Journal of the International Neuropsychological Society (2020), **26**, 954–962

doi:10.1017/S1355617720000454

Memory Performance and Quantitative Neuroimaging Software in Mild Cognitive Impairment: A Concurrent Validity Study

Laura Glass Umfleet[1,*], Alissa M. Butts[1], Julie K. Janecek[1], Katherine Reiter[1], Mohit Agarwal[1], Benjamin L. Brett[1], Joseph J. Ryan[2], James Reuss[3], Andrew Klein[1], Anthony N. Correro II[1] and Malgorzata Franczak[1]

[1]Medical College of Wisconsin, 9200 W Wisconsin Ave, Milwaukee, WI 53226, USA
[2]School of Nutrition, Kinesiology, and Psychological Science, University of Central Missouri, Warrensburg, MO 64093, USA
[3]Prism Clinical Imaging, Inc., 890 Elm Grove Rd #209, Elm Grove, WI 53122, USA

(Received October 8, 2019; Final revision March 6, 2020; Accepted March 20, 2020; First Published Online April 28, 2020)

Abstract

Objective: This study examined the relationship between patient performance on multiple memory measures and regional brain volumes using an FDA-cleared quantitative volumetric analysis program – Neuroreader™. **Method:** Ninety-two patients diagnosed with mild cognitive impairment (MCI) by a clinical neuropsychologist completed cognitive evaluations and underwent MR Neuroreader™ within 1 year of testing. Select brain regions were correlated with three widely used memory tests. Regression analyses were conducted to determine if using more than one memory measures would better predict hippocampal *z*-scores and to explore the added value of recognition memory to prediction models. **Results:** Memory performances were most strongly correlated with hippocampal volumes than other brain regions. After controlling for encoding/Immediate Recall standard scores, statistically significant correlations emerged between Delayed Recall and hippocampal volumes (*r*s ranging from .348 to .490). Regression analysis revealed that evaluating memory performance across multiple memory measures is a better predictor of hippocampal volume than individual memory performances. Recognition memory did not add further predictive utility to regression analyses. **Conclusions:** This study provides support for use of MR Neuroreader™ hippocampal volumes as a clinically informative biomarker associated with memory performance, which is a critical diagnostic feature of MCI phenotype.

Keywords: Hippocampal volumes, Logical Memory, Visual Reproduction, HVLT-R, Alzheimer's disease, MRI

INTRODUCTION

It has been well-established that episodic memory loss is a core deficit of Alzheimer's disease (AD) that often co-occurs with hippocampal atrophy thought to be secondary to the neurodegeneration associated with the disease process. Medial temporal lobe atrophy is correlated with episodic memory deficits and both features progress with disease advancement (e.g., Marra et al., 2011; Scahill et al., 2002). Mild cognitive impairment (MCI) is a heterogenous syndrome that is often considered to be the transitional state between typical cognitive changes associated with advancing age and dementia, wherein cognitive decline is evident without functional impairment (e.g., Petersen et al., 1999). Approximately half of those diagnosed with MCI convert to dementia within 5 years, with an annual conversion rate of about 10% in clinical environments (Gauthier et al., 2006; Mitchell & Shiri-Feshki, 2009). Patients with amnestic MCI (aMCI) are more likely to convert to AD than persons with non-amnestic MCI (naMCI; e.g., Busse, Hensel, Guhne, Angermeyer, & Riedel-Heller, 2006; Fischer et al., 2007; Petersen et al., 1999; Tabert et al., 2006). Brain magnetic resonance imaging (MRI) studies show that compared to healthy controls, those with aMCI have gray matter loss in the hippocampus and several regions of the medial, anterior, and lateral temporal lobe, which is consistent with known structural involvement in Alzheimer's disease (Whitwell et al., 2008). Medial temporal volume and global volumetric data have been found to accurately classify MCI and AD at percentages ranging from 77 to 82 (Bottino et al., 2002; Du et al., 2001; Wolf et al., 2004).

The standard of care in diagnosing AD includes obtaining cognitive evaluation and brain MRI (Knopman et al., 2001), in part to rule out cognitive decline secondary to other

*Correspondence and reprint requests to: Laura Glass Umfleet, Department of Neurology, Division of Neuropsychology, Medical College of Wisconsin, 9200 W Wisconsin Ave, Milwaukee, WI 53226, USA. Phone: +1 414 955 0664, Fax: +1 414 955 0076. E-mail: lumfleet@mcw.edu

