HHS Public Access

Author manuscript

J Int Neuropsychol Soc. Author manuscript; available in PMC 2021 March 18.

Published in final edited form as:

J Int Neuropsychol Soc. 2021 March ; 27(3): 211–226. doi:10.1017/S1355617720000752.

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[a]

Nikki H. Stricker, **Teresa J. Christianson**, **Emily S. Lundt**, **Eva C. Alden**, **Mary M. Machulda**, **Julie A. Fields**, **Walter K. Kremers**, **Clifford R. Jack Jr**, **David S. Knopman**, **Michelle M. Mielke**, **Ronald C. Petersen**

Department of Psychiatry and Psychology, Division of Neurocognitive Disorders (Drs Stricker, Alden, Machulda, and Fields), Department of Health Sciences Research (Mss Christianson and Lundt, and Drs Kremers, Mielke, and Petersen), Department of Radiology (Dr Jack), and Department of Neurology (Drs Knopman and Petersen), Mayo Clinic, Rochester, Minnesota

Abstract

Objective: Rey's Auditory Verbal Learning Test (AVLT) is a widely used word list memory test. We update normative data to include adjustment for verbal memory performance differences between men and women and illustrate the effect of this sex-adjustment and the importance of excluding participants with mild cognitive impairment (MCI) from normative samples.

Method: This study advances the Mayo's Older Americans Normative Studies (MOANS) by using a new population-based sample through the Mayo Clinic Study of Aging, which randomly samples residents of Olmsted County, Minnesota, from age- and sex-stratified groups. Regression-based normative T-score formulas were derived from 4,428 cognitively unimpaired adults aged 30 to 91 years. Fully adjusted T-scores correct for age, sex, and education. We also derived T-scores that correct for 1) age or 2) age and sex. Test-retest reliability data are provided.

Results: From raw score analyses, sex explained a significant amount of variance in performance above and beyond age (8%-10%). Applying original age-adjusted MOANS norms to the current sample resulted in significantly fewer-than-expected participants with low delayed recall performance, particularly in women. After application of new T-scores adjusted only for age, even in normative data derived from this sample, these age-adjusted T-scores showed scores < 40 T occurred more frequently among men and less frequently among women relative to T-scores that also adjusted for sex.

[a]Table 1 and Appendix used with permission of Mayo Foundation for Medical Education and Research; all rights reserved.

Reprints: Nikki H. Stricker, PhD, LP, Division of Neurocognitive Disorders, Mayo Clinic, 200 First St SW, Rochester, MN 55905 (stricker.nikki@mayo.edu Phone: 507-284-2649 Fax: 507-284-4158).

Portions of this manuscript were presented as a poster at the 17th Annual American Academy of Clinical Neuropsychology, Chicago, Illinois, June 5-8, 2019. Portions of this manuscript were presented at the 39th Annual National Academy of Neuropsychology Conference, San Diego, California, November 14, 2019. Portions of this manuscript and the figures were presented at the internal Mayo Clinic Neuropsychology Multi-Site Seminar, April 8, 2020, Rochester, Minnesota.

Conclusions: Our findings highlight the importance of using normative data that adjust for sex with measures of verbal memory and provide new normative data that allow for this adjustment for the AVLT.

Keywords

cognitive aging; memory; mild cognitive impairment; neuropsychology; reliability and validity; psychometrics

Mayo's Older Americans Normative Studies (MOANS) have served as an important resource for clinicians for nearly three decades, providing valuable normative data for older adults (Ivnik et al., 1992a, 1992b, 1992c; Malec et al., 1992), especially those older than 75 years, for whom normative data are historically limited (Smith & Ivnik, 2003). These norms are freely available, are provided in published normative textbooks (Strauss, Sherman, Spreen, & Spreen, 2006), and have been widely applied in clinical and research contexts. However, normative data benefit from updates to help account for historical cohort differences (Hessel, Kinge, Skirbekk, & Staudinger, 2018). Updates are also needed because prior MOANS studies only included older adults, which limits broader application of this important normative resource.

Despite the rigor of the original MOANS study procedures at the time of data collection, which excluded persons with a dementia diagnosis, data collection began in the late 1980s, before the development and widespread use of *mild cognitive impairment* (MCI) as a construct (Petersen, 2004; Petersen et al., 2009; Petersen et al., 1999). Thus, a sizeable proportion of those included in the original MOANS cohort most likely had undetected MCI, particularly in older age groups, because the prevalence of MCI increases substantially with age (Petersen et al., 2018). Updated norms that exclude persons with MCI are needed because their inclusion will artificially lower the sensitivity of normative data when applied to patient groups. Word list memory tests consistently demonstrate high diagnostic accuracy for distinguishing persons who are cognitively unimpaired (CU) from those with clinically diagnosed MCI or Alzheimer's dementia, often outperforming other memory measures; the sensitivity of memory measures, however, is lower for MCI than Alzheimer's dementia (Weissberger et al., 2017).

Over the human life span, women perform better than men on tests of verbal memory (Aartsen, Martin, Zimprich, & Longitudinal Aging Study, 2004; Kramer, Delis, Kaplan, O'Donnell, & Prifitera, 1997). This female advantage in verbal memory is maintained even when women have the same levels of neuropathologic burden (cortical amyloid β deposition or hippocampal atrophy) relative to men (Sundermann et al., 2017; Sundermann et al., 2016). As a result of this advantage in verbal memory, women may have a delay in the diagnosis of amnestic MCI when tests that do not adjust for sex are used (Roberts et al., 2012). Sundermann et al., (2019) show that using standard, non–sex-adjusted norms underestimates MCI in 10% of women and overestimates it in 10% of men. They further show that sex-adjusted normative data improve the diagnostic accuracy of memory measures for distinguishing the presence or absence of MCI based on positivity rates of cerebrospinal fluid levels of phospho-tau/amyloid β and cortical amyloid β deposition (by florbetapir F 18

positron emission tomography) and APOE-ε4 frequency in misclassified persons. This implies that failure to use sex-adjusted normative data may contribute to missed diagnosis of MCI in women and an observed higher prevalence of MCI in men (Di Carlo et al., 2007; Petersen et al., 2010).

Thus, the availability of sex-adjusted normative data for clinical use is critical for minimizing sex-related health disparities by helping to prevent underdiagnosis in women and overdiagnosis in men. The original MOANS norms for the Auditory Verbal Learning Test (AVLT) did not adjust for sex or education because of a decision not to correct for demographic variables accounting for less than 10% of shared variance. Subsequent MOANS norms for recognition percentage correct, however, found that the observed degree of sex differences warranted separate recognition norms for men and women (Harris, Ivnik, & Smith, 2002).

The current study aimed to 1) demonstrate the need for updated AVLT normative data by examining base rates of low delayed recall performances between men and women when applying original MOANS norms; 2) illustrate the effect of adjusting for additional demographic variables other than age; and 3) provide updated normative data for the AVLT. These new norms are referred to as the Mayo Normative Studies (MNS) to differentiate them from the MOANS. These normative data will improve upon past MOANS norms by including adjustment for age, sex, and education, expanding the age range and sample size through use of a new population-based sample, and excluding persons with MCI. We hypothesized that application of original MOANS norms to an independent sample would result in lower-than-expected base rates of low performance, particularly in women, and that updated AVLT norms adjusting for sex would provide more balanced base rates of low scores across men and women relative to updated norms that adjust only for age.

Method

Participants

The current study follows in the tradition of the MOANS, but the study sample is independent of previously published MOANS norms (Ivnik et al., 1992a, 1992b, 1992c). This study uses the extensive epidemiologic data collected through the Mayo Clinic Study of Aging (MCSA), which provides an excellent basis for the development of normative standards. The MCSA was initiated in 2004 as a population-based study of aging. Participants were randomly sampled from Olmsted County, Minnesota, using the Rochester Epidemiology Project Medical Records linkage system (St Sauver, Grossardt, Yawn, Melton, & Rocca, 2011); 97% of Olmsted County residents agree to the use of their medical records for research. More than 60% of contacted residents enroll in the MCSA. The details of the study design and sampling procedures have been previously published; enrollment follows an age- and sex-stratified random sampling design to ensure that men and women are equally represented in each 10-year age stratum (Roberts et al., 2008). The MCSA initially enrolled residents aged 70 to 89 years. Enrollment was extended in 2012 to include ages 50 to 69 years and was again extended in 2015 to include ages 30 to 49 years, following the same sampling methods. The study protocols were approved by the Mayo Clinic and

Olmsted Medical Center institutional review boards. All participants provided written informed consent.

Study visits included a neurologic evaluation, interview by a study coordinator, and neuropsychological testing conducted by a psychometrist (Roberts et al., 2008). Neurologic evaluation was completed by a physician and included medical history review, complete neurologic examination, and administration of the Short Test of Mental Status (Kokmen, Smith, Petersen, Tangalos, & Ivnik, 1991). The study coordinator collected demographic information, medical history, and information about subjective memory and daily functioning from both the participant and informant by using the Clinical Dementia Rating (CDR®) instrument (Morris, 1993). The neuropsychological battery included nine tests covering four domains, as previously described (Roberts et al., 2008; Wennberg et al., 2018). To decrease the possibility of circularity or incorporation bias, the neuropsychologist's diagnostic impression was not considered for diagnosis for this study, which varies from the consensus diagnosis typically used in the MCSA. Instead, a determination of *cognitively unimpaired* by both the interviewing study coordinator and examining physician (who were blind to neuropsychological data) was required to be eligible for study inclusion (Petersen, 2004; Petersen et al., 2010; Roberts et al., 2008).

The current study included cognitively unimpaired persons aged 30 years or older who completed the AVLT at a baseline study visit and were test naïve (ie, participants were excluded if they had previously participated in the Mayo Clinic Alzheimer's Disease Research Center and were administered the AVLT). Additional exclusion criteria were having a terminal illness or being in hospice. Baseline AVLT data were available for 4,428 cognitively unimpaired adults aged 30 to 91 years, a sample size 8 times larger than that for previously published MOANS AVLT data (Ivnik et al., 1992a, 1992b, 1992c; Malec et al., 1992). Participants are predominantly White (see Table 1). The mean (SD) age of the overall sample was 68.3 (13.1) years, and the mean (SD) education level was 14.7 (2.6) years. Results by decade for demographics and all AVLT variables are provided in Supplemental Table 1 to further characterize the sample.

