



## Research Article

# Mayo normative studies: regression-based normative data for ages 30–91 years with a focus on the Boston Naming Test, Trail Making Test and Category Fluency

Aimee J. Karstens<sup>1</sup> , Teresa J. Christianson<sup>2</sup>, Emily S. Lundt<sup>2</sup>, Mary M. Machulda<sup>1</sup> , Michelle M. Mielke<sup>3,4</sup>, Julie A. Fields<sup>1</sup> , Walter K. Kremers<sup>2</sup>, Jonathan Graff-Radford<sup>5</sup>, Prashanthi Vemuri<sup>6</sup>, Clifford R. Jack Jr.<sup>6</sup>, David S. Knopman<sup>5</sup>, Ronald C. Petersen<sup>5</sup> and Nikki H. Stricker<sup>1</sup> 

<sup>1</sup>Division of Neurocognitive Disorders, Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, USA, <sup>2</sup>Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA, <sup>3</sup>Department of Epidemiology and Prevention, Wake Forest University School of Medicine, Winston-Salem, NC, USA, <sup>4</sup>Department of Gerontology and Geriatric Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA, <sup>5</sup>Department of Neurology, Mayo Clinic, Rochester, MN, USA and <sup>6</sup>Department of Radiology, Mayo Clinic, Rochester, MN, USA

### Abstract

**Objective:** Normative neuropsychological data are essential for interpretation of test performance in the context of demographic factors. The Mayo Normative Studies (MNS) aim to provide updated normative data for neuropsychological measures administered in the Mayo Clinic Study of Aging (MCSA), a population-based study of aging that randomly samples residents of Olmsted County, Minnesota, from age- and sex-stratified groups. We examined demographic effects on neuropsychological measures and validated the regression-based norms in comparison to existing normative data developed in a similar sample. **Method:** The MNS includes cognitively unimpaired adults  $\geq 30$  years of age ( $n = 4,428$ ) participating in the MCSA. Multivariable linear regressions were used to determine demographic effects on test performance. Regression-based normative formulas were developed by first converting raw scores to normalized scaled scores and then regressing on age, age<sup>2</sup>, sex, and education. Total and sex-stratified base rates of low scores ( $T < 40$ ) were examined in an older adult validation sample and compared with Mayo's Older Americans Normative Studies (MOANS) norms. **Results:** Independent linear regressions revealed variable patterns of linear and/or quadratic effects of age ( $r^2 = 6\text{--}27\%$  variance explained), sex (0–13%), and education (2–10%) across measures. MNS norms improved base rates of low performance in the older adult validation sample overall and in sex-specific patterns relative to MOANS. **Conclusions:** Our results demonstrate the need for updated norms that consider complex demographic associations on test performance and that specifically exclude participants with mild cognitive impairment from the normative sample.

**Keywords:** Cognitive aging; mild cognitive impairment; neuropsychology; neuropsychological tests; psychometrics; base rates; executive function; animal fluency

(Received 17 May 2023; final revision 6 October 2023; accepted 10 October 2023; First Published online 28 November 2023)

### Introduction

Normative data are fundamental to the clinical interpretation of neuropsychological test performance. Often, normative data are developed within a target population and demographic adjustments are derived statistically to define stratified distributions. Co-normed datasets allow for cross-domain test comparisons that improve interpretation. However, many of the widely used adult lifespan multitest datasets for English speakers were published 15–23 years ago, with some data collection occurring over 50 years ago (Casaletto & Heaton, 2017; Collins & Riley, 2016). For

example, Heaton normative data for the Halsted Reitan battery and other measures were collected over the course of 25 years before being published, including data collection from earlier norms published in 1991 (Heaton et al., 1991; Heaton et al., 2004). These datasets remain gold standard clinical tools despite several limitations that may reduce sensitivity of normative data, including the influence of population-level changes in cognitive performance (e.g., Flynn effect on IQ) and improved methodological approaches (Bildler & Reise, 2019; Heaton et al., 2004; Hiscock, 2007). In addition, intergenerational sociopolitical, linguistic, and cultural

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

A paper presentation of a portion of this work was presented at the International Neuropsychological Society conference (February 2021).

**Cite this article:** Karstens A.J., Christianson T.J., Lundt E.S., Machulda M.M., Mielke M.M., Fields J.A., Kremers W.K., Graff-Radford J., Vemuri P., Jack C.R., Jr., Knopman D.S., Petersen R.C., & Stricker N.H. (2024) Mayo normative studies: regression-based normative data for ages 30–91 years with a focus on the Boston Naming Test, Trail Making Test and Category Fluency. *Journal of the International Neuropsychological Society*, 30: 389–401, <https://doi.org/10.1017/S1355617723000760>

© Mayo Foundation for Medical Education and Research, 2023. Published by Cambridge University Press on behalf of International Neuropsychological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.

differences influence the salience of test and item construction (Beatty et al., 2017). Factors that also limit the applicability of normative data include recruitment of convenience samples that are not representative of local demographics, ill-defined exclusion criteria, and lack of or limited demographic corrections (Mitrushina et al., 2005; Tombaugh et al., 1999; Tombaugh, 2004).

Various methods have been employed to control for the effects of demographic factors including the use of percentiles, overlapping cells, and various regression-based corrections. Many norms do not control for sex and/or education (Benedict & Brandt, 2001; Benedict, 1997; Lucas et al., 1998; Mitrushina et al., 2005; Wechsler, 1997, 2009). When norms do control for age, sex, and education, demographic bins with small sample sizes may misrepresent select groups or be underpowered. In addition, access to norms with additional demographic corrections may require specialized software (Delis et al., 2017). Importantly, the effects of age, sex, and education or other relevant premorbid proxies (e.g., IQ, reading) vary by test/construct within populations and the degree of variance varies between populations (Avila et al., 2020; Avila et al., 2019; Werry et al., 2019). Relatively small attributable variance can result in high false positive/negative rates at the population level, for example, when age-corrected norms of verbal memory that do not additionally adjust for sex are used to detect mild cognitive impairment (Edmonds et al., 2016; Stricker et al., 2021; Sundermann et al., 2021). This may be further exacerbated in older adults as numerous normative datasets do not explicitly exclude individuals with mild cognitive impairment (MCI) (Heaton et al., 2004; Ivnik et al., 1996; Lucas et al., 1998; Mitrushina et al., 2005; Tombaugh et al., 1999; Tombaugh, 2004).

While co-normed datasets are useful for interpretation, outdated norms without appropriate demographic adjustments may inflate Type I or Type II error. Early Alzheimer's disease (AD) related cognitive changes, for example, could in part explain why some studies have shown greater prevalence of MCI in males even though more women develop AD dementia (Au et al., 2017; Nebel et al., 2018; Petersen et al., 2010) and AD pathology is equally prevalent in men and women (Jack et al., 2019). Numerous older adult datasets have been developed in tandem with NIH-funded aging studies or for research purposes (Clark et al., 2016; Fine et al., 2012; Holtzer et al., 2008; Miller et al., 2015; Pedraza et al., 2010; Steinberg et al., 2005; Wang et al., 2021; Zec et al., 2007), and have improved upon prior methods for recruitment, exclusion criteria, and statistical approaches. However, the time and resources necessary to develop norms has often precluded this work in lifespan samples. For example, extensive resources needed for normative data collection likely limits the expansion of these data to younger age groups or to include more representative (vs convenience) samples. As a result, limitations of existing lifespan normative datasets often go unaddressed. Characterizing the effect of biological, social, and combined factors on test performances using sufficiently powered samples is necessary to improve the utility of neuropsychological tests.

A primary goal of the Mayo Normative Studies (MNS) is to address limitations in currently available normative data with an updated population-based cohort and advanced methods. For example, we previously published MNS Rey Auditory Verbal Learning Test (AVLT) data that provides several enhancements relative to the Mayo's Older Americans Normative Studies (MOANS; Ivnik et al., 1996; Lucas et al., 1998) through an expanded age range, exclusion of persons with MCI, updated normative methods using a regression-based approach adjusting for age, sex, and education, and a publically available, user-friendly

calculator (Stricker et al., 2021). We found that the prevalence of low test scores (e.g., base rates of scores  $<-1$  SD below the mean) was lower-than-expected when MOANS AVLT norms were applied to a cognitively unimpaired validation sample, but that application of fully-corrected MNS AVLT norms yielded base rates of low test performance that were within expectation (Brooks & Iverson, 2010; Ivnik et al., 1992b; Stricker et al., 2021). To expand upon this work, the current study developed norms for additional measures administered in the Mayo Clinic Study of Aging, with an interpretative focus on measures of processing speed/executive function and language. Specifically, in a population-based sample excluding individuals with MCI, we examined effects of demographic variables on test performances, developed regression-based norms correcting for key demographic variables, and validated the norms by comparing base rates of low test scores in older adults using the MNS and MOANS.

## Methods

The MNS leverages data from the Mayo Clinic Study of Aging (MCSA), a longitudinal population-based study of cognitive aging initiated in 2004. MCSA participants were recruited using a random sampling method in the Rochester Epidemiology Project Medical Records linkage system (St. Sauver et al., 2011) in Olmsted County, Minnesota. 97% of Olmsted County residents agreed to the use of their medical records for research. Over 60% of contacted residents enrolled in the MCSA using an age- and sex-stratified random sampling design to ensure equal representation of women and men between 70 and 89 years in each 10-year age strata (Roberts et al., 2008). Extended enrollment periods included younger ages (50- to 69-year-olds added in 2012; 30- to 49-year-olds added in 2015). MCSA participants are followed longitudinally at 15-month intervals. Full MCSA sampling and detailed study procedures have been published previously (Roberts et al., 2008).

This study was completed in accordance with the Helsinki Declaration. The study protocols were approved by the Mayo Clinic and Olmsted Medical Center Institutional Review Boards. All participants provided written informed consent.

Participants were included in the current retrospective study if they were 30 years or older, cognitively unimpaired, naïve to the neuropsychological testing battery (i.e., only baseline visit used and excluded if previously had testing due to other research participation) and were not terminally ill or receiving hospice care. Due to study sampling procedures that limits recruitment of MCSA participants to individuals living in Olmsted County, this results in a predominantly White sample from the midwest region of the United States. Participant study visits include a medical record review and neurological evaluation, including administration of the Short Test of Mental Status that is similar to the Mini Mental Status Exam (Kokmen et al., 1991). A specific cutoff on the Short Test of Mental Status was not applied but performance on this measure informed the Neurologist's diagnosis. Neuropsychological testing was conducted by a trained psychometrist and included nine tests covering four domains (Kokmen et al., 1991; Roberts et al., 2008; Wennberg et al., 2018). Participants and their informants underwent a structured interview with a study coordinator to collect additional demographic information, medical history, subjective memory, and daily functioning assessments using the Clinical Dementia Rating (CDR<sup>®</sup>) instrument (Morris, 1993). A CDR cutoff was not applied but informed the study coordinator diagnosis. As previously described (Stricker et al., 2021), participants were determined to be

cognitively unimpaired by both the physician and study coordinator, who were blind to neuropsychological test results as opposed to the typical MCSA approach of a consensus diagnosis (Petersen et al., 2010; Petersen, 2004; Roberts et al., 2008). This minimized bias or circularity of using the neuropsychologist's impression based on neuropsychological data to define new norms.