etiologies such as tumor or stroke. Hippocampal atrophy is a well-established neuroimaging biomarker for AD (e.g., Jack, Petersen, O'Brien, & Tangalos, 1992). Unfortunately, visual assessment of early hippocampal atrophy at the individual patient level by neuroradiologists has low sensitivity (i.e., 27%) for detection of possible AD (Ringman, Pope, & Salamon, 2010), thus limiting the utility of this biomarker in typical clinical settings. The relatively recent development of quantitative neuroimaging software programs offers an opportunity to quantify hippocampal volumes and detect potentially subtle patterns of atrophy earlier than solely clinical visual assessment reads by neuroradiologists. This biomarker information can be integrated into clinical care and, to some extent, applied to inform differential diagnosis, potentially even at the early stages of disease onset. Research reveals that quantitative approaches are superior to visual ratings for differentiating healthy controls from individuals with MCI and for following the progression of MCI to AD (e.g., Varon et al., 2015). Kovacevic, Rafii, and Brewer (2009) analyzed data on MCI participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (included only the subjects who had baseline MRI and both baseline and 6-month follow-up cognitive evaluations – 269 total) and provided correlational data between NeuroQuant™ (the first FDA-cleared, clinically available automated volumetric analysis tool; CorTechs Labs Inc, La Jolla, CA) structures and Wechsler Memory Scale-Revised (WMS-R) Logical Memory (LM) II and Rey Auditory Verbal Learning Test (AVLT). NeuroQuant™ hippocampal volumes were significantly correlated with LM II ($r = .22$), as well as AVLT Learning ($r = .24$) and Delayed Recall ($r = .34$). They also found that mesial temporal lobe volumes were correlated with the rate of cognitive decline. Ahdidan et al. (2016) found that manual segmentation by "master tracers" and automated segmentation via Neuroreader™ (the second FDA-cleared, clinically available automated volumetric analysis program; Brainreader Aps, Horsens, Denmark) were equivalent using healthy controls and persons with MCI or AD from the ADNI study. Ahdidan and colleagues concluded that Neuroreader™ seems comparable to Freesurfer and Neuroquant™. This conclusion was made based on reviewing previous studies using labor-intensive manual tracings, Freesurfer, and/or Neuroquant™ (e.g., Ochs, Ross, Zannoni, Abildskov, & Bigler, 2015). A separate study compared hippocampal volumes to predict MCI to AD conversion using Neuroreader™ and Neuroquant™ programs, and area under the curve values did not significantly differ between programs (Tanpitukpongse, Mazurowski, Ikhena, & Petrella, 2017). Overall, to better assess abnormal hippocampal atrophy, clinically available MRI analysis tools that compare patient MRI data to normative controls will yield a more reliable practice to discriminate between normal and abnormal aging. Both Neuroquant™ and Neuroreader™ are automated segmentation software that can be used as part of routine clinical work-up and both compare regional volumes to a normative database and provide *z*-score statistics.

The emergence of these segmentation and quantification tools, and their application to the clinical setting, is an important response to the call to integrate biomarkers into routine clinical practice to improve diagnostic accuracy. While this technology holds great potential, there are multiple challenges to integrating known biomarkers into clinical care (e.g., standardizing collection and costs and lack of easy-to-interpret data; Sperling & Johnson, 2013), and it is still relatively new, with the necessary validity of its clinical application still building. One of the promising, clinically available software programs is Neuroreader™, which provides an easy-to-read PDF file with a summary of metrics for over 30 cortical and subcortical brain regions. Some of the key metrics include raw regional volumes for the individual patient, as well as *z*-scores and percentiles based on comparison to an age- and gender-matched healthy control sample. Up until recently, volumetric studies on neurodegenerative conditions have largely relied on research protocols which vary across studies, but the standardized metrics generated from this clinically available tool allow for a greater opportunity to integrate MRI biomarker into routine practice. Examining the validity of Neuroreader™ in a clinical sample, particularly as it relates to concomitant clinical measures (i.e., neuropsychological measures of memory), is needed to determine its diagnostic value and to evaluate its translational science potential. Integrating rapidly advancing quantitative volumetric tools such as Neuroreader™ into routine clinical care will help elucidate the relationship between brain atrophy patterns and clinical phenotype of neurodegenerative diseases and progression.

Therefore, the current study set out to evaluate the relationship between memory performance (i.e., verbal and visual) and regional brain volumes in a clinical MCI sample using the FDA-cleared fully automated quantitative volumetric analysis program – Neuroreader™. This is the first known study to measure the relationship between Neuroreader™ volumes and multiple memory measures in a convenience sample of patients referred for neuropsychological evaluation as part of their clinical care. This clinical sample included patients with either aMCI or naMCI. We have two main hypotheses. First, we hypothesized that hippocampal *z*-scores would yield the most significant correlations with memory measures (i.e., word list, story, and visual memory tests) compared to other brain regions. Second, to further demonstrate the validity of Neuroreader™ hippocampal volumes, we hypothesized that using more than one memory measure would better predict hippocampal *z*-scores. It is expected that the use of two or more memory measures would yield higher correlations considering previous studies that compared different cutoffs for MCI diagnosis (e.g., Petersen/Winblad MCI criteria increase the risk of false-positive errors) and research that showed a high base rate of one abnormal neuropsychological test score in an otherwise normal profile (e.g., Brooks, Iverson, Holdnack, & Feldman, 2008; Clark et al., 2013; Edmonds et al., 2015; Jack et al., 2016; Schretlen, Testa,

https://doi.org/10.1017/S1355617720000454 Published online by Cambridge University Press

Table 1. Sample characteristics of the HVLT-R and HVLT-R plus groups

Variable	HVLT-R group $N = 92$		HVLT-R plus group $n = 56$	
	Freq.		Freq.	
Gender				
Women	46		27	
Men	46		29	
MCI subtype				
aMCI	71		44	
naMCI	21		12	
	M	*SD*	*M*	*SD*
Age	73.21	7.48	72.96	7.63
Education	15.03	2.75	15.29	2.74
Word Reading	103.89	16.07	104.76	17.22
HVLT-R Standard Scores				
Total Recall	80.62	14.62	83.25	15.79
Delayed Recall	74.51	17.74	76.56	18.99
Percent Retention	78.77	21.31	80.36	21.21
Recognition	78.53	18.59	78.43	18.81
WMS-IV LM Scaled Scores				
Immediate Recall			6.33	3.17
Delayed Recall			5.71	3.50
Recognition			*Freq.*	
<2			11	
3–9			6	
10–16			8	
17–25			4	
26–50			12	
51–75			7	
>75			8	
			M	*SD*
WMS-IV VR Scaled Scores				
Immediate Recall			7.67	3.07
Delayed Recall			6.76	2.82
Recognition			Freq.	
<2			3	
3–9			8	
10–16			5	
17–25			10	
26–50			15	
51–75			9	
>75			6	

Winicki, Pearlson, & Gordon, 2008) Lastly, we explored that added value of recognition memory to prediction models.