Auditory Verbal Learning Test

The AVLT is a widely used word list memory test that is also known as the Rey Auditory Verbal Learning Test (RAVLT) (Rey, 1964). Participants are read a list of 15 words (list A) and asked to recall as many as possible over five learning trials, after one intervening list, and again 30 minutes later. A written recognition test is then administered. The procedures for test administration are the same as those used in the original MOANS studies (Ivnik et al., 1992a, 1992b, 1992c; Malec et al., 1992). Ferman et al (2005) provide details about the specific form used, including the full word lists, recognition items, and verbatim instructions.

The four primary variables of interest for the current study were 1) the total recall score for trials 1 through 5, 2) 30-minute delayed recall, 3) sum of trials (trials 1-5 total + trial 6 + 30-minute recall), and 4) recognition percentage correct. Sum of trials is sensitive to early changes in memory (Jack et al., 2015). When recognition measures have an equal number of targets and distractors, simple methods correlate robustly with more complex methods such

as d-prime (Delis et al., 2000). Thus, we used the same method as in prior MOANS studies to calculate recognition percentage correct (Ferman et al., 2005; Harris et al., 2002). Secondary variables were collected to allow for further interpretation and to facilitate investigation of multiple aspects of memory performance; however, use of the primary variables is encouraged for clinical decision making. See Appendix (Table A1) for a full list of AVLT indices and relevant formulas. The full 5 trial version of the AVLT is recommended, but Trials 1-3 total is included to support use of an abbreviated administration when needed. If an abbreviated version of the AVLT with only 3 learning trials is administered, normative data cannot be applied to any other AVLT measures beyond Trial 3. One secondary measure, the memory efficiency score, can be calculated across different versions of the AVLT. It combines measures of encoding and retention and has demonstrated good sensitivity and specificity in distinguishing persons with Alzheimer's dementia from those with behavioral variant frontotemporal dementia and cognitively unimpaired persons (Ricci, Graef, Blundo, & Miller, 2012). Data for repetition errors were not available. Item-level data were not available; thus, serial position effects were not included.

Statistical Methods

To demonstrate the need for updated normative data that exclude participants with MCI and adjust for sex, we applied original age-adjusted MOANS norms (Ivnik et al., 1992a, 1992b, 1992c; Malec et al., 1992) for AVLT 30-minute delayed recall to cognitively unimpaired participants aged 56 years or older. We used one-sample tests of proportions comparing observed rates of low test performance with rates expected on the basis of a normal distribution of performance for the full sample and when stratified by sex. These scaled scores are expected to approximate a normal distribution (mean [SD], 10 [3]) due to the forced normalization and smoothing process applied in MOANS (Ivnik et al., 1992a, 1992b, 1992c; Malec et al., 1992). On this basis, we can compute the expected percentage of participants performing below conventional cutoff values based on a normal distribution and compare them against the observed percentage in our sample. Analyses focused on rates of low test performance using a <−1 SD cut-off. Descriptors applied in clinical practice, including the recently published American Academy of Clinical Neuropsychology Consensus Statement (Guilmette et al., 2020), often include several descriptor categories to indicate degree of departure from average performance. Our focus on a <−1 SD cut-off captures all performances below that cut-off together, thus collapsing across several descriptor categories. For this reason, we describe all performances below −1 SD as "low."

Normative Data Development

In contrast to cell-based methods used in prior MOANS studies, we used a regression-based normative approach. First, each test score distribution was normalized similarly to approaches described by other groups (Casaletto et al., 2015; Heaton, 2004). Raw test scores were converted to normalized scaled scores by calculating their percentile rank based on the cumulative frequency distributions of raw scores and scaling them to a mean (SD) of 10 (3). Converting to scaled scores allows for easier comparison between measures by putting all measures on the same scale and led to improved distribution properties for a few non-normally distributed variables.

Demographic Corrections/Derivation of Fully Corrected T-Scores

Selecting Demographic Variables—The normalized test scores were regressed on age, age squared, sex, and education. Given the broad age range of the sample, we decided a priori to include a nonlinear age term, on the basis of past research (Salthouse, 2010; Van der Elst, van Boxtel, van Breukelen, & Jolles, 2005; Verhaeghen & Salthouse, 1997) and visual inspection of our data. Expected correlations of age, sex and education with AVLT performance were confirmed in our sample. Traditional stepwise procedures to evaluate the need for additional predictors based on p-values were overly-sensitive given our large sample size. To avoid overfitting the model, we considered the need for additional predictors using 1% incremental variance explained where the R-squared from a model including the additional predictor was required to be improved by 1% or better. None of the terms for non-linear education (quadratic, cubic), cubic age, or two-way interactions met this requirement.

Derivation of Fully Corrected T-Scores—Linear regression models were fit on each primary and secondary scaled score and adjusted for age, age squared, sex, and education. To assess normality of standardized residuals, Q-Q plots were reviewed. Each participant's covariate-adjusted score was converted to a T-score by rescaling the residuals ($e_i = Y_i - Y_{pred}$) to mean (SD) of 50 (10). We computed mean (SD) values by age group (30-59 years, 60-69 years, 70-79 years, and ≥80 years), sex, and years of education (8-12, 13-15, 16, and 17-20) and compared them against expected criteria. If means were within 3 T-score points of the expected value of 50 (47 to 53) (Heaton, 2004) and if the SD was 9.4 to 10.6 T-score points, then we made no further adjustments for mean and SD. Four variables had SDs outside the specified range, which indicated the need for smoothing. Smoothing by age and by age squared were considered, with age squared performing best. To continuously estimate the variances, we fit a new linear regression between absolute value of residuals and age squared. The T-score equations for measures list B, short- and long-term percentage retention, and memory efficiency score involve adjusted means and SDs based on these fits. A scaled score look-up table and T-score formulas were created for the generation of age, age squared, sex, and education-corrected T-scores (Appendix; Tables A2 and A3).

Derivation of Alternative T-Scores—Use of fully corrected T-scores is recommended to understand a person's level of performance relative to their expected level of performance compared with an average peer of the same age, sex, and education. This is most helpful in determining whether performance represents a decline from an estimated baseline of performance. However, in some circumstances a different standard of comparison may be helpful. For example, age-adjusted scores may be useful when comparing memory performance with other neuropsychological measures that only adjust for age. An option to calculate an age- and sex-adjusted score may be helpful if applying these norms to a person with an atypical educational background. For these circumstances, and to help facilitate future comparison of the effect of these differing levels of demographic adjustments, we also present regression equations adjusting for age alone and age and sex for our four primary variables (Appendix).

Cumulative Percentiles—Intrusion errors are extra-list intrusions during recall trials (ie, inserting an incorrect word not from the presented list during recall). Because total

intrusions were highly skewed, there were too few positive observations to be able to use the normative approach described above, and demographic variables explained very little variance (<1% combined; Table 2), we provide cumulative percentiles for the entire sample without stratifying by demographic variables for total intrusions (Appendix; Table A4). Inclusion of intrusions in these normative data is important given recent evidence that total intrusion errors on AVLT contributed to predicting progression from normal cognition to MCI or mild dementia, whereas learning slope and interference scores did not (Thomas et al., 2018).

Results

Need for Updated Norms

Application of original age-adjusted MOANS norms to cognitively unimpaired participants aged 56 years or older (n=3,603) resulted in significantly fewer-than-expected participants demonstrating low performance on delayed recall. By using a lenient cutoff value for low performance (<−1 SD, which is a scaled score <7), only 7.3% of participants demonstrated low performance (vs 15.9% expected; $P<.001$), and women were 3 times less likely to show low performance than men (3.5% vs 11.1%). Furthermore, only 0.6% of participants demonstrated performance at a cutoff of <−2 SD (scaled score <4), none of whom were aged 80 to 91 years (2.3% expected; $P<.001$).

Effects of Demographic Characteristics on AVLT Performance

Figure 1 and Table 2 demonstrate the robust and important effect of age on AVLT performance across nearly all variables. Age effects were most prominent for recall performances and explained 17% to 26% of the variance in performance across most recall variables, except for trial 1 recall, which showed relatively less but still significant variance explained by age. Measures reflecting memory storage, including percentage retention (short and long delay), recognition percentage correct, and the memory efficiency score, which combines retention and recognition performances, showed a less robust but significant relationship with age and explained 4% to 5% of the variance in performance.

Sex effects were also robust across numerous AVLT response variables. Sex explained 6% to 9% of variance across most recall variables (except trial 1 and list B). Sex explained more variance (7%) than age (5%) for recognition percentage correct.

Education effects were significant but explained less variance than both age and sex. Education explained 4% to 6% of variance across most immediate-recall variables. Less variance was explained by education for delayed recall performances (2%-3%), and education explained less than 1% of variance for retention and recognition measures.

Assessing Performance Characteristics of Fully Adjusted T-Scores

For every test variable, visual inspection of Q-Q plots was used to ensure normality. We inspected several plots to determine whether T-score corrections performed in the expected manner across the different levels of age, sex, and education for all variables. Means and 95% CIs for all groups by age (30-59 years, 60-69 years, 70-79 years, and 80-91 years),

years of education by bins (8-12, 13-15, 16, and 17-20), and sex were inspected (Supplemental Figure 1). All observed means were within 3 points of the expected value of 50 (Heaton, 2004) and SDs were within 0.6 units of 10 (except where specified above). The derived fully corrected T-scores had a mean of approximately 50 across all ages (Figure 2), and the average T-score for persons with 12 or more years of education was approximately 50 (Supplemental Figure 2). Subtle deviations from 50 can be seen for those with less than 12 years of education, but the mean is still within our a priori specified range. The sample size for participants with less than a high school degree was small relative to our other education groups (Supplemental Table 2), which may be related to the method of defining education in the MCSA that counts a General Educational Development degree as 12 years of education (Appendix). Fully adjusted T-scores effectively removed relationships to demographic variables as desired (all Pearson r<.003; all P>.84).

Assessing Performance Characteristics of Alternative T-Scores: Effect of Sex

Derived alternative T-scores had a mean of approximately 50 across all ages. Supplemental Figure 1 shows the effect of adjusting for sex and education, in addition to age alone, for the primary variables. When stratifying by age group, sex, and education level, T-scores deviated widely from the expected mean of 50 if only age adjustment was used, narrowing substantially when sex was added to the model and to a lesser degree with the addition of education adjustment. Figure 3 shows the effect on base rates of low performance between men and women when sex adjustments are applied. When applying T-scores that adjust only for age, even in normative data derived from this sample, rates of low performance in women (T<40) are approximately 3 times lower (6%-7%) than in men (19%-23%) and all are significantly different than expected base rates of low performance (95% CIs did not include the expected 14.7%). Application of fully-corrected T-scores yielded expected base rates of low test performance (see Supplementary Table 3).