### Neuropsychological battery

The MCSA neuropsychological testing battery included 9 measures of 4 cognitive domains (memory, language, attention/executive, visuospatial), with test administration procedures consistent with those in the original MOANS. The current manuscript provides regression-based norms for all tests given in the MCSA except for the AVLT, which was the focus of our prior work (Stricker et al., 2021). This manuscript focuses on measures of language and processing speed/executive functioning. We also include Logical Memory immediate (LMI) and delayed (LMII) recall and Visual Reproduction immediate (VRI) and delayed (VRII recall) from the Wechsler Memory Scale-Revised and Digit Symbol, Picture Completion and Block Design subtests from the Wechsler Adult Intelligence Scales-Revised (Wechsler, 1981). However, these tests are given less emphasis in this manuscript because they have undergone two additional revisions since these measures were introduced into the MCSA battery (WMS-III, WMS-IV, WAIS-III, WAIS-IV). The WAIS-R/WMS-R versions were used due to the longitudinal needs of this study, as the Mayo Clinic Study of Aging was an update to the Alzheimer's Disease Patient Registry study that began in 1986 and was a primary source of prior MOANS. Although WAIS-R/WMS-R are outdated, we chose to present results on these measures in order to contrast against the current gold standard of MOANS norms and inform the clinically relevant question of how normative sample composition can influence performance of norms when applied to an independent sample. Given that updated WAIS-R/WMS-R measures are similar to these earlier versions, lessons learned from the data remain of interest even though we do not recommend use of the WAIS-R/WMS-R versions of these measures clinically. Language measures include confrontation naming (Boston Naming Test (BNT); Kaplan et al., 1983) and semantic fluency (Category Fluency) (Strauss et al., 2006); reported as total fluency and individual categories (Animals, Fruits, Vegetables). Note that administration of the BNT noose item was omitted starting in 2017 due to its violent racist origins and subsequently a point has been credited automatically for the item (Eloi et al., 2021). Attention/executive measures include visuomotor scanning (Trail Making Test A (TMTA); Reitan, 1958) and cognitive flexibility (Trail Making Test B (TMTB); Reitan, 1958); scores of these tests were inverted prior to norming (180-TMTA raw; 300-TMTB raw). Updated MNS regression-based normative data are currently available for the AVLT that guided a priori decisions about norms development for the current study. The norms presented in this manuscript are added to that excel file and available at: <https://www.mayo.edu/research/centers-programs/alzheimers-disease-research-center/research-activities/mayo-clinic-study-aging/researchers/data-sharing-resources>.

### Statistical approach

#### Examining effects of demographic variables

Quantitative (e.g.,  $r^2$  = percent variance explained via independent linear regressions) and visual inspection methods (stratified predicted scores) were used to compare effects of demographic

variables across tests. Multivariable regression models examined the independent and interactive effects of age, age<sup>2</sup>, sex, and education on scores as further described below.

#### Regression-based demographically corrected norms

The current study applied the same quantitative and qualitative approaches used with the MNS AVLT data to evaluate regression-based norms and to determine the need for smoothing of variables (Stricker et al., 2021). Regression-based normative formulas were developed by first converting raw scores to normalized scaled scores ( $M = 10$ ,  $SD = 3$ ) using percentile ranks within frequency distributions and then regressing on age, age<sup>2</sup>, sex and education. Standardized scores were used to minimize skew for tests that are not normally distributed. As described previously (Stricker et al., 2021), stepwise procedures were overly sensitive given our large sample size, and additional predictors were considered for inclusion if at least 1% incremental variance was explained in the models beyond a priori predictors (age, age<sup>2</sup>, sex, education). While significant, the variance explained by models when adding non-linear education (quadratic, cubic), cubic age, or two-way interaction terms of all a priori predictor variables was less than 1% and thus not included in normative models (data not shown). More complex curvilinear relationships were considered by applying spline transformations but were determined not to be needed for modeling. We additionally examined whether race/ethnicity (White non-Hispanic vs. all other individuals) met this criteria and found that this variable also explained less than 1% variance beyond age, age<sup>2</sup>, sex and education for all measures.

As previously described, Q-Q plots of standardized residuals were reviewed by rescaling ( $e_i = Y_i - Y_{pred}$ ) raw and covariate adjusted (age, age<sup>2</sup>, sex, education) scores scaled to mean (SD) of 50 (10). We also calculated the difference between observed mean (SD)  $T$ -scores and the expected mean (SD)  $T$ -scores by levels of age (30–59 years, 60–69 years, 70–79 years, and 80 years), sex, and education (8–12, 13–15, 16, and 17–20 years) to determine whether smoothing was indicated based on an absolute mean difference greater than 3  $T$ -score points and  $SD$  outside of the range 9.4–10.6 (Heaton et al., 2004) criteria. If scores were within the range, variables were included as is. If outside of the range, smoothing was applied and reexamined and the smoothing approach that allowed for the least amount of deviation from the criteria across bins was applied. The Appendix provides the information needed for normative data derivation using these MNS norms; this same information is also provided via an excel file at the link provided. Tables of unadjusted scaled scores are the first step in the norming process; raw scores are converted to unadjusted scaled scores and then  $T$ -score formulas are applied (unadjusted scaled scores in isolation are not recommended for clinical use). Fully-adjusted regression-based  $T$ -scores are recommended for clinical interpretation.

#### Application of norms to examine rates of low test performance

We used the independent validation sample and methods previously described (Stricker et al., 2021) to examine rates of low test performance defined as performances below  $-1$   $SD$  when applying MOANS and MNS norms. Rates are significantly different than expected when 95% confidence intervals (CIs) do not include the expected 14.7% base rate value. We similarly examined application of MNS norms in the same sample used to derive the norms to ensure models performed as expected.

## Results

### Participants

Baseline neuropsychological data were available for 4,428 cognitively unimpaired adults, aged 30–91 years (mean age 68.3,  $SD = 13.1$ ), 50.1% female, 97.9% White, mean education 14.7 ( $SD = 2.6$ ). All available test data were used for each measure, with the total  $N$ 's varying slightly by test (see Table 1 for full participant characteristics for the normative sample and  $n$ 's by test; also see Supplemental Table 1). Table 1 also demonstrates that inclusion criteria for this normative sample are broad and result in a highly generalizable sample with regard to health status/medical history. The inclusion requirement that individuals must be judged to be "cognitive unimpaired" by the study physician and study coordinator administering the CDR helps to ensure exclusion of individuals with clinically relevant cognitive impairment related to current or past medical history.

### Effects of demographic variables

The percent variance of test performance explained by each demographic variable independently are reported in Table 2. The variance ( $r^2$ ) explained by demographic variables ranged from 5.7–33.8% for age, 0.0–13.1% for sex, and 2.6–9.8% for education. Combined, these demographic variables explained 13.5–42.5% of variance in test performance. Table 2 also presents the incremental variance explained by each predictor, above and beyond other predictors in the model. Table 3 presents a correlation matrix to show the amount of overlap among predictors. Line plots showing model-predicted scores for age, age<sup>2</sup>, education (20, 16, 12, and 8 years), and sex of select measures are depicted in Figure 1 for language and attention/executive tests and illustrate robust effects. Results from multivariable regression models for all measures are provided in Supplemental Table 2.

### Regression-based demographically corrected norms

Regression-based norms corrected for age, age<sup>2</sup>, sex, and education. Based on our a priori criteria, variables that required smoothing (and the smoothing applied) included BNT (age<sup>2</sup>), vegetable fluency ( $\sqrt{\text{education}}$ ), picture completion ( $\sqrt{\text{age}}$ ), LMI (age + age<sup>2</sup> + age<sup>3</sup>), VRI (age + age<sup>2</sup> + age<sup>3</sup>), and VRII (age). All other  $T$ -scores fell in the appropriate range within age, sex and education bins without smoothing needed. Fully corrected  $T$ -scores had a mean of approximately 50 across all age values, education values and sex. The  $SD$  of nearly all fully corrected  $T$ -scores also fell within the desired range for each age, sex and education bin except for WMS-R VRI ( $SD = 9.27$  for the 60–69 age bin), but this was the best option of several smoothing strategies. Fully adjusted  $T$ -scores effectively removed relationships to demographic variables as desired (all Pearson  $|\rho| < .003$ ; all  $p$ 's  $> .84$ ).

### Cumulative percentiles

We provide cumulative percentiles for the entire sample without stratifying by demographic variables for total errors on Trails A and on Trails B (Table 4). Because total errors were highly skewed,

**Table 1.** Participant characteristics.

	Total ( $n = 4,428$ )
Age, (years), max $n$ , %	
30–39	214, 4.8%
40–49	210, 4.7%
50–59	610, 13.8%
60–69	916, 20.7%
70–79	1655, 37.4%
80–91	823, 18.6%
Education (years), max $n$ , %	
8–12	1257, 28.4%
13–15	1263, 28.5%
16	922, 20.8%
17–20	986, 22.3%
Sex, max $n$ male, %	2211, 49.9%
Race, max $n$ , %	
White, non-Hispanic	4317, 97.5%
White, Hispanic	16, 0.36%
Black/African American	22, 0.50%
Asian	29, 0.65%
Native Hawaiian or Other Pacific Islander	0, 0.0%
American Indian/Alaska native	4, 0.09%
Multiethnic/racial	25, 0.56%
Unknown	15, 0.34%
Neuropsychological measures, $M$ ( $SD$ )	
Boston Naming Test <sup>1</sup>	55.1 (4.2)
Category fluency total <sup>2</sup>	46.4 (10.5)
Animals fluency <sup>3</sup>	20.2 (5.3)
Fruits fluency <sup>3</sup>	13.5 (3.8)
Vegetables fluency <sup>3</sup>	12.8 (3.8)
Trail Making Test, Part A <sup>4</sup>	35.9 (14.3)
Trail Making Test, Part B <sup>5</sup>	89.3 (46.2)
WAIS-R Digit Symbol <sup>6</sup>	48.3 (12.3)
WAIS-R Block Design <sup>7</sup>	26.6 (9.6)
WAIS-R Picture Completion <sup>8</sup>	13.6 (3.1)
WMS-R Logical Memory I <sup>9</sup>	23.6 (6.9)
WMS-R Logical Memory II <sup>10</sup>	18.5 (7.7)
WMS-R Visual Reproduction I <sup>11</sup>	30.5 (5.4)
WMS-R Visual Reproduction II <sup>12</sup>	23.4 (8.8)
Medical history characterization, $n$ , % <sup>13</sup>	
Cancer history <sup>14</sup>	860, 18.7%
History of myocardial infarction	514, 11.7%
History of diabetes, definite or possible <sup>15</sup>	664, 15.1%
History of stroke(s)	130, 3.0%
History hypertension, without treatment	322, 7.3%
History hypertension, with treatment	2,272, 51.8%

<sup>1</sup> $N = 4,329$ .