METHOD

Participants

The current retrospective study included 92 patients (46 women and 46 men) diagnosed with MCI by a clinical neuropsychologist at a Midwestern US academic medical center and who underwent MR Neuroreader™ within 1 year of cognitive testing ($M = 1.94$ months, $Med = 1.00$ months, $Range = 0$–12). Means for age and education were 73.21 ($SD = 7.48$) and 15.03 ($SD = 2.76$) years, respectively (see Table 1 for additional sample characteristics). Patients were evaluated either in the general neuropsychology clinic or in a specialized memory disorder clinic. Psychometric assessment is more targeted (i.e., briefer) in the specialty clinic and the battery included only one memory measure, whereas two or more memory measures were administered in the general neuropsychology clinic. Of the 92 patients, 56 were seen in the general neuropsychological clinic and were administered 3 memory measures, whereas 36 patients were seen in the specialty clinic and had only a word-list memory test administered. All 92 patients completed a word-list memory test (see materials below). Patients were excluded if they had a history of focal brain pathology (e.g., brain tumor), severe mental illness, and diagnosis of dementia.

Table 2. Brain regions included in correlational analyses

MR Neuroreader™ Region
Whole brain
Gray matter
Hippocampus
Right hippocampus
Left hippocampus
Temporal lobe
Right temporal lobe
Left temporal lobe
Amygdala
Putamen
Thalamus
Ventral diencephalon
Pallidum
Caudate
Brainstem
Frontal lobe
Parietal lobe
Occipital lobe
Cerebellum
Lateral ventricle

Procedure and Materials

All procedures were approved by the institutional review board in accordance with the Helsinki Declaration. This study is a retrospective chart review of patients who fit the criteria listed above (e.g., completed a neuropsychological evaluation and MR Neuroreader™ within 1 year of cognitive testing). Neuroreader™ is a quantitative volumetric analysis program and has an FDA-cleared normative database and processing algorithms in order to compare patient MRI data to a healthy normative sample. Over 30 cortical and subcortical brain volumes are transformed into *z*-scores. This program has been described in further detail elsewhere (see Ahdidan et al., 2016). Listed in Table 2 are the brain regions used in correlational analyses described later. As part

https://doi.org/10.1017/S1355617720000454 Published online by Cambridge University Press

of a neuropsychological evaluation, all patients were administered the Hopkins Verbal Learning Test-Revised (HVLT-R; Brandt & Benedict, 2001) and will be referred to as the *HVLT-R group (N = 92)*. The HVLT-R is a word-list memory task. Four scores were derived from this measure; Immediate Recall is the sum of 3 learning trials; Delayed Recall is the number of words freely recalled after an approximately 20–25-min delay; Percent Retention = Delayed Recall/number of words recalled on trial 2 or 3 of learning trials × 100; Recognition = total number of words correctly identified minus false positives. As part of a general neuropsychological clinic evaluation, a group of patients completed the HVLT-R, as well as the Logical Memory (story memory; LM I = total number of recalled story units on Immediate Recall; LM II = total number of story units freely recalled after approximately a 20–30 min delay; Recognition = total correctly recognized target and nontarget story details) and Visual Reproduction (a visual memory task, VR I = total number of design units freely recalled on immediate trials, VR II = total number of design units freely recalled after approximately a 20–30 min delay; Recognition = correct item selected out of 6 total options) subtests from the Wechsler Memory Scale – Fourth Edition (WMS-IV; Wechsler, 2009) according to standardized instructions. This will be referred to as the *HVLT-R plus* group ($n = 56$). We decided to conduct data analyses on the HVLT-R group as the sample size is larger, and a word-list memory measure is commonly used in published studies looking at memory functioning in patients with MCI.

Statistical Analysis

Controlling for encoding/immediate memory scores, partial correlations were calculated to show the relationship between memory tests and Neuroreader™ *z*-scores (see Table 2 for brain regions included in analyses for both the HVLT-R and HVLT plus groups). Parametric correlations (i.e., Pearson's *r*) were calculated for continuous data, and nonparametric correlations (i.e., Spearman's Rho) were conducted to examine ordinal-level data of the WMS-IV LM and VR recognition tasks. In order to control for false positives associated with multiple comparisons, a false-discovery rate (FDR) correction, the Benjamini–Hochberg procedure, was performed. Hierarchical regression analysis (i.e., manual entry) was carried out for the HVLT-R plus group to determine if additional memory measures explained significantly more variance in hippocampal *z*-scores. To control for any possible variance explained by Immediate Recall performance, entered in step 1 was immediate memory scores across all 3 tests, and HVLT-R Delayed Recall was entered for step 2 as it is a common practice to administer a word-list memory task and has been shown to be sensitive to differentiating MCI from normal aging (e.g., Rabin et al., 2009). LM II and VR II were then included as the third entry of the model. Three separate hierarchical multiple regression analyses (i.e., manual entry) were also performed to determine if recognition memory for each memory measure explained significantly more variance in bilateral hippocampal *z*-scores, after controlling for immediate and free Delayed Recall standard scores. Specifically, for each memory measure Immediate Recall, Delayed Recall, and recognition were entered into the models as steps 1, 2, and 3, respectively. Of note, variance inflation factors (VIF) were calculated in regression models and all were less than 5 for all Immediate Recall, Delayed Recall, and recognition measures. VIF is metric that reflects multicollinearity among the predictors in a regression model (i.e., the degree to which the different independent variables are linearly related to one another). Lower VIF values reflect less shared explained variance among predictors.