Validation in Independent Sample

To validate results in an independent sample, we replicated several analyses using individuals newly enrolled in the MCSA after the time of data freeze used for this normative sample (March 2018) by including individuals age 56 or older to facilitate comparison to MOANS. Based on diagnosis per physician and study coordinator we had 261 CU participants with mean age 73.3 (8.0 SD), mean education 15.2 (2.4) and 50.2% were male. We applied MOANS and the new MNS to 30-minute delayed recall. In CU individuals, Wilcoxon signed rank with continuity correction for dependent sample analyses showed there were significant differences when we have the same participant with the same raw score but two evaluation systems (MOANS age-corrected scaled scores and MNS fully-corrected T-scores) for the overall group and among women ($p < 0.001$). There was no difference among men ($p > .05$). Figure 4 illustrates the percentage of participants showing low test performance across MOANS and MNS systems (age-adjusted and fully-adjusted T-scores), with results replicating analyses within the normative sample (also see Supplemental Table 4).

Test-Retest Reliability

For a subset of participants (80.3%) with one follow-up AVLT visit available, we assessed test-retest reliability of raw scores by computing single-rating, absolute-agreement, two-way mixed intraclass correlation coefficients (ICC; Shrout & Fleiss, 1979). We also report Pearson correlation coefficients (Table 3). The average time between the baseline and follow-up test administrations was 16.7 months (*SD* = 4.5; range= 8.1 to 37.7). The mean age of the subsample at baseline was 68.6 (12.5 SD), mean education was 14.8 (2.6) and 50.2% were male. Sum of trials showed high reliability. Several other recall measures showed adequate reliability (Trials 1-5 total, Trials 1-3 total, Trial 5, Trial 6, 30-minute recall). Single-trial immediate recall scores (e.g., Trial 1, List B, etc.) and measures of encoding (recognition percent correct, percent retention scores, memory efficiency scores) showed marginal to low reliability.

Discussion

This study demonstrates a clear need for updated AVLT normative data that exclude persons with MCI, in addition to dementia, and that adjust for sex. Consistent with our first hypothesis, application of the original MOANS to this independent sample results in lower-than-expected base rates of low performance, with 7% of participants having a delayed recall AVLT scaled score less than 7, and only 0.6% having a scaled score less than 4. This suggests that despite screening participants for dementia, the original MOANS normative cohort may have included persons with MCI because data collection predated the introduction of formal MCI diagnostic criteria (Petersen et al., 1999). The normative samples of many test publishers similarly do not exclude persons with MCI and therefore may similarly have lower sensitivity than expected when applied to patients (Delis, Kramer, Kaplan, & Ober, 2017). Our results demonstrate that this issue is amplified in women: Only 3.5% of women older than 55 years in our cohort had a delayed recall MOANS AVLT scaled score less than 7, compared with 11.1% of men.

These findings are particularly salient given that many clinicians view a −1 SD cutoff as a minimum expectation for mild levels of cognitive impairment, and the Jak/Bondi comprehensive criteria for MCI diagnosis require 2 scores lower than a −1 SD cutoff in one domain (Jak et al., 2009). In addition, although participants in this sample are cognitively unimpaired on the basis of consensus diagnosis between the study physician and study coordinator, a portion of the sample would still be expected to show low performance on normative scores based on an expected normal distribution of test scores. It is notable that none of our participants aged 80 years or older had a delayed recall MOANS AVLT scaled score less than 4 (ie, <2 SD below the mean). This highlights the importance of the updated norms we provide, particularly for older adults and women. Results suggest that normative data collected without careful exclusion criteria and participant screening most likely result in decreased sensitivity of normative scores and lower-than-expected base rates of low performance in cognitively unimpaired participants.

In agreement with previous studies, we found that sex accounts for variance in AVLT performance above and beyond age (8%-10%), with women consistently outperforming men on AVLT measures. For example, Van der Elst et al. (2005), in a large sample of 1,855

healthy participants aged 24 to 81 years, showed significant effects of sex on all AVLT measures, with women consistently outperforming men. Similarly, Sundermann et al. (2017; 2016) found better immediate and delayed recall performance on the AVLT in female versus male controls; they further demonstrated that this female advantage held for AVLT delayed recall in persons with amnestic MCI despite comparable temporal lobe glucose metabolic rates and cortical amyloid β deposition. In a cognitively unimpaired cohort, Caselli et al. (2015) also found that women had significantly better long-term memory on the AVLT than men regardless of APOE-ε4 status. Although the female advantage in verbal episodic memory may vary depending on task, it is not unique to the AVLT and has been observed consistently on other verbal memory measures (Asperholm et al., 2019).

This female advantage in verbal memory despite comparable levels of neuropathologic burden might imply preserved functioning, but evidence from our study suggests that base rates of cognitive impairment may be higher than previously thought. Consistent with our second hypothesis, adjusting for sex increased the base rate of low scores to expected levels relative to adjusting for age only. This indicates that sex-adjusted norms provide a better balance of sensitivity and specificity than the original MOANS norms and other normative sources that do not adjust for sex. These data support the increasing evidence that applying demographic adjustments to neuropsychological test norms is important for enhancing diagnostic accuracy.

In support of this, Sundermann et al. (2019) recently demonstrated higher sensitivity in women and higher specificity in men by using sex-specific norms (compared with non–sex-specific norms) when applying the algorithmic Jak/Bondi psychometric criteria for amnestic MCI diagnosis to data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), as opposed to the typical clinical diagnosis applied in ADNI (Bondi et al., 2014; Jak et al., 2009). Specifically, they found that when applying sex adjustments to AVLT performance, 10% of women previously identified as normal were classified as amnestic MCI; conversely, 10% of men with a previous diagnosis of amnestic MCI were classified as normal, with results supported by biomarker status. Both underestimating and overestimating cognitive impairment have treatment implications, one resulting in missed treatment opportunities and the other resulting in stress and unnecessary treatment. Both types of misdiagnosis may negatively affect quality of life. We anticipate that our updated norms will help decrease misdiagnoses and that inclusion of alternative T-scores will facilitate future investigations into the effects of sex on diagnostic accuracy across various neurologic disorders.

Although associations between memory performance and education are less robust than age and sex, our results show that education explains a significant amount of variance for most AVLT variables. This finding is consistent with evidence that those with higher education perform better on memory measures (Uchiyama et al., 1995; Van der Elst et al., 2005). Application of our fully adjusted T-score formulas for persons with 8 versus 20 years of education results in a 10-point T-score difference for men (11 for women) on the score for trials 1 to 5 total. Thus, the 6% variance explained by education has important clinical influence for persons with high and low levels of education. Some studies have also demonstrated that adjusting for estimated or observed intelligence quotient (IQ) accounts for variance otherwise explained by education (Steinberg, Bieliauskas, Smith, Ivnik, & Malec,

2005), but these estimates are not always readily available or use test versions that may differ from those originally used in studies that apply an estimated IQ adjustment. Thus, adjusting for education affords greater flexibility than adjusting for IQ. Overall, adjusting for education remains a reasonable method to help determine the extent to which a score represents a decline from premorbid estimates or is atypical for an individual examinee.

There are several strengths of the current normative data. Whereas prior MOANS studies only included older adults, the current sample includes persons aged 30 years and older, thereby expanding the AVLT normative data to younger adults. In addition, the current study used a large population-based sample. Whereas similar large-scale research studies exist and could be used for development of normative data, the majority include only older adults and mirror a memory clinic in their recruitment designs. For instance, in the ADNI (Weiner et al., 2017), participant recruitment was designed to match that of clinical trials; thus, a high proportion of participants enroll because they have a memory concern or family history of dementia. Our population-based sample is representative of typical aging adults in the general community; participants are not seen in a clinical context. This has several important implications. For instance, the normative data presented here can be applied to a wide range of clinical populations besides those with memory concerns or risk of dementia, because inclusion criteria for the MCSA are broad. Test-retest reliability data are provided for a nearly 1.5 year average follow-up interval, which approximates what is often seen in clinical practice. Clinicians and researchers are encouraged to prioritize AVLT measures with at least adequate reliability coefficients (Slick, 2006), and should note that process scores and recognition have low to adequate reliability.

The AVLT is sensitive to verbal memory deficits across numerous neurologic conditions (Strauss et al., 2006), and the AVLT has demonstrated better sensitivity than the California Verbal Learning Test (CVLT) to left temporal abnormality (Loring et al., 2008). These updated norms will facilitate continued widespread use of the AVLT and provide an important alternative to calculation of z-scores based on cell means and SDs, as is currently available via the AVLT meta-norms (Schmidt, 1996). Also, unlike a typical research design using convenience sampling, our epidemiologic study design prospectively sampled persons to match the census of Olmsted County. Therefore, the current study is less susceptible to selection bias, which further improves the generalizability of these results.

The current study also has limitations. First, our sample is predominantly white and not of Hispanic origin. We recommend using available Mayo's Older African Americans Normative Studies norms for African American persons (Ferman et al., 2005). Second, the overall study sample remains heavily weighted toward older adults, with 76% of the current sample older than 60 years. However, because the sample is large (N=4,428), representation in the younger adult age ranges is sufficient for derivation of normative data. To put our sample size in perspective relative to other commonly used normative datasets, AVLT meta-norms (Strauss et al., 2006) provide a sample of 2,699 persons, only 359 of whom are older than 57 years. The CVLT, 2nd edition, norms were based on 1,087 adults aged 16 to 89 years, with 397 older than 60 years (Delis & Psychological Corporation., 2000). The recently released CVLT, 3rd edition, included 700 examinees, with even distribution across the following age groups: 16-19, 20-29, 30-44, 45-59, 60-69, 70-79, and 80-90 years (Delis

et al., 2017). In comparison, our older adult sample is 8.5 to 11 times larger, with 3,375 participants aged 60 years or older. Third, the way education data were collected coded individuals with a GED and individuals who graduated from high school as both having 12 years of education. There may be important differences in the educational experiences of these two groups that are not captured by these norms. Finally, some neuropsychologists have questioned whether age should be adjusted in normative data (Sliwinski, Hofer, Hall, Buschke, & Lipton, 2003), and growing evidence suggests that a large proportion of late-life cognitive decline is driven by clinically silent common neuropathologic processes (Harrington et al., 2018; Sliwinski et al., 2003; Yu et al., 2015). Some of these processes can now be measured in vivo (Jack et al., 2016), whereas a majority can only be measured at autopsy or remain unknown (Boyle et al., 2017; Boyle et al., 2018). Although we recommend fully corrected T-scores for clinical use, neuropsychologists should remain aware of this issue and understand that a normal score for a person in a particular age group does not guarantee the absence of preclinical pathology-related cognitive change. For researchers interested in conceptualizing cognitive decline as a reflection of cumulative damage to the brain with increasing age, the use of age-corrected scores may confound the desired effect of interest (Jack et al., 2015).