<sup>2</sup> $N = 4,387$ .

<sup>3</sup> $N = 4,286$ .

<sup>4</sup> $N = 4,350$ .

<sup>5</sup> $N = 4,341$ .

<sup>6</sup> $N = 4,338$ .

<sup>7</sup> $N = 4,335$ .

<sup>8</sup> $N = 4,360$ .

<sup>9</sup> $N = 4,415$ .

<sup>10</sup> $N = 4,412$ .

<sup>11</sup> $N = 4,368$ .

<sup>12</sup> $N = 4,367$ .

<sup>13</sup>Medical history variables were abstracted based on thorough review of the medical record by a nurse abstractor.

<sup>14</sup>The most common cancer types were prostate cancer ( $N = 251$  men, 11.4% of men), breast cancer ( $N = 172$ , 3.9%), melanoma ( $N = 103$ , 2.3%), colon cancer ( $N = 70$ , 1.6%), uterine cancer ( $N = 48$ , 1.1%), and bladder cancer ( $N = 47$ , 1.1%); other cancer types were present in <1% of the sample. This excludes non-melanoma skin cancer.

<sup>15</sup> $N = 46$  (1.0%) possible diabetes,  $N = 618$  (14.1%) definite diabetes,  $N = 10$  with Type 1 diabetes,  $N = 165$  (3.7%) on insulin.

Note. WAIS-R = Wechsler Adult Intelligence Scales-Revised. WMS-R = Wechsler Memory Scale-Revised. All participants completed the Auditory Verbal Learning Test (AVLT) as previously reported (Stricker et al., 2021). Subsamples reported here indicate slight variations in sample size by measure. 93% of participants in the original AVLT sample have all other measures listed here. Table used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

**Table 2.** Individual and incremental percentage variance explained ( $R^2 \times 100$ ) for each demographic variable and the full regression model (combined).

Measure (raw scores)	Individual variable $R^{2a}$				Incremental (partial) $R^{2b}$					Combined $R^2$
	Age	Age squared	Sex	Educ	Age	Age squared	Sex	Educ	Shared <sup>c</sup>	All
Category Fluency Total	16.25	16.93	5.33	7.83	12.30	0.75	7.36	5.21	2.61	28.23
Animals	14.68	15.26	0.10	8.75	10.73	0.54	0.01	4.61	4.14	20.03
Fruits	12.01	12.31	9.95	4.13	9.60	0.32	12.03	3.04	1.05	26.04
Vegetables	5.98	6.43	13.08	2.64	4.59	0.79	15.10	2.53	0.02	23.03
Boston Naming Test	5.74	7.03	1.76	9.76	3.22	4.12	0.79	5.99	3.81	17.93
TMT-A seconds, reversed	24.58	26.46	0.22	3.58	21.71	3.10	0.53	0.76	2.81	28.91
TMT-A errors	0.00	0.00	0.29	0.17	0.02	0.00	0.36	0.26	0.00	0.55
TMT-B seconds, reversed	25.06	27.36	0.00	7.30	20.56	4.14	0.14	2.66	4.63	32.13
TMT-B errors	5.04	5.54	0.12	2.51	3.81	0.93	0.01	1.14	1.38	7.27
AVLT Sum of Trials <sup>d</sup>	25.50	26.19	9.39	5.21	21.60	0.75	11.52	2.61	2.89	39.37
WAIS-R Digit Symbol	33.18	33.79	4.32	8.04	27.74	0.57	6.10	3.71	4.39	42.51
WAIS-R Block Design	23.07	23.49	2.24	8.83	18.24	0.30	1.26	3.22	5.57	28.59
WAIS-R Picture Completion	12.33	13.62	2.50	7.86	8.93	2.34	1.50	3.46	4.40	20.63
WMS-R Logical Memory I	7.85	8.36	0.08	7.37	5.15	0.62	0.44	4.80	2.56	13.57
WMS-R Logical Memory II	10.13	10.63	0.26	6.86	7.17	0.55	0.76	4.16	2.69	15.33
WMS-R Visual Reprod. I	16.61	17.95	0.00	6.04	13.17	2.18	0.08	2.48	3.56	21.47
WMS-R Visual Reprod. II	26.69	28.37	0.05	7.12	22.07	2.36	0.33	2.51	4.61	31.88

<sup>a</sup>Individual variable (e.g., univariate) variance explained, which reflects the amount of variance explained when a single predictor is in the model. These  $R^2 \times 100$  values reported are equivalent to Pearson Correlation Coefficients, Squared. The majority of  $P$  values for Pearson correlation coefficients (before squaring) are  $p < .001$ , except as follows: associations with age differed from  $p < .001$  for TMT-A errors ( $p = 0.752$ ); associations with age squared differed from  $p < .001$  for TMT-A errors ( $p = 0.763$ ); associations with sex differed from  $p < .001$  for Animal fluency ( $p = .039$ ), TMT-A seconds ( $p = .002$ ), TMT-B seconds ( $p = .999$ ), TMT-B errors ( $p = 0.027$ ), LM-I ( $p = .059$ ), VR-I ( $p = .787$ ), and VR-II ( $p = .139$ ); associations with education differed from  $p < .001$  for TMT-A errors ( $p = 0.007$ ).

<sup>b</sup>We performed a series of hierarchical multiple regressions for each test variable in which all but one demographic predictor was included in step one (e.g., age, age squared and sex) and the remaining variable (e.g., education) is entered in a second step. Thus, the incremental (i.e., marginal) variance explained is the amount of variance accounted for by each variable (e.g., education) beyond that explained by the other variables. This allows us to understand the incremental variance accounted for by each predictor, which is the partial  $R^2$ .

<sup>c</sup>Shared = overlapping variance explained by a combination of all 4 model predictors simultaneously; this is calculated as combined variance explained – sum of incremental variance explained for all 4 predictors. For example, shared variance for category fluency total =  $28.23 - (12.30 + 0.75 + 7.36 + 5.21) = 2.61$ .

<sup>d</sup>Auditory Verbal Learning Test Sum of Trials (Trials 1–5 + Trial 6 Short-Delay + 30-minute delayed recall) was included here to provide the incremental variance explained data for this primary AVLT variable that was the focus of our prior work (Stricker et al., 2021) using the same sample.

Note. Table used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

**Table 3.** Correlations between demographic variables.

Measure (raw scores)	Pearson correlation coefficients			
	Age	Age squared	Sex (male)	Education
Age	–	0.992 <sup>a</sup>	–0.024	–0.217*
Age squared	–	–	–0.027	–0.221*
Sex (male)	–	–	–	0.135*
Education	–	–	–	–

<sup>a</sup> $p$ -value not reported because a correlation is expected given age squared is a transformation of age.

\* $p < .001$  (remaining  $p$ -values  $> .05$ ). Table used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

Note. Individuals with  $< 12$  years of education tended to be among the oldest participants.

there were too few positive observations to be able to use the normative approach described above.

### Base rates

#### Normative sample

In the total normative sample, fully adjusted (age, sex, education)  $T$ -scores had a typical distribution of low performances (see Supplemental Table 3). When fully adjusted  $T$ -scores were stratified by sex, the base rates of low performances were greater than expected in males for Fruit Fluency and Vegetable Fluency. Other sex-stratified  $T$ -score base rates were within expectation.

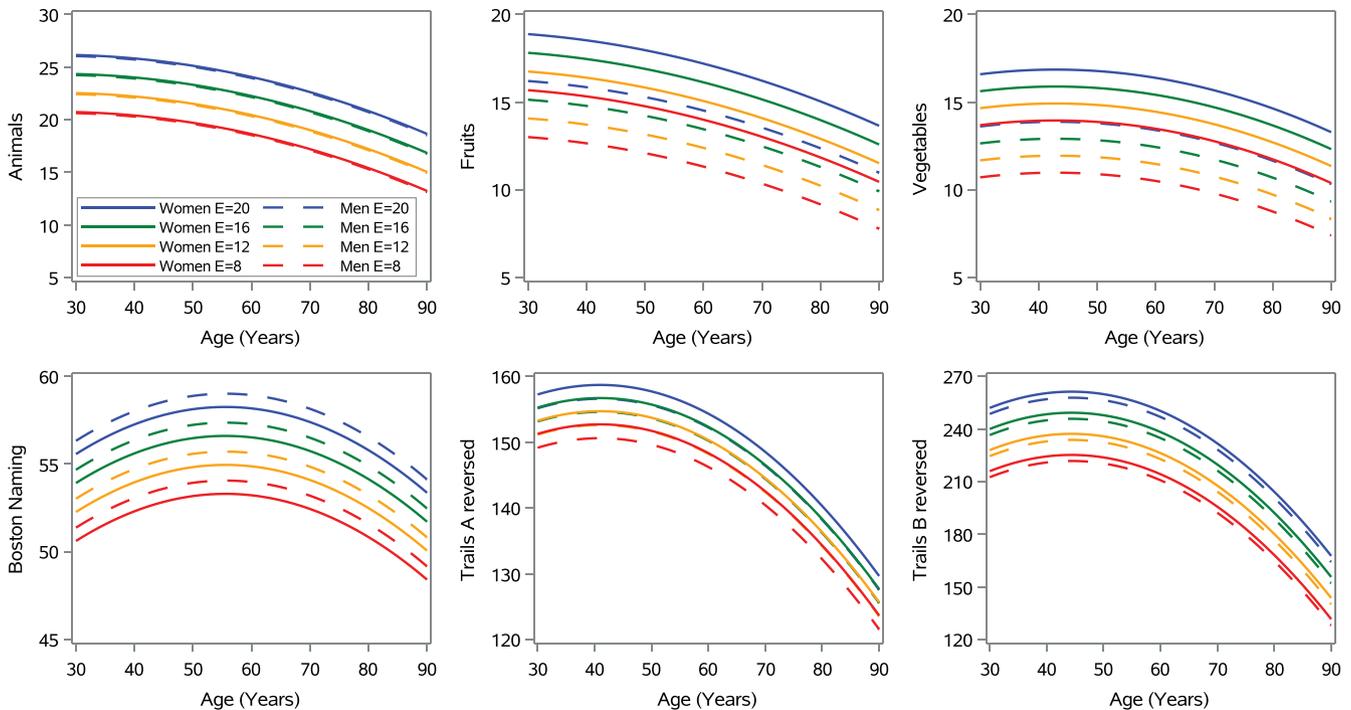
#### Validation sample, all participants

In an independent validation sample of 261 cognitively unimpaired participants aged 56 and older who enrolled in the MCSA after the freeze date for the normative sample (as also described in

Stricker et al., 2021), the application of age-adjusted MOANS norms showed lower-than-expected base rates of low test performance for all measures except LMI and LMII. Thus, lower-than-expected base rates were seen for BNT, Category Fluency Total, Trails A and B, Digit Symbol, Block Design, Picture Completion, and VRI and VRII (see Figure 2 and Supplemental Table 4). Application of age and education-adjusted MOANS norms (not available for all measures) improved the base rates of low test performance, though base rate low performances remained significantly lower-than-expected when collapsing across males and females. Application of fully-adjusted (age, sex, education) MNS norms showed a normal proportion of base rate low performances for all measures except LMI, which showed a higher-than-expected base rate of low performances.