RESULTS

All patients had at least 10 years of education ($M = 15.03$, $SD = 2.75$) and a Word Reading subtest score of >70 standard score points ($M = 103.89$, $SD = 16.07$). Patients were diagnosed by a clinical neuropsychologist with aMCI ($n = 71$) or naMCI ($n = 21$). Presented in Table 1 are sample characteristics for the HVLT-R group ($N = 92$) and for the patients who completed two additional memory measures (HVLT plus group, $n = 56$). It should be noted that correlations between Immediate Recall scores across all measures and hippocampal *z*-scores were nonsignificant (all *p* values > .05), though the HVLT-R total Immediate Recall significantly correlated with the regional left temporal lobe *z*-score and VR I correlated with total amygdala *z*-score.

Correlational Analyses

Provided in Table 3 are the statistically significant correlations between memory scores and MR Neuroreader™ *z*-scores for the HVLT-R group. Visual inspection of this table showed that the largest correlations were between the hippocampus *z*-scores and HVLT-R Delayed Recall standard scores and HVLT-R % Retention standard scores (*r*s ranged from .297 to .366, corrected *p*-values ≤ .03). After correcting for multiple correlations, there was only one non-hippocampal structure that correlated with HVLT % Retention (frontal lobe, $r = -.302$, $p = .02$). HVLT-R Recognition standard scores were not significantly correlated with volumetric *z*-scores (*p*s > .05). Detailed in Table 4 are statistically significant findings for the HVLT-R plus group. Correlations between the HVLT-R and the hippocampus were, again, statistically significant. HVLT-R Recognition and hippocampal *z*-scores were, again, not significantly correlated. Both LM II ($r = .483$) and VR II ($r = .490$) were correlated with hippocampal *z*-scores. VR II was also significantly correlated with temporal lobe (right and left), amygdala, and lateral ventricle *z*-scores. LM II and VR II Recognition scores were not significantly correlated with any brain region, after correcting for multiple correlations. Of mention, for significant correlations between delayed memory scores and hippocampal *z*-scores, visual inspection revealed a trend with slightly stronger correlations with the right relative to left hippocampus.

https://doi.org/10.1017/S1355617720000454 Published online by Cambridge University Press

Table 3. Partial correlations (controlling for HVLT-R Total Recall standard score) between Neuroreader™ Region and HVLT-R Delayed Recall Metrics

Variable	HVLT group $N = 92$		
	Pearson *r*	*p* value	Corrected *p*
HVLT Delayed Recall SS			
Whole brain volume	−.252	.017	.06
Hippocampus	.348	.001	.01
Right hippocampus	.366	<.001	<.01
Left hippocampus	.297	.005	.03
Putamen	−.209	.048	.12
HVLT % Retention SS			
Whole brain volume	−.235	.026	.09
Hippocampus	.348	.001	.01
Right hippocampus	.358	.001	.01
Left hippocampus	.304	.004	.02
Brain stem	−.220	.037	.11
Frontal lobe	−.302	.004	.02
Cerebellum	−.252	.017	.07
HVLT Discrimination SS			
Putamen	−.241	.043	.43

Table 4. Partial correlations (controlling for HVLT-R Total Recall, Logical Memory I, or Visual Reproduction I standard scores) between Neuroreader™ Region and Delayed Recall Metrics

Variable	HVLT-R plus group $n = 56$		
	r	*p* value	Corrected *p*
HVLT Delayed Recall SS			
Hippocampus	.414	.002	.02
Right hippocampus	.430	.001	.02
Left hippocampus	.359	.008	.05
Frontal lobe	−.280	.040	.20
HVLT % Retention SS			
Hippocampus	.425	.001	.01
Right hippocampus	.441	.001	.01
Left hippocampus	.366	.006	.04
Amygdala	.271	.048	.24
HVLT Discrimination SS			
Cerebellum	−.275	.044	.75
WMS-IV LM II			
Hippocampus	.483	.001	<.01
Right hippocampus	.542	<.001	<.01
Left hippocampus	.373	.003	.04
Caudate	−.322	.017	.09
WMS-IV LM Recognition			
Right hippocampus	.323	.017	.10
Amygdala	.322	.018	.10
Occipital lobe	−.270	.048	.59
WMS-IV VR II			
Hippocampus	.490	<.001	<.01
Right hippocampus	.503	<.001	<.01
Left hippocampus	.422	.002	.01
Temporal lobe	.439	.001	.01
Right temporal lobe	.473	<.001	<.01
Left temporal lobe	.333	.014	.04
Amygdala	.407	.002	.01
Lateral ventricle	−.321	.018	.05
WMS-IV VR Recognition			
Right hippocampus	.283	.036	.17
Amygdala	.396	.003	.17
Caudate	−.347	.009	.70