In summary, the current study provides updated normative data for the AVLT for an expanded age range, excludes persons with MCI in addition to dementia, and adjusts for important demographic variables. Results highlight the importance of adjusting for sex; this adjustment will help to account for clinically meaningful variance in AVLT performance between men and women and provide a better balance of sensitivity and specificity than the original MOANS norms and other normative sources that do not adjust for sex. Results from the current study continue the MOANS tradition of providing normative data and will help keep the AVLT as an important alternative measure for clinicians and research centers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We 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, and R01 AG041851, RF1 AG55151), a grant from the Alzheimer's Association (AARG-17-531322), the Robert Wood Johnson Foundation, The Elsie and Marvin Dekelboum Family Foundation, Alzheimer's Association, and the Mayo Foundation for Education and Research.

N.H.S. and M.M.Mi. serve as consultants to Biogen and Lundbeck. D.S.K. 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. R.C.P. has served as a consultant for Hoffman-La Roche Inc., Merck Inc., Genentech Inc., Biogen Inc., Eisai, Inc., and GE Healthcare. The authors report no conflicts of interest.

Appendix

All materials in the Appendix used with permission of Mayo Foundation of Medical Education and Research. An excel file that automates T-scores calculations is available by request through the Mayo Clinic Study of Aging website at the following link: https://

www.mayo.edu/research/centers-programs/alzheimers-disease-research-center/research-activities/mayo-clinic-study-aging/for-researchers/data-sharing-resources

Table A1.

AVLT Measures.

AVLT measure	Formula / description
Primary	
Trials 1-5 total	trial 1 recall + trial 2 recall + trial 3 recall + trial 4 recall + trial 5 recall
30-min recall	30-minute recall total correct (delayed recall)
Sum of trials	trials 1-5 total + trial 6 + 30-minute recall
Recognition % correct	{[(recognition hits + (15 − recognition FP errors)] / 30} x 100
Secondary	
Trial 1	trial 1 recall total correct
Trial 2	trial 2 recall total correct
Trial 3	trial 3 recall total correct
Trial 4	trial 4 recall total correct
Trial 5	trial 5 recall total correct
Trials 1-3 total [a]	trial 1 recall + trial 2 recall + trial 3 recall
List B	list B recall total correct
Trial 6	trial 6 recall total correct (short delayed recall)
Short-term % retention	100 x (trial 6 / trial 5)
Long-term % retention	100 x (30 minute delay / trial 5)
Total intrusions	intrusions summed across trial 1 + trial 2 + trial 3 + trial 4 + trial 5 + list B + trial 6 + 30-min recall
Memory efficiency score	{[(30 min delay / 15) / (trials 1-5 total / 75)] + [(recognition hits / 15) − (recognition FP errors / 15)]}

Abbreviation: AVLT, Auditory Verbal Learning Test. %, percentage. FP = false positive.

[a]We recommend using the 5 trial version of the AVLT. If an abbreviated version of the AVLT with only 3 learning trials is administered, note that normative data cannot be applied to any other AVLT measures beyond Trial 3.

Table A2.

Table for converting raw scores to unadjusted scaled scores for primary variables. [a]

SS	Trials 1-5 Total	30-Min Recall	Sum of Trials	Recognition PC
0	0-14	-	0-17	0-23
1	15-17	-	18-21	27-47
2	18-20	-	22-24	50-57
3	21-23	0	25-27	60-63
4	24-25	1	28-31	67
5	26-27	2	32-35	70-73
6	28-30	3	36-39	77
7	31-33	4	40-44	80
8	34-36	5-6	45-49	83-87
9	37-40	7	50-55	90

SS	Trials 1-5 Total	30-Min Recall	Sum of Trials	Recognition PC
10	41-43	8	56-61	-
11	44-47	9	62-66	93
12	48-50	10-11	67-72	97
13	51-54	12	73-78	-
14	55-57	13	79-83	-
15	58-61	14	84-88	100
16	62-63	-	89-92	-
17	64-65	15	93-95	-
18	66-67	-	96-96	-
19	68	-	97-97	-
20	69-75	-	98-105	-

[a]Scaled scores are provided only as a step in determining the demographically-corrected T-scores using the equations below. These scaled scores are not adjusted for any demographic variables and should not be used for clinical practice. Use of the fully-adjusted T-scores is recommended.

Note. 30-Min Recall = 30-Minute Delayed Recall; PC = Percentage Correct; Sum of Trials = Trials 1-5 total + trial 6 + 30-min recall.

Table A3.

Table for converting raw scores to unadjusted scaled scores for secondary variables. [a]

SS	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trials 1-3 Total	List B	Trial 6	STPR	LTPR	Memory Efficiency Score
0	0	0-1	0-1	0-2	0-2	0-7	-	-	-	-	−1.0 - −0.10
1	-	2	2	3	3	8	-	-	-	-	−0.09 - 0.03
2	1	-	3	4	4	9	0	0	0	-	0.04 - 0.27
3	2	3	4	-	5	10-11	-	-	7-13	0	0.28 - 0.46
4	-	4	-	5	-	12	1	1-2	14-33	7-18	0.47 - 0.71
5	-	-	5	6	6	13	-	3	36-44	20-33	0.72 - 0.94
6	3	5	6	-	7	14-15	2	4	45-55	36-45	0.95 - 1.19
7	-	-	-	7	8	16-17	3	5	56-62	46-56	1.20 - 1.38
8	4	6	7	8	9	18	-	6	64-70	57-64	1.39 - 1.56
9	-	7	8	9	10	19-20	4	7	71-77	67-71	1.57 - 1.71
10	5	-	9	10	11	21-22	-	8	78-82	73-79	1.72 - 1.85
11	-	8	10	11	12	23-24	5	9-10	83-89	80-86	1.86 - 1.97
12	6	9	11	12	13	25-26	-	11	90-92	87-92	1.98 - 2.07
13	-	10	12	13	-	27-28	6	12	93	93	2.08 - 2.15
14	7	-	-	-	14	29-30	7	13	100	-	2.16 - 2.22
15	8	11	13	14	-	31-32	-	14	-	100	2.23 - 2.29
16	9	12	14	-	15	33-34	8	15	-	-	2.30 - 2.35
17	-	13	-	15	-	35-36	9	-	-	-	2.36 - 2.43
18	10	14	15	-	-	37	10	-	-	-	2.44 - 2.50

SS	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trials 1-3 Total	List B	Trial 6	STPR	LTPR	Memory Efficiency Score
19	11	-	-	-	-	38	11	-	-	-	2.51 - 2.60
20	12-15	15	-	-	-	39-45	12-15	-	-	-	2.61 - 76.0

[a]Scaled scores are provided only as a step in determining the demographically-corrected T-scores using the equations below. These scaled scores are not adjusted for any demographic variables and should not be used for clinical practice. Use of the fully-adjusted T-scores is recommended.

Note. STPR = Short-Term Percentage Retention; LTPR = Long-Term Percentage Retention

T Score Formulas

Age, sex, and education-adjusted T scores for a subject's raw score(s) can be calculated with the formulas below.

SS = scaled score: determined from look-up tables above.

Sex: 0 = Female, 1 = male

Education level determination rules were as follows:

- If < 12 years, each full year of school completed is counted
- Vocational/Trade school years completed are counted
- GED=12 years
- High School Diploma=12 years
- College without degree: years completed are counted (13-15 years)
- 4 or more years of college, with no degree=15 years
- Bachelor's Degree=16 years
- Bachelor's Degree plus some graduate school=17 years
- Master's Degree=18 years
- Master's Degree plus some doctoral level courses=19 years
- Attorneys and Priests=19 years
- Doctoral degree=20 years

Note. Caution is suggested when interpreting performance in individuals with 8-11 years of education. Application of the fully demographically corrected normative formulas for individuals with age or education levels outside of the observed ranges is not recommended.

Equations for fully-adjusted T-Scores:

Primary variables—Trials 1-5 Total T-Score =round(50+((((AVLTSum5SS −(10.2048820335+(Age* 0.0696731708)+(Male*−2.0691847063)+(EDUC* 0.2076286782)+(Age**2 *−0.0014410120)))/1)+0.0000000637336)/0.23569807));

30-Min Recall T-Score =round(50+((((AVDSS −(12.4118437425+(Age*−0.0016432817)+(Male*−1.8612455591)+(EDUC* 0.1380628944)+(Age**2 *−0.0007027918)))/1)+0.0000001024411)/0.25299505));

Sum of Trials T-Score =round(50+((((AVLTsum5p6DSS−(10.8349191766+(Age* 0.0514562686)+(Male*−2.0670904968)+(EDUC* 0.1915793153)+(Age**2 * −0.0012694294)))/1)−0.0000001038205)/0.23673872));

Recognition Percentage Correct T-Score =round(50+((((AVRecPCSS −(10.7915054797+(Age* 0.0163995950)+(Male*−1.8832719513)+(EDUC* 0.1180746912)+(Age**2 * −0.0005488200)))/1)+0.0000001925238)/0.29155771));

Secondary variables—Trial 1 T-Score =round(50+((((AV1SS −(10.5554207904+(Age* 0.0361599800)+(Age**2 *−0.0009181852)+(Male*−1.2432854518)+(EDUC*0.1518778446)))/1)−0.0000001305326)/0.26867547));

Trial 2 T-Score =round(50+((((AV2SS −(10.2054872384+(Age* 0.0513655747)+(Age**2 *−0.0011932848)+(Male*−1.7651639080)+(EDUC*0.1919280336)))/1)−0.0000000448036)/0.25613384));

Trial 3 T-Score =round(50+((((AV3SS −(10.6083066798+(Age* 0.0436932539)+(Age**2 *−0.0011424483)+(Male*−1.7605746863)+(EDUC*0.1870822095)))/1)−0.0000001720072)/0.24470110));

Trial 4 T-Score =round(50+((((AV4SS −(10.3271703981+(Age* 0.0637213583)+(Age**2 *−0.0013266346)+(Male*−1.9598250972)+(EDUC*0.1859304791)))/1)−0.0000002083300)/0.24650622));

Trial 5 T-Score =round(50+((((AV5SS −(9.9952872306+(Age* 0.0622550674)+(Age**2 * −0.0012837374)+(Male*−1.9754640301)+(EDUC*0.1870905944)))/1)−0.0000000482041)/0.24516225));

Trials 1-3 Total T-Score =round(50+((((AVLTSum3SS −(10.4083623066+(Age* 0.0537576200)+(Age**2 *−0.0012667176)+(Male*−1.8439550539)+(EDUC*0.1987705165)))/1)−0.0000000560410)/0.23872421));

