



Research Article

Sex-specific differences in neuropsychological profiles of mild cognitive impairment in a hospital-based clinical sample

Aimee J. Karstens^{1,4} , Taylor R. Maynard^{2,3}  and Geoffrey Tremont^{2,4}

¹Mayo Clinic, Department of Psychiatry and Psychology, Rochester, MN, USA, ²Neuropsychology Program, Rhode Island Hospital, Providence, RI, USA,

³Northwestern University, Feinberg School of Medicine, Department of Psychology, Chicago, IL, USA and ⁴Department of Psychiatry, Brown University, Alpert School of Medicine, Department of Psychology, Providence, RI, USA

Abstract

Objective: Mild cognitive impairment (MCI) is an etiologically nonspecific diagnosis including a broad spectrum of cognitive decline between normal aging and dementia. Several large-scale cohort studies have found sex effects on neuropsychological test performance in MCI. The primary aim of the current project was to examine sex differences in neuropsychological profiles in a clinically diagnosed MCI sample using clinical and research diagnostic criteria. **Method:** The current study includes archival data from 349 patients (age $M = 74.7$; $SD = 7.7$) who underwent an outpatient neuropsychological evaluation and were diagnosed with MCI. Raw scores were converted to z -scores using normative datasets. Sex differences in neurocognitive profiles including severity, domain-specific composites (memory, executive functioning/information processing speed, and language), and modality-specific learning curves (verbal, visual) were examined using Analysis of Variance, Chi-square analyses, and linear mixed models. *Post hoc* analyses examined whether sex effects were uniform across age and education brackets. **Results:** Females exhibit worse non-memory domain and test-specific cognitive performances compared to males with otherwise comparable categorical MCI criteria and global cognition measured via screening and composite scores. Analysis of learning curves showed additional sex-specific advantages (visual Males > Females; verbal Females > Males) not captured by MCI subtypes. **Conclusions:** Our results highlight sex differences in a clinical sample with MCI. The emphasis of verbal memory in the diagnosis of MCI may result in diagnosis at more advanced stages for females. Additional investigation is needed to determine whether these profiles confer greater risk for progressing to dementia or are confounded by other factors (e.g., delayed referral, medical comorbidities).

Keywords: MCI; female; dementia; criteria; diagnosis; verbal memory

(Received 22 June 2022; final revision 19 December 2022; accepted 29 December 2022; First Published online 3 March 2023)

Introduction

Mild cognitive impairment (MCI) is a stage of cognitive decline falling between normal aging and dementia. A subtype of MCI known as amnesic MCI (aMCI), presents with a marked decline in verbal memory and may reflect early manifestations of a neurodegenerative disease process like Alzheimer's disease (AD; Eppig et al., 2017; Petersen et al., 1999). Though etiologies can be inferred based on patterns of cognitive impairment, comprehensive medical/psychosocial history, and additional neurodiagnostic testing, MCI is etiologically nonspecific. Various diagnostic criteria have been developed and revised in attempts to best capture MCI patterns that define a pre-dementia syndrome (versus stable or remitting MCI) despite its inherent heterogeneity (Clark et al., 2013; Clark et al., 2016; Edmonds et al., 2016; Grundman et al., 2004; Rountree et al., 2007). MCI subtypes have been further characterized (e.g., aMCI, dysexecutive, amnestic) by employing conventional Peterson/Winblad ("single-test") criteria as well as actuarial neuropsychological ("multi-test") criteria by Jak and Bondi. The Jak/Bondi criteria categorizes specific cognitive phenotypes that are differentially associated with relatively high rates of the

APOE $\epsilon 4$ allele, abnormal CSF levels of AD biomarkers, and 49% AD conversion rates over 7 years of follow-up (Bondi et al., 2014).

In the last decade, there has been a growing research interest in the role of sex on cognition in MCI. For instance, several large-scale studies show a higher incidence and prevalence of aMCI in men (Petersen et al., 2010; Roberts et al., 2012), though a recent meta-analysis of population and community-based studies around the world shows no significant sex differences in prevalence or incidence of aMCI. In contrast, there is a higher prevalence, but not incidence, of non-aMCI in females compared to males (Au et al., 2017). Moreover, evidence suggests that women tend to be diagnosed with MCI at a later disease stage, and progress more quickly from MCI to dementia than men (Holland et al., 2013; Sundermann et al., 2019). Across the lifespan, females exhibit better verbal memory performance, especially on list-learning memory tasks (Sundermann et al., 2019).

The "female verbal advantage" for memory performance (Stricker et al., 2021; Sundermann et al., 2016b; Sundermann et al., 2017) is posited to play a role in false-negative diagnostic

Corresponding author: Geoffrey Tremont, email: gtremont@lifespan.org

Cite this article: Karstens A.J., Maynard T.R., & Tremont G. (2023) Sex-specific differences in neuropsychological profiles of mild cognitive impairment in a hospital-based clinical sample. *Journal of the International Neuropsychological Society*, 29: 821–830, <https://doi.org/10.1017/S1355617723000085>

Copyright © INS. Published by Cambridge University Press, 2023. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

errors and delayed diagnosis of MCI in women. Consistent with cognitive reserve theory (Stern et al., 1994; Stern, 2002), the “female verbal advantage” may mask the milder early memory changes emphasized in aMCI. For example, when typical Jak/Bondi cut scores for aMCI diagnosis were applied to Rey Auditory Verbal Learning Test (RAVLT) participant scores from the Alzheimer’s Disease Neuroimaging Initiative and the Mayo Clinic Study of Aging, results showed a 20% diagnostic error rate (10% false positives for men; 10% false negatives for women), but when sex-specific aMCI criteria were applied, this effect was diminished (Sundermann et al., 2019). Further, female older adults with positive AD-related markers of neurodegeneration (i.e., moderate hippocampal atrophy) remained cognitively intact on verbal memory assessments (Sundermann et al., 2016a), suggesting a delay in verbal memory decline and a sex-specific expression of cognitive reserve. Thus, when AD biomarkers are employed as the gold standard, sex-specific cut points improve diagnostic accuracy (Sundermann et al., 2016a; Sundermann et al., 2019).

Participants meeting MCI criteria that was initially developed to detect earlier indicators of cognitive impairment are not comparable to outpatient samples meeting clinical MCI criteria that more heavily weighs subjective concerns and relative preservation of independent living (Albert et al., 2011; Petersen et al., 2018). Additional research is needed to determine if sex differences in memory profiles are also present in clinical samples with MCI (Stricker et al., 2021) to guide the development of sex-specific diagnostic tools (e.g., criteria, norms). Further, other key cognitive domains in MCI (e.g., language, executive functioning/information processing speed) must be better characterized in the context of the female verbal advantage. The aim of the current project was to determine sex differences in neuropsychological profiles of individuals who were clinically diagnosed with MCI using both research and clinical criteria. Specifically, we examined severity of global and domain-specific cognition (memory, language, executive functioning/ information processing speed), and material-specific memory profiles (verbal vs. visual). We also examined age and education as potential moderators of significant differences. We hypothesized that females would exhibit evidence of (1) specific subtypes including nonamnestic or multi-domain MCI; (2) broader or more severe global or non-memory domain impairments; and (3) have a relative verbal memory advantage compared to males.