Regression Analyses

Immediate Recall scores for the three memory measures were not significantly associated with hippocampal *z*-scores, $F(3, 51) = .68$, $p = .57$, $R^2 = .04$. After controlling for Immediate Recall scores, the HVLT-R Delayed Recall was associated with a significant increase in explained variance of hippocampal *z*-scores, $\Delta F(4, 50) = 10.63$ $p < .01$, $\Delta R^2 = .17$ (full model; $F(4, 50) = 3.27$, $p = .02$, $R^2 = .207$). LM II and VR II, together, significantly predicted hippocampal *z*-scores $F(6, 48) = 5.16$, $p < .001$, $R^2 = .392$ and explained additional variance in hippocampal *z*-scores, $\Delta F(2, 48) = 7.29$ $p < .01$, $\Delta R^2 = .19$.

Three separate linear regression analyses were conducted for the HVLT-R plus group to determine if recognition memory scores for the three different memory measures would significantly predict hippocampal volumes after controlling for Immediate Recall and Delayed Recall scores. While Delayed Recall and recognition independently significantly predicted hippocampal volumes, recognition memory did not account for statistically significantly more variance in hippocampal volumes beyond Delayed Recall. For all three measures, only Delayed Recall indices, but not Immediate Recall and recognition trials, uniquely explained variance in bilateral, hippocampal *z*-scores (Table 5). Of mention, reordering the entry of memory tests with recognition memory entered into the step before free Delayed Recall provided further support of the robustness of the aforementioned regression analyses as we found that free Delayed Recall explained additional variance in hippocampal volumes in all analyses.

DISCUSSION

The current study showed that spontaneous Delayed Recall scores across three different memory measures were most strongly correlated with hippocampal *z*-scores using an FDA-cleared, clinically available, quantitative MRI volumetric program (Neuroreader™) in a sample diagnosed with MCI. However, corrected correlational values for brain region and memory testing failed to reach statistical significance for recognition memory metrics. Correlational values between the right *versus* left hippocampus and memory were only modestly discrepant, and findings are considered to reflect a slight trend in the direction of higher correlational values with memory and the right hippocampus. Our regression

https://doi.org/10.1017/S1355617720000454 Published online by Cambridge University Press

Table 5. Regression analyses for each memory test

	F	R^2	p-value	ΔF	ΔR^2	p-value
HVLT-R ($N = 92$)						
Immediate Recall	2.76	0.03	0.1			
Delayed Recall	7.60	0.15	0.001	12.1	0.12	0.001
Recognition Trial	5.28	0.15	0.002	0.69	0.01	0.41
WMS-IV LM ($n = 56$)						
Immediate Recall	1.26	0.02	0.27			
Delayed Recall	8.74	0.25	0.001	15.86	0.23	<0.001
Recognition Trial	6.13	0.27	0.001	0.93	0.01	0.34
WMS-IV VR ($n = 56$)						
Immediate Recall	0.14	0.00	0.71			
Delayed Recall	8.30	0.24	0.001	16.41	0.24	<0.001
Recognition Trial	5.55	0.25	0.002	0.28	<0.01	0.6

analyses indicate that considering more than one memory measure will improve the prediction of hippocampal z-scores. However, we found that recognition memory of any of the three measures did not add additional predictive value.

It is not surprising that the most robust correlations were between the hippocampus and Delayed Recall, given that the medial temporal lobe is the primary region involved in memory consolidation and storage (e.g., Squire, Stark, & Clark, 2004). The finding that Delayed Recall across three memory measures had the highest correlational values with the right hippocampus has been reported by others studying this relationship in MCI cohorts (Barbeau et al., 2008; deToledo-Morrell et al., 2004; Momenan, Rawlings, Fong, Knutson, & Hommer, 2004), whereas left greater than right hippocampal atrophy occurs in AD compared to aMCI (Chetelat et al., 2002; Karas et al., 2004). Visual inspection of the correlations between hippocampal volumes and the three memory measures in the current study reveal higher correlation coefficients with the LM II subtest compared to those found by Kovacevic et al. (2009) using NeuroQuant™ and verbal memory measures (i.e., AVLT and WMS-R LM – strongest correlation coefficient was .34 between hippocampal volumes and AVLT Delayed Recall). The average age and education of our sample and the ADNI sample were early 70s and 15 years, respectively, but our sample is a small, clinical sample. Also, we used the WMS-IV LM subtest, which has undergone revisions since the WMS-R LM (e.g., WMS-IV includes an older adult version with one of the paragraphs being completely different from previous WMS LM subtests), and the WMS standardization samples have also changed with revisions. It is possible that our clinical sample is further along in the progression of neurodegeneration compared to the ADNI sample. Future research using Neuroreader™ is needed to further investigate and better characterize the above-discussed hippocampal asymmetry changes at the dementia stage. It is possible that the significant negative correlations between certain memory variables and non-hippocampal regions (i.e., HVLT-R % Retention and frontal lobe z-scores) reflect less atrophy in certain regions compared to the hippocampus in our predominantly aMCI cohort. This conclusion is possibly consistent with research on neuroimaging in AD which has shown that volume loss is more notable in the posterior insula, temporal lobes, and parietal lobes compared to frontal lobes (Baron et al., 2001; Busatto et al., 2003; Hirata et al., 2005; Rombouts, Barkhof, Witter, & Scheltens, 2000).

