), iPad and web-based administration options. Most participants completed their first Cogstate session in clinic on a PC or iPad (see Stricker, et al. [27] for details about Cogstate procedures). Less than 1% completed their baseline Cogstate session at home through web-based administration on a PC, thus the small differences in performance at home versus in clinic that we previously reported on measures of accuracy are unlikely to impact results presented [28]. The ability to reliably complete and adhere to the requirements of each task was determined by completion checks as previously described [27]. All data values with a failed completion flag were removed from subsequent analyses.

The CBB includes the following tasks: (1) Detection (DET): A simple reaction time paradigm that measures psychomotor function. (2) Identification (IDN): A choice reaction time paradigm that measures visual attention. (3) One Card Learning (OCL): A continuous visual recognition learning task that assesses visual recognition memory and attention. (4) One Back (ONB): A task that assesses working memory and attention. Detection and Identification previously showed limited diagnostic accuracy for MCI and were not included in analyses [17]. Accuracy data (OCL and ONB) were transformed using arcsine transformation to normalize the variables.

The CBB Lrn/WM composite was the primary measure of interest. It is derived using the average of age corrected z-scores for One Card Learning Accuracy and One Back Accuracy. Normative data were provided by Cogstate [29] in the form of means and SDs by age groups for arcsine transformed OCL accuracy and ONB accuracy scores. Secondary exploratory analyses also used only One Card Learning accuracy as an alternative primary outcome variable since measures of episodic memory are typically more sensitive to early decline in preclinical AD relative to measures of working memory [1, 2], and to allow direct comparison to raw-score analyses. One Card Learning z-scores and arcsine transformed raw scores were used in this analysis.

Auditory Verbal Learning Test—The AVLT is a 15-item word list memory test that is part of the neuropsychological battery used at each MCSA study visit. Mayo's Older Americans Normative Studies (MOANS) age-corrected scaled scores [21] for 30-minute delayed recall was the primary measure of interest; these scaled scores have a mean of 10 and SD of 3. Because many participants were already enrolled in the MCSA and followed longitudinally at the time of the baseline Cogstate session that is the focus of this study, a portion of participants had prior exposure to the AVLT (see Results). The normative data applied [21] does not take prior test exposure into account, which likely decreases the sensitivity of this measure as used in the current study, particularly in the subgroups with greater frequency of prior exposure.

Subtle Objective Cognitive Impairment (sOBJ)—sOBJ was defined using a conventional cut-off of performance of less than or equal to 1 standard deviation below the mean (-1 SD) based on age-corrected normative scores. This translates to an age-corrected z-score of -1 for Cogstate measures [29] and an age-corrected MOANS scaled score of 7 for AVLT 30-minute delayed recall [21].

Statistical Methods

Demographic and clinical differences between groups were assessed by linear model ANOVA test for means and Chi-squared test for frequencies. Effect size (Hedge's g) was computed using a weighted and pooled standard deviation. AUROC analyses were conducted to assess the ability of measures to differentiate two groups. The Youden index method was used to identify cut points [30]. This method defines the optimal cut-point as the point maximizing the Youden function, which is the difference between the true positive rate (TPR) and false positive rate (FPR) over all possible cut-point values. We also directly tested

biomarker group-wise discrimination as summarized by the AUROC for select comparisons [31]. Frequency of sOBJ using a conventional cut-off was compared across groups.

For participants with biomarker data, we also used a multivariate logistic regression model to predict group membership (A-T- v A+T+, A-T- v A+T-) and to compute odds ratios (OR) and 95% confidence intervals (CI). Multivariate logistic regression models were adjusted for age, sex, and years of education and included a single cognitive measure of interest as the predictor. This allowed us to compute a demographics-adjusted OR for the Lrn/WM composite and AVLT delayed recall. In order to better compare to Lrn/WM, AVLT delayed recall was rescaled to z-score units. For every 1.0 z-score less, the odds of being in the biomarker positive group increases by (OR-1)%.

Secondary analyses were performed with select raw scores to ensure results were not driven by use of age-corrected scores. Secondary analyses also included the STMS to facilitate direct comparison of the CBB results to a public domain, traditional person-administered screening measure.

RESULTS

CU participants with biomarker data

Demographics and group mean comparisons—A+T+ and A+T- participants were older than A-T- participants and were comparable on education and sex (see Table 2). Performance on Lrn/WM across A+T- and A-T- groups was comparable ($p = .27$, Hedge's $g = 0.17$). Lrn/WM performance in the A+T+ group was marginally lower ($p = .07$) than in the A-T- group, with a small to medium effect size (Hedge's $g = 0.35$). Both of the biomarker positive groups (i.e., A+T+ and A+T-) showed significantly lower AVLT delayed recall performance (p 's $< .01$) relative to the A-T- group, with medium effect sizes for both the A+T+ (Hedge's $g = .60$) and A+T- (Hedge's $g = 0.54$) comparisons. Of note, 40.6% of participants with biomarker data had prior exposure to the AVLT overall, with increasing frequency of prior exposure in the biomarker positive groups (see Table 2). Although direct comparison of the A+T+ and A+T- groups was not part of our a priori planned hypotheses, we did include some comparisons of these groups for the interested reader (see Table 2). Mean comparisons of the A+T+ and A+T- groups showed comparable performances on Lrn/WM (Hedge's $g = 0.11$) and AVLT delayed recall (Hedge's $g = 0.08$; both p 's $> .05$).

sOBJ frequency—For Lrn/WM, the A+T+ group showed a significantly higher frequency of sOBJ relative to the A-T- group (see Table 2). Despite this, sensitivity was very low, with 18% of A+T+ participants meeting sOBJ criteria on Lrn/WM. Only 5% of CU A-T- participants met this cut-off, thus specificity of Lrn/WM was excellent (95%). A+T- and A-T- groups showed a comparable frequency of sOBJ on Lrn/WM. For AVLT delayed recall, both of the biomarker positive groups showed a significantly higher frequency of sOBJ relative to the A-T- group. AVLT sensitivity was 45% for the A+T+ group, 25% for the A+T- group, and only 12% of A-T- participants met the sOBJ cut-off (88% specificity).

Diagnostic accuracy—Overall diagnostic accuracy of Lrn/WM for differentiating both A+T+ and A+T- from A-T- participants was very low and not better than chance (see Table 3).