Method

Participants

The current study includes archival data from patients ($n = 349$) who underwent a neuropsychological evaluation through the Rhode Island Hospital Neuropsychology Program (RIH-NP) and were clinically diagnosed with MCI or a diagnostic equivalent. The RIH-NP archival database was developed with entry of clinical data on a secure server, REDCap (Research Electronic Data Capture; Harris et al., 2009). From 2006 to 2019, all English-speaking outpatients that met diagnostic criteria from the Diagnostic and Statistical Manual of Mental Disorders Fourth and Fifth Editions (DSM-IV; DSM-5; American Psychiatric Association, 1994, 2013, respectively) along the MCI spectrum (e.g., cognitive disorder not otherwise specified, minor neurocognitive disorder) were entered into the database. For the current study, additional data from clinical reports were extracted from the larger medical record as needed to determine eligibility and code additional variables of interest. The current study included

people over the age of 55 and excluded for a history of severe mental illness like bipolar disorder, schizophrenia (depression, post-traumatic stress disorder, and anxiety were not exclusionary unless there was a history of psychiatric hospitalization), attempted suicide, human immunodeficiency virus or other brain infections, encephalitis, moderate/ severe traumatic brain injury, stroke (included lacunes or transient ischemic attack), epilepsy (included isolated seizure or possible seizure), multiple sclerosis, brain surgery, Parkinson’s disease (included Parkinsonism/Essential Tremor), normal pressure hydrocephalus, alcohol or drug dependence, current cancer treatment, blindness, prior developmental disorder diagnosis (e.g., learning disorder, attention deficit hyperactivity disorder), or <75% of pre-specified neuropsychological measures available for analysis. Each participant’s neuropsychological evaluation, including medical history and medication lists indicated in the report, were reviewed by at least two raters (predoctoral clinical psychology intern, trained research assistant). This project was approved by the Rhode Island Hospital Institutional Review Board and conducted in accordance with the 1975 Declaration of Helsinki, as revised in 2008.

Of the 531 participants in the database, 182 were excluded due to the following: 2 for being 55 years of age or younger, 89 for neurological diagnoses/issues other than MCI and/or neurosurgical intervention, 29 for psychiatric diagnoses, 19 for substance use history, 8 for developmental disorders, 5 did not have an MCI diagnosis upon further review, 8 were missing a report or raw data in the report, 4 were not administered at least 9 tests necessary for inclusion, 2 were administered with an interpreter, and 16 met multiple exclusionary criteria related to history, mode of assessment, or data availability. The final sample used for analyses included 349 participants.

Flexible neuropsychological batteries were selected for clinical purposes by licensed clinical neuropsychologists and administered under their supervision by trained psychometricians, graduate-level practicum and clinical psychology interns, and/or postdoctoral neuropsychology fellows. To allow for inclusion of participants with slight variations in their clinical battery, we identified comparable cognitive measures to act as substitutes and widely used normative data to interpret scores within and across constructs. This allowed us to examine comparable test paradigms within broader cognitive domains without limiting our sample size or potentially introducing sample bias. Most participants had all 12 test scores/substitutes ($n = 274$), 1 missing test ($n = 57$), or 2 missing tests ($n = 12$). A z -score of -1.5 was used as a cut point for “low” or “impaired” performance to be conservative given that MCI clinical samples typically are more severe than research samples.

MCI criteria

MCI criteria included the Petersen/Winblad criteria defined as normative scores equivalent to ≤ 1.5 standard deviations on any of the 9 tests or comparable equivalents and the Jak/Bondi criteria defined as any two (or more) tests equivalent to ≤ 1 SD (z -score ≤ -1) within a domain across the 9 tests or designated comparable equivalents (Jak et al., 2009). We further defined groups into single domain amnestic (memory domain impaired alone), single domain nonamnestic (1 domain impaired other than memory), multi-domain amnestic (memory plus at least one other domain) and multi-domain nonamnestic (multiple domains impaired, not memory) based on the -1.5 SD cut point. All participants were evaluated as referrals to the RIH-NP, and therefore, subjective cognitive impairment or concern about cognitive changes was

inherently present by the patient, the patient's loved one(s) and/or a medical provider.

Neuropsychological measures

The memory domain included total learning trials 1-3 and delayed recall using the Hopkins Verbal Learning Test-Revised (HVLT-R; Benedict & Brandt, 2001) and the Brief Visuospatial Memory Test-Revised (BVMT-R; Benedict, 1997), and immediate and delayed prose passage recall using the Wechsler Memory Scale Third or Fourth Edition (Wechsler, 1997) Logical Memory subtest (LMI and LMII) or the Neuropsychological Assessment Battery (NAB; White & Stern, 2003) Story Memory subtest. Executive functioning and information processing speed was measured via graphomotor sequencing, and set-shifting using Trail Making Tests A and B (TMT-A, TMT-B; Reitan & Wolfson, 1985), lexical fluency (letters F, A, and S) using the Controlled Oral Word Association Test (COWAT; Benton et al., 1994), and visuomotor processing/coding using the Wechsler Adult Intelligence Scale-Third or Fourth Edition (Wechsler, 2009) Coding subtest or the Symbol Digit Modalities Test (SDMT; Smith, 1982). The language domain included confrontation naming using the Boston Naming Test (BNT; Goodglass et al., 1983) or the NAB Naming subtest (White & Stern, 2003), and semantic fluency using the category Animals (Benton et al., 1994).

Normative data with the demographic adjustments used for specified tests are listed under each cognitive domain in Supplemental Table I, along with the available *ns* for each measure. Normative scores were converted to *z*-scores for ease of interpretation and to create composite scores, and herein are referred to as *z*-scores. For the few participants that were administered multiple tests within one category (e.g., BNT and NAB Naming), standardized scores were averaged for the domain composites and performance within that test category was coded as "low" if at least one of the two test *z*-scores was ≤ -1.5 . The multi-test categories were renamed once converted to *z*-scores (e.g., "visuomotor coding" for Coding or SDMT) to reflect one construct variable, regardless of test. When raw scores were examined, the more frequently administered test in this sample was used. Discontinued TMT-A and TMT-B scores were coded as impaired; analyses of raw and *z*-scores included the maximum time (300") and the lowest possible *z*-score (-4.10).

Global severity

Severity was measured by calculating the proportion (%) of impaired scores (*z*-scores ≤ -1.5) on tests of interest administered overall and by creating a global cognitive composite of all tests (averaged *z*-scores).

Domain-specific patterns of impairment

Patterns of domain-specific impairment were examined using continuous (*z*-score) and categorical ($z \leq -1.5$ cut-point American Academy of Clinical Neuropsychology [AACN] descriptors; Board of Directors, 2007; Guilmette et al., 2020) variables for memory, executive functioning/information processing speed, and language. Domain-specific composites with significant sex effects were examined *post hoc* to (1) determine sex differences in raw scores of individual measures and (2) explore whether sex effects were uniform across age and education. Composites for memory ($\alpha = .76$) and executive functioning/information processing speed ($\alpha = .66$) were reliable. As expected, language

was less reliable ($\alpha = .48$), given that one of only two tests was not normally distributed (BNT). We allowed for lower reliability in the language composite as this was meant to guide targeted, data-driven analysis. The distribution of participants falling into the AACN descriptive categories by sex (Guilmette et al., 2020) was visually represented. An additional category was added to the descriptors to show greater differentiation in the lowest performing groups (*z*-scores ≤ -2) and average and higher groups were collapsed into one category given the limited above average performances in this sample.

Memory profiles

Memory profiles and sex differences in learning curves were examined via trial-by-trial recall (Trial 1, Trial 2, Trial 3, and Delayed recall) or learning curves on the HVLT-R and BVMT-R, as well as through visual representation modality-specific (HVLT-R, BVMT-R, Story) performances. Patterns of modality-specific impairments were illustrated using Venn diagrams for learning/immediate and delayed recall impairment across sex (www.meta-chart.com; Figure 1).

Statistical analyses

Analysis of variance (ANOVA) and Chi-square analyses were used to examine sex differences in participant characteristics, eligibility and main outcome variables of interest including the standard criteria for MCI (Jak/Bondi criteria and subdomains), and severity across and within domains using SPSS Statistical Package Version 25 (IBM). Significance was determined by $p < .05$ and *post hoc* analyses applied Bonferroni correction. When exploring *post hoc* analyses, age and education were entered as categorical variables (age: ≤ 65 , 66–75, 76–85, ≥ 86 ; education: ≤ 11 , 12–15, ≥ 16), but were set as continuous variables when included as covariates. Analyses of normative-corrected scores did not control for age or education, as many of the norms already control for those variables; however, all raw score analyses adjusted for age and education in the models.