List B T-Score =round(50+((((AVBSS −(8.9167820377+(Age* 0.0780069203)+(Age**2 * −0.0013677187)+(Male*−1.1375184278)+(EDUC*0.1914262059)))/(2.3996448266+(Age**2 *−0.0000533322)))+0.00000967380900)/0.12287159));

Trial 6 T-Score =round(50+((((AV6SS −(11.7981182251+(Age* 0.0154689603)+(Age**2 *−0.0008517340)+(Male*−1.6396477808)+(EDUC*0.1436500033)))/1)+0.0000000285607)/0.25404381));

Short-Term Percentage Retention T-Score =round(50+((((AVSTPRSS −(12.3521586103+(Age*−0.0320618335)+(Age**2 *−0.0001238422)+(Male*−0.7356439106)+(EDUC*0.0493228458)))/(1.7628871897+(Age**2 * 0.0001168168)))+0.00000072036751)/0.12110756));

Long-Term Percentage Retention T-Score =round(50+((((AVLTPRSS −(13.3565938006+(Age*−0.0448177870)+(Age**2 *−0.0000815093)+(Male*−1.3287251835)+(EDUC*0.0567713694)))/(2.0060197485+(Age**2 * 0.0000707217)))+0.00000042085768)/0.12427314));

Memory Efficiency Score T-Score =round(50+((((AVMemEffScSS −(11.6072102613+(Age* 0.0018532956)+(Age**2 *−0.0004164258)+(Male*−1.6379708166)+(EDUC*0.0747602251)))/(1.8870824513+(Age**2 * 0.0000241449)))−0.00000100593800)/0.14018809));

Alternative T-score formulas for primary variables

Equations for age- and sex-adjusted T-Scores:

Trials 1-5 Total T-Score =round(50+((((AVLTSum5SS −(13.3975677899+(Age* 0.0739281419)+(Age**2 *−0.0015449437)+(Male*−1.9300702911)))/1)−0.000000215137900)/0.24136506));

30-Min Recall T-Score =round(50+((((AVDSS −(14.5348232863+(Age* 0.0011860656)+(Age**2 *−0.0007719013)+(Male*−1.7687412914)))/1)−0.000000142809100)/0.25534659));

Sum of Trials T-Score =round(50+((((AVLTsum5p6DSS−(13.7813144042+(Age* 0.0553751285)+(Age**2 *−0.0013653039)+(Male*−1.9391251867)))/1)+0.000000188774220)/0.24155121));

Recognition Percentage Correct T-Score =round(50+((((AVRecPCSS −(12.6060146804+(Age* 0.0188551657)+(Age**2 *−0.0006082048)+(Male*−1.8039811735)))/1)+0.000000150739090)/0.29305345));

Equations for age-adjusted T-Scores:

Trials 1-5 Total T-Score =round(50+((((AVLTSum5SS −(12.7156031925+(Age* 0.0621605724)+(Age**2 *−0.0014370025)))/1)−0.000000053633100)/0.25992588));

30-Min Recall T-Score =round(50+((((AVDSS −(13.9098621394+(Age*−0.0095978869)+(Age**2 *−0.0006729826)))/1)+0.000000048148911)/0.27021437));

Sum of Trials T-Score =round(50+((((AVLTsum5p6DSS−(13.0971814680+(Age* 0.0435109130)+(Age**2 *−0.0012564840)))/1)−0.000003235023000)/0.26026677));

Recognition Percentage Correct T-Score =round(50+((((AVRecPCSS −(11.9632533397+(Age* 0.0080540559)+(Age**2 *−0.0005089992)))/1)+0.000000096505809)/0.30660879));

Table A4.

Total number of intrusions across all trials (trials 1-5, list B, trial 6, and 30 minute recall). The second column includes the observed cumulative percent. The third column provides ranges for suggested bins for clinical decision-making.

Intrusions	Observed Cumulative Percent	Cumulative Percentile Range
0	100	≥ 14
1	68.4	
2	45.3	
3	30.2	
4	20.5	
5	14.0	
6	9.7	2-10
7	6.8	
8	4.6	
9	3.3	
10	2.4	
11	1.6	< 2
12	1.2	
13+	0.9	

Abbreviations

ADNI	Alzheimer’s Disease Neuroimaging Initiative
AVLT	Auditory Verbal Learning Test
CVLT	California Verbal Learning Test
IQ	intelligence quotient
MCI	mild cognitive impairment
MCSA	Mayo Clinic Study of Aging
MNS	Mayo Normative Studies
MOANS	Mayo’s Older Americans Normative Studies

References

Aartsen MJ, Martin M, Zimprich D, & Longitudinal Aging Study, A. (2004). Gender differences in level and change in cognitive functioning. Results from the Longitudinal Aging Study Amsterdam. Gerontology, 50(1), 35–38. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14654725. doi:10.1159/000074387 [PubMed: 14654725]

Asperholm M, Nagar S, Dekhtyar S, Herlitz A (2019) The magnitude of sex differences in verbal episodic memory increases with social progress: Data from 54 countries across 40 years. PLoS ONE 14(4): e0214945. 10.1371/journal.pone.0214945. [PubMed: 31009465]

Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, … Salmon DP (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. Journal of Alzheimer's Disease, 42(1), 275–289. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24844687. doi:10.3233/JAD-140276

Boyle PA, Yang J, Yu L, Leurgans SE, Capuano AW, Schneider JA, … Bennett DA (2017). Varied effects of age-related neuropathologies on the trajectory of late life cognitive decline. Brain, 140(3), 804–812. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28082297. doi:10.1093/brain/aww341 [PubMed: 28082297]

Boyle PA, Yu L, Wilson RS, Leurgans SE, Schneider JA, & Bennett DA (2018). Person-specific contribution of neuropathologies to cognitive loss in old age. Annals of Neurology, 83(1), 74–83. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29244218. doi:10.1002/ana.25123 [PubMed: 29244218]

Casaletto KB, Umlauf A, Beaumont J, Gershon R, Slotkin J, Akshoomoff N, & Heaton RK (2015). Demographically Corrected Normative Standards for the English Version of the NIH Toolbox Cognition Battery. Journal of the International Neuropsychological Society, 21(5), 378–391. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26030001. doi:10.1017/S1355617715000351 [PubMed: 26030001]

Caselli RJ, Dueck AC, Locke DE, Baxter LC, Woodruff BK, & Geda YE (2015). Sex-based memory advantages and cognitive aging: a challenge to the cognitive reserve construct? Journal of the International Neuropsychological Society, 21(2), 95–104. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25665170. doi:10.1017/S1355617715000016 [PubMed: 25665170]

Delis DC, Kramer J, Kaplan E, & Ober BA (2017). California Verbal Learning Test third edition (CVLT3) (Third edition ed.). San Antonio, TX: The Psychological Corporation.

Delis DC, Kramer J, Kaplan E, & Ober BA (2000). CVLT-II : California verbal learning test (2nd ed.). San Antonio, TX: Psychological Corp. Harcourt Brace Corp.

Di Carlo A, Lamassa M, Baldereschi M, Inzitari M, Scafato E, Farchi G, & Inzitari D (2007). CIND and MCI in the Italian elderly: frequency, vascular risk factors, progression to dementia. Neurology, 68(22), 1909–1916. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17536047. doi:10.1212/01.wnl.0000263132.99055.0d [PubMed: 17536047]

Ferman TJ, Lucas JA, Ivnik RJ, Smith GE, Willis FB, Petersen RC, & Graff-Radford NR (2005). Mayo's Older African American Normative Studies: Auditory Verbal Learning Test norms for African American elders. Clinical Neuropsychologist, 19(2), 214–228. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16019705. doi:10.1080/13854040590945300

Guilmette TJ, Sweet JJ, Hebben N, Koltai D,Mahone EM, Spiegler BJ, Stucky K, Westerveld M & Conference Participants (2020): American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores, The Clinical Neuropsychologist, doi:10.1080/13854046.2020.1722244

Harrington KD, Schembri A, Lim YY, Dang C, Ames D, Hassenstab J, … Group, A. R. (2018). Estimates of age-related memory decline are inflated by unrecognized Alzheimer's disease. Neurobiology of Aging, 70, 170–179. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/30015036. doi:10.1016/j.neurobiolaging.2018.06.005 [PubMed: 30015036]

Harris ME, Ivnik RJ, & Smith GE (2002). Mayo's Older Americans Normative Studies: expanded AVLT Recognition Trial norms for ages 57 to 98. Journal of Clinical and Experimental Neuropsychology, 24(2), 214–220. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11992204. doi:10.1076/jcen.24.2.214.995 [PubMed: 11992204]

Heaton RK (2004). Revised comprehensive norms for an expanded Halstead-Reitan battery : demographically adjusted neuropsychological norms for african american and caucasian adults : professional manual. Odessa, FL: Psychological Assessment Resources.

Hessel P, Kinge JM, Skirbekk V, & Staudinger UM (2018). Trends and determinants of the Flynn effect in cognitive functioning among older individuals in 10 European countries. Journal of Epidemiology and Community Health, 72(5), 383–389. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29440306. doi:10.1136/jech-2017-209979 [PubMed: 29440306]

Ivnik RJ (2005). Normative psychology: a professional obligation? Clinical Neuropsychologist, 19(2), 159–161. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16019701. doi:10.1080/13854040590945247

Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC, Kokmen E, & Kurland LT (1992a). Mayo's older americans normative studies: Updated AVLT norms for ages 56 to 97. Clinical Neuropsychologist, 6(sup001), 83–104. Retrieved from 10.1080/13854049208401880. doi:10.1080/13854049208401880

Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC, Kokmen E, & Kurland LT (1992b). Mayo's older americans normative studies: WAIS-R norms for ages 56 to 97. Clinical Neuropsychologist, 6(sup001), 1–30. Retrieved from 10.1080/13854049208401877. doi:10.1080/13854049208401877

Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC, Kokmen E, & Kurland LT (1992c). Mayo's older americans normative studies: WMS-R norms for ages 56 to 94. Clinical Neuropsychologist, 6(sup001), 49–82. Retrieved from 10.1080/13854049208401879. doi:10.1080/13854049208401879

Jack CR Jr., Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, … Dubois B (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology, 87(5), 539–547. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27371494. doi:10.1212/WNL.0000000000002923 [PubMed: 27371494]

Jack CR Jr., Wiste HJ, Weigand SD, Knopman DS, Vemuri P, Mielke MM, … Petersen RC (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. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25775353. doi:10.1001/jamaneurol.2014.4821 [PubMed: 25775353]