#### Validation sample, sex stratified

Sex-specific differences emerged when stratifying the older adult validation sample by sex (see Figure 3). When age-adjusted MOANS norms were applied, VRII had a lower-than-expected base rate of low performance for females, but not males. Conversely, block design had lower-than-expected base rates of low performance for males, but not females when both age-adjusted and age and education-adjusted MOANS norms were applied. When age and education-adjusted MOANS norms were applied, females had lower-than-expected base rates of low performance for Category Fluency and Trails B, whereas males did not (see Supplemental Table 4). Other sex-stratified results were similar to the total validation sample for MOANS norms. When fully adjusted MNS norms were applied to the sex-stratified validation sample, base rates were within expectation for all measures except for Trails B (female base rate of low performances



**Fig. 1.** Predicted scores from models show 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 each category fluency trial (top row) and for Boston Naming Test, Trails A seconds reversed and Trails B seconds reversed (bottom row). Figure used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

**Table 4.** Observed cumulative percentile for total number of errors on Trail Making Test Part A and Part B.

Total errors	Trails A observed		Trails B observed	
	Total errors	cumulative percentile	Total errors	cumulative percentile
0	0	100	0	100
1	1	11.5	1	31.0
2	2	0.9	2	10.7
3+	3	0.2	3	3.8
-	4	-	4	1.2
-	5	-	5	0.5
-	6+	-	6+	0.2

*Note.* Because Trail Making Test errors were highly skewed, there were too few positive observations to be able to use the normative approach described above, thus we provide cumulative percentiles. Table used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

remained just below expectation, with the upper CI 0.2 below the cutoff).

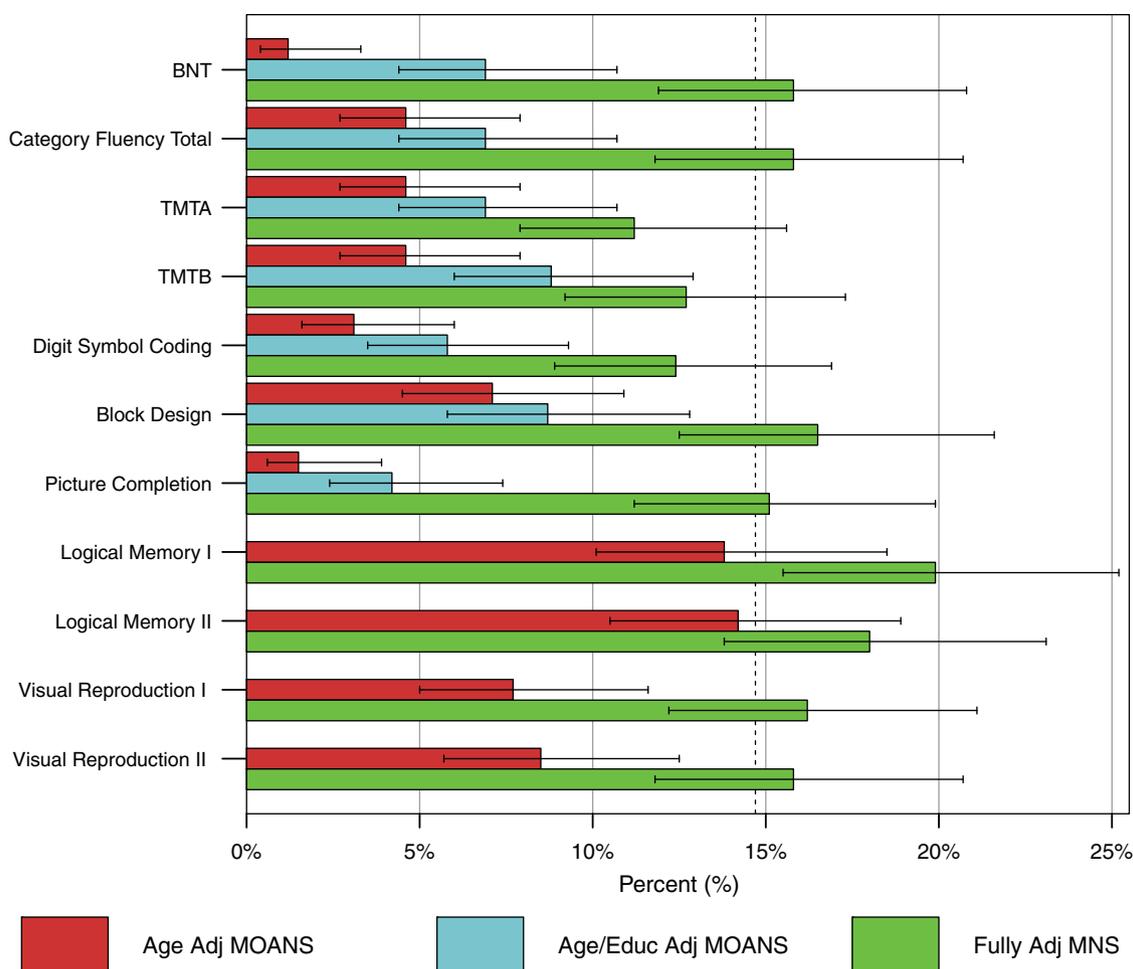
**Discussion**

The MNS aim to develop updated normative data to improve the utility and sensitivity of available clinical tools. The current study reports demographic effects on multiple cognitive measures, provides new MNS regression-based normative data, and examines base rates of low performances relative to MOANS norms in an independent validation sample. We closely examined the different contributions of demographic variables and the patterns of independent variance for each test and performed quality checks on psychometric properties at each step of the regression-based norms approach. In the validation sample, the MNS norms consistently outperformed the MOANS norms that, like many other normative datasets, do not control for sex or

exclude participants with MCI. Our results contribute to a larger discussion of how demographic variables contribute to cognition via biological entities (e.g. brain aging, sex hormones, innate intelligence) and complex social constructs (i.e., generational differences, gender norms, cognitive reserve/socioeconomic resources), and highlight the need for more data-driven and culturally informed normative data to control for these proxy variables.

Quantitative and qualitative analyses show variable patterns of linear and/or quadratic age ( $r^2$ 's 5.74–27.36), sex ( $r^2$ 's 0.00–13.08), and education ( $r^2$ 's 2.64–9.76) associations across measures. The nuances of the effects of demographic variables are key for developing norms that appropriately adjust for variables in a target population. Age accounted for the greatest proportion of independent variance across all but two language measures (BNT, Vegetable Fluency); however, the quadratic age associations varied considerably between measures (see Figure 1 for visual differences in curves across measures). Previous work suggests that the relationship between biological age and cognitive performance is domain specific (Zahodne et al., 2011). Across the adult lifespan, fund of knowledge is projected to increase, whereas fluid abilities including efficient processing and retrieval are predicted to decrease (Salthouse, 2010). As expected, age had a robust negative effect on tasks that require visuomotor speed (Digit Symbol, Trails A, Trails B, Block Design;  $r^2$ 's 23.07–33.79). Regarding memory measures, age accounted for greater variance in delayed recall relative to immediate recall, and design recall relative to story recall.

The curvilinear effect of age on BNT performance suggests that age may also be confounded by generational effects, likely due to decreased salience of BNT items in individuals born within the last 3-to-4 decades. For example, item-level error analysis by Martielli and colleagues revealed error rates from 20–49% on 11 items and