Linear mixed models were conducted using the PROC MIXED procedure in SAS Enterprise Edition release 9.04.01M4P11092016 (SAS Institute Inc, Cary NC) to determine sex differences in HVLT-R and BVMT-R raw score learning curves (Trial 1, Trial 2, Trial 3, and Delayed recall). Interactive effects of sex and trial (sex*time, sex*time²) were examined in two ways for separate HVLT-R and BVMT-R models. First, the data were structured with time (i.e., trial, Level 1) nested within participants (Level 2). Random intercepts were specified at the test and participant levels, and random slopes were specified for rectilinear and quadratic time. Second, we examined the interaction between sex and a categorical trial variable to contrast performance at each trial. All models were adjusted for age and education and were fit to the data using maximum likelihood estimation, assuming incomplete data as missing at random. Unstructured covariance and Kenward-Roger degrees of freedom approximation were used.

Results

Participant characteristics are reported in Table 1 by sex and for the total sample. There were no significant sex differences in age, ethnorracial background, dementia screening raw and normative performance (Mini-Mental State Examination [MMSE]; Folstein et al., 1975; Mattis Dementia Rating Scale [DRS]; Mattis, 1988), MCI subtypes, depression severity, or proportion of multilingual

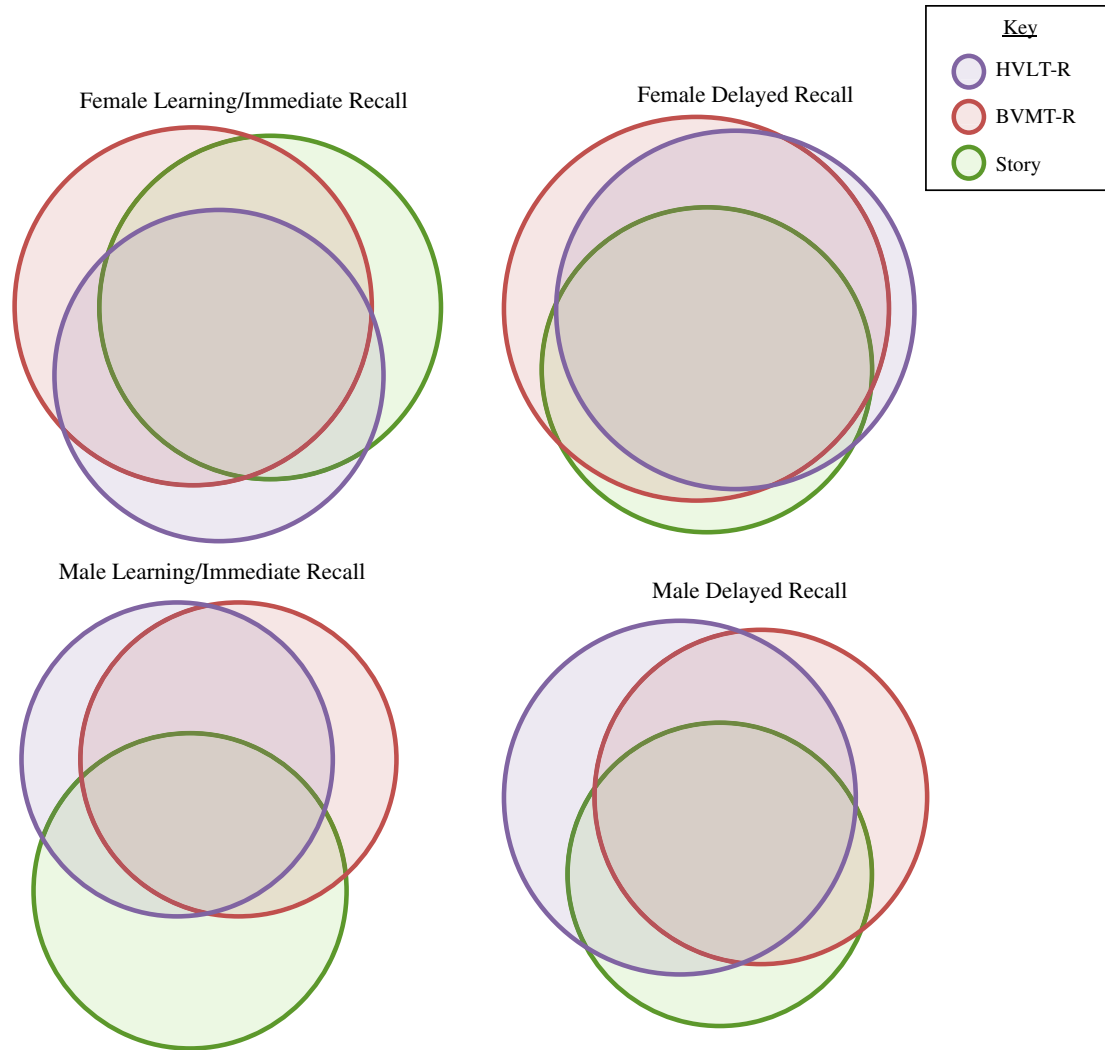


Figure 1. Venn Diagrams of Modality Specific Learning and Memory Impairment. *Note.* Female immediate recall/learning: 4.78% HVLt-R only, 9.04% BVMT-R only, 7.98% Story only, 12.23% HVLt-R+BVMT-R only, 45.21% HVLt-R+Story, 13.29% BVMT-R+Story, 37.77% HVLt-R+BVMT-R+Story; Female delay: 3.19% HVLt-R only, 9.57% BVMT-R only, 3.19% Story only, 20.21% HVLt-R+BVMT-R only, 2.13% HVLt-R+Story, 6.91% BVMT-R+Story, 50.0% HVLt-R+BVMT-R+Story; Male immediate recall/learning: 13.70% HVLt-R only, 8.22% BVMT-R only, 4.11% Story only, 16.44% HVLt-R+BVMT-R only, 10.27% HVLt-R+Story, 7.53% BVMT-R+Story, 33.56% HVLt-R+BVMT-R+Story; Male delay: 12.33% HVLt-R only, 9.59% BVMT-R only, 0.68% Story only, 28.08% HVLt-R+BVMT-R only, 3.42% HVLt-R+Story, 0.0% BVMT-R+Story, 41.78% HVLt-R+BVMT-R+Story.

Table 1. Participant characteristics

	Female (n = 199)	Male (n = 150)	Total (n = 349)
Age, M(SD)	75.3 (7.9)	73.9 (7.4)	74.7 (7.7)
<65 : 66-75 : 76-85 : ≥86**	22 : 75 : 83 : 19	17 : 75 : 46 : 12	39 : 150 : 129 : 31
White, n, %	192, 97.0%	145, 98.6%	337, 97.7%
Education, M(SD)**	13.50 (2.94)	14.91 (3.28)	14.11 (3.17)
<12 : 12 : 13-15 : ≥16**	27 : 69 : 34 : 69	15 : 28 : 24 : 83	42 : 97 : 58 : 152
Other primary language	5, 3.7%	5, 3.9%	10, 3.8%
MMSE, M(SD)	27.5 (6.2)	27.1 (2.3)	27.3 (4.9)
Global cognition	(n = 190)	(n = 143)	(n = 333)
DRS raw, M(SD)	129.4 (7.5)	130.4 (6.8)	129.8 (7.2)
DRS scaled score, M(SD)	7.6 (2.5)	7.8 (2.5)	7.7 (2.5)
MCI subtypes			
Single domain amnesic	122, 61.3%	104, 69.3%	226, 64.8%
Single domain nonamnesic	4, 2.0%	2, 1.3%	6, 1.7%
Multi-domain amnesic	60, 30.2%	37, 24.7%	97, 27.8%
Multi-domain nonamnesic	0, 0.0%	0, 0.0%	0, 0.0%
No subtype	13, 6.5%	7, 4.7%	20, 5.7%
Depression	(n = 174)	(n = 131)	(n = 301)
BDI, M(SD)	9.3(7.9)	8.2(7.1)	8.8(7.6)

Note. BDI: Beck Depression Inventory; DRS: Dementia Rating Scale; MMSE: Mini-Mental Status Examination. Jak & Bondi MCI criteria applied. Scaled scores: M = 10, SD = 3. *p < .01.