Diagnostic accuracy for the AVLT was low, but AVLT delayed recall discriminated both biomarker positive (A+T+ and A+T-) groups from the A-T- group better than chance (CI's did not include 0.50). Comparing total AUC values for both measures, AVLT delayed recall differentiated A+T- and A-T- groups significantly better than Lrn/WM Composite ($p = .03$). However, both measures comparably differentiated A+T+ and A-T- groups ($p = .40$).

Logistic regression analyses controlling for age, sex, and education confirmed that these diagnostic accuracy results cannot be fully explained by age differences across groups (see Table 4). AVLT delayed recall significantly separates A+T- and A-T- groups, above and beyond age, education and sex. For differentiation of A+T+ and A-T- groups, the pattern of results remained consistent with frequency and diagnostic accuracy analyses for AVLT delayed recall, but the p -value was at trend level ($p = .08$). Lrn/WM did not significantly separate groups, confirming the above analyses.

To investigate whether the pattern of results would change if raw scores were used and to ensure that the sensitivity of Lrn/WM was not artificially decreased by including a measure of working memory in the composite, we completed secondary analyses using One Card Learning accuracy raw scores and z -scores, as well as AVLT delayed recall raw scores (see Figure 1 and Table 5). Total AUC values were similar between the Lrn/WM Composite and OCL z -score, between OCL z -score and raw score, and between AVLT delayed recall scaled score and raw score. In secondary analyses, the STMS did not differentiate A-T- and A+T+ groups better than chance. Overall diagnostic accuracy of the STMS for differentiating A+T- from A-T- participants was low, but better than chance (CI did not include 0.50). Sensitivity and specificity were low to poor for the optimally derived cut-off (A+T- vs. A-T-), suggesting limited clinical utility.

CU participants with follow-up data

Demographics and group mean comparisons—Converter and stable CU groups were comparable on age, education, sex and years of follow-up available (see Table 6). Time to conversion was 2.7 years after baseline Cogstate, on average. The CU converter group showed lower performance relative to the stable CU group (p 's < .001) on Lrn/WM (Hedge's $g = 0.59$) and AVLT delayed recall (Hedge's $g = 0.72$). Across CU participants with follow-up data, 57.0% of participants had prior AVLT exposure overall, with comparable frequency of prior exposure across stable and converter groups.

sOBJ frequency—The CU converter group showed a significantly higher frequency of sOBJ on both Lrn/WM and AVLT relative to the CU stable group (see Table 6). Using a conventional -1 SD cut-off, Lrn/WM showed 25% sensitivity for predicting conversion. Only 12% of CU stable participants met the sOBJ cut-off for Lrn/WM, thus specificity was good (88%). AVLT delayed recall had slightly higher sensitivity (38%) for predicting conversion. Only 14% of CU stable participants met the sOBJ cut-off for AVLT delayed recall, thus specificity was good (86%).

Prognostic accuracy for predicting incident MCI/dementia—Prognostic accuracy of baseline memory performance for differentiating CU stable and CU converter participants

was moderate based on total AUC and better than chance (total AUC CI's did not include 0.5) for both Lrn/WM and AVLTL (see Table 3).

DISCUSSION

The main finding of this study is that sOBJ, defined based on a single administration of a traditional neuropsychological memory test (AVLT), may help identify CU individuals with (1) preclinical Alzheimer's disease as defined by neuroimaging biomarkers of amyloid and tau, and (2) incident MCI/dementia. In accordance with the primary study aim to determine whether baseline CBB performance can help identify CU participants with preclinical AD, a single CBB assessment was not useful in predicting elevated brain amyloid either alone or in combination with tau (A+T- or A+T+). However, CBB Lrn/WM Composite significantly contributed to predicting conversion from CU at baseline to MCI/dementia over follow-up

Our finding that a single baseline memory test performance could identify 45% of CU individuals with a biological diagnosis of AD based on neuroimaging biomarkers of amyloid and tau (A+T+) and 25% with AD pathologic change alone (A+T-) was somewhat unexpected. Prior studies have shown that individuals with AD pathologic change (A+) but unknown T status do not show impairment in verbal memory when assessed on a single occasion [32–34]. Consistent with our hypothesis, we demonstrated that CU individuals with a biological diagnosis of AD based on positive amyloid and tau biomarkers (A+T+) showed a higher frequency of sOBJ relative to CU A-T- individuals on AVLTL delayed recall. However, despite our prediction to the contrary, CU individuals with Alzheimer pathologic change alone (A+T-) also showed a higher frequency of sOBJ on AVLTL delayed recall. Prior reports have largely shown small effect sizes in studies comparing CU individuals with and without positive amyloid status, with meta-analytic results ranging from -0.14 [2] to -0.17[1]. Based on a smaller pool of 7 studies using the 2011 preclinical AD framework [35] a medium effect size (-.47) was demonstrated for Stage 2 (A+ and evidence of either positive neurodegeneration or tau) vs. Stage 0 [1]. Our findings suggest a medium effect size on AVLTL delayed recall for both CU individuals with AD (A+T+) and Alzheimer pathologic change (A+T-) relative to CU individuals without AD biomarkers (A-T-).

Inconsistent with our hypothesis, the total AUC was significantly greater for AVLTL delayed recall than for Lrn/WM Composite for differentiating A+T- and A-T- CU groups. Statistical comparison of the total AUC suggested that both AVLTL delayed recall and the Lrn/WM Composite comparably differentiated A+T+ and A-T- groups. However, overall diagnostic accuracy of the Lrn/WM Composite was not better than chance and only 18% of the A+T+ group met the sOBJ criterion, suggesting very limited clinical utility. Thus, when considering single assessment results, the AVLTL shows greater clinical utility than the CBB for predicting biomarker positive status. Word list memory tests typically have greater sensitivity to MCI and AD dementia relative to other memory testing paradigms, particularly recognition memory [36]. Thus, this difference in clinical utility may simply reflect the type of memory measure used (word list delayed recall versus visual recognition memory) and does not necessarily imply a difference between traditional and computerized cognitive testing. Importantly, the CBB was designed with a focus on sensitivity to detecting change versus sensitivity for use at a single time point. However, given the CBB's burgeoning use as

a cognitive screening measure for clinical trials and for patient care, a better understanding of its clinical utility based on one time point is important. Further, although the CBB has shown some promise in identifying risk of elevated brain amyloid based on memory decline over time on OCL accuracy [15], the initial data presented in this regard suffered from some circularity in study design and there has not been a direct comparison of the added benefit of longitudinal CBB assessment relative to baseline CBB or baseline traditional memory measures. We plan to directly compare the sensitivity of cross-sectional and longitudinal approaches to defining sOBJ and subtle objective cognitive decline in future work once additional follow-up data have been collected. Relatively few studies have used true longitudinal change over time to define objective subtle cognitive decline [15, 16, 33, 37] even though intra-individual decline is expected to provide a more sensitive means for detecting preclinical cognitive changes.