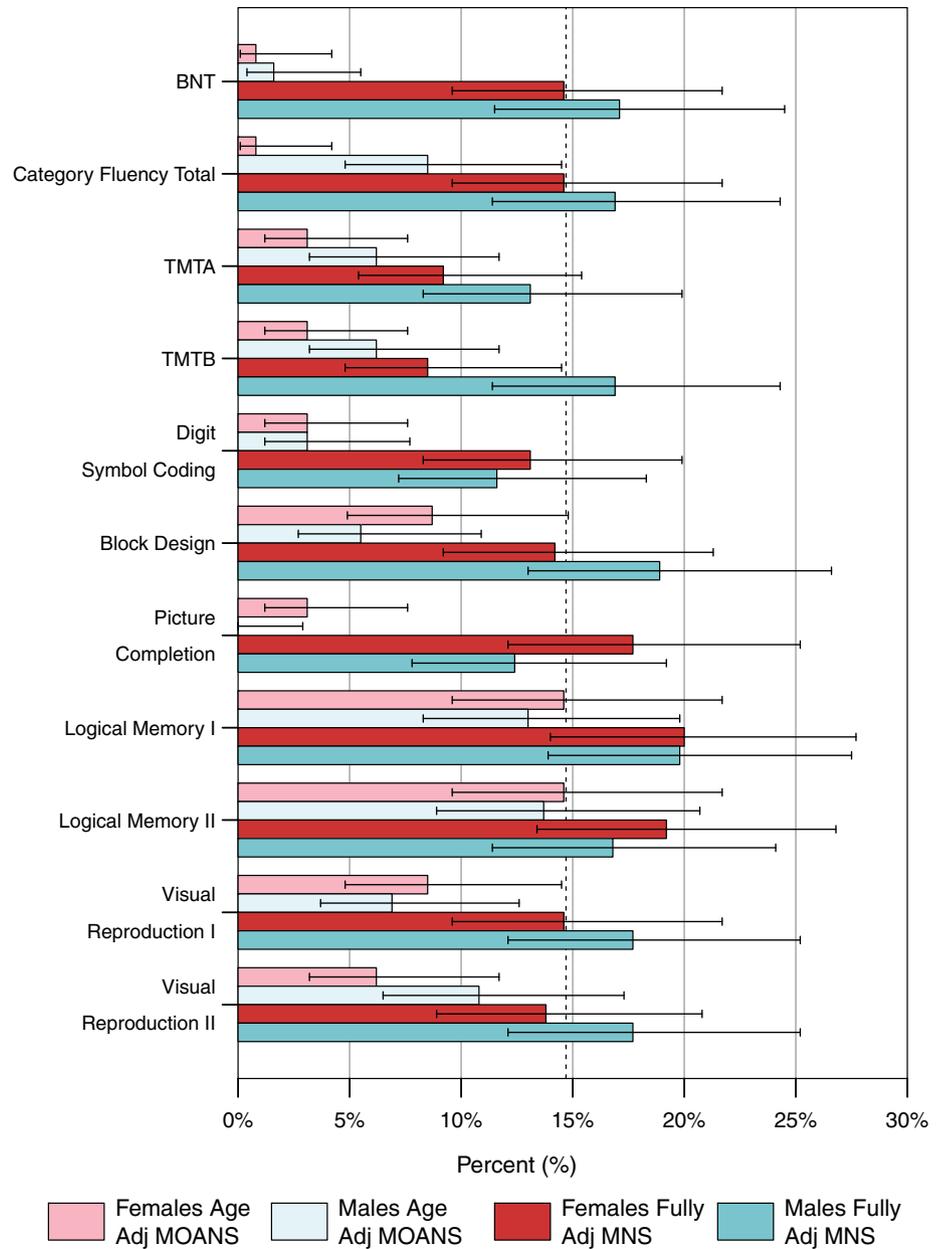
**Fig. 2.** Observed proportions of the validation sample ( $N = 261$ ) showing low test performance ( $SS < 7$  for age-corrected MOANS;  $SS < 7$  for age and education-corrected MOANS;  $T < 40$  for age, sex and education-corrected MNS) with 95% Confidence Intervals (CIs). CIs that do not contain the 14.7% expected base rate value (vertical dashed line) are significantly different than expected.

Note. Adj = adjusted. BNT = Boston Naming Test. MNS = Mayo Normative Studies. MOANS = Mayo's Older Americans Normative Studies. TMTA = Trail Making Test Part A. TMTB = Trail Making Test Part B. When both age-adjusted and age and education-adjusted MOANS norms are available, both are provided above. Logical Memory and Visual Reproduction MOANS are only adjusted for age (Ivnik et al., 1992a). Fully-adjusted MNS adjusts for age, age squared, sex and education. Numeric values corresponding to this figure are available in Supplemental Table 4. Figure used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

50–91% on 5 items, suggesting that limited item familiarity confounds object naming performance in older adolescents falling within the same generation as the lower age bracket of this sample (Martelli & Blackburn, 2016). The popularity of specific words changes over time and is quantifiable through examination of word corpora or a cursory search through Google ngrams (Beattley et al., 2017). Similarly, there are total performance and item-level differences cross-culturally (Li et al., 2022). Despite these issues, clinicians and researchers may be reluctant to adopt alternatives to the BNT, which remains a widely used measure. The few available alternatives are often not available for clinical use, do not have validated norms, or have limited sensitivity (Durant et al., 2021; Loring et al., 2008; Stasenko et al., 2019). Age is the most commonly adjusted-for variable in normative datasets. Biological age is susceptible to noise from environmental phenomena that may systematically vary by population-specific risk factors, recruitment approaches (epidemiological vs aging research samples), generational history, values, and exposure to test paradigms/stimuli. While there is a need to innovate via new test development, updating normative data for existing tests is an important interim step to address the impact of changing

demographics and sociocultural contexts on the existing standards of practice.

A pattern emerged where education contributed greater relative variance in models where sex minimally contributed to the models (e.g., for measures where sex explained <6% of variance). Education accounted for a greater proportion of independent variance than sex across all measures except Vegetable Fluency and Fruit Fluency. These results are broadly consistent with literature exploring demographic effects on cognitive domains as well as individual scores and composites (Vonk et al., 2020; Werry et al., 2019; Zahodne et al., 2011; Zec et al., 2007). While paradigms/test stimuli that are influenced by semantic knowledge base (BNT, Fluency) are intuitively influenced by years of education and other sociocultural factors, education effects in our results were more consistent for memory (6.0–7.4%) and visuospatial (8.0–8.8%) measures. The effect of education on speeded executive/information processing speed appeared to increase with greater complexity of the task/stimuli. These results are an important reminder that visually mediated tasks are not culture-free (Goh & Park, 2009). Of the language measures, BNT and Animal Fluency had a greater proportion of variance attributed to education (8.8–9.8%)



**Fig. 3.** Observed proportions of the validation sample by sex ( $n = 130$  females;  $n = 131$  males) showing low test performance ( $SS < 7$  for age-corrected MOANS;  $T < 40$  for age, sex and education-corrected MNS) with 95% Confidence Intervals (CIs). CIs that do not contain the 14.7% expected base rate value (vertical dashed line) are significantly different than expected. Note. Adj = adjusted. BNT = Boston Naming Test. MNS = Mayo Normative Studies. MOANS = Mayo’s Older Americans Normative Studies. TMTA = Trail Making Test Part A. TMTB = Trail Making Test Part B. Fully-adjusted MNS adjusts for age, age squared, sex and education. Only age-adjusted MOANS norms are presented in this figure for simplicity, but age and education-adjusted MOANS norms are available in Supplemental Table 4 for measures where those are available. Figure used with permission of Mayo Foundation for Medical Education and Research, all rights reserved.

compared to Fruit Fluency and Vegetable Fluency (2.6–4.1%). For these language measures, the pattern of variance shows a tradeoff between sex and education. This duality is not surprising, as sex (or as a social construct, the gender binary) and educational attainment are complex and historically intertwined constructs. Figure 1 illustrates how demographics differentially affect efficient semantic retrieval depending on the stimuli: Animal Fluency (Age > Education > Sex), Fruit Fluency (Age > Sex > Education) and Vegetable Fluency (Sex > Age > Education). On visual inspection, a female with 8 years of education has comparable predicted Vegetable Fluency performance as a male with 20 years of education. Males showed slightly higher-than-expected base rates of low performances for Fruit and Vegetable Fluency demographic adjustments as well. These results highlight how differences in task demands may alter the impact of demographic variables, including paradigms with verbal or visual stimuli.

Results revealed robust sex differences across verbal fluency measures (female advantage for Total, Fruit, Vegetable, but not Animal), with males performing lower on Fruit and Vegetable Fluency. Unlike verbal memory (e.g., “female verbal advantage”) that has evidence of sexual dimorphism in brain structure and biomarker data (Sundermann et al., 2020; Sundermann et al., 2016), after early childhood, language-based differences in cognition, lateralization, and morphometry do not differ between sexes (Wallentin, 2009). Discrepancies in sex effects between fluency categories have been repeatedly observed across samples within the US and from different countries. Specifically, Animal Fluency appears to have minimal-to-no sex difference, whereas females or males more often show category-specific advantages (e.g., fruit/food/supermarket for females, tools and vehicles for males) (Ardila, 2020; Mathuranath et al., 2003; McCarrey et al., 2016; Rivera et al., 2019; St-Hilaire et al., 2016). The salience or

lexicon of semantic knowledge may be influenced indirectly by social norms, resulting in differences that are not necessarily driven by biological sex (Laws, 2004). It is possible that these differences emerge from early language exposures, as age of word acquisition predicts more efficient word retrieval for object naming, verbal fluency, and memory (Morrison et al., 1992). These differences may be mitigated in contexts with fewer socially constructed roles and systemic inequities for females (Gerlach & Gainotti, 2016). Thus, the inconsistent results across fluency categories suggest that the sex differences in verbal fluency are best contextualized as gender differences that are the result of sociocultural norms and experiences.

Our analyses were also powered to reveal additional sex differences, including female advantages on visuomotor speed, cognitive flexibility, and memory measures and male advantages on confrontation naming and visuospatial measures. Digit symbol showed a significant female advantage equating to over 6 points higher than males. This surprising difference underscores the importance of investigating these effects in normative datasets. While women were slightly faster on Trails A, the effect was less clinically meaningful relative to other studies (Munro et al., 2012). The BNT showed a slight advantage for males that similarly may be influenced more by item-level characteristics than naming abilities. The literature is mixed regarding sex differences on the BNT, with a number of studies showing a similar result suggestive of a slight male advantage (Zhang et al., 2017) and others showing no difference (McCarrey et al., 2016). Regarding normative data, adjusting for even subtle differences may be particularly relevant for clinical interpretation on measures that are not normally distributed such as Trails or the BNT. The confluence of biological variables and social constructs that influence demographic effects in these models are population specific and also susceptible to shifts over time with changes in access to resources and sociocultural factors.

In addition to informing the need for demographic adjustments, our results support the need for updating normative data to improve test sensitivity in older adults. MOANS norms, developed in the same geographic region, did not exclude participants with MCI. Accordingly, application of age-adjusted MOANS norms showed lower-than-expected base rates of low test performance ranging from 0.8% to 8.8% on most non-memory measures. In contrast, MOANS norms applied to LMI and LMII had normal total and sex-stratified base rates, VRII had normal base rates for males, and Block Design showed normal base rates for females. The MNS norms detect low performances ( $T < 40$ ; base rate CI's contain 14.7%) within expectation based on a normal distribution in our older adult validation sample. The exception to this is when MNS norms were applied to Trails B performances in females, which had a slightly lower-than-expected base rate and LMI that had an elevated base rate in the overall sample. Our findings raise important points about demographic adjustments to address complex construct/stimuli-related performance variability and the need for updating normative data for older adults that has not previously used stringent inclusion/exclusion criteria. Sex differences may be less robust in contexts where other demographic factors or policies drive equity/inequity (e.g., socioeconomic status, systemic racism, parental leave). Further, the lack of data to support whether sex or gender drive differences in specific cognitive functions limits the ability to serve transgender and gender nonconforming individuals. However, our interpretation would suggest that determinations about what norms to use should

emphasize an individual's lived experience based on their insights and identities.