As previously mentioned, we found that adding LM II and VR II to regression analyses explained additional variance in hippocampal volumes. This is consistent with what we hypothesized and should be expected considering previous reported findings of combining more than one measure of Delayed Recall enhances the prediction of conversion from MCI to AD (e.g., Rabin et al., 2009) and smaller hippocampal volumes emerge between MCI groups with one abnormal memory score *versus* two abnormal memory scores (Jak et al., 2009). Depending on one memory measure to diagnosis MCI offers relatively limited interpretive value to reduce the risk of Type 1 errors, as it is common for normal individuals to have one or more abnormal score in a neuropsychological profile (e.g., Brooks et al., 2008). Relying on one impaired score, cognitive screeners, and questionnaires of activities of daily living is common practice in research and in clinical practice by physicians or other clinicians and could help explain the inaccuracy and instability of MCI and dementia diagnoses. The superiority of using multiple neuropsychological measures to diagnose MCI and predict progression to dementia has been demonstrated. Our findings offer further support that use of multiple psychometrically sound neuropsychological measures within one domain, when possible, yields stronger relationships with MRI biomarkers. When time allows, clinicians are encouraged to use more than one test in each cognitive domain assessed to reduce false-positive errors, increase diagnostic certainty, and possibly improve the relationship between biomarkers and clinical presentation (Bondi & Smith, 2014; Clark et al., 2013; Jak et al., 2009).

We were somewhat surprised that recognition memory was not significantly correlated with hippocampal volumes

https://doi.org/10.1017/S1355617720000454 Published online by Cambridge University Press

(corrected correlational values) or did not explain additional variance in regression analyses. There are a variety of plausible reasons including (1) selection of a heterogenous clinical sample that included patients with multidomain MCI and those with naMCI, (2) different underlying etiologies of abnormal cognitive profiles were not accounted for, and (3) failure to analyze false positive *versus* true positives. Neuropsychologists commonly interpret memory profiles by attending to the pattern of performance on immediate and delayed free recall trials as well as recognition trials. If a patient performs poorly on both Delayed Recall and recognition, this can be useful in discriminating AD from other conditions such as frontotemporal dementia and Parkinson's disease that typically benefit from recognition cues. However, a patient with poor Delayed Recall and largely normal recognition may still be on the trajectory of developing AD. Relatedly, Lange et al. (2002) showed that while multiple aspects of episodic memory (including learning trials, short- and long-Delayed Recall, recognition, and retention) decline in AD, retention performances had a faster rate of decline in the predementia phase.

The majority of our clinical sample produced variable memory profiles instead of consistent deficits in consolidation and storage across memory measures. This is not an uncommon clinical situation, and others have documented that a small percentage of patients with MCI presenting to a memory clinic are purely amnestic (e.g., Nordlund et al., 2005). We were unable to perform group-level analyses to compare pattern of memory performance (e.g., persons with consistent impairments in Delayed Recall and recognition *vs.* consistent deficits in Delayed Recall but intact recognition) across memory measures and brain atrophy patterns due to limited power across sample sizes. Longitudinal research looking at recognition tasks in predicting conversion from aMCI to AD *versus* non-AD dementia may be useful.

Additional study limitations include the use of a convenience sample with administration of a clinical test battery that was flexible and yielded a variety of standard scores available for analysis. However, this could also be considered a strength as our sample is more representative of persons who present to clinic and a flexible-battery/hypothesis-driven assessment approach is common practice in clinical neuropsychology. It would also have been informative to conduct separate analyses for patients with naMCI and aMCI. These entities have different clinical presentations and may also produce different MRI findings. However, this was not possible in the present investigation due to small sample sizes. Limitations of Neuroreader™ include currently providing *z*-scores for only regional volumes (e.g., temporal lobe, hippocampus, frontal lobe, etc.) and not for atrophy in specific gyri or other mesial-temporal structures (e.g., rhinal cortices), which limits clinical utility for aiding in the diagnosis of MCI – this is exemplified by a recent study that analyzed MRI data with FreeSurfer v6.0 of patients with MCI compared to controls and found that a model comprised of regions including the superior and middle temporal cortices, superior temporal sulcus, subiculum, insula, and others had a group classification accuracy of 93.75% (DeVivo et al., 2019). Nevertheless, considering that hippocampal atrophy can emerge years before clinical presentation of AD (e.g., Chetelat et al., 2005), using MR Neuroreader™ hippocampal *z*-scores to help determine risk of conversion from normal to MCI and MCI to AD would provide valuable clinical information. Exploring the diagnostic utility of MR Neuroreader™ hippocampal and other regional brain volume *z*-scores in neurodegenerative conditions due to different etiologies is also recommended. Overall, while hippocampal atrophy in aMCI and AD has been well-established, this study is novel in that it examined the validity of a clinically available MRI automated analysis program in a patient sample diagnosed with MCI who completed widely used memory measures as part of a larger neuropsychological battery. This MRI analysis program could be a valuable and practical adjunctive tool for clinicians who provide neurodiagnostic services to older populations. The results of this study contribute to clinical translational science as it evaluated MRI biomarker data collected as part of clinical care and compared these data to commonly administered memory measures as part of a clinical neuropsychological evaluation. Sperling and Johnson (2013) discussed the need for improving integration of established biomarkers that are easy to interpret into clinical care. Clinical integration of automated MRI software programs that provide normative comparison data for easy interpretation offers a needed standardized analysis approach, though further refinement of these programs will be needed as technology continues to progress.