Table 2. Sex differences in research criteria: MCI criteria, subtypes, & severity

	Females (n = 195)		Males (n = 148)		Total (n = 343)	
	M	SD	M	SD	M	SD
Total low scores (%)**	45.81	20.86	41.27	19.71	43.86	20.47
Global composite (z)**	-1.36	0.60	-1.20	0.55	-1.29	0.58
Memory composite (z)	-1.86	0.80	-1.72	0.69	-1.80	0.75
Language composite (z)**	-0.76	0.93	-0.41	0.76	-0.61	0.88
Naming (z)**	-0.46	0.82	-0.04	0.67	-0.28	0.79
Category Fluency (z)*	-1.05	1.41	-0.77	1.16	-0.93	1.31
EF/PS composite (z-score)	-0.85	0.81	-0.75	0.72	-0.80	0.77
Domain-specific impairment	n	%	n	%	n	%
Memory domain	182	91.5%	141	94.0%	323	92.6%
Language domain*	19	9.5%	5	3.3%	24	6.9%
Executive domain*	103	51.8%	60	40.0%	163	46.7%

Note. Domains impaired: percentage of 3 domains where at least 2 tests are impaired (z-scores ≤ -1.5); Jak/Bondi criteria: 2 tests within a domain $< 1SD$; Language composite: average z-score for Animals category fluency and confrontation naming of either Neuropsychological Assessment Battery Naming subtest or the Boston Naming Test; Memory composite: average z-score for learning trials total and delayed recall from the Brief Visual Memory Test-Revised and the Hopkins Verbal Learning Test-Revised and immediate and delayed recall of either Neuropsychological Assessment Battery story memory subtest or the Wechsler Memory Scale Logical Memory subtest; Multi-domain amnesic: memory and at least one other domain (z-scores ≤ -1.5); Multi-domain nonamnesic: (z-scores ≤ -1.5) impaired in multiple domains that are not memory; Single domain amnesic: only memory impaired (z-scores ≤ -1.5); Single domain nonamnesic: 1 domain other than memory (z-scores ≤ -1.5); Executive functioning/Speeded processing (EF/PS): average z-score for Trail Making Tests A and B, FAS phonemic fluency and coding from either digit symbol modalities test or the Wechsler Adult Intelligence scale coding subtest; Tests impaired: percentage of tests (z-scores ≤ -1.5). * $p < .05$, ** $p < .01$.

Table 3. Sex differences in stratified age and education groups

	Females		Males		Results
	M	SD	M	SD	
Language Composite (z)					
≤ 65	-1.08	1.57	-.093	.759	$F(1,37) = 5.6, p = 0.02, \eta_p^2 = 0.13$
66-75	-.641	.994	-.50	.774	$F(1,148) = 0.96, p = 0.33, \eta_p^2 = 0.01$
76-85	-.818	.642	-.377	.786	$F(1, 127) = 11.9, p < .001, \eta_p^2 = 0.09$
≥ 86	-.589	.712	-.417	.509	$F(1,29) = .51, p = 0.48, \eta_p^2 = 0.02$
≤ 11	-1.18	1.29	-.284	.725	$F(1,40) = 6.1, p = 0.02, \eta_p^2 = 0.13$
12-15	-.758	.901	-.389	.660	$F(1,153) = 6.8, p = 0.01, \eta_p^2 = 0.04$
≥ 16	-.591	.758	-.443	.831	$F(1,150) = 1.3, p = 0.26, \eta_p^2 = 0.01$
BNT Raw					
≤ 65	49.9	7.78	53.5	5.47	$F(1,36) = 2.7, p = 0.11, \eta_p^2 = 0.07$
66-75	47.7	9.16	52.2	7.2	$F(1,145) = 10.8, p < .001, \eta_p^2 = 0.07$
76-85	43.3	10.5	49.4	8.9	$F(1,125) = 10.6, p < .001, \eta_p^2 = 0.08$
≥ 86	42.6	9.58	46.7	9.6	$F(1,29) = 1.3, p = 0.26, \eta_p^2 = 0.04$
≤ 11	38.3	11.4	47.8	6.0	$F(1,37) = 8.3, p = 0.01, \eta_p^2 = 0.18$
12-15	45.4	9.57	51.6	7.4	$F(1,153) = 16.7, p < .001, \eta_p^2 = 0.01$
≥ 16	48.5	8.46	51.2	8.53	$F(1,147) = 3.8, p = 0.05, \eta_p^2 = 0.03$
Animals Total Raw					
≤ 65	16.1	9.74	18.7	4.96	$F(1,37) = 1.0, p = 0.32, \eta_p^2 = 0.03$
66-75	13.9	6.4	13.6	4.89	$F(1,148) = 0.09, p = 0.76, \eta_p^2 = 0.00$
76-85	11.2	3.51	13.4	4.38	$F(1,126) = 9.8, p < 0.05, \eta_p^2 = 0.07$
≥ 86	11.6	4.18	11.7	2.71	$F(1,29) = .004, p = 0.95, \eta_p^2 = 0.00$
≤ 11	11.6	7.86	13.5	4.14	$F(1,39) = 0.80, p = 0.38, \eta_p^2 = 0.02$
12-15	12.5	6.27	13.7	4.64	$F(1,153) = 1.5, p = 0.26, \eta_p^2 = 0.01$
≥ 16	13.6	4.32	14.2	5.21	$F(1,150) = 0.53, p = 0.47, \eta_p^2 = 0.00$

Note. Language composite: average z-score for Animals category fluency and confrontation naming of either Neuropsychological Assessment Battery Naming subtest or the Boston Naming Test. $†p < .10$, * $p < .05$, ** $p < .01$.

individuals. Males had greater years of education on average than females, [$F(1, 342) = 18.57, p < .001$], with categorical differences indicating a greater proportion of females had 12 years of education and a greater proportion of males had at least 16 years of education [$X^2(3, 349) = 17.2, p = .001$].

Global severity

Females had a greater proportion of impaired tests at 1.5 SD, [$F(1, 347) = 4.24, p = 0.04, \eta_p^2 = 0.01$] and lower global composite z-scores [$F(1, 347) = 6.91, p = .009, \eta_p^2 = 0.02$] compared to males (Table 2).

Domain-specific patterns of impairment

A greater proportion of females met domain-specific impairment in language [$X^2(1, 349) = 5.2, p = .02, \phi = 0.12$] and executive functioning/ information processing speed [$X^2(1, 349) = 4.8, p = .03, \phi = 0.12$] than males. No sex differences were observed for the memory domain, MCI criteria or subtypes (p -values $> .05$). See Tables 2-3 for full results. ANOVAs revealed that females performed worse on language composites [$F(1, 342) = 13.97, p < .001, \eta_p^2 = 0.04$], but only marginally worse on memory composites [$F(1, 347) = 3.04, p = 0.08, \eta_p^2 = 0.01$]. There were no significant sex differences seen within executive functioning/