The current normative study has many strengths including a large population-based sample that allows for a regression-based approach to demographic adjustments. It is important to note that this approach will look different for different populations where proxies of cognitive reserve and other variables that help estimate "normal" performance are bound to the local resources, risk factors, and culture. Limitations of this dataset should be considered when applying normative data. Importantly, the homogeneity of education (e.g., governmental regulation of school attendance, curriculums, quality and quantity) within this sample is representative of the local population and should be considered when applying the norms to individuals. Further, different approaches have been taken to assigning years of education (e.g., Neuropsychological Assessment Battery uses 11 years of education for those obtaining a GED). The MCSA (and prior MOANS) codes education as 12 years for individuals with a GED or who graduated from high school; there is no way to separate out those completing a GED in this retrospective study. The MCSA, these norms and prior MOANS also count 1-year vocational or trade certificate as 13 years of education. These educational coding differences could yield lower T-scores for individuals with a GED or with vocational training than other normative systems and clinicians should be aware of this potential limitation. While Olmsted County is predominantly White and is not broadly representative of the US population (St. Sauver et al., 2012), the MNS AVLT norms have been validated in a more diverse urban sample (Loring et al., 2022). Given that 97.5% of participants in this normative sample are White and non-Hispanic, significant caution is needed when applying these norms to individuals who are not well represented in this normative sample and future studies are needed to expand these normative data to include better representation of individuals from other racial and ethnic groups and/or empirically test performance of the current norms in these groups. In our study, the battery is fixed to allow for longitudinal continuity and has not been updated for select tests with later iterations. Thus, we focused our interpretation on the publically available tests that continue to be widely used, but we also report results for the WAIS-R/WMS-R measures that have more updated version available to provide a larger context of results and for limited use when relevant (e.g., fixed research batteries, retrospective data analysis).

In conclusion, the MNS improves upon earlier normative studies by making use of available population-based research data with a large sample of test-naïve adults ranging from ages 30-91 years that reflects the demographics of Olmsted County, excludes individuals with MCI, and allows for correction of demographic variables (age, sex, and education). Our sample size is much larger than other frequently used normative datasets, particularly for older adults. For example, the sample sizes for MOANS for individuals 80 and older ( $n = 49$  for TMT,  $n = 236$  for Category Fluency,  $n = 232$  for BNT) (Ivnik et al., 1996; Lucas et al., 1998) is notably smaller than for the MNS norms ( $n > 800$  for individuals 80+) described here. Similarly, the MNS sample size is significantly larger than that of the Heaton norms for the White participant sample; while details about specific  $n$ 's by age bins are lacking from that technical manual, for measures in the Halsted Reitan Battery that included TMT, there were 634 total White participants and 121 participants over the age of 64 years, and there were 350 total White participants for the BNT (Heaton et al., 2004).

Our findings highlight the importance of evaluating updated normative data to adjust for key variables that may increase sensitivity for low cognitive performance. Further, we provide a clinical tool that may be useful in neuropsychological evaluations or research. Future work will expand on initial work providing AVLT norms for follow-up visits (Alden et al., 2022), examine the impact of biomarker-negative normative data for older adults, and expand to include other populations.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S1355617723000760>

**Acknowledgments.** The authors report no competing interests related to the content of this manuscript. We acknowledge Ryan Frank, M.S., for his assistance with figure preparation. We thank the participants and staff at the Mayo Clinic Study of Aging.

**Financial statement.** This work was supported by the Rochester Epidemiology Project (R01 AG034676), the National Institutes of Health (grant numbers P50 AG016574, P30 AG062677, U01 AG006786, R01 AG041851, RF1 AG55151, R21 AG073967), an Alzheimer's Disease Research Center Development Award, the Robert Wood Johnson Foundation, The Elsie and Marvin Dekelboum Family Foundation, GHR, Alzheimer's Association, and the Mayo Foundation for Medical Education and Research.

## References

- Alden, E. C., Lundt, E. S., Twohy, E. L., Christianson, T. J., Kremers, W. K., Machulda, M. M., Jack Jr., C. R., Knopman, D. S., Mielke, M. M., Petersen, R. C., & Stricker, N. H. (2022). Mayo normative studies: A conditional normative model for longitudinal change on the Auditory Verbal Learning Test and preliminary validation in preclinical Alzheimer's disease. *Alzheimer's & Dementia*, 14(1), e12325. <https://doi.org/10.1002/dad2.12325>
- Ardila, A. (2020). A cross-linguistic comparison of category verbal fluency test (ANIMALS): A systematic review. *Archives of Clinical Neuropsychology*, 35(2), 213–225. <https://doi.org/10.1093/arclin/acz060>
- Au, B., Dale-McGrath, S., & Tierney, M. C. (2017). Sex differences in the prevalence and incidence of mild cognitive impairment: A meta-analysis. *Ageing Research Reviews*, 35, 176–199. <https://doi.org/10.1016/j.arr.2016.09.005>
- Avila, J. F., Renteria, M. A., Witkiewitz, K., Verney, S. P., Vonk, J. M. J., & Manly, J. J. (2020). Measurement invariance of neuropsychological measures of cognitive aging across race/ethnicity by sex/gender groups. *Neuropsychology*, 34(1), 3–14. <https://doi.org/10.1037/neu0000584>
- Avila, J. F., Vonk, J. M. J., Verney, S. P., Witkiewitz, K., Arce Renteria, M., Schupf, N., Mayeux, R., & Manly, J. J. (2019). Sex/gender differences in cognitive trajectories vary as a function of race/ethnicity. *Alzheimer's & Dementia*, 15(12), 1516–1523. <https://doi.org/10.1016/j.jalz.2019.04.006>
- Beattey, R. A., Murphy, H., Cornwell, M., Braun, T., Stein, V., Goldstein, M., & Bender, H. A. (2017). Caution warranted in extrapolating from Boston Naming Test item gradation construct. *Applied Neuropsychology: Adult*, 24(1), 65–72. <https://doi.org/10.1080/23279095.2015.1089505>
- Benedict, R. (1997). *Brief Visuospatial Memory Test-revised*. Psychological Assessment Resources's, Inc.
- Benedict, R., & Brandt, J. (2001). *Hopkins Verbal Learning Test-Revised (HVLT-R): Professional manual*. Psychological Assessment Resources.
- Bilder, R. M., & Reise, S. P. (2019). Neuropsychological tests of the future: How do we get there from here? *The Clinical Neuropsychologist*, 33(2), 220–245. <https://doi.org/10.1080/13854046.2018.1521993>
- Brooks, B. L., & Iverson, G. L. (2010). Comparing actual to estimated base rates of “abnormal” scores on neuropsychological test batteries: Implications for interpretation. *Archives of Clinical Neuropsychology*, 25(1), 14–21. <https://doi.org/10.1093/arclin/acp100>
- Casaleto, K. B., & Heaton, R. K. (2017). Neuropsychological assessment: Past and future. *Journal of the International Neuropsychological Society*, 23(9–10), 778–790. <https://doi.org/10.1017/S1355617717001060>
- Clark, L. R., Kosciak, R. L., Nicholas, C. R., Okonkwo, O. C., Engelman, C. D., Bratzke, L. C., Hogan, K. J., Mueller, K. D., Bendlin, B. B., Carlsson, C. M., Asthana, S., Sager, M. A., Hermann, B. P., & Johnson, S. C. (2016). Mild cognitive impairment in late middle age in the Wisconsin registry for Alzheimer's prevention study: Prevalence and characteristics using robust and standard neuropsychological normative data. *Archives of Clinical Neuropsychology*, 31(7), 675–688. <https://doi.org/10.1093/arclin/acw024>
- Collins, F. S., & Riley, W. T. (2016). NIH's transformative opportunities for the behavioral and social sciences. *Science Translational Medicine*, 8(366), 366ed314. <https://doi.org/10.1126/scitranslmed.aai9374>
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2017). *California verbal learning test manual* (3rd ed.). NCS Pearson, Inc.
- Durant, J., Berg, J. L., Banks, S. J., Kaylegian, J., & Miller, J. B. (2021). Comparing the Boston naming test with the neuropsychological assessment battery-naming subtest in a neurodegenerative disease clinic population. *Assessment*, 28(5), 1256–1266. <https://doi.org/10.1177/1073191119872253>
- Edmonds, E. C., Delano-Wood, L., Jak, A. J., Galasko, D. R., Salmon, D. P., Bondi, M. W., & Alzheimer's Disease Neuroimaging Initiative (2016). “Missed” mild cognitive impairment: High false-negative error rate based on conventional diagnostic criteria. *Journal of Alzheimer's Disease*, 52(2), 685–691. <https://doi.org/10.3233/JAD-150986>
- Eloi, J. M., Lee, J., Pollock, E. N., Tayim, F. M., Holcomb, M. J., Hirst, R. B., Tocco, C., Towns, S. J., Lichtenstein, J. D., & Roth, R. M. (2021). Boston Naming Test: Lose the noose. *Archives of Clinical Neuropsychology*, 36(8), 1465–1472. <https://doi.org/10.1093/arclin/acab017>
- Fine, E. M., Kramer, J. H., Lui, L.-Y., Yaffe, K., & Study of Osteoporotic Fractures (2012). Normative data in women aged 85 and older: Verbal fluency, digit span, and the CVLT-II short form. *Clinical Neuropsychologist*, 26(1), 18–30. <https://doi.org/10.1080/13854046.2011.639310>
- Gerlach, C., & Gainotti, G. (2016). Gender differences in category-specificity do not reflect innate dispositions. *Cortex*, 85, 46–53. <https://doi.org/10.1016/j.cortex.2016.09.022>
- Goh, J. O., & Park, D. C. (2009). Culture sculpts the perceptual brain. *Progress in Brain Research*, 178, 95–111. [https://doi.org/10.1016/S0079-6123\(09\)17807-X](https://doi.org/10.1016/S0079-6123(09)17807-X)
- Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an expanded Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications*. Psychological Assessment Resources.
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Psychological Assessment Resources.
- Hiscock, M. (2007). The Flynn effect and its relevance to neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, 29(5), 514–529. <https://doi.org/10.1080/13803390600813841>
- Holtzer, R., Goldin, Y., Zimmerman, M., Katz, M., Buschke, H., & Lipton, R. B. (2008). Robust norms for selected neuropsychological tests in older adults. *Archives of Clinical Neuropsychology*, 23(5), 531–541. <https://doi.org/10.1016/j.acn.2008.05.004>
- Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E., & Petersen, R. C. (1996). Neuropsychological tests' norms above age 55: COWAT, BNT, MAE token, WRAT-R reading, AMNART, STROOP, TMT, and JLO. *The Clinical Neuropsychologist*, 10(3), 262–278. <https://doi.org/10.1080/13854049608406689>
- Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E., Petersen, R. C., Kokmen, E., & Kurland, L. T. (1992a). Mayo's Older Americans Normative Studies: WMS-R norms for ages 56 to 94. *The Clinical Neuropsychologist*, 6(Supplement), 49–82.
- Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E., Petersen, R. C., Kokmen, E., & Kurland, L. T. (1992b). Mayo's Older Americans Normative Studies: Updated AVLT norms for ages 56 to 97. *The Clinical Neuropsychologist*, 6(Supplement), 83–104. <https://doi.org/10.1080/13854049608406689>
- Jack Jr., C. R., Therneau, T. M., Weigand, S. D., Wiste, H. J., Knopman, D. S., Vemuri, P., Lowe, V. J., Mielke, M. M., Roberts, R. O., Machulda, M. M., Graff-Radford, J., Jones, D. T., Schwarz, C. G., Gunter, J. L., Senjem, M. L., Rocca, W. A., & Petersen, R. C. (2019). Prevalence of biologically vs clinically defined Alzheimer spectrum entities using the National Institute on