Although this study presents data regarding the diagnostic and prognostic accuracy of memory measures, memory measures are non-specific with regards to etiology. Low memory performance can be due to a number of factors including numerous other biological etiologies, demographic variables and transient situational factors. Clinical diagnosis remains the gold standard for describing clinical syndromes. However, the need for data-driven and efficient approaches to identifying risk of preclinical AD based on cognitive measures or other methods is paramount to ongoing clinical trial efforts, and for identifying individuals who will most benefit from treatment once a viable treatment is identified. Predictive algorithms also show some promise in helping to predict the likelihood of elevated brain amyloid. However, in addition to data from in-person cognitive assessments that are typically included in these algorithms, additional information such as genetic testing (i.e., APOE genotype) and/or knowledge of other medical conditions [38, 39] is also often required, and thus may not be as simple and efficient as cognitive-based methods alone.

Our results showed that AVLT delayed recall alone could help predict, better than chance, which CU individuals had elevated brain amyloid. Although the total AUC values of 66% for A+T+ and 64% for A+T- are “low” per traditional standards, these results are promising for discriminating among CU individuals. For example, a recent study reported an AUC of 66% for plasma A β 42 alone and 68% for plasma A β 42/A β 40 ratio for discriminating subjects with normal vs. abnormal amyloid PET scans in CU individuals with subjective cognitive decline [40]. In addition, the sensitivity of the AVLT in our study is likely artificially lowered in two ways. First, a portion of the sample had prior exposure to the AVLT, thus unaccounted for practice effects likely led to slightly improved performance for the assessment used for these participants [41]. Also, AVLT performance was considered for diagnosis in a way that would decrease the sensitivity of this measure, as AVLT performance was considered during consensus diagnosis conference and participants were still given a CU diagnosis. Within ADNI, Kandel et al. [42] also demonstrated that AVLT performance has been shown to help predict CSF amyloid- β (A β)₁₋₄₂ among patients with MCI, with predictive utility comparable to neuroimaging biomarkers (FDG-PET and hippocampal volume).

A secondary aim of this study was to determine whether performance on the CBB Lrn/WM Composite at baseline could predict conversion from CU to MCI/dementia. As a reference,

we also examined AVLT delayed recall performance. Despite the circularity resulting from the AVLT being considered in the diagnosis of MCI, the results were encouraging as both measures achieved moderate overall diagnostic accuracy, with total AUC of 70%. Other cognitive screening measures have similar predictive accuracy values. Both the Montreal Cognitive Assessment (MoCA) and STMS performed similarly for detecting incident MCI with AUCs of 0.70 and 0.71 respectively [43]. Our secondary analyses confirm that the STMS performed similarly within this sample, with a total AUC of 0.74. However, STMS performance is considered as part of the consensus diagnosis. Rizk-Jackson and colleagues [44] found that MRI volumes and FDG PET measures alone showed 65% overall accuracy to predict conversion to MCI within two years, although a model combining MRI and FDG-PET measures achieved 81% total accuracy. Our results are particularly notable given that prior research in this area demonstrated similar AUC values when employing numerous clinical variables. Prediction of incident MCI in CU individuals from the NACC database using 14 non-invasive, clinical variables including performance on various neuropsychological measures resulted in 75% total AUC [45]. Despite our encouraging AUC results for predicting incident MCI/dementia with a single administration of the CBB, derived optimal cut-offs for the CBB fall well within the range of normal performance, complicating meaningful clinical application.

Although this study provides a unique look at the frequency of sOBJ in CU individuals with a biological diagnosis of Alzheimer's disease, AD pathologic change, and incident MCI/dementia, there are also important limitations. Most notably, biomarker positive groups were significantly older than biomarker negative groups. Although age-corrected normative scores help account for the substantial age differences across groups, this may have had unintended consequences. For example, biomarker status is known to be significantly associated with age, which precluded use of an age-matched sample for biomarker groups. We aimed to further address this important confound by performing logistic regression equations that controlled for age, education and sex, and the pattern of results did not change. Also of note, although we report analyses for 3:1 age, sex and education-matched groups for identification of CU stable vs. CU converter individuals, unreported preliminary results without this matching procedure showed equivalent findings. Finally, because the current study's primary focus was on the CBB, a sizeable portion of AVLT performances were derived from individuals who had already been exposed to the AVLT. Application of normative data based on baseline performance therefore likely reduced the sensitivity of AVLT results by failing to take into account the known influence of practice effects [46, 47]. Of note, this may be particularly relevant for the biomarker subgroups, as both the A+T+ and A+T- groups were more likely to have had prior exposure to the AVLT relative to the A-T- group. Prior work by our group showed that among CU individuals, biomarker subgroups may be differentially sensitive to practice effects, with A-N- and A+N- subgroups both showing continued practice effects on memory measures across multiple visits, whereas A+N+ and A-N+ showed relatively stable performances over time [48]. Future work considering the important role of practice effects when optimizing definitions of sOBJ and OBJ is needed.

Despite these limitations, this study contributes to the rapidly growing body of research focused on how to best define sOBJ in preclinical AD and fosters further research in the area. For instance, results suggest that normative data with better sensitivity to preclinical

AD are needed. Alternative cut-offs may also need to be considered. For example a -0.5 SD cut-off may be appropriate for sOBJ [49, 50]. Knopman and colleagues [51] showed that individuals with sOBJ based on a -0.5 SD cutoff have an increased risk of incident MCI/dementia. Use of norms that adjust for sex and education, in addition to age, is recommended for future studies but was not an option for the independent normative data used in this study from the MOANS and Cogstate [21, 29]. Intra-individual definitions of objective subtle cognitive decline may also enhance the sensitivity of cognitive measures to preclinical AD [37] and represents an important future direction.