ACKNOWLEDGMENTS

No funding sources were used to conduct this study.

CONFLICT OF INTEREST

Authors of this manuscript have no potential conflicts.

References

Ahdidan, J., Raji, C.A., DeYoe, E.A., Mathis, J., Noe, K., Rimestad, J., . . . Lopez, O. (2016). Quantitative neuroimaging software for clinical assessment of hippocampal volumes on MR Imaging. *Journal of Alzheimer's Disease*, *49*, 723–732.

Barbeau, E.J., Ranjeva, J.P., Didic, M., Confort-Gouny, S., Felician, O., Soulier, E., . . . Poncet, M. (2008). Profile of memory impairment and gray matter loss in amnestic mild cognitive impairment. *Neuropsychologia*, *46*, 1009–1019.

Baron, J., Chetelat, G., Desgranges, B., Perchey, G., Landeau, B., de la Sayette, V., & Eustache, F. (2001). In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. *Neuroimage*, *14*, 298–309.

Brandt, J. & Benedict, R.H.B. (2001). *Hopkins Verbal Learning Test-Revised*. Lutz, FL: Psychological Assessment Resources.

Bondi, M.W. & Smith, G.E. (2014). Mild cognitive impairment: A concept and diagnostic entity in need of input from neuropsychology. *Journal of the International Neuropsychological Society*, *20*, 129–134.

https://doi.org/10.1017/S1355617720000454 Published online by Cambridge University Press

Bottino, C.M, Castro, C.C., Gomes, R.L., Buchpiguel, C.A., Marchetti, R.L., & Neto, M.R. (2002). Volumetric MRI measurements can differentiate Alzheimer's disease, mild cognitive impairment, and normal aging. *International Psychogeriatrics*, *14*, 59–72.

Brooks, B.L., Iverson, G.L., Holdnack, J.A., & Feldman, H.H. (2008). Potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International Neuropsychological Society*, *14*, 463–478.

Busatto, G., Garrido, G., Almeida, O., Castro, C., Camargo, C., Cid, C., . . . Bottino, C.M. (2003). A voxel-based morphometry study of temporal lobe gray matter reductions in Alzheimer's disease. *Neurobiology of Aging*, *24*, 221–231.

Busse, A., Hensel, A., Guhne, U, Angermeyer, M.C., & Riedel-Heller, S.G. (2006). Mild cognitive impairment: Long-term course of four clinical subtypes. *Neurology*, *67*, 2176–2785.

Chetelat, G., Desgranges, B., De La Sayette, V., Viader, F., Eustache, F., & Baron, J-C. (2002). Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. *Neuroreport*, *13*, 1939–1943.

Chetelat, G., Landeau, B., Eustache, F., Mezenge, F., Viader, F., de la Sayette, V., . . . Baron, J. C. (2005). Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *Neuroimage*, *27*, 934–946.

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.

deToledo-Morrell, L., Stoub, T., Bulgakova, M., Wilson, R., Bennett, D., Leurgans, S., . . . Turner, D.A. (2004). MRI-derived entorhinal volume is a good predictor of conversion from MCI to AD. *Neurobiology of Aging*, *25*, 1197–1203.

DeVivo, R., Zajac, L., Mian, A., Cervantes-Arslanian, A., Steinberg, E., Alosco, M., . . . Killany, R. (2019). Differentiating between healthy control participants and those with mild cognitive impairment using volumetric MRI data. *JINS*, *25*, 800–810.

Du, A.T., Schuff, N., Amend, D., Lasskso, M.P., Hsu, Y.Y., Jagust, W.J., . . . Weiner, M.W. (2001). Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *71*, 441–447.

Edmonds, E.C., Delano-Wood, L., Clark, L.R., Jak, A.J., Nation, D.A., McDonald, C.R. . . . for the Alzheimer's Disease Neuroimaging Initiative. (2015). Susceptibility of the conventional criteria for MCI to false positive diagnostic errors. *Alzheimer's & Dementia*, *11*, 415–424.

Fischer, P., Jungwirth, S., Zehetmayer, S., Weissgram, S., Hoeningschnabl, S., Gelpi, E., & Tragl, K. H. (2007). Conversion from subtypes of mild cognitive impairment to Alzheimer's dementia. *Neurology*, *68*, 288–291.

Gauthier S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., . . . Winblad, B. (2006). Mild cognitive impairment. *The Lancet*, *367*, 1262–1270.

Hirata, Y., Matsuda, H., Nemoto, K., Ohnishi, R., Hirao, K., Yamashita, F., . . . Samejima, H. (2005). Voxel-based morphometry to discriminate early Alzheimer's disease from controls. *Neuroscience Letters*, *382*, 269–274.

Jack, C.R., Petersen, R.C., O'Brien, P.C., & Tangalos, E.G. (1992). MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease, *Neurology*, *42*, 183–188.

Jack, C.R., Therneau, R.M., Wiste, H.J., Weigand, S.D., Knopman, D.S., Lowe, V.J . . . .Petersen, R.C. (2016). Transition rates between amyloid and neurodegeneration biomarker states and to dementia: A population-based, longitudinal cohort study. *Lancet Neurology*, *15*, 56–64.

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*, *2*, 1–8.