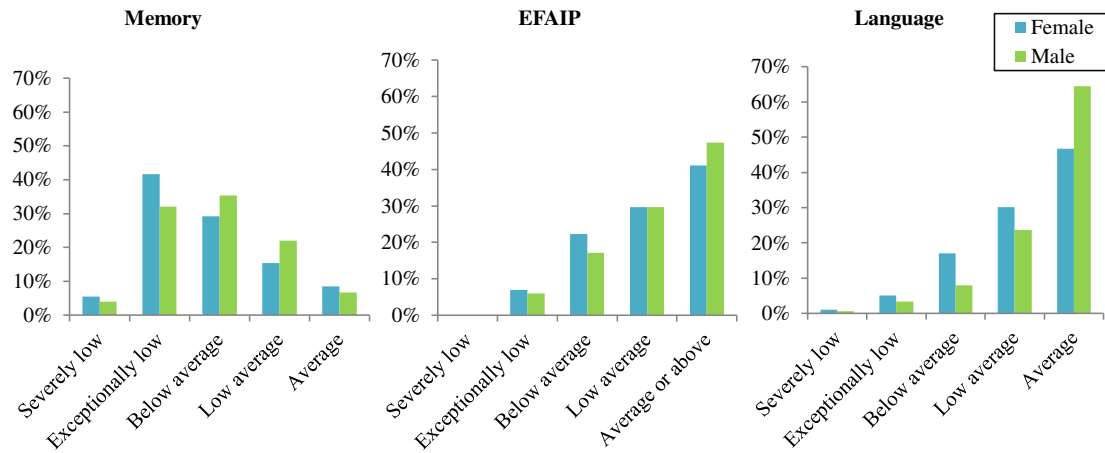
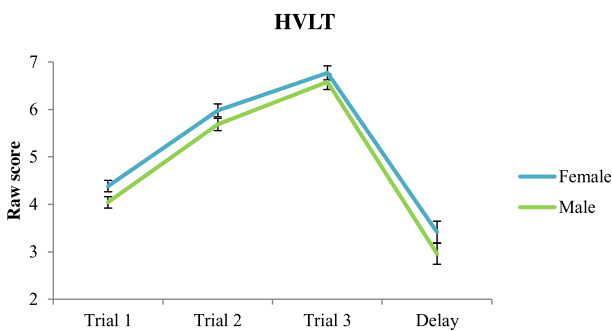
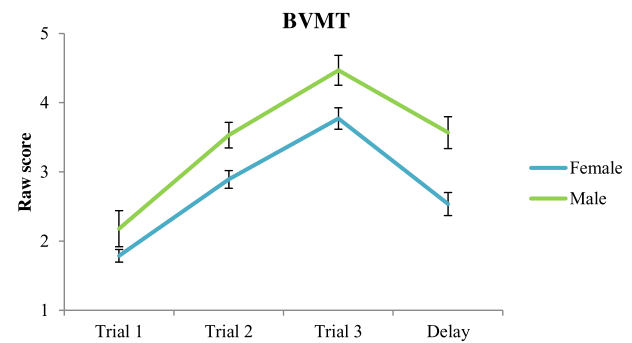


Figure 2. AACN Descriptors for composite domains.

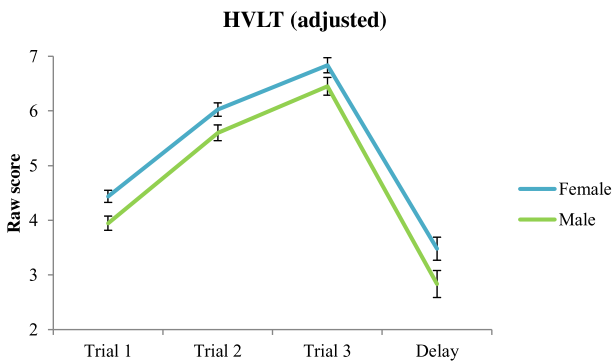
(a) Unadjusted HVLT Trial Performances by Sex.



(c) Unadjusted BVMT Trial Performances by Sex.



(b) Age and Education Adjusted HVLT Trial Performances by Sex.



(d) Age and Education Adjusted BVMT Trial Performances by Sex.

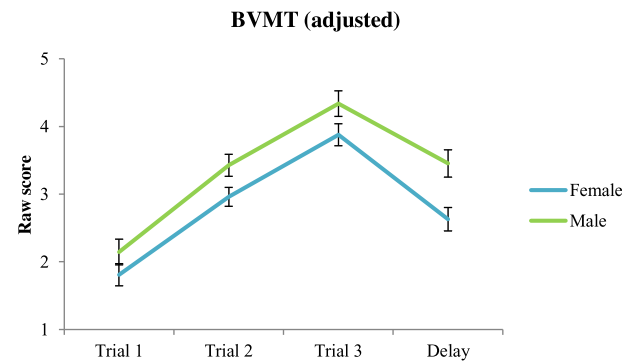


Figure 3. (a) Unadjusted HVLT Trial performances by sex. (b) Age and education adjusted HVLT trial performances by sex. (c) Unadjusted BVMT trial performances by sex. (d) Age and education adjusted BVMT trial performances by sex.

information processing speed composites [$F(1, 347) = 1.46, p = .23, \eta_p^2 = .004$]. *Post hoc* analyses indicated that performance was worse for females on both object naming, [$F(1, 346) = 26.03, p < .001, \eta_p^2 = 0.07$], and animal fluency [$F(1, 346) = 4.09, p = 0.04, \eta_p^2 = 0.012$], but only object naming survived Bonferroni correction. When controlling for age and education, there was a significant sex effect on BNT raw scores [Female $M = 45.59, SD = 9.91$, Male $M = 51.02, SD = 7.96, F(3, 339) = 17.64, p < .001, \eta_p^2 = 0.05$], but not animal fluency raw scores [Female $M = 12.79, SD = 5.92$, Male $M = 13.97, SD = 4.90, F(3, 344) = 0.77, p = .38, \eta_p^2 = .002$].

Stratified age analyses ($\leq 65, 66-75, 76-85, \geq 86$) revealed that the sex effect on language was maintained in the 76-85 year-old total composite [$F(1, 127) = 11.85, p = .001, \eta_p^2 = .09$], object naming [$F(1, 127) = 9.32, p = .003, \eta_p^2 = .07$], and animal fluency performances [$F(1, 127) = 7.86, p = .006, \eta_p^2 = 0.06$]. Stratified education ($\leq 11, 12-15, \geq 16$) analyses revealed a pattern of consistent significant or marginal effects of sex across all education groups, except for the language composite and animal fluency performance in individuals with at least college education. See Table 3 for *Ms* and *SDs*. Nonsignificant *post hoc* results are not shown. Bar graphs show the proportion of females vs. males falling

Table 4. Linear mixed models of sex differences across learning curves and memory

Variable	Recall			
	Model 1: HVLt-R		Model 1: BVMT-R	
	b (SE)	p-value	b (SE)	p-value
sex	0.81(0.40)	0.04	-0.19(0.56)	0.74
age	-0.05(0.01)	<0.001	-0.06(0.01)	<0.001
education	0.08(0.03)	0.003	0.06(0.03)	0.07
trial	6.35(0.28)	<0.001	3.32(0.34)	<0.001
trial ²	-1.32(0.06)	<0.001	-0.56(0.06)	<0.001
sex*trial	-0.33(0.40)	0.38	-0.01(0.45)	0.98
sex*trial ²	0.07(0.08)	0.39	-0.04(0.08)	0.62
Variable	Model 2: HVLt-R			
	Model 2: HVLt-R		Model 2: BVMT-R	
	b (SE)	p-value	b (SE)	p-value
sex	0.65(0.96)	<0.001	-0.85(0.25)	0.0006
age	-0.06(0.01)	<0.001	-0.08(0.01)	<0.001
education	0.08(0.03)	0.008	0.06(0.03)	0.07
sex*trial 1	0.57(0.24)	0.02	-0.17(0.25)	0.48
sex*trial 2	0.52(0.24)	0.03	-0.42(0.25)	0.09
sex*trial 3	0.42(0.24)	0.08	-0.48(0.25)	0.05
sex*trial 4	0.65(0.23)	0.0062	-0.85(0.25)	0.0006

Note. Models included raw scores from trials 1–3 and delayed recall (coded 1, 2, 3, 4). Sex was dummy coded female = 1 male = 0. Model 1 includes trial as a continuous variable and model 2 includes trial as a categorical variable. Fixed effects are reported for sex, age, and education and least square mean differences are reported for sex effects at each trial.

into descriptor categories based on overall domain and specific test z-scores (Figure 2).

Sex differences: Memory profiles

Learning curves illustrating adjusted and unadjusted sex differences on trials 1–3 and delayed recall performances are graphed in Figures 3 a–d and linear mixed model results are reported in Table 4. There were no effects of sex on learning slopes or rate of decline in retention (sex*time, sex*trial²) in models using trial as a continuous variable (Table 4). When trial was entered as a categorical variable, females recalled more HVLt-R items on Trial 1, Trial 2, and Trial 4 (Delayed Recall) than males (p 's < .05). In contrast, females recalled fewer BVMT-R figure details than males overall, with significant differences on Trial 3 and Trial 4 (Delayed Recall; p 's < .05). Figure 1 illustrates a Venn diagram of females vs. males and learning/immediate vs. delayed recall for low score profiles across the three test modalities (HVLt-R, BVMT-R, and Story).