In summary, results suggest that a single CBB session may show some utility in predicting conversion to MCI/dementia over follow-up, but is unlikely to be helpful in predicting which CU individuals are more likely to have Alzheimer's pathologic change or Alzheimer's disease based on AD neuroimaging biomarkers of amyloid and tau. In contrast, a traditional word list memory test shows promise in this regard, particularly if normative data can be refined to enhance sensitivity.

Acknowledgements

The authors wish to thank the participants and staff at the Mayo Clinic Study of Aging. This work was supported by the Rochester Epidemiology Project (R01 AG034676), the National Institutes of Health (grant numbers P50 AG016574, P30 AG062677, U01 AG006786, R37 AG011378, R01 AG041851, RF1 AG55151), a grant from the Alzheimer's Association (AARG-17-531322), Zenith Award from the Alzheimer's Association, the Robert Wood Johnson Foundation, The Elsie and Marvin Dekelboum Family Foundation, Alexander Family Alzheimer's Disease Research Professorship of the Mayo Clinic, Liston Award, Schuler Foundation, GHR Foundation, AVID Radiopharmaceuticals, and the Mayo Foundation for Education and Research. We would like to greatly thank AVID Radiopharmaceuticals, Inc., for their support in supplying AV-1451 precursor, chemistry production advice and oversight, and FDA regulatory cross-filing permission and documentation needed for this work. NHS and MMMi serve as consultants to Biogen and Lundbeck. DSK serves on a Data Safety Monitoring Board for the DIAN-TU study and is an investigator in clinical trials sponsored by Lilly Pharmaceuticals, Biogen, and the University of Southern California. RCP has served as a consultant for Hoffman-La Roche Inc., Merck Inc., Genentech Inc., Biogen Inc., Eisai, Inc. and GE Healthcare.

References

- [1]. Duke Han S, Nguyen CP, Stricker NH, Nation DA (2017) Detectable Neuropsychological Differences in Early Preclinical Alzheimer's Disease: A Meta-Analysis. *Neuropsychol Rev* 27, 305–325. [PubMed: 28497179]
- [2]. Baker JE, Lim YY, Pietrzak RH, Hassenstab J, Snyder PJ, Masters CL, 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. [PubMed: 28239636]
- [3]. Edmonds EC, Delano-Wood L, Galasko DR, Salmon DP, Bondi MW (2015) Subtle Cognitive Decline and Biomarker Staging in Preclinical Alzheimer's Disease. *J Alzheimers Dis* 47, 231–242. [PubMed: 26402771]
- [4]. Thomas KR, Bangen KJ, Weigand AJ, Edmonds EC, Wong CG, Cooper S, Delano-Wood L, Bondi MW, Alzheimer's Disease Neuroimaging I (2019) Objective subtle cognitive difficulties predict future amyloid accumulation and neurodegeneration. *Neurology*.
- [5]. Hassenstab J, Ruvolo D, Jasielec MS, Xiong C, Grant E, Morris JC (2015) Absence of practice effects in preclinical Alzheimer's Disease. *Neuropsychology* 29, 940–948. [PubMed: 26011114]
- [6]. Jack CR, Knopman DS, Weigand SD, Wiste HJ, Vemuri P, Lowe V, Kantarci K, Gunter JL, Senjem ML, Ivnik RJ, Roberts RO, Rocca WA, Boeve BF, Petersen RC (2012) An operational approach to National Institute on Aging–Alzheimer's Association criteria for preclinical Alzheimer disease. *Ann Neurol* 71, 765–775. [PubMed: 22488240]

- [7]. Vos SJB, Xiong C, Visser PJ, Jasielec MS, Hassenstab J, Grant EA, Cairns NJ, Morris JC, Holtzman DM, Fagan AM (2013) Preclinical Alzheimer's disease and its outcome: a longitudinal cohort study. *Lancet Neurol* 12, 957–965. [PubMed: 24012374]
- [8]. Jack CR Jr., Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E, Molinuevo JL, Montine T, Phelps C, Rankin KP, Rowe CC, Scheltens P, Siemers E, Snyder HM, Sperling R (2018) NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* 14, 535–562. [PubMed: 29653606]
- [9]. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI (2012) Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *Arch Clin Neuropsychol* 27, 362–373. [PubMed: 22382386]
- [10]. Zygouris S, Tsolaki M (2015) Computerized cognitive testing for older adults: a review. *Am J Alzheimers Dis Other Demen* 30, 13–28. [PubMed: 24526761]
- [11]. Maruff P, Thomas E, Cysique L, Brew B, Collie A, Snyder P, Pietrzak RH (2009) Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Arch Clin Neuropsychol* 24, 165–178. [PubMed: 19395350]
- [12]. Mackin RS, Insel PS, Truran D, Finley S, Flenniken D, Nosheny R, Ulbright A, Comacho M, Bickford D, Harel B, Maruff P, Weiner MW (2018) Unsupervised online neuropsychological test performance for individuals with mild cognitive impairment and dementia: Results from the Brain Health Registry. *Alzheimers Dement (Amst)* 10, 573–582. [PubMed: 30406176]
- [13]. Weiner MW, Nosheny R, Camacho M, Truran-Sacrey D, Mackin RS, Flenniken D, Ulbright A, Insel P, Finley S, Fockler J, Veitch D (2018) The Brain Health Registry: An internet-based platform for recruitment, assessment, and longitudinal monitoring of participants for neuroscience studies. *Alzheimers Dement* 14, 1063–1076. [PubMed: 29754989]
- [14]. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Morris JC, Petersen RC, Salazar J, Saykin AJ, Shaw LM, Toga AW, Trojanowski JQ (2017) The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia* 13, 561–571.
- [15]. Darby DG, Brodtmann A, Pietrzak RH, Fredrickson J, Woodward M, Villemagne VL, Fredrickson A, Maruff P, Rowe C (2011) Episodic memory decline predicts cortical amyloid status in community-dwelling older adults. *J Alzheimers Dis* 27, 627–637. [PubMed: 21876253]
- [16]. Darby DG, Pietrzak RH, Fredrickson J, Woodward M, Moore L, Fredrickson A, Sach J, Maruff P (2012) Intraindividual cognitive decline using a brief computerized cognitive screening test. *Alzheimers Dement* 8, 95–104. [PubMed: 22404851]
- [17]. Maruff P, Lim YY, Darby D, Ellis KA, Pietrzak RH, Snyder PJ, Bush AI, Szoeki C, Schembri A, Ames D, Masters CL (2013) Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC Psychology* 1, 30. [PubMed: 25566378]
- [18]. Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, Ivnik RJ, Tangalos EG, Petersen RC, Rocca WA (2008) The Mayo Clinic Study of Aging: Design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology* 30, 58–69. [PubMed: 18259084]
- [19]. Kokmen E, Smith GE, Petersen RC, Tangalos E, Ivnik RC (1991) The short test of mental status: Correlations with standardized psychometric testing. *Arch Neurol* 48, 725–728. [PubMed: 1859300]
- [20]. Morris JC (1993) The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* 43, 2412–2414.
- [21]. Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC, Kokmen E, Kurland LT (1992) Mayo's Older Americans Normative Studies: WAIS-R, WMS-R and AVLT norms for ages 56 through 97. *The Clinical Neuropsychologist* 6, 1–104.
- [22]. Petersen RC, Roberts RO, Knopman DS, Geda YE, Cha RH, Pankratz VS, Boeve B, Tangalos EG, Ivnik RJ, Rocca WA (2010) Prevalence of mild cognitive impairment is higher in men: The Mayo Clinic Study of Aging. *Neurology* 75, 889–897. [PubMed: 20820000]