Karas, G.B., Scheltens, P., Rombouts, S.A., Visser, P.J., van Schijndel, R.A., Fox, N.C., & Barkhof, N.C. (2004). Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *Neuroimage*, *23*, 708–716.

Knopman, D.S., DeKosky, S.T., Cummings, J.L., Chui, H., Corey-Bloom, J., Relkin, N., . . . Stevens, J.C. (2001). Practice parameter: Diagnosis of dementia (an evidence based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, *56*, 1143–1153.

Kovacevic, S., Rafii, M.S., Brewer, J.B., & the Alzheimer's Disease Neuroimaging Initiative. (2009). High-throughput, fully-automated volumetry for prediction of MMSE and CDR decline in mild cognitive impairment. *Alzheimer's Disease & Associated Disorders*, *23*, 139–145.

Lange, K.L., Bondi, M.W., Salmon, D.P., Galasko, D., Delis, D.C., Thomas, R.G., & Thal, L. J. (2002). Decline in verbal memory during preclinical Alzheimer's disease: Examination of the effect of APOE genotype. *Journal of the International Neuropsychological Society*, *8*, 943–955.

Marra, C., Ferraccioli, M., Vita, M.G., Quaranta, D., & Gainotti, G. (2011). Patterns of cognitive decline and rates of conversion to dementia in patients with degenerative and vascular forms of MCI. *Current Alzheimer Research*, *8*, 24–31.

Mitchell, A.J. & Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia – meta-analysis of 41 robust inception cohort studies. *Acta Psychiatrica Scandinavica*, *119*, 252–265.

Momenan, R., Rawlings, R., Fong, G., Knutson, B., & Hommer, D. (2004). Voxel-based homogeneity probability maps of gray matter in groups: Assessing the reliability of functional effects. *Neuroimage*, *21*, 965–972.

Nordlund, A., Rolstad, S., Hellstrom, P., Sjogren, M., Hansen, S., & Wallin, A. (2005). The Goteborg MCI study: Mild cognitive impairment is a heterogeneous condition. *Journal of Neurology, Neurosurgery, and Psychiatry*, *76*, 1485–1490.

Ochs, A.L., Ross, D.E., Zannoni, M.D., Abildskov, T.J., & Bigler, E.D. (2015). Comparison of automated brain volume measures obtained with NeuroQuant(R) and FreeSurfer. *Journal of Neuroimaging*, *25*, 721–727.

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.

Rabin, L., Paré, N., Saykin, A., Brown, M.J., Wishart, H.A., Flashman, L.A., & Santulli, R. B. (2009). Differential memory test sensitivity for diagnosing amnestic mild cognitive impairment and predicting conversion to Alzheimer's disease. *Aging, Neuropsychology, and Cognition*, *16*, 357–376.

Ringman, J.M., Pope, W., & Salamon, N. (2010). Insensitivity of visual assessment of hippocampal atrophy in familial Alzheimer's disease. *Journal of Neurology*, *257*, 839–842.

Rombouts, S., Barkhof, F., Witter, M., & Scheltens, P. (2000). Unbiased whole-brain analysis of gray matter in Alzheimer's disease. *Neuroscience Letters*, *285*, 231–233.

Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N., & Fox, N.C. (2002). Mapping the evolution of regional atrophy in Alzheimer's disease: Unbiased analysis of fluid – registered serial MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 4703–4707.

Schretlen, D.J., Testa, S.M., Winicki, J.M., Pearlson, G.D., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society*, *14*, 436–445.

Sperling, R. & Johnson, K. (2013). Biomarkers of Alzheimer's disease: Current and future applications to diagnostic criteria. *Continuum*, *19*, 325–338.

Squire, L.R., Stark, C.E.L., & Clark, R.E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, *27*, 279–306.

Tabert, M.H., Manly, J.J., Liu, X., Pelton, G.H., Rosenblum, S., Jacobs, M., . . . Devanand, D. P. (2006). Neuropsychological prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Archives of General Psychiatry* , *63*, 913–924.

Tanpitukpongse, T.P., Mazurowski, M.A., Ikhena, J., Petrella, J.R., & for the Alzheimer's Disease Neuroimaging Initiative (2017). Predictive utility of marketed volumetric software tools in subjects at risk for Alzheimer's disease: Do regions outside the hippocampus matter? *AJNR*, *38*, 546–552.

Varon, D., Barker, W., Loewenstein, D., Greig, M., Bohorquez, A., Santos, I., . . . Alzhiemer's Disease Neuroimaging Initiative. (2015). Visual rating and volumetric measurement of medial temporal atrophy in the Alzhiemer's Disease Neuroimaging Initiative (ADNI) cohort: Baseline diagnosis and the prediction of MCI outcome. *International Journal of Geriatric Psychiatry*, *30*, 192–200.

Wechsler, D. (2009). *Wechsler Memory Scale–Fourth Edition* (4th ed.). San Antonio, TX: Pearson.

Whitwell, J.L., Shiung, M.M., Przybelski, S.A., Weigand, S.D., Knopman, D.S., Boeve, B.F., . . . Jack, C.R. (2008). MRI patterns of atrophy associated with progression to AD in amnestic mild cognitive impairment. *Neurology*, *70*, 512–520.

Wolf, H., Hensel, A., Kruggel, F., Riedel-Heller, S.G., Arendt, T., Wahlund, L.-O., & Gertz, H.-J. (2004). Structural correlates of mild cognitive impairment. *Neurobiology of Aging*, *25*, 913–924.