J Prev Alz Dis 2023;2(10):236-243
Published online January 28, 2023, http://dx.doi.org/10.14283/jpad.2023.19

Original Research

Validity of Normative Volumetric Estimates from Open Access Software in Amnestic Mild Cognitive Impairment

S. Fountain-Zaragoza[1], O. Horn[2], K.E. Thorn[2], A.Z. Kraal[3], A. Benitez[2]

1. Department of Health Sciences and Research, Medical University of South Carolina, Charleston, SC, USA; 2. Department of Neurology, Medical University of South Carolina, Charleston, SC, USA; 3. Department of Neurology and Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Irving Medical Center, New York, NY, USA

Corresponding Author: Andreana Benitez, 96 Jonathan Lucas St, MSC 606, Charleston, SC, 29425, USA, benitez@musc.edu; Tel (843) 876-2479

Abstract

BACKGROUND: Neurodegeneration in Alzheimer's disease (AD) is typically assessed through brain MRI. Although proprietary software can provide normative estimates of regional atrophy, such tools can be cost-prohibitive for research settings. Free software for generating normative estimates has recently been released but has yet to be validated in the context of amnestic mild cognitive impairment (aMCI).

OBJECTIVES: Determine whether normative morphometric estimates generated from open-source software replicate established patterns of neurodegeneration in aMCI, and whether these metrics correlate with episodic memory performance.

DESIGN: Observational study of brain MRI and cognition in aging and aMCI with two identical study visits occurring approximately 1.2 years apart.

SETTING: Participants were recruited from the local community and outpatient clinical settings.

PARTICIPANTS: Adults ages 60-85 with aMCI (n = 25) and cognitively normal controls (CN; n = 74). A subset returned for follow-up (aMCI n = 11, CN n = 52).

MEASUREMENTS: Participants completed brain MRI and two neuropsychological tests of verbal episodic memory. FreeSurfer v6.0 and Normative Morphometry Image Statistics were used to generate normative morphometric estimates for AD-relevant regions (hippocampus, parahippocampus, entorhinal cortex, amygdala) and control regions (cuneus, lingual gyrus, pericalcarine gyrus), adjusting for age, sex, head size, scanner manufacturer, and field strength. We tested for baseline group differences in ROI volumes and memory and assessed their within-group associations. We also evaluated changes in ROI volumes over time and tested whether these changes corresponded to declines in memory.

RESULTS: At baseline, the aMCI group exhibited poorer memory and smaller volumes in AD-relevant regions than the CN group. There were no group differences in control region volumes. Memory was associated with volumes in AD-relevant regions in the aMCI group only. The aMCI group exhibited greater declines than the CN group in hippocampal volume (17% vs. 8% annual decline) and entorhinal volume (54% vs. 5% annual decline). Decrease in hippocampal volume was marginally associated with decline in memory for the aMCI group.

CONCLUSIONS: Normative morphometric values generated from freely available software demonstrated expected patterns of group differences in AD-related volumes and associations with memory. Significant effects were localized to AD-relevant brain regions and only occurred in the aMCI group. These findings support the validity of these free tools as reliable and cost-effective alternatives to proprietary software for use in research settings.

Key words: Neurodegeneration, mild cognitive impairment, memory.

Introduction

Neurodegeneration is a hallmark of Alzheimer's disease (AD; 1) that is detectable in the prodromal stage called mild cognitive impairment (MCI). Individuals with the amnestic subtype of MCI (aMCI) are more likely to progress to AD, at rates of 10-15% annually (2, 3), than those with non-amnestic MCI (4), and demonstrate preferential atrophy in medial temporal lobe structures (i.e., hippocampus, entorhinal cortex, parahippocampus, amygdala). Medial temporal lobe atrophy is associated with poorer memory performance (5–7) and portends risk for future memory decline (8, 9), loss of functional status, and conversion to AD (10–12).

Morphometric estimates are typically obtained from brain MRI using automated segmentation software. Existing proprietary software offer streamlined pipelines for producing regional volumetric estimates that are adjusted for important confounding variables, including demographic information based on normative samples. The two most commonly used software packages are NeuroQuant®, which was the first FDA-approved automated segmentation software and has been widely used primarily in clinical settings since 2007, and Neuroreader®, which has been FDA approved since 2015 and works similarly to NeuroQuant®. While useful, the proprietary nature of these packages presents several barriers to use in research settings. In addition to often being cost-prohibitive, barring access for many researchers, important information about the processing pipelines (e.g., details about the normative samples and how normative data were generated), is not publicly available, which poses challenges for scientific reporting and reproducibility.