- [23]. Jack CR Jr., Lowe VJ, Senjem ML, Weigand SD, Kemp BJ, Shiung MM, Knopman DS, Boeve BF, Klunk WE, Mathis CA, Petersen RC (2008) 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnesic mild cognitive impairment. *Brain* 131, 665–680. [PubMed: 18263627]
- [24]. Jack CR Jr., Wiste HJ, Weigand SD, Therneau TM, Lowe VJ, Knopman DS, Gunter JL, Senjem ML, Jones DT, Kantarci K, Machulda MM, Mielke MM, Roberts RO, Vemuri P, Reyes DA, Petersen RC (2017) Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers Dement* 13, 205–216. [PubMed: 27697430]
- [25]. Vemuri P, Lowe VJ, Knopman DS, Senjem ML, Kemp BJ, Schwarz CG, Przybelski SA, Machulda MM, Petersen RC, Jack CR Jr., (2017) Tau-PET uptake: Regional variation in average SUVR and impact of amyloid deposition. *Alzheimers Dement (Amst)* 6, 21–30. [PubMed: 28138510]
- [26]. Jack CR Jr., Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, Hampel H, Jagust WJ, Johnson KA, Knopman DS, Petersen RC, Scheltens P, Sperling RA, Dubois B (2016) A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology* 87, 539–547. [PubMed: 27371494]
- [27]. Stricker NH, Lundt ES, Edwards KK, Machulda MM, Kremers WK, Roberts RO, Knopman DS, Petersen RC, Mielke MM (2019) Comparison of PC and iPad administrations of the Cogstate Brief Battery in the Mayo Clinic Study of Aging: Assessing cross-modality equivalence of computerized neuropsychological tests. *The Clinical neuropsychologist* 33, 1102–1126. [PubMed: 30417735]
- [28]. Stricker NH, Lundt ES, Alden EC, Albertson SM, Machulda MM, Kremers WK, Knopman DS, Petersen RC, Mielke MM (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, 21–28.
- [29]. Cogstate (2018) Cogstate Pediatric and Adult Normative Data, Cogstate, Inc, New Haven, CT.
- [30]. Perkins NJ, Schisterman EF (2005) The Youden Index and the optimal cut-point corrected for measurement error. *Biom J* 47, 428–441. [PubMed: 16161802]
- [31]. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 837–845. [PubMed: 3203132]
- [32]. Lim YY, Ellis KA, Harrington K, Kamer A, Pietrzak RH, Bush AI, Darby D, Martins RN, Masters CL, Rowe CC, Savage G, Szoeki C, Villemagne VL, Ames D, Maruff P (2013) Cognitive consequences of high A β amyloid in mild cognitive impairment and healthy older adults: Implications for early detection of Alzheimer's disease. *Neuropsychology* 27, 322–332. [PubMed: 23688214]
- [33]. Baker JE, Lim YY, Jaeger J, Ames D, Lautenschlager NT, Robertson J, Pietrzak RH, Snyder PJ, Villemagne VL, Rowe CC, Masters CL, Maruff P (2018) Episodic Memory and Learning Dysfunction Over an 18-Month Period in Preclinical and Prodromal Alzheimer's Disease. *J Alzheimers Dis* 65, 977–988. [PubMed: 30103330]
- [34]. Insel PS, Donohue MC, Mackin RS, Aisen PS, Hansson O, Weiner MW, Mattsson N, Alzheimer's Disease Neuroimaging I (2016) Cognitive and functional changes associated with Abeta pathology and the progression to mild cognitive impairment. *Neurobiol Aging* 48, 172–181. [PubMed: 27710807]
- [35]. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr., Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, Phelps CH (2011) 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. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 7, 280–292.
- [36]. Weissberger GH, Strong JV, Stefanidis KB, Summers MJ, Bondi MW, Stricker NH (2017) Diagnostic Accuracy of Memory Measures in Alzheimer's Dementia and Mild Cognitive Impairment: a Systematic Review and Meta-Analysis. *Neuropsychol Rev* 27, 354–388. [PubMed: 28940127]