Jak AJ, Bondi MW, Delano-Wood L, Wierenga C, Corey-Bloom J, Salmon DP, & Delis DC (2009). Quantification of five neuropsychological approaches to defining mild cognitive impairment. American Journal of Geriatric Psychiatry, 17(5), 368–375. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19390294. doi:10.1097/JGP.0b013e31819431d5

Kokmen E, Smith GE, Petersen RC, Tangalos E, & Ivnik RC (1991). The short test of mental status. Correlations with standardized psychometric testing. Archives of Neurology, 48(7), 725–728. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/1859300. doi:10.1001/archneur.1991.00530190071018 [PubMed: 1859300]

Kramer JH, Delis DC, Kaplan E, O'Donnell L, & Prifitera A (1997). Developmental sex differences in verbal learning. Neuropsychology, 11(4), 577–584. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9345701. doi:10.1037//0894-4105.11.4.577 [PubMed: 9345701]

Loring DW, Strauss E, Hermann BP, Barr WB, Perrine K, Trenerry MR, … Bowden SC (2008). Differential neuropsychological test sensitivity to left temporal lobe epilepsy. Journal of the International Neuropsychological Society, 14(3), 394–400. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18419838. doi:10.1017/S1355617708080582 [PubMed: 18419838]

Malec JF, Ivnik RJ, Smith GE, Tangalos EG, Petersen RC, Kokmen E, & Kurland LT (1992). Mayo's older americans normative studies: Utility of corrections for age and education for the WAIS-R. Clinical Neuropsychologist, 6(sup001), 31–47. Retrieved from 10.1080/13854049208401878. doi:10.1080/13854049208401878

Morris JC (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology, 43(11), 2412–2414. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/8232972. doi:10.1212/wnl.43.11.2412-a

Petersen RC (2004). Mild cognitive impairment as a diagnostic entity. Journal of Internal Medicine, 256(3), 183–194. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15324362. doi:10.1111/j.1365-2796.2004.01388.x [PubMed: 15324362]

Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, … Rae-Grant A (2018). Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. Neurology, 90(3), 126–135. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29282327. doi:10.1212/WNL.0000000000004826 [PubMed: 29282327]

Petersen RC, Roberts RO, Knopman DS, Boeve BF, Geda YE, Ivnik RJ, … Jack CR Jr. (2009). Mild cognitive impairment: ten years later. Archives of Neurology, 66(12), 1447–1455. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20008648. doi:10.1001/archneurol.2009.266 [PubMed: 20008648]

Petersen RC, Roberts RO, Knopman DS, Geda YE, Cha RH, Pankratz VS, … Rocca WA (2010). Prevalence of mild cognitive impairment is higher in men. The Mayo Clinic Study of Aging. Neurology, 75(10), 889–897. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20820000. doi:10.1212/WNL.0b013e3181f11d85 [PubMed: 20820000]

Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, & Kokmen E (1999). Mild cognitive impairment: clinical characterization and outcome. Archives of Neurology, 56(3), 303–308. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10190820. doi:10.1001/archneur.56.3.303 [PubMed: 10190820]

Rey A (1964). L' examen clinique en psychologie (2. éd. ed.). Paris,: Presses universitaires de France.

Ricci M, Graef S, Blundo C, & Miller LA (2012). Using the Rey Auditory Verbal Learning Test (RAVLT) to differentiate alzheimer's dementia and behavioural variant fronto-temporal dementia. Clinical Neuropsychologist, 26(6), 926–941. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22809061. doi:10.1080/13854046.2012.704073

Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, … Rocca WA (2008). The Mayo Clinic Study of Aging: design and sampling, participation, baseline measures and sample characteristics. Neuroepidemiology, 30(1), 58–69. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18259084. doi:10.1159/000115751 [PubMed: 18259084]

Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, … Petersen RC (2012). The incidence of MCI differs by subtype and is higher in men: the Mayo Clinic Study of Aging. Neurology, 78(5), 342–351. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/22282647. doi:10.1212/WNL.0b013e3182452862 [PubMed: 22282647]

Salthouse TA (2010). Selective review of cognitive aging. Journal of the International Neuropsychological Society, 16(5), 754–760. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20673381. doi:10.1017/S1355617710000706 [PubMed: 20673381]

Schmidt M (1996). Rey Auditory Verbal Learning Test. Los Angeles, CA: Western Psychological Services.

Sliwinski MJ, Hofer SM, Hall C, Buschke H, & Lipton RB (2003). Modeling memory decline in older adults: the importance of preclinical dementia, attrition, and chronological age. Psychology and Aging, 18(4), 658–671. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14692855. doi:10.1037/0882-7974.18.4.658 [PubMed: 14692855]

Shrout PE, & Fleiss JL (1979), Intraclass correlation: uses in assessing rater reliability. Psychological Bulletin, 86, 420–428. [PubMed: 18839484]

Slick DJ (2006). Psychometrics in neuropsychological assessment. In Strauss E, Sherman EMS & Spreen O (Eds.), A compendium of neuropsychological tests: administration, norms, and commentary (3rd ed.) (pp. 3–31). New York, NY : Oxford University Press.

Smith GE, & Ivnik RJ (2003). Normative neuropsychology. In Petersen RC (Ed.), Mild cognitive impairment : aging to Alzheimer's disease (pp. 63–88). Oxford: Oxford University Press.

St Sauver JL, Grossardt BR, Yawn BP, Melton LJ 3rd, & Rocca WA (2011). Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project. American Journal of Epidemiology, 173(9), 1059–1068. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21430193. doi:10.1093/aje/kwq482 [PubMed: 21430193]

Steinberg BA, Bieliauskas LA, Smith GE, Ivnik RJ, & Malec JF (2005). Mayo's Older Americans Normative Studies: Age- and IQ-Adjusted Norms for the Auditory Verbal Learning Test and the Visual Spatial Learning Test. Clinical Neuropsychologist, 19(3-4), 464–523. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/16120537. doi:10.1080/13854040590945193

Strauss E, Sherman EMS, Spreen O, & Spreen O (2006). A compendium of neuropsychological tests : administration, norms, and commentary (3rd ed.). Oxford ; New York: Oxford University Press.

Sundermann EE, Biegon A, Rubin LH, Lipton RB, Landau S, Maki PM, & Alzheimer's Disease Neuroimaging, I. (2017). Does the Female Advantage in Verbal Memory Contribute to Underestimating Alzheimer's Disease Pathology in Women versus Men? Journal of Alzheimer's

Disease, 56(3), 947–957. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28106548. doi:10.3233/JAD-160716

Sundermann EE, Biegon A, Rubin LH, Lipton RB, Mowrey W, Landau S, … Alzheimer's Disease Neuroimaging, I. (2016). Better verbal memory in women than men in MCI despite similar levels of hippocampal atrophy. Neurology, 86(15), 1368–1376. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26984945. doi:10.1212/WNL.0000000000002570 [PubMed: 26984945]

Sundermann EE, Maki P, Biegon A, Lipton RB, Mielke MM, Machulda M, … Alzheimer's Disease Neuroimaging, I. (2019). Sex-specific norms for verbal memory tests may improve diagnostic accuracy of amnestic MCI. Neurology. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/31597708. doi:10.1212/WNL.0000000000008467

Thomas KR, Eppig J, Edmonds EC, Jacobs DM, Libon DJ, Au R, … Alzheimer's Disease Neuroimaging, I. (2018). Word-list intrusion errors predict progression to mild cognitive impairment. Neuropsychology, 32(2), 235–245. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29528684. doi:10.1037/neu0000413 [PubMed: 29528684]

Uchiyama CL, D'Elia LF, Dellinger AM, Becker JT, Selnes OA, Wesch JE, … Miller EN (1995). Alternate forms of the Auditory-Verbal Learning Test: issues of test comparability, longitudinal reliability, and moderating demographic variables. Archives of Clinical Neuropsychology, 10(2), 133–145. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/14589735. [PubMed: 14589735]

Van der Elst W, van Boxtel MP, van Breukelen GJ, & Jolles J (2005). Rey's verbal learning test: normative data for 1855 healthy participants aged 24-81 years and the influence of age, sex, education, and mode of presentation. Journal of the International Neuropsychological Society, 11(3), 290–302. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15892905. doi:10.1017/S1355617705050344 [PubMed: 15892905]

Verhaeghen P, & Salthouse TA (1997). Meta-analyses of age-cognition relations in adulthood: estimates of linear and nonlinear age effects and structural models. Psychological Bulletin, 122(3), 231–249. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9354147. doi:10.1037/0033-2909.122.3.231 [PubMed: 9354147]

Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, … Alzheimer's Disease Neuroimaging, I. (2017). The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. Alzheimers Dement, 13(5), 561–571. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27931796. doi:10.1016/j.jalz.2016.10.006 [PubMed: 27931796]

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. Neuropsychology Review, 27(4), 354–388. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28940127. doi:10.1007/s11065-017-9360-6 [PubMed: 28940127]

Wennberg AMV, Hagen CE, Machulda MM, Hollman JH, Roberts RO, Knopman DS, … Mielke MM (2018). The association between peripheral total IGF-1, IGFBP-3, and IGF-1/IGFBP-3 and functional and cognitive outcomes in the Mayo Clinic Study of Aging. Neurobiology of Aging, 66, 68–74. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29547749. doi:10.1016/j.neurobiolaging.2017.11.017 [PubMed: 29547749]

Yu L, Boyle PA, Segawa E, Leurgans S, Schneider JA, Wilson RS, & Bennett DA (2015). Residual decline in cognition after adjustment for common neuropathologic conditions. Neuropsychology, 29(3), 335–343. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25495832. doi:10.1037/neu0000159 [PubMed: 25495832]