Discussion

The primary objective of the present study was to examine sex differences in cognitive profiles of individuals diagnosed with MCI. We applied both research criteria and clinical methods to characterize our sample and interpret sex differences. Our findings show differences in language and executive functioning/information processing speed in females with MCI compared to males. Analysis of learning curves through statistical modeling as well as visual representation of descriptors and memory modalities showed additional nuances not captured by standard MCI subtypes. Despite these differences in cognitive profiles, there were no sex differences in standard MCI research criteria (i.e., Jak/Bondi MCI subtypes) and severity of global cognition via screeners. Detailed analysis and the use of conservative cut scores are strengths of this work. These results expand the growing literature examining sex differences in MCI to a clinical sample using archival neuropsychological evaluation data.

Despite having comparable global cognition (MMSE, DRS) scores, females had a greater proportion of low scores ($z \leq -1.5$ SD) than males, equivalent to 60% of females having 1 more impaired score out of 9–12 tests. This finding was driven by a proportion of impaired performances in non-memory domains that perhaps reflect greater severity not captured by cognitive screeners or existing memory paradigms. The female verbal advantage may mask the degree of impairment captured by cognitive screeners and isolated memory test scores that are often the primary diagnostic tool of determining MCI or dementia status, or the outcome variable in large-scale research. This is in line with sex-stratified analyses showing a discrepancy in presence/severity of AD biomarkers with better verbal learning and memory performance in females compared to males (Koran, et al., 2017; Sundermann et al., 2016a; Sundermann et al., 2017). The “female verbal advantage” has implications for clinical practice, particularly with respect to global cognitive screenings and referrals for neuropsychological evaluation. If cut scores on cognitive screenings reflect this verbal advantage, particularly in women who are high functioning at baseline, there could be a marked reduction in false-negative error rates in MCI diagnosis. Preclinical Alzheimer's Cognitive Composites are possible alternatives; however, further investigation is needed to establish normative data and reliable cut points (Randolph, 2020).

Summation of memory performances appeared to wash out the pattern of modality-specific memory differences observed using linear mixed models. Across the MCI spectrum, subtle verbal versus visual memory advantages were observed for females and males, respectively. According to our mixed models, this was not due to strengths in learning curves or weaknesses in rapid forgetting, but overall trends of higher or lower recall across trials. The complexity of this heterogeneity of memory performance is illustrated by the Venn diagrams (Figure 1) and the AACN descriptor tables (Figure 2). Results from our clinical sample support the need for sex adjusted normative data (Stricker et al., 2021) and sex-stratified analyses in aging research (Nebel et al., 2018). These data have implications for recruitment into clinical

trials, which often rely on brief screening or memory testing results for inclusion. Visuospatial impairments were also more pronounced in females (data not shown), though our analyses focused on tests and domains more commonly used in research and clinical samples. There is a need to examine the utility of cognitive domains examined in existing MCI criteria in heterogeneous clinical samples likely to include AD, vascular, and Lewy body dementias (Taylor et al., 2020; Tensil et al., 2018; Wolters & Ikram, 2019), as well as premorbid or comorbid conditions affecting baseline cognitive performance or rate of progression (Albai et al., 2019; Mubashir et al., 2019). While the learning curves showed small sex effects, the larger literature allude to the potential public health consequences of relative sex and/or gender differences at different stages of AD (Bloomberg et al., 2021; Shang et al., 2021; Udeh-Momoh & Watermeyer, 2021).

Contrary to prior cohort studies (Bondi et al., 2014; Edmonds et al., 2015; Edmonds et al., 2016; Eppig et al., 2017) that have found concerning false positive (31.3%–34.2%) and false-negative (7.1%) rates of MCI, the Jak/Bondi MCI criteria had minimal false-negative rates (0.9–2.0%) in our clinical sample. Given that cohort study algorithms are primarily designed to detect emergent MCI in more homogenous and high functioning research participants at enrollment, our findings are not surprising. A strength of the current study is that our sample was more likely to have multiple comorbid factors that commonly affect cognition in older adults (e.g., vascular risk, obstructive sleep apnea, anticholinergic medications), similar to epidemiological research samples with fewer exclusion criteria (Roberts et al., 2008). However, clinicians may be more conservative than research algorithms in diagnosing MCI in individuals with few low scores that could alternatively be attributed to baseline weaknesses or chance (Brooks et al., 2007). While we excluded various medical and psychiatric factors often associated with significant cognitive impairment, clinical samples may inevitably be seen at more advanced stages when subjective cognitive impairment is more concerning or beginning to cause mild functional changes (Holland et al., 2013). Likewise, it is possible that females are referred for neuropsychological testing at an older age due to their ability to compensate in daily life with verbal strengths. This is one possible explanation for evidence of more severe MCI or faster progression to AD from MCI in females than males (Holland et al., 2013; Sundermann et al., 2019).

The interpretation of MCI criteria and domain-specific low scores/severity are subject to the normative data and cut points used, a constant consideration for clinicians. While all normative datasets we included adjust for age, the characteristics of the normative samples, parameters of age corrections/stratification, and additional variables adjusted for vary by dataset. For example, the NAB [Story Immediate and Delayed Recall $n = 58$; Naming $n = 6$] and Heaton norms [TMT-A $n = 342$; TMT-B $n = 343$] additionally adjusted for sex and education. The domain-specific effect is unlikely to influence our sex-specific results given that the variance explained by sex for LMI and LMII ($r^2 = 0.001$ – 0.003), TMT-A and TMT-B ($r^2 = 0.00$ – 0.002), and the BNT ($r^2 = 0.0176$, less sensitive than the NAB Naming measure) are minimal (Karstens et al., 2021). It is possible that our conservative cut point, in combination with norms that did not adjust for sex (e.g., on tests with known sex effects such as list-learning/recall), underestimated impairment in our sample. Although sex-specific normative data have recently been published for the AVLT (Stricker et al., 2021), this is only one of many verbal memory tests employed in research and clinical settings that historically failed to control for sex differences (Gale et al., 2007). Thus, we conducted

exploratory analyses examining raw scores and stratifying them by age groups and years of education; sex effects on language measures were evident in 76–85 year-olds ($n = 127$) and individuals with less than a college education (≤ 11 , 12–15; $n = 193$). Although our sample sizes are somewhat limited in these groups, these results are likely related to evidence for increased cognitive decline and MCI between the ages of 70 to 80 (Legdeur et al., 2018) and suggest that other social factors related to cognitive reserve via education or occupational demands may mitigate the sex effect. Little to no variance ($< 1\%$) is typically attributed to sex for Animal fluency and confrontation naming tasks in similar populations (Tombaugh et al., 1999; Karstens et al., 2021). Previous studies have found slightly improved naming with increasing age, while other studies with less stringent inclusion criteria (Zec et al., 2005) showed a decline in performance after age 60 (Zec et al., 2005). Education typically accounts for considerable variance in naming/fund of knowledge type tasks (e.g., 18.6% of variance; Karstens et al., 2021; Tombaugh et al., 1999). Importantly, these data suggest that sex effects on MCI cognitive profiles may be influenced by other factors (e.g., cognitive reserve, age).

There are many strengths of this project including comprehensive hospital-based neuropsychological evaluations and our detailed approach to data coding and analysis. However, there are limitations to address. First, this data is cross-sectional. While many patients have repeat follow-ups, these data have not yet been incorporated for longitudinal purposes. Further, we attempted to use the earliest data available to avoid introducing practice effects, but some data were not available, or the evaluations were completed outside of the RIH-NP. Future follow-up work including the inclusion of other clinical and research data (e.g., PET imaging results) is needed to characterize the diagnostic outcomes of the sex differences we observed. This clinical dataset represents the patients evaluated by the RIH-NP at an academic medical center that serves predominantly English-speaking patients and is not representative of the ethnoracial makeup of the United States. More work is needed to characterize sex differences in cognitive decline, MCI, and dementia in diverse cultural and ethnic minority groups. Finally, as a clinical dataset, not all patients were given identical test batteries, potentially limiting our ability to detect impairments in some cases. Few of the test z -scores were derived from co-normed datasets, which could influence the reliability of our findings, but nevertheless reflects clinical issues we probe with this research question and dataset. However, the benefit of using this large, clinical dataset outweighs the minor variability in testing for a portion of participants. Additional review of patient charts in the future may allow for further characterization related to subjective impairments, medications, new onset depression or other psychiatric symptoms, family history, vascular risk factors, and other factors relevant to cognitive impairment. Of note, sex and gender differences cannot be entirely disentangled in this sample or others like it. Thus, the severity of these implications may vary by generation or with other elements (e.g., systemic racism, cultural-linguistic factors) that are not captured in this dataset. This speaks to a general limitation of our field in determining the diagnostic implications of key variables that influence neuropsychological test performance.