- [37]. Nation DA, Ho JK, Dutt S, Han SD, Lai MHC (2019) Neuropsychological Decline Improves Prediction of Dementia Beyond Alzheimer's Disease Biomarker and Mild Cognitive Impairment Diagnoses. *J Alzheimers Dis* 69, 1171–1182. [PubMed: 31104015]
- [38]. Lee JH, Byun MS, Yi D, Sohn BK, Jeon SY, Lee Y, Lee JY, Kim YK, Lee YS, Lee DY (2018) Prediction of Cerebral Amyloid With Common Information Obtained From Memory Clinic Practice. *Front Aging Neurosci* 10, 309. [PubMed: 30337868]
- [39]. Carr SJA, Jaeger J, Bian S, He P, Maserejian N, Wang W, Maruff P, Enayetallah A, Wang Y, Chen Z, Lerner A, Tatsuoka C, Australian Imaging B, Lifestyle Flagship Study of A (2019) Associating Cognition With Amyloid Status Using Partially Ordered Set Analysis. *Front Neurol* 10, 976. [PubMed: 31572291]
- [40]. Verberk IMW, Slot RE, Verfaillie SCJ, Heijst H, Prins ND, van Berckel BNM, Scheltens P, Teunissen CE, van der Flier WM (2018) Plasma Amyloid as Prescreener for the Earliest Alzheimer Pathological Changes. *Ann Neurol* 84, 648–658. [PubMed: 30196548]
- [41]. Knopman DS, Caselli RJ (2012) Appraisal of cognition in preclinical Alzheimer's disease: a conceptual review. *Neurodegener Dis Manag* 2, 183–195. [PubMed: 22798965]
- [42]. Kandel BM, Avants BB, Gee JC, Arnold SE, Wolk DA, Alzheimer's Disease Neuroimaging I (2015) Neuropsychological Testing Predicts Cerebrospinal Fluid Amyloid- β in Mild Cognitive Impairment. *Journal of Alzheimer's disease : JAD* 46, 901–912. [PubMed: 25881908]
- [43]. Townley RA, Syrjanen JA, Botha H, Kremers WK, Aakre JA, Fields JA, Machulda MM, Graff-Radford J, Savica R, Jones DT, Knopman DS, Petersen RC, Boeve BF (2019) Comparison of the Short Test of Mental Status and the Montreal Cognitive Assessment Across the Cognitive Spectrum. *Mayo Clin Proc* 94, 1516–1523. [PubMed: 31280871]
- [44]. Rizk-Jackson A, Insel P, Petersen R, Aisen P, Jack C, Weiner M (2013) Early indications of future cognitive decline: stable versus declining controls. *PLoS One* 8, e74062.
- [45]. Lin M, Gong P, Yang T, Ye J, Albin RL, Dodge HH (2018) Big Data Analytical Approaches to the NACC Dataset: Aiding Preclinical Trial Enrichment. *Alzheimer Dis Assoc Disord* 32, 18–27. [PubMed: 29227306]
- [46]. Machulda MM, Pankratz VS, Christianson TJ, Ivnik RC, Mielke MM, Roberts RO, Knopman DS, Boeve BF, Petersen RC (2013) Practice effects and longitudinal cognitive change in normal aging vs. incident mild cognitive impairment and dementia in the Mayo Clinic Study of Aging. *The Clinical Neuropsychologist* 27, 1247–1264. [PubMed: 24041121]
- [47]. Elman JA, Jak AJ, Panizzon MS, Tu XM, Chen T, Reynolds CA, Gustavson DE, Franz CE, Hatton SN, Jacobson KC, Toomey R, McKenzie R, Xian H, Lyons MJ, Kremen WS (2018) Underdiagnosis of mild cognitive impairment: A consequence of ignoring practice effects. *Alzheimers Dement (Amst)* 10, 372–381. [PubMed: 30003138]
- [48]. Machulda MM, Hagen CE, Wiste HJ, Mielke MM, Knopman DS, Roberts RO, Vemuri P, Lowe VJ, Jack CR Jr., Petersen RC (2017) Practice effects and longitudinal cognitive change in clinically normal older adults differ by Alzheimer imaging biomarker status. *Clin Neuropsychol* 31, 99–117. [PubMed: 27724156]
- [49]. Buckley RF, Sparks KP, Papp KV, Dekhtyar M, Martin C, Burnham S, Sperling RA, Rentz DM (2017) Computerized Cognitive Testing for Use in Clinical Trials: A Comparison of the NIH Toolbox and Cogstate C3 Batteries. *The journal of prevention of Alzheimer's disease* 4, 3–11.
- [50]. Machulda MM, Lundt ES, Albertson SM, Kremers WK, Mielke MM, Knopman DS, Bondi MW, Petersen RC (2019) Neuropsychological subtypes of incident mild cognitive impairment in the Mayo Clinic Study of Aging. *Alzheimer's & Dementia* 15, 878–887.
- [51]. Knopman DS, Beiser A, Machulda MM, Fields J, Roberts RO, Pankratz VS, Aakre J, Cha RH, Rocca WA, Mielke MM (2015) Spectrum of cognition short of dementia: Framingham heart study and Mayo Clinic study of aging. *Neurology* 85, 1712–1721. [PubMed: 26453643]

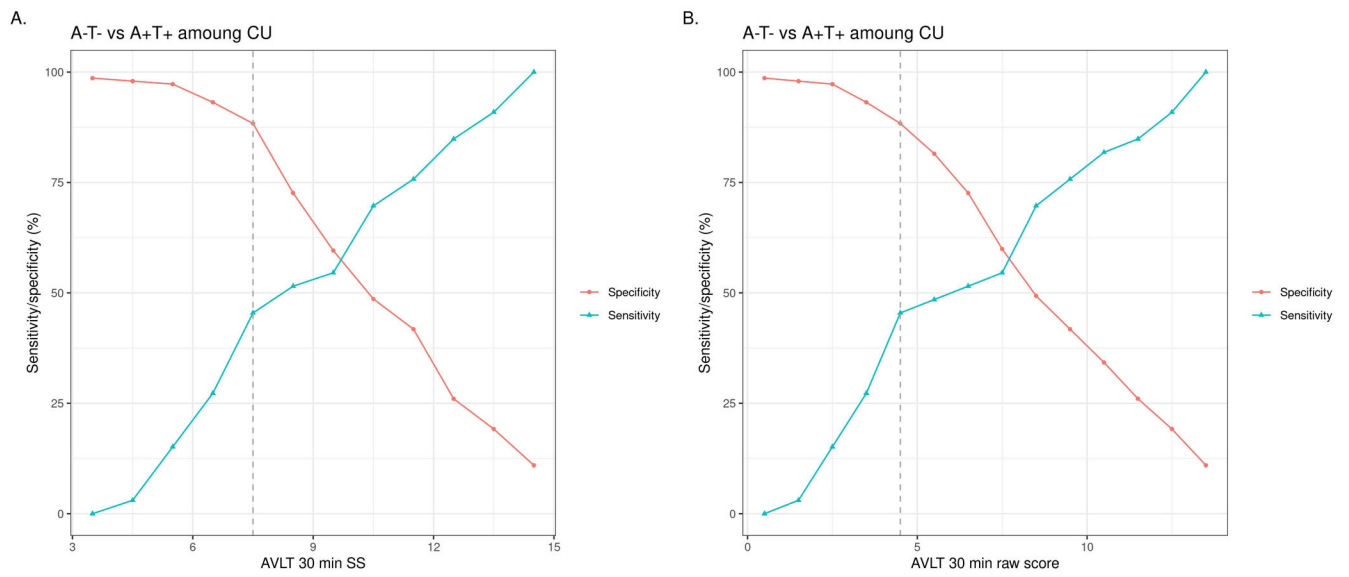


Figure 1. AVLT 30 minute delayed recall sensitivity/specificity plot for differentiating CU A-T- and CU A+T+ groups. Panel A shows age-corrected scaled scores and panel B shows raw scores. The dotted line indicates the derived optimal cut-off per Youden's index, which jointly maximizes the sum of sensitivity and specificity.