- Aging-Alzheimer's Association Research Framework. *JAMA Neurology*, 76(10), 1174. <https://doi.org/10.1001/jamaneurol.2019.1971>
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test* (2nd ed.). Lea & Febiger.
- Kokmen, E., Smith, G. E., Petersen, R. C., Tangalos, E., & Ivnik, R. C. (1991). The short test of mental status: Correlations with standardized psychometric testing. *Archives of Neurology*, 48(7), 725–728. <https://doi.org/10.1001/archneur.1991.00530190071018>
- Laws, K. R. (2004). Sex differences in lexical size across semantic categories. *Personality and Individual Differences*, 36(1), 23–32. [https://doi.org/10.1016/S0191-8869\(03\)00048-5](https://doi.org/10.1016/S0191-8869(03)00048-5)
- Li, Y., Qiao, Y., Wang, F., Wei, C., Wang, R., Jin, H., Xie, B., You, J., Jia, J., & Zhou, A. (2022). Culture effects on the Chinese version Boston Naming Test performance and the normative data in the native Chinese-speaking elders in mainland China. *Frontiers in Neurology*, 13, 866261. <https://doi.org/10.3389/fneur.2022.866261>
- Loring, D. W., Saurman, J. L., John, S. E., Bowden, S. C., Lah, J. J., & Goldstein, F. C. (2022). The Rey Auditory Verbal Learning Test: Cross-validation of Mayo Normative Studies (MNS) demographically corrected norms with confidence interval estimates. *Journal of the International Neuropsychological Society*, 29(4), 397–405. <https://doi.org/10.1017/S1355617722000248>
- Loring, D. W., Strauss, E., Hermann, B. P., Barr, W. B., Perrine, K., Trenerry, M. R., Chelune, G., Westerveld, M., Lee, G., Meador, K. G., & Bowden, S. C. (2008). Differential neuropsychological test sensitivity to left temporal lobe epilepsy. *Journal of the International Neuropsychological Society*, 14(3), 394–400. <https://doi.org/10.1017/S1355617708080582>
- Lucas, J. A., Ivnik, R. J., Smith, G. E., Bohac, D. L., Tangalos, E. G., Graff-Radford, N. R., & Petersen, R. C. (1998). Mayo's older Americans normative studies: Category fluency norms. *Journal of Clinical and Experimental Neuropsychology*, 20(2), 194–200. <https://doi.org/10.1076/jcen.20.2.194.1173>
- Martelli, T. M., & Blackburn, L. B. (2016). When a funnel becomes a martini glass: Adolescent performance on the Boston Naming Test. *Child Neuropsychology*, 22(4), 381–393. <https://doi.org/10.1080/09297049.2015.1014899>
- Mathuranath, P. S., George, A., Cherian, P. J., Alexander, A., Sarma, S. G., & Sarma, P. S. (2003). Effects of age, education and gender on verbal fluency. *Journal of Clinical and Experimental Neuropsychology*, 25(8), 1057–1064. <https://doi.org/10.1076/jcen.25.8.1057.16736>
- McCarrey, A. C., An, Y., Kitner-Triolo, M. H., Ferrucci, L., & Resnick, S. M. (2016). Sex differences in cognitive trajectories in clinically normal older adults. *Psychology and Aging*, 31(2), 166–175. <https://doi.org/10.1037/pag0000070>
- Miller, I. N., Himali, J. J., Beiser, A. S., Murabito, J. M., Seshadri, S., Wolf, P. A., & Au, R. (2015). Normative data for the cognitively intact oldest-old: The Framingham Heart Study. *Experimental Aging Research*, 41(4), 386–409. <https://doi.org/10.1080/0361073X.2015.1053755>
- Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment* (2nd ed.). Oxford University Press.
- Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, 43(11), 2412–2414. <https://doi.org/10.1212/WNL.43.11.2412-a>
- Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory and Cognition*, 20(6), 705–714. <https://doi.org/10.3758/bf03202720>
- Munro, C. A., Winicki, J. M., Schretlen, D. J., Gower, E. W., Turano, K. A., Muñoz, B., Keay, L., Bandeen-Roche, K., & West, S. K. (2012). Sex differences in cognition in healthy elderly individuals. *Development, and Cognition. Section B: Aging, Neuropsychology and Cognition*, 19(6), 759–768. <https://doi.org/10.1080/13825585.2012.690366>
- Nebel, R. A., Aggarwal, N. T., Barnes, L. L., Gallagher, A., Goldstein, J. M., Kantarci, K., Mallampalli, M. P., Mormino, E. C., Scott, L., Yu, W. H., Maki, P. M., & Mielke, M. M. (2018). Understanding the impact of sex and gender in Alzheimer's disease: A call to action. *Alzheimers & Dementia*, 14(9), 1171–1183. <https://doi.org/10.1016/j.jalz.2018.04.008>
- Pedraza, O., Lucas, J. A., Smith, G. E., Petersen, R. C., Graff-Radford, N. R., & Ivnik, R. J. (2010). Robust and expanded norms for the Dementia Rating Scale. *Archives of Clinical Neuropsychology*, 25(5), 347–358. <https://doi.org/10.1093/arclin/acq030>
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3), 183–194. <https://doi.org/10.1111/j.1365-2796.2004.01388.x>
- Petersen, R. C., Roberts, R. O., Knopman, D. S., Geda, Y. E., Cha, R. H., Pankratz, V. S., Boeve, B. F., Tangalos, E. G., Ivnik, R. J., & Rocca, W. A. (2010). Prevalence of mild cognitive impairment is higher in men: The Mayo Clinic Study of Aging. *Neurology*, 75(10), 889–897. <https://doi.org/10.1212/WNL.0b013e3181f11d85>
- Reitan, R. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8(3), 271–276. <https://doi.org/10.2466/pms.1958.8.3.271>
- Rivera, D., Olabarrieta-Landa, L., Van der Elst, W., Gonzalez, I., Rodriguez-Agudelo, Y., Aguayo Arellis, A., Rodriguez-Irizarry, W., García de la Cadena, C., & Arango-Lasprilla, J. C. (2019). Normative data for verbal fluency in healthy Latin American adults: Letter M, and fruits and occupations categories. *Neuropsychology*, 33(3), 287–300. <https://doi.org/10.1037/neu0000518>
- Roberts, R. O., Geda, Y. E., Knopman, D. S., Cha, R. H., Pankratz, V. S., Boeve, B. F., Ivnik, R. J., Tangalos, E. G., Petersen, R. C., & Rocca, W. A. (2008). The Mayo Clinic Study of Aging: Design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*, 30(1), 58–69. <https://doi.org/10.1159/000115751>
- Salthouse, T. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, 16(5), 754–760. <https://doi.org/10.1017/S1355617710000706>
- St-Hilaire, A., Hudon, C., Vallet, G. T., Bherer, L., Lussier, M., Gagnon, J.-F., Simard, M., Gosselin, N., Escudier, F., Rouleau, I., & Maccoir, J. (2016). Normative data for phonemic and semantic verbal fluency test in the adult French-Quebec population and validation study in Alzheimer's disease and depression. *Clinical Neuropsychologist*, 30(7), 1126–1150. <https://doi.org/10.1080/13854046.2016.1195014>
- St. Sauer, J. L., Grossardt, B. R., Leibson, C. L., Yawn, B. P., Melton III, L. J., & Rocca, W. A. (2012). Generalizability of epidemiological findings and public health decisions: An illustration from the Rochester Epidemiology Project. *Mayo Clinic Proceedings*, 87(2), 151–160. <https://doi.org/10.1016/j.mayocp.2011.11.009>
- St. Sauer, J. L., Grossardt, B. R., Yawn, B. P., Melton III, L. J., & Rocca, W. A. (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. <https://doi.org/10.1093/aje/kwq482>
- Stasenko, A., Jacobs, D. M., Salmon, D. P., & Gollan, T. H. (2019). The Multilingual Naming Test (MINT) as a measure of picture naming ability in Alzheimer's disease. *Journal of the International Neuropsychological Society*, 25(8), 821–833. <https://doi.org/10.1017/S1355617719000560>
- Steinberg, B. A., Bieliauskas, L. A., Smith, G. E., Langelotti, C., & Ivnik, R. J. (2005). Mayo's older Americans normative studies: Age- and IQ-adjusted norms for the Boston Naming Test, the MAE Token Test, and the Judgment of Line Orientation Test. *Clinical Neuropsychologist*, 19(3-4), 280–328. <https://doi.org/10.1080/13854040590945229>
- Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.
- Stricker, N. H., Christianson, T. J., Lundt, E. S., Alden, E. C., Machulda, M. M., Fields, J. A., Kremers, W. K., Jack Jr., C. R., Knopman, D. S., Mielke, M. M., & Petersen, R. C. (2021). Mayo normative studies: Regression-based normative data for the auditory verbal learning test for ages 30–91 years and the importance of adjusting for sex. *Journal of the International Neuropsychological Society*, 27(3), 211–226. <https://doi.org/10.1017/S1355617720000752>
- Sundermann, E. E., Barnes, L. L., Bondi, M. W., Bennett, D. A., Salmon, D. P., & Maki, P. M. (2021). Improving detection of amnesic mild cognitive impairment with sex-specific cognitive norms. *Journal of Alzheimer's Disease*, 84(4), 1763–1770. <https://doi.org/10.3233/JAD-215260>

Sundermann, E. E., Maki, P. M., Reddy, S., Bondi, M. W., Biegion, A., & Alzheimer's Disease Neuroimaging Initiative (2020). Women's higher brain metabolic rate compensates for early Alzheimer's pathology. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 12(1), e12121. <https://doi.org/10.1002/dad2.12121>