MCI is a heterogeneous diagnosis that is prognostically limited without additional strong etiological evidence (e.g., biomarker data, neuropsychiatric syndromes). Our findings suggest that examining profiles that are predominately memory-based without appropriate sex-specific corrections may limit characterization of

impairment. The brain is a highly integrated network; thus, localizing to the medial temporal lobe (e.g., entorhinal cortex, hippocampus), or other dysfunction in mild disorder/disease states is complex. Neuropsychological testing, while the gold standard in this endeavor, has its limitations. First, tests themselves and normative data are biased and infrequently revised or reinvented. The same test may not localize to brain functions similarly between individuals, especially for memory functions that may involve numerous aspects of cognition. Thus, the test itself does not determine the weighted integration of brain resources an individual employs to complete a task. However, patterns within a task, across multiple tests, and over time intervals help to provide an integrated picture of an individual's functioning and cognitive trajectory. Quantifying these differences and refining our tools for capturing memory impairment for research purposes is a challenge but may be worthwhile to better understand sex-based risk algorithms and profiles.

Taken together, females may benefit from earlier referrals for neuropsychological evaluation to capture MCI before it progresses to dementia, although development of new, more sensitive tests is also necessary. Future work following this cohort will be valuable to determine whether language and/or executive changes are possible signs of medial temporal lobe dysfunction in individuals with relative memory preservation (i.e., whether this pattern predicts the onset of dementia due to possible/probable AD). It is feasible that characteristics attributed to performance variance in healthy individuals may alter the pattern of detectable functional changes in MCI. Thus, by the time comparable and even slightly advantaged verbal impairment is observed in females, other domains may have begun to progress. These results underscore the importance of more nuanced cognitive assessment and characterization of MCI in stratified groups and highlight the need for multidisciplinary care and earlier intervention.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S1355617723000085>.

Acknowledgments. The authors do not have any sources of financial support or conflicts of interest to report.

Funding statement. None

Conflicts of interest. None

References

- Albai, O., Frandes, M., Timar, R., Roman, D., & Timar, B. (2019). Risk factors for developing dementia in type 2 diabetes mellitus patients with mild cognitive impairment. *Neuropsychiatric disease and treatment*, *15*, 167–175. <https://doi.org/10.2147/NDT.S189905>
- Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, *7*, 270–279. <https://doi.org/10.1016/j.jalz.2011.03.008>
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Arlington, Virginia: American Psychiatric Association.
- American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. Arlington, Virginia: American Psychiatric Association.
- 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/https://doi.org/10.1016/j.arr.2016.09.005>
- Benedict, R., & Brandt, J. (2001). *Hopkins verbal learning test-revised (HVLTR): Professional manual*. Lutz, FL: Psychological Assessment Resources.
- Benedict, R. H. (1997). *Brief visuospatial memory test-revised*. PAR.
- Benton, A. L., Hamsher de S., K., & Sivan, A. B. (1994). *Multilingual aphasia examination*. AJA associates.
- Bloomberg, M., Dugravot, A., Dumurgier, J., Kivimaki, M., Fayosse, A., Steptoe, A., . . . Severine, S. (2021). Sex differences and the role of education in cognitive ageing: analysis of two UK-based prospective cohort studies. *The Lancet Public Health*, *6*, E106–E115. [https://doi.org/10.1016/S2468-2667\(20\)30258-9](https://doi.org/10.1016/S2468-2667(20)30258-9)
- Board of Directors (2007) American academy of clinical neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation, *The Clinical Neuropsychologist*, *21*, 209–231. <https://doi.org/10.1080/13825580601025932>
- Bondi, M. W., Edmonds, E. C., Jak, A. J., Clark, L. R., Delano-Wood, L., McDonald, C. R., . . . Salmon, D. P. for the Alzheimer's Disease Neuroimaging, I. (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *Journal of Alzheimer's Disease*, *42*, 275–289. <https://doi.org/10.3233/JAD-140276>
- Brooks, B. L., Iverson, G. L., and White, T. (2007). Substantial risk of "Accidental MCI" in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *13*, 490–500. <https://doi.org/10.1017/S1355617707070531>
- Clark, L. R., Delano-Wood, L., Libon, D. J., McDonald, C. R., Nation, D. A., Bangen, K. J., . . . Bondi, M. W. (2013). Are empirically-derived subtypes of mild cognitive impairment consistent with conventional subtypes? *Journal of the International Neuropsychological Society*, *19*, 635–645. <https://doi.org/10.1017/S1355617713000313>
- Clark, L. R., Koscik, R. L., Nicholas, C. R., Okonkwo, O. C., Engelman, C. D., Bratzke, L. C., . . . 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*, 675–688. <https://doi.org/10.1093/arclin/acw024>
- Edmonds, E. C., Delano-Wood, L., Clark, L. R., Jak, A. J., Nation, D. A., McDonald, C. R., . . . Bondi, M. W. (2015). Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. *Alzheimer's & Dementia*, *11*, 415–424. <https://doi.org/10.1016/j.jalz.2014.03.005>
- Edmonds, E. C., Delano-Wood, L., Jak, A. J., Galasko, D. R., Salmon, D. P., & Bondi, M. W. (2016). "Missed" mild cognitive impairment: High false-negative error rate based on conventional diagnostic criteria. *Journal of Alzheimer's Disease*, *52*, 685–691. <https://doi.org/10.3233/JAD-150986>
- Eppig, J. S., Edmonds, E. C., Campbell, L., Sanderson-Cimino, M., Delano-Wood, L., & Bondi, M. W. (2017). Statistically derived subtypes and associations with cerebrospinal fluid and genetic biomarkers in mild cognitive impairment: A latent profile analysis. *J Int Neuropsychol Soc*, *23*, 564–576. <https://doi.org/10.1017/s135561771700039x>
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, *12*, 189–198. <https://www.sciencedirect.com/science/article/abs/pii/0022395675900266?via%3Dihub>
- Gale, S. D., Baxter, L., Connor, D. J., Herring, A., & Comer, J. (2007). Sex differences on the rey auditory verbal learning test and the brief visuospatial memory test-revised in the elderly: Normative data in 172 participants. *Journal of Clinical and Experimental Neuropsychology*, *29*, 561–567. <https://doi.org/10.1080/13803390600864760>
- Goodglass, H., Kaplan, E., & Weintraub, S. (1983). *Boston naming test*. Lea & Febiger Philadelphia, PA.
- Grundman, M., Petersen, R. C., Ferris, S. H., Thomas, R. G., Aisen, P. S., Bennett, D. A., . . . Thal, L. J. (2004). Mild cognitive impairment can be distinguished from alzheimer disease and normal aging for clinical trials. *Archives of Neurology*, *61*, 59–66. <https://doi.org/10.1001/archneur.61.1.59>
- Guilmette, T. J., Sweet, J. J., Hebben, N., Koltai, D., Mahone, E. M., Spiegler, B. J., . . . & Conference Participants. (2020). American academy of clinical neuropsychology consensus conference statement on uniform labeling of performance test scores. *The Clinical Neuropsychologist*, *34*, 437–453.

- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42, 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Holland, D., Desikan, R. S., Dale, A. M., & McEvoy, L. K. (2013). Higher rates of decline for women and apolipoprotein E $\epsilon 4$ carriers. *American Journal of Neuroradiology*, 34, 2287–2293. <https://doi.org/10.3174/ajnr.A3601>
- Jak, A. J., Urban, S., McCauley, A., Bangen, K. J., Delano-Wood, L., Corey-Bloom, J., . . . Bondi, M. W. (2009). Profile of hippocampal volumes and stroke risk varies by neuropsychological definition of mild cognitive impairment. *Journal of the International Neuropsychological Society : JINS*, 15, 890–897. <https://doi.org/10.1017/S1355617709090638>
- Karstens, A. J., Christianson, T. J., Lundt, E. S., Machulda, M. M., Kremers, W., Jack, C. R., Knopman, D. S., Peterson, R., & Stricker, N. H. (2021). Mayo normative studies: regression-based normative data for category fluency and boston naming test for ages 30–91 years and the varying impact of demographic variables across measures. Paper presentation at the International Neuropsychological Society Annual Meeting.
- Koran, M. E. I., Wagener, M., Hohman, T. J., & Alzheimer's Neuroimaging, I. (2017). Sex differences in the association between AD biomarkers and cognitive decline. *Brain imaging and behavior*, 11, 205–213. <https://doi.org/10.1007/s11682-016-9523-8>
- Legdeur, N., Heymans, M. W., Comijs, H. C., Huisman, M., Maier, A. B., & Visser, P. J. (2018). Age dependency of risk factors for cognitive decline. *BMC Geriatrics*, 18, 187. <https://doi.org/10.1186/s12877-018-0876-2>
- Mattis, S. (1988). *Dementia rating scale: professional manual*. Psychological Assessment Resources, Incorporated.
- Mubashir, T., Abrahamyan, L., Niazi, A., Piyasena, D., Arif, A. A., Wong, J., . . . Chung, F. (2019). The prevalence of obstructive sleep apnea in mild cognitive impairment: A systematic review. *BMC Neurology*, 19, 195. <https://doi.org/10.1186/s12883-019-1422-3>
- Nebel, R. A., Aggarwal, N. T., Barnes, L. L., Gallagher, A., Goldstein, J. M., Kantarci, K., . . . Mielke, M. M. (2018). Understanding the impact of sex and gender in Alzheimer's disease: A call to action. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 14, 1171–1183. <https://doi.org/10.1016/j.jalz.2018.04.008>
- Petersen, R. C., Lopez O., Armstrong M. J., Getchius T. S. D., Ganguli, M., Gloss, D., . . . 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, 126–135. <https://doi.org/10.1212/WNL.00000000000004826>
- Petersen, R. C., Roberts, R. O., Knopman, D. S., Geda, Y. E., Cha, R. H., Pankratz, V. S., . . . Rocca, W. A. (2010). Prevalence of mild cognitive impairment is higher in men. *The Mayo Clinic Study of Aging*, 75, 889–897. <https://doi.org/10.1212/WNL.0b013e3181f11d85>
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, 56, 303–308. <https://doi.org/10.1001/archneur.56.3.303>
- Randolph, C. (2020). Composite neurocognitive endpoints in Alzheimer's disease clinical trials: A commentary. *Alzheimer's & Dementia*, 12, e12010. <https://doi.org/10.1002/dad2.12010>
- Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (Vol. 4). Reitan Neuropsychology.
- Roberts, R. O., Geda, Y. E., Knopman, D. S., Cha, R. H., Pankratz, V. S., Boeve, B. F., . . . Petersen, R. C. (2012). The incidence of MCI differs by subtype and is higher in men. *The Mayo Clinic Study of Aging*, 78, 342–351. <https://doi.org/10.1212/WNL.0b013e3182452862>
- Roberts, R. O., Geda, Y. E., Knopman, D. S., Cha, R. H., Pankratz, V. S., Boeve, B. F., . . . Rocca, W. A. (2008). The Mayo Clinic study of aging: Design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*, 30, 58–69. <https://doi.org/10.1159/000115751>
- Rountree, S. D., Waring, S. C., Chan, W. C., Lupo, P. J., Darby, E. J., & Doody, R. S. (2007). Importance of subtle amnesic and nonamnesic deficits in mild cognitive impairment: Prognosis and conversion to dementia. *Dementia and Geriatric Cognitive Disorders*, 24, 476–482. <https://doi.org/10.1159/000110800>
- Shang, D., Wang, L., Klionsky, D. J., Cheng, H., & Zhou, R. (2021). Sex differences in autophagy-mediated diseases: toward precision medicine. *Autophagy*, 17, 1065–1076. <https://doi.org/10.1080/15548627.2020.1752511>
- Smith, A. (1982). *Symbol digit modality test manual*. Los Angeles: Western Psychological Services.
- Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, 8, 448–460.
- Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., & Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA*, 271, 1004–1010. <https://doi.org/10.1001/jama.1994.03510370056032>
- Stricker, N. H., Christianson, T. J., Lundt, E. S., Alden, E. C., Machulda, M. M., Fields, J. A., . . . 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, 211–226. <https://doi.org/10.1017/S1355617720000752>
- Sundermann, E. E., Biegon, A., Rubin, L. H., Lipton, R. B., Landau, S., & Maki, P. M. (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, 947–957. <https://doi.org/10.3233/JAD-160716>
- Sundermann, E. E., Biegon, A., Rubin, L. H., Lipton, R. B., 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, 1368–1376. <https://doi.org/10.1212/WNL.0000000000002570>
- Sundermann, E. E., Maki, P., Biegon, A., Lipton, R. B., Mielke, M. M., Machulda, M., . . . Bondi, M. W. (2019). Sex-specific norms for verbal memory tests may improve diagnostic accuracy of amnesic MCI. *Neurology*, 93, e1881–e1889. <https://doi.org/10.1212/wnl.00000000000008467>
- Sundermann, E. E., Maki, P. M., Rubin, L. H., Lipton, R. B., Landau, S., & Biegon, A. (2016). Female advantage in verbal memory. *Evidence of sex-specific cognitive reserve*, 87, 1916–1924. <https://doi.org/10.1212/wnl.00000000000003288>
- Taylor, J. P., McKeith, I. G., Burn, D. J., Boeve, B. F., Weintraub, D., Bamford, C., . . . O'Brien, J. T. (2020). New evidence on the management of Lewy body dementia. *The Lancet Neurology*, 19, 157–169. [https://doi.org/10.1016/S1474-4422\(19\)30153-X](https://doi.org/10.1016/S1474-4422(19)30153-X)
- Tensil, M., Hessler, J. B., Gutmiedl, M., Riedl, L., Grimmer, T., & Diehl-Schmid, J. (2018). Sex Differences in Neuropsychological Test Performance in Alzheimer's Disease and the Influence of the ApoE Genotype. *Alzheimer disease and associated disorders*, 32, 145–149. <https://doi.org/10.1097/wad.0000000000000229>
- 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, 167–177. [https://doi.org/10.1016/S0887-6177\(97\)00095-4](https://doi.org/10.1016/S0887-6177(97)00095-4)
- Udeh-Momoh, C. & Watermeter, T. (2021). Female specific risk factors for the development of Alzheimer's disease neuropathology and cognitive impairment: Call for a precision medicine approach. *Ageing Research Reviews*, 71, 101459. <https://doi.org/10.1016/j.arr.2021.101459>
- Wechsler, D. (1997). *WAIS-3, WMS-3: Wechsler adult intelligence scale, Wechsler memory scale: Technical manual*. Psychological Corporation.
- Wechsler, D. (2009). *Subtest Administration and Scoring. WAIS-IV: Administration and Scoring Manual*. San Antonio, TX: The Psychological Corporation, 87–93.
- White, T., & Stern, R. A. (2003). *NAB, neuropsychological assessment battery: Psychometric and technical manual*. Psychological Assessment Resources.
- Wolters, F. J., & Ikram, M. A. (2019). Epidemiology of Vascular Dementia. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 39, 1542–1549. <https://doi.org/10.1161/ATVBAHA.119.311908>
- Zec, R. F., Markwell, S. J., Burkett, N. R., Larsen, D. L. (2005). A longitudinal study of confrontation naming in the “normal” elderly. *Journal of the International Neuropsychological Society*, 11, 716–726. <https://doi.org/10.1017/S1355617705050897>