Table 1.

Study Definitions.

Consensus Diagnosis: Cognitively Unimpaired (CU)	
Grouped by Future Diagnosis	Grouped by Biomarker Profile (AT) ¹
CU Stable ² (n = 450)	CU A-T- (n = 146)
CU Converter ³ (n = 150)	CU A+T- (n = 61)
	CU A+T+ (n = 33)
sOBJ	
Exploring the utility of possible definitions of cross-sectional subtle objective cognitive impairment (sOBJ)	
1. Cogstate Brief Battery Lm/WM Composite ⁴ -1 SD	
2. AVLT 30-Minute Delay ⁵ -1 SD	

¹ Biomarker profile defined by neuroimaging biomarkers of amyloid and tau within two years of baseline Cogstate.

² CU at baseline Cogstate and remained CU at all available follow-up visits. N=51 participants with biomarker data were also represented in this group (n=29 CU A-T-, n=14 CU A+T-, n = 8 CU A+T+).

³ CU at baseline Cogstate and converted to MCI/dementia at any follow-up visit. N=18 participants with biomarker data were also represented in this group (n=5 CU A-T-, n=5 CU A+T-, n = 8 CU A+T+).

⁴ Cogstate is independent of consensus diagnosis; -1 SD cut-off is equivalent to an age-corrected z-score of -1 derived from normative data provided by Cogstate.

⁵ AVLT is considered as part of consensus diagnosis; -1 SD cut-off is equivalent to a Mayo's Older Americans Normative Studies age-corrected scaled score of 7.

Note. CU = Cognitively Unimpaired; A = amyloid; T = tau; sOBJ = subtle objective cognitive impairment; AVLT = Auditory Verbal Learning Test 30-Minute Delayed Recall; Lm/WM = Learning/Working Memory.

Table 2. Demographic characteristics and mean performance on memory measures among cognitively unimpaired (CU) participants.

	CU A-T- 146	CU A+T+ 33	CU A-T- 61	p-value - vs A+T+	p-value vs A+T-	p-value CU A +T- vs A+T+
Age Mean (SD)	66.32 (12.06)	79.52 (6.27)	77.78 (9.41)	< 0.001	< 0.001	0.341
Education Mean (SD)	15.24 (2.44)	14.79 (3.08)	14.77 (2.81)	0.362	0.229	0.978
Sex (% Male)	78 (53.4%)	20 (60.6%)	37 (60.7%)	0.454	0.340	0.996
Short Test of Mental Status [†]	36.18 (1.99)	35.69 (2.12)	35.08 (2.31)	0.214	< 0.001	0.220
AVLT Delay MOANS SS [†]	10.58 (2.92)	8.82 (3.05)	9.05 (2.58)	0.002	< 0.001	0.699
Prior AVLT exposure (%)	26.7%	72.7%	56.7%	< 0.001	< 0.001	0.126
CBB Learning/Working Memory Composite z-score	0.22 (0.78)	-0.07 (0.98)	0.06 (1.27)	0.072	0.274	0.617
CBB Learning/Working Memory Composite				0.008	0.083	0.369
Normal z > -1	138 (95.2%)	27 (81.8%)	54 (88.5%)			
sOBJ z -1	7 (4.8%)	6 (18.2%)	7 (11.5%)			
AVLT 30 minute delay [†]				< 0.001	0.016	0.043
Normal SS > 7	129 (88.4%)	18 (54.5%)	45 (75.0%)			
sOBJ SS 7	17 (11.6%)	15 (45.5%)	15 (25.0%)			

Note. P-values represent linear model ANOVAs for mean comparisons or Pearson's Chi-squared test for frequency comparisons of amyloid positive groups with the biomarker negative group. A = amyloid; T = tau; CBB = Cogstate Brief Battery; AVLT = Auditory Verbal Learning Test 30-Minute Delayed Recall; MOANS = Mayo's Older Americans Normative Studies; SS = scaled score; sOBJ = objective subtle cognitive impairment.

[†] Performance is considered as part of consensus diagnosis.

Diagnostic accuracy of memory measures based on derived, optimal cut-off scores and conventional cut-off scores (± 1 SD) indicative of subtle objective cognitive impairment (sOBJ).

Table 3.

Measure	Group Comparison	AUC (95% CI)	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
CBB Lm/WM composite z-score	A-T- vs A+T- among CU	0.51 (0.42,0.60)	-0.23 z	0.38 (0.26,0.51)	0.75 (0.68,0.82)
	A-T- vs A+T+ among CU	0.59 (0.47,0.72)	Conventional	0.11 (0.05,0.20)	0.95 (0.92,0.99)
	CU stable vs converter	0.70 (0.65,0.74)	Conventional	0.55 (0.36,0.73)	0.68 (0.61,0.76)
AVLT delay MOANS Scaled Score ¹	A-T- vs A+T- among CU	0.64 (0.57,0.72)	11.5 SS	0.18 (0.06,0.33)	0.95 (0.91,0.99)
	A-T- vs A+T+ among CU	0.66 (0.56, 0.77)	Conventional	0.77 (0.70,0.83)	0.58 (0.53, 0.62)
	CU stable vs converter	0.70 (0.65, 0.75)	Conventional	0.25 (0.18,0.32)	0.88 (0.85,0.91)
AVLT delay MOANS Scaled Score ¹	A-T- vs A+T- among CU	0.64 (0.57,0.72)	11.5 SS	0.83 (0.73,0.92)	0.42 (0.34,0.50)
	A-T- vs A+T+ among CU	0.66 (0.56, 0.77)	Conventional	0.25 (0.15,0.37)	0.88 (0.83,0.93)
	CU stable vs converter	0.70 (0.65, 0.75)	Conventional	0.45 (0.27, 0.64)	0.88 (0.83, 0.93)
AVLT delay MOANS Scaled Score ¹	A-T- vs A+T- among CU	0.64 (0.57,0.72)	11.5 SS	0.45 (0.27, 0.61)	0.88 (0.83, 0.93)
	A-T- vs A+T+ among CU	0.66 (0.56, 0.77)	Conventional	0.70 (0.63, 0.77)	0.61 (0.56, 0.65)
	CU stable vs converter	0.70 (0.65, 0.75)	Conventional	0.38 (0.31, 0.46)	0.86 (0.83, 0.89)