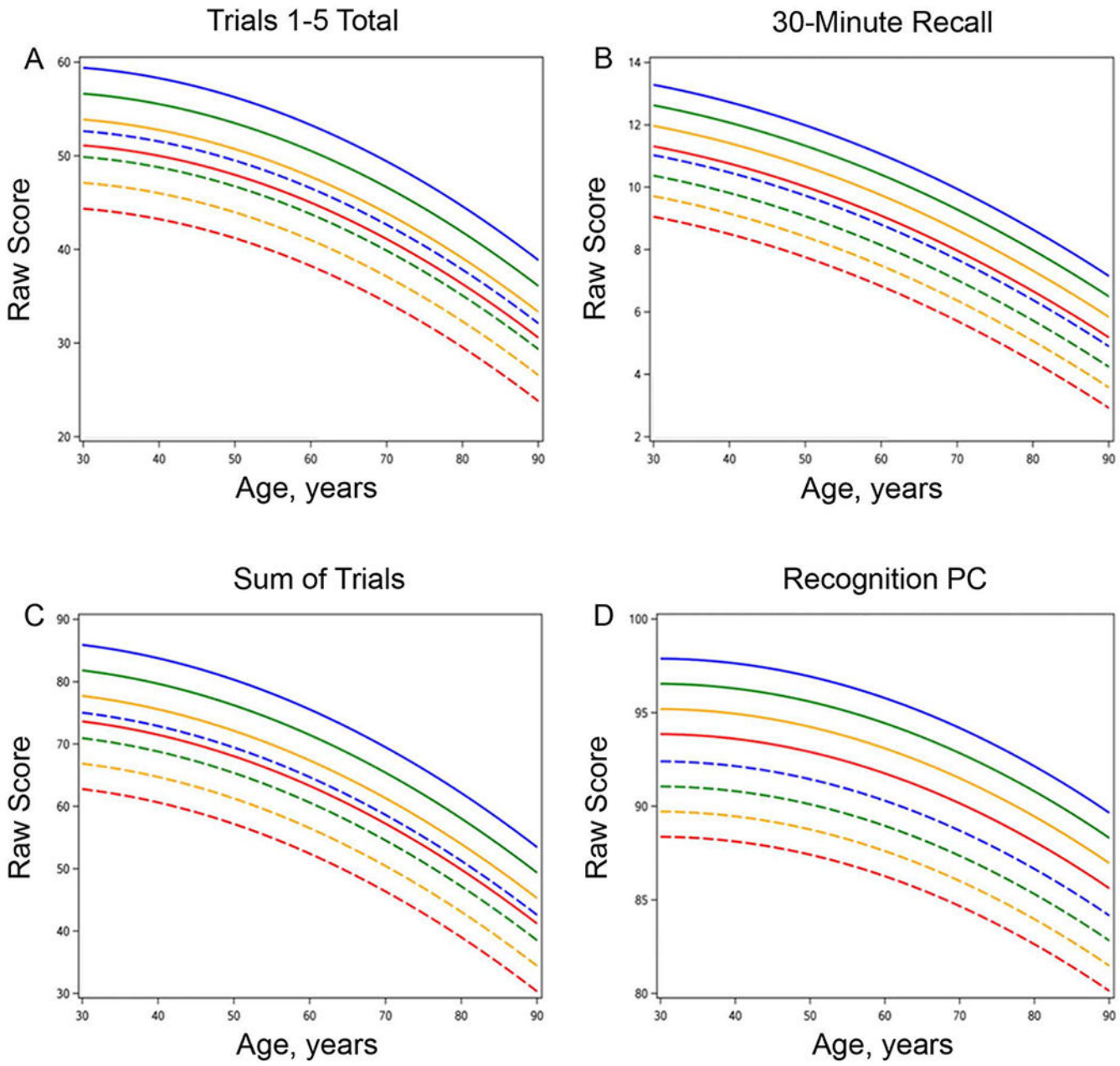

Figure 1. Regression Models.

Raw Auditory Verbal Learning Test scores were modeled on age/sex/education/age squared. Each graph shows the effect of age, age squared, sex (women, solid lines; men, dashed lines), and years of education (blue, 20 years; green, 16 years; orange, 12 years; red, 8 years on raw scores for trials 1-5 total (A), 30-minute recall (B), sum of trials (C), and recognition percentage correct (Recognition PC) (D).

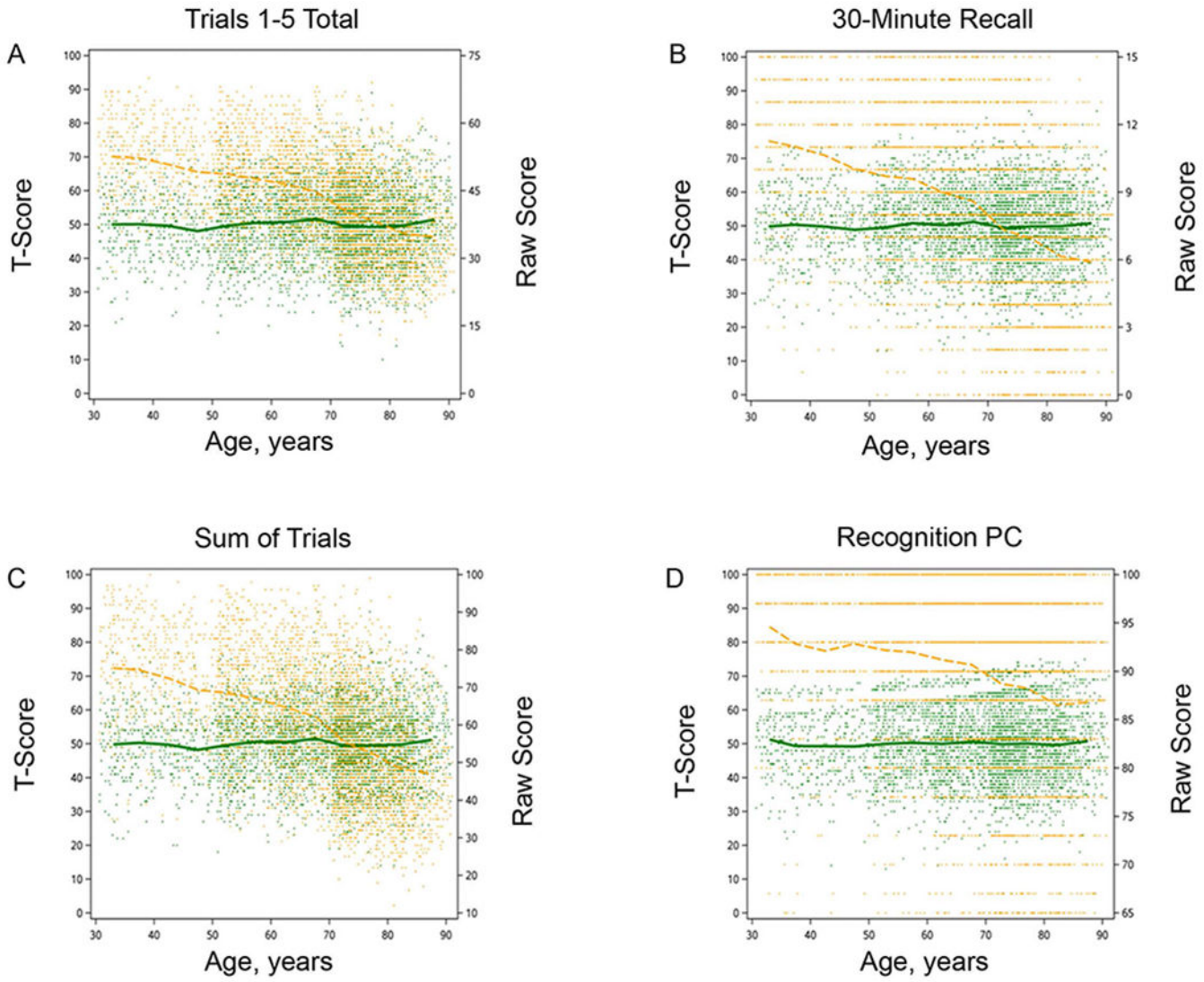

Figure 2. Auditory Verbal Learning Test (AVLT) T-Scores Correct for Age Effects.

Raw AVLT scores (orange, right side of Y-axis) and demographically corrected (for age/sex/education/age^2) AVLT T-scores (green, left side of Y-axis) demonstrating that T-scores correct for age effects. Corrected and raw AVLT T-scores are plotted vs age (lines represent mean scores within age quinquennia). Scores are for trials 1-5 total (A), 30-minute recall (B), sum of trials (C), and recognition percentage correct (Recognition PC) (D). For recognition percentage correct, the raw scores had a skewed distribution, with the majority of participants scoring between 65% and 100%. However, a small number of participants (n=86) scored in the range of 10% to 63%. For improved visualization, those values were plotted at 65%. This does not affect the computation of the mean (orange) line.

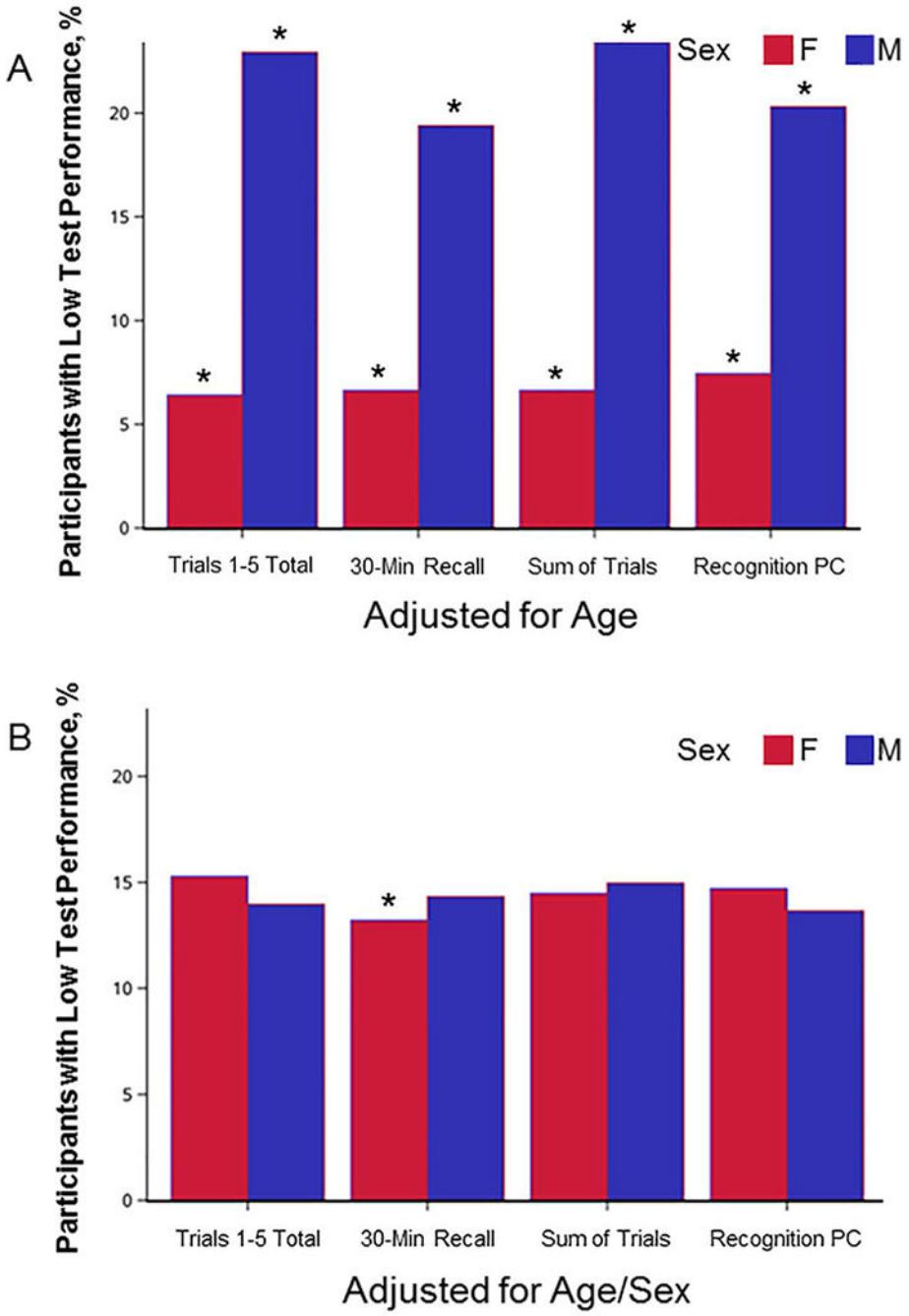

Figure 3. Percentage of Participants Showing Low Test Performance (T-score <40).