Sundermann, E. E., Maki, P. M., Rubin, L. H., Lipton, R. B., Landau, S., Biegion, A., & Alzheimer's Disease Neuroimaging Initiative (2016). Female advantage in verbal memory: Evidence of sex-specific cognitive reserve. *Neurology*, 87(18), 1916–1924. <https://doi.org/10.1212/WNL.0000000000003288>

Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, 19(2), 203–214. [https://doi.org/10.1016/S0887-6177\(03\)00039-8](https://doi.org/10.1016/S0887-6177(03)00039-8)

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, 14(2), 167–177. <https://www.ncbi.nlm.nih.gov/pubmed/14590600>

Vonk, J. M. J., Higby, E., Nikolaev, A., Cahana-Amitay, D., Spiro, A., Albert, M. L., Obler, L. K., & Taler, V. (2020). Demographic effects on longitudinal semantic processing, working memory, and cognitive speed. *Journals of Gerontology. Series B: Psychological Sciences and Social Sciences*, 75(9), 1850–1862. <https://doi.org/10.1093/geronb/gbaa080>

Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and Language*, 108(3), 175–183. <https://doi.org/10.1016/j.bandl.2008.07.001>

Wang, C., Katz, M. J., Chang, K. H., Qin, J., Lipton, R. B., Zwerling, J. L., Sliwinski, M. J., Derby, C. A., Rabin, L. A., & Abner, E. (2021). UDSNB 3.0 neuropsychological test norms in older adults from a diverse community: Results from the Einstein Aging Study (EAS). *Journal of Alzheimer's Disease*, 83(4), 1665–1678. <https://doi.org/10.3233/JAD-210538>

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale (WAIS-III)* (3rd ed.). The Psychological Corporation.

Wechsler, D. (2009). *Subtest administration and scoring. WAIS-IV: Administration and scoring manual*. The Psychological Corporation.

Wechsler, D. A. (1981). *Wechsler Adult Intelligence Scale-Revised*. The Psychological Corporation.

Wennberg, A. M. V., Lesnick, T. G., Schwarz, C. G., Savica, R., Hagen, C. E., Roberts, R. O., Knopman, D. S., Hollman, J. H., Vemuri, P., Jack Jr., C. R., Petersen, R. C., & Mielke, M. M. (2018). Longitudinal association between brain amyloid-beta and gait in the Mayo Clinic Study of Aging. *Journals of Gerontology. Series A: Biological Sciences and Medical Sciences*, 73(9), 1244–1250. <https://doi.org/10.1093/gerona/glx240>

Werry, A. E., Daniel, M., & Bergstrom, B. (2019). Group differences in normal neuropsychological test performance for older non-Hispanic White and Black/African American adults. *Neuropsychology*, 33(8), 1089–1100. <https://doi.org/10.1037/neu0000579>

Zahodne, L. B., Glymour, M. M., Sparks, C., Bontempo, D., Dixon, R. A., MacDonald, S. W., & Manly, J. J. (2011). Education does not slow cognitive decline with aging: 12-year evidence from the victoria longitudinal study. *Journal of the International Neuropsychological Society*, 17(6), 1039–1046. <https://doi.org/10.1017/S1355617711001044>

Zec, R. F., Burkett, N. R., Markwell, S. J., & Larsen, D. L. (2007). A cross-sectional study of the effects of age, education, and gender on the Boston Naming Test. *Clinical Neuropsychologist*, 21(4), 587–616. <https://doi.org/10.1080/13854040701220028>

Zhang, J., Zhou, W., Wang, L., Zhang, X., & Harvard Aging Brain, S. (2017). Gender differences of neuropsychological profiles in cognitively normal older people without amyloid pathology. *Comprehensive Psychiatry*, 75, 22–26. <https://doi.org/10.1016/j.comppsy.2017.02.008>

**Appendix**

All materials in the Appendix used with permission of Mayo Foundation of Medical Education and Research, all rights reserved. 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.** Table for converting raw scores to unadjusted scaled scores for language and attention/executive measures.<sup>a</sup>

SS	Category Fluency Total			Animals fluency	Fruits fluency	Veg. fluency	TMTA	TMTB	SS
	BNT								
0	0–30	0–18	0–6	0–2	0–3	122–180	–	0	
1	31–36	19–21	7	3–4	–	107–121	–	1	
2	37–39	22–23	8	5	4	93–106	300	2	
3	40–42	24–25	9–10	6	5	80–92	260–299	3	
4	43–45	26–29	11	7	6	69–79	203–259	4	
5	46–48	30–31	12	8	7	60–68	164–202	5	
6	49–50	32–34	13–14	9	8	51–59	135–163	6	
7	51–52	35–37	15	10	9	45–50	113–134	7	
8	53–54	38–40	16–17	11	10	40–44	97–112	8	
9	55	41–43	18	12	11	36–39	84–96	9	
10	56	44–47	19–20	13	12–13	32–35	73–83	10	
11	57	48–50	21–22	14–15	14	28–31	64–72	11	
12	58	51–54	23–24	16	15	25–27	56–63	12	
13	59	55–58	25–26	17–18	16–17	23–24	50–55	13	
14	–	59–62	27–28	19	18	21–22	44–49	14	
15	–	63–67	29–30	20	19–20	19–20	39–43	15	
16	60	68–72	31–33	21–22	21	17–18	35–38	16	
17	–	73–75	34–35	23–24	22–23	15–16	31–34	17	
18	–	76–80	36–38	25–26	24–25	14	29–30	18	
19	–	81–95	39	27–28	26	12–13	26–28	19	
20	–	≥96	≥40	≥29	≥27	≤11	≤25	20	

<sup>a</sup>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. See Supplementary Material for WAIS-R/WMS-R measures; these are not included here because they are not recommended for clinical use given the availability of updated versions of these tests.  
 Note. BNT = Boston Naming Test. Category Fluency Total = animals + fruits + vegetables. SS = scaled score. TMTA = Trail Making Test Part A. TMTB = Trail Making Test Part B.

**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.

**Table A2.** Education level determination rules.

Education completed	Years of Education Assigned
Below High School Diploma/GED: each full year completed in formal education is counted	0-11
High School Diploma/GED	12
1 or more years of Vocational/Trade School	13
1 or more full-time years of Associate's program without degree	13
Associate degree	14
1 full-time year of Bachelor's program without degree	13
2 full-time years of Bachelor's program without degree	14
3 or more full-time years of Bachelor's program without degree	15
Bachelor's Degree	16
1 or more full-time years of Master's program without degree	17

(Continued)

**Table A2.** (Continued)

Education completed	Years of Education Assigned
Master's Degree	18
1 or more full-time years of Doctoral program without degree	19
Attorneys and Priests	19
Doctoral degree	20
Continuing education/certifications: no additional years are counted	

*Note.* 12 years of education includes individuals with a GED as well as individuals who graduated from high school with a high school diploma. These data were coded the same and thus could not be differentiated. Caution is suggested when interpreting performance in individuals with 8-11 years of education, as this group was less represented in the normative sample ( $n = 131$  or 2.96% of the overall normative sample vs. 1257 with 12 years of education as defined above or 28.39% of the overall sample; see Supplemental Table 2 from Stricker et al. (2021) for  $n$ 's by each level 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

$$\text{TScoreBNT} = \text{rounde}(50 + (((\text{BNTSS} - (-3.65527301238070 + (\text{Age} * 0.33738281984054) + (\text{Age}^{**}2 * -0.00300164411145) + (\text{Male} * 0.56914319383324) + (\text{EDUC} * 0.32502825187670)))/ (1.9590575887 + (\text{Age}^{**}2 * 0.0000462471))) + 0.000001603684499390)/0.124674264647337))$$

$$\text{TScoreCFT} = \text{rounde}(50 + (((\text{CFTSS} - (7.25285833912243 + (\text{Age} * 0.09590239915042) + (\text{Age}^{**}2 * -0.00141757357343) + (\text{Male} * -1.66433267090259) + (\text{EDUC} * 0.26864406243851)))/1) + 0.000000000019639730)/0.256098211730331))$$

$$\text{TScoreCFA} = \text{rounde}(50 + (((\text{CFASS} - (6.87650447809604 + (\text{Age} * 0.08074860956743) + (\text{Age}^{**}2 * -0.00124305406725) +$$

$$(\text{Male} * -0.05459425948988) + (\text{EDUC} * 0.24868871078244)))/1) - 0.000000000007342960)/0.266972646133392))$$

$$\text{TScoreCFF} = \text{rounde}(50 + (((\text{CFFSS} - (9.40596330525577 + (\text{Age} * 0.04460703211948) + (\text{Age}^{**}2 * -0.00093878501696) + (\text{Male} * -2.18633365107317) + (\text{EDUC} * 0.21067777851466)))/1) + 0.000000000003846940)/0.266854710793310))$$

$$\text{TScoreCFV} = \text{rounde}(50 + (((\text{CFVSS} - (7.16653336860412 + (\text{Age} * 0.10966408817646) + (\text{Age}^{**}2 * -0.00127359943400) + (\text{Male} * -2.40105493155836) + (\text{EDUC} * 0.18750751554837)))/ (1.3818783104 + (\text{Educ}^{**}0.5 * 0.1861728454))) + 0.000215448881490048)/0.126225678816071))$$

$$\text{TScoreTMA} = \text{rounde}(50 + (((\text{TMASS} - (11.02075160126570 + (\text{Age} * 0.07794875829736) + (\text{Age}^{**}2 * -0.00161923144463) + (\text{Male} * -0.43860175484993) + (\text{EDUC} * 0.11486703322575)))/1) + 0.000000000019508923)/0.246976245644101))$$

$$\text{TScoreTMB} = \text{rounde}(50 + (((\text{TMBSS} - (8.85967150966154 + (\text{Age} * 0.10252077910469) + (\text{Age}^{**}2 * -0.00183370870098) + (\text{Male} * -0.34041423893386) + (\text{EDUC} * 0.21407076945498)))/1) - 0.000000000007059292)/0.234111975138344))$$

*Note.* BNT = Boston Naming Test. CFT = Category Fluency Total. CFA = Animal Fluency. CFF = Fruit Fluency. CFV = Vegetable Fluency. Male = indicates male is coded as 1, female is coded as 0. Rounde = signifies the specific round function used in Statistical Analysis Software (SAS) Version 9.4. SS = unadjusted scaled score. TMA = Trail Making Test Part A. TMB = Trail Making Test Part B. See Supplementary Material for WAIS-R/WMS-R measures; these are not included here because they are not recommended for clinical use given the availability of updated versions of these tests.