Note. A = amyloid; T = tau; CU = Cognitively Unimpaired; CBB = Cogstate Brief Battery; Lm/WM = Learning/Working Memory; AVLT = Auditory Verbal Learning Test 30-Minute Delayed Recall; MOANS = Mayo's Older Americans Normative Studies.

¹ Performance is considered as part of consensus diagnosis.

Table 4.

Odds ratios from logistic regression models after adjusting for age, education and sex.

Groups compared	term	OR (95% CI)	p
CU A-T- vs. CU A+T+	CBB Lrm/Wm composite z 1 unit less	1.25 (0.78, 2.05)	0.35
CU A-T- vs. CU A+T-	CBB Lrm/Wm composite z 1 unit less	1.03 (0.74, 1.46)	0.85
CU A-T- vs. CU A+T+	AVLT Delay z 1 unit less [/]	1.54 (0.96, 2.53)	0.08
CU A-T- vs. CU A+T-	AVLT Delay z 1 unit less [/]	1.58 (1.07, 2.39)	0.03

Note. For logistic regression analyses, AVLT delayed recall scaled score was rescaled to z-score units to aid interpretation. For every 1.0 z-score less, the odds of being in the biomarker positive group increases by the OR reported above. A = amyloid; T = tau; CU = Cognitively Unimpaired; CBB = Cogstate Brief Battery; AVLT = Auditory Verbal Learning Test 30-Minute Delayed Recall.

[/] Performance is considered as part of consensus diagnosis.

Table 5.

Diagnostic accuracy of secondary measures.

	Description	Threshold (Optimal)	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
OCL accuracy (z-score)	A-T- vs A+T- among CU	0.18	0.56 (0.44, 0.69)	0.58 (0.50, 0.66)	0.53 (0.44, 0.62)
	A-T- vs A+T+ among CU	-0.21	0.45 (0.27, 0.64)	0.77 (0.70, 0.83)	0.60 (0.49, 0.71)
	CU stable vs converter	-0.13	0.63 (0.55, 0.70)	0.67 (0.63, 0.72)	0.68 (0.63, 0.72)
OCL accuracy (Arcsine transformed raw score)	A-T- vs A+T- among CU	0.66	0.48 (0.34, 0.61)	0.68 (0.61, 0.76)	0.55 (0.46, 0.64)
	A-T- vs A+T+ among CU	0.66	0.55 (0.39, 0.70)	0.68 (0.60, 0.75)	0.62 (0.51, 0.73)
	CU stable vs converter	0.64	0.60 (0.51, 0.68)	0.70 (0.65, 0.74)	0.67 (0.62, 0.72)
AVLT delay (raw score) [†]	A-T- vs A+T- among CU	9.50	0.83 (0.73, 0.92)	0.42 (0.34, 0.49)	0.64 (0.57, 0.72)
	A-T- vs A+T+ among CU	4.50	0.45 (0.30, 0.64)	0.88 (0.83, 0.93)	0.67 (0.56, 0.78)
	CU stable vs converter	7.50	0.70 (0.63, 0.78)	0.61 (0.56, 0.65)	0.70 (0.65, 0.75)
STMS (raw score) [†]	A-T- vs A+T- among CU	36.5	0.67(0.56,0.79)	0.58(0.50,0.66)	0.67(0.59,0.74)
	A-T- vs A+T+ among CU	35.5	0.41(0.25,0.56)	0.71(0.64,0.78)	0.57(0.46,0.68)
	CU stable vs converter	34.5	0.62(0.55,0.70)	0.75(0.71,0.79)	0.74(0.69,0.78)

Note. A = amyloid; T = tau; OCL = One Card Learning; AVLT = Auditory Verbal Learning Test 30-Minute Delayed Recall. STMS = Short Test of Mental Status.

[†] Performance is considered as part of consensus diagnosis.

Table 6. Demographic characteristics and mean performance on memory measures among CU stable and converter participants.

	CU Stable <i>M</i> (<i>SD</i>) (n = 450)	CU Converter <i>M</i> (<i>SD</i>) (n = 150)	<i>p</i> -value
Age Mean (SD)	75.437 (10.345)	75.502 (10.493)	0.947
Education Mean (SD)	14.171 (2.528)	13.940 (2.499)	0.331
Sex (% Male)	240 (53.3%)	80 (53.3%)	1.00
Time to conversion (years)	-	2.723 (1.517)	
Total follow up (years)	3.664 (1.691)	3.848 (1.801)	0.257
Short Test of Mental Status [†]	35.563 (1.980)	33.624 (2.431)	< 0.001
AVLT 30 MOANS SS [†]	10.282 (2.771)	8.245 (2.918)	< 0.001
Prior AVLT exposure (%)	56.5%	58.5%	0.676
Learning/Working Memory Composite (z-score)	0.014 (0.974)	-0.552 (0.873)	< 0.001
Learning/Working Memory Composite			< 0.001
Normal z > -1	396 (88.4%)	113 (75.3%)	
sOBJ z -1	52 (11.6%)	37 (24.7%)	
AVLT 30 minute delay [†]			< 0.001
Normal SS > 7	382 (86.0%)	91 (61.9%)	
sOBJ SS 7	62 (14.0%)	56 (38.1%)	

Note. P-values represent linear model ANOVAs for mean comparisons or Pearson's Chi-squared test for frequency comparisons. MOANS = Mayo's Older Americans Normative Studies; SS = scaled score; sOBJ = objective subtle cognitive impairment.

[†] Performance is considered as part of consensus diagnosis.