Percentage of participants showing low test performance for age-adjusted (A) and age- and sex-adjusted (B) T-scores for the primary variables shown. The expected base rate is 14.7%. * Confidence Interval does not contain the 14.7% expected base rate value. F indicates female; M, male; PC, percentage correct.

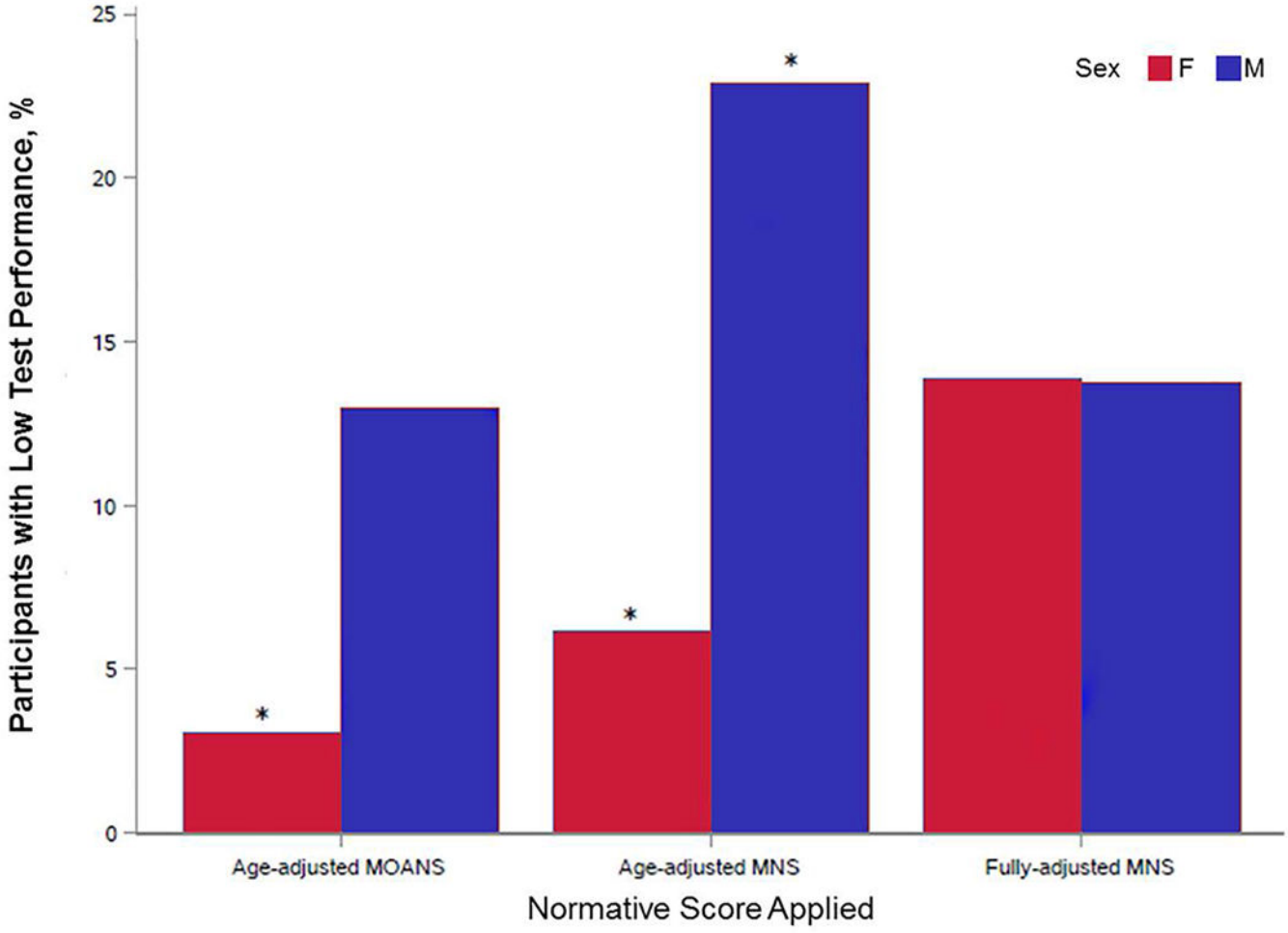

Figure 4. Percentage of Participants in Validation Sample Showing Low Test Performance on 30-minute Recall (< −1 SD).

Percentage of participants showing low test performance on 30-minute delayed recall when applying age-adjusted MOANS scaled scores (SS < 7), age-adjusted MNS T-scores (T-score < 40), and fully-adjusted MNS T-scores (T-score < 40). The expected base rate is 14.7%. * Confidence Interval does not contain the 14.7% expected base rate value. F indicates female; M, male. MOANS = Mayo's Older Americans Normative Studies. MNS = Mayo Normative Studies.

Table 1.

Demographic Characteristics

Characteristic	No. of Participants (%) (N=4,428)
Age, years	
30-39	214 (4.8)
40-49	210 (4.7)
50-59	610 (13.8)
60-69	916 (20.7)
70-79	1,655 (37.4)
80-91	823 (18.6)
Sex (Male)	2,211 (49.9)
Education, years	
8-12	1,257 (28.4)
13-15	1,263 (28.5)
16	922 (20.8)
17-20	986 (22.3)
Race	
White	4,333 (97.9)
Black/African American	22 (0.5)
Asian	29 (0.7)
American Indian/Alaska native	4 (0.1)
More than 1	25 (0.5)
Unknown	15 (0.3)
Ethnicity other than Hispanic	4,391 (99.2)

Table used with permission of Mayo Foundation for Medical Education and Research.

Table 2.

Percentage Variance Explained (Pearson Correlation Coefficient, Squared) for Each Demographic Variable and the Full Regression Model (Combined)[a]

AVLT Measure (Raw)	Age	Age Squared	Sex	Education	Combined
Primary					
Trials 1-5 total	26.56	27.41	9.08	5.93	40.74
30-Min recall (long delay)	17.28	17.56	8.50	2.75	28.31
Sum of trials[b]	25.50	26.19	9.39	5.21	39.37
Recognition percentage correct[c]	4.73	4.91	7.17	0.97	13.37
Secondary					
Trial 1	14.10	14.36	3.27	3.38	19.70
Trial 2	20.58	21.09	6.49	4.74	30.77
Trial 3	22.41	23.04	6.93	5.06	33.54
Trial 4	22.25	23.09	8.45	4.72	35.07
Trial 5	20.79	21.67	8.80	4.54	34.04
Trials 1-3 total	24.83	25.45	7.29	5.72	36.03
List B	18.87	19.58	2.31	5.30	24.51
Trial 6 (short delay)	18.23	18.61	6.39	3.33	27.36
Short-term percentage retention[d]	4.89	4.95	1.33	0.48	6.65
Long-term percentage retention[e]	5.76	5.82	4.30	0.45	10.29
Total intrusions[f]	0.53	0.55	0.00	0.29	0.71
Memory efficiency score[g]	5.56	5.70	6.74	0.60	13.34

Abbreviation: AVLT, Auditory Verbal Learning Test.

[a] Pearson Correlation Coefficients, Squared is equivalent to R^2. All *P* values for Pearson correlation coefficients (before squaring) are $P<.001$, except for intrusions which are not significantly related to sex (Pearson or Spearman; $P>.66$).

[b] Trials 1-5 total + trial 6 + 30-min recall.

[c] ([Recognition hits + {15−recognition false positive errors}]/30) × 100.

[d] 100 × (trial 6/trial 5).

[e] 100 × (30-min delay/trial 5).

[f] Intrusions summed across trial 1 + trial 2 + trial 3 + trial 4 + trial 5 + list B + trial 6 + 30-min recall.

[g] {[(30-min delay/15)/(trials 1-5 total/75)] + [(recognition hits/15) − (recognition false positive errors/15)]}.

Table 3.

Reliability coefficients: Pearson's rho and ICC for test-retest sample (N = 3,555)

AVLT Measure (Raw)	r	ICC	ICC 95% CI
Primary			
Trials 1-5 total	0.7982	0.7901	(0.7774, 0.8021)
30-Min recall (long delay)	0.7611	0.7519	(0.7373, 0.7659)
Sum of trials	0.8312	0.8216	(0.8107, 0.8320)
Recognition percentage correct	0.5652	0.5622	(0.5395, 0.5845)
Secondary			
Trial 1	0.4968	0.4853	(0.4603, 0.5105)
Trial 2	0.6153	0.6066	(0.5857, 0.6272)
Trial 3	0.6858	0.6807	(0.6628, 0.6980)
Trial 4	0.6950	0.6920	(0.6746, 0.7089)
Trial 5	0.7191	0.7158	(0.6995, 0.7316)
Trials 1-3 total	0.7320	0.7208	(0.7047, 0.7363)
List B	0.5066	0.5066	(0.4821, 0.5310)
Trial 6 (short delay)	0.7367	0.7293	(0.7136, 0.7444)
Short-term percentage retention	0.4197	0.4161	(0.3892, 0.4435)
Long-term percentage retention	0.5062	0.5025	(0.4779, 0.5270)
Total intrusions	0.3838	0.3860	(0.3577, 0.4150)
Memory efficiency score	0.6361	0.6320	(0.6120, 0.6515)

Abbreviation: AVLT, Auditory Verbal Learning Test.

J Int Neuropsychol Soc. Author manuscript; available in PMC 2021 March 18.