Received October 24, 2022
Accepted for publication December 18, 2022

However, these barriers may be circumvented by the increasing availability of free tools for generating norm-adjusted morphometric estimates. The open-source software package FreeSurfer (13) has been shown to perform comparably in generating volumetric estimates to proprietary software, demonstrating good to excellent inter-method reliability with NeuroQuant® (14–17). An important recent advance is the release of the Normative Morphometry Image Statistics (NOMIS) tool for use with FreeSurfer outputs (18), which produces normative scores adjusted for subject characteristics (e.g., age, sex, head size) and scanner/image information (e.g., resolution, contrast-to-noise ratio) based on a large comparison sample of nearly 7,000 healthy adults. An initial validation study of this tool in patients with AD found expected volumetric differences from a healthy sample, characterized by lower volumes in frontal, temporal (most prominently hippocampus and entorhinal cortex), and parietal regions as well as enlarged ventricles (18). As NOMIS has only recently been released, no studies have yet validated its use in aMCI.

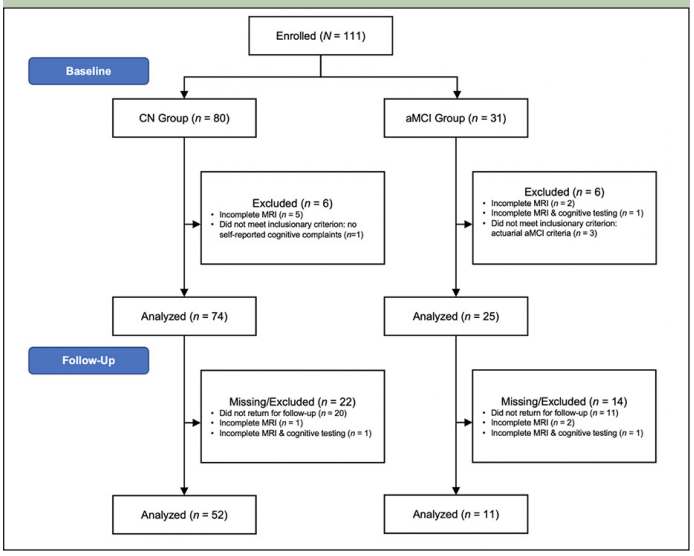
Thus, the goal of this study was to validate the normative volumetric values produced by the freely available software, FreeSurfer and NOMIS, in the context of research on aMCI. Our primary aims were to 1) compare volumes of AD-relevant regions (hippocampus, parahippocampus, entorhinal cortex, amygdala) and control regions (cuneus, lingual gyrus, pericalcarine gyrus) between patients with aMCI and cognitively normal controls (CN) and 2) examine the associations between regional volumes and episodic memory in each group. We expected to replicate the well-established pattern of greater atrophy in AD-relevant regions, but not in control regions, in the aMCI group, but not in the CN group. We also expected that lower volumes in AD-relevant regions would be associated with poorer episodic memory in the aMCI group. Given the availability of some longitudinal data in this study, our secondary aim was to explore how these volumetric estimates change over time in a sub-sample of participants, and to relate these changes with change in episodic memory. This study will inform the use of FreeSurfer and NOMIS for research purposes, but our results will not apply to clinical practice as neither software was created for clinical or commercial use as a medical device.

Methods

Participants

Adults ages 60-85 years were recruited from the local community and outpatient clinical settings for an observational study of MRI and neuropsychological measures in aging and MCI. Participation involved two identical study visits over approximately 1-2 years; baseline visits were conducted between 2013 and 2018 and follow-up visits were conducted between 2015 and 2020. Participant exclusion/inclusion across timepoints is detailed in the flow chart in Figure 1. It is important to note that the low retention rate at follow-up is largely attributable to the fact that this study was not initially designed to be longitudinal; one source of funding initiated the baseline study focused on CN, but upon receipt of another source of funding the study was continued and more CN were recruited along with aMCI patients; both groups returned for follow-up if they were able.

Figure 1. Flow Chart of Participant Exclusion Across Timepoints



CN = cognitively normal control group; aMCI = amnestic mild cognitive impairment group

All enrolled participants met the following eligibility criteria (described previously; 19): English as a first/primary language; no prior diagnosis of a significant neurologic disease (e.g., stroke, epilepsy, dementia), serious mental illness (e.g., schizophrenia, bipolar disorder), or other poorly controlled or intractable disease (apart from aMCI) with known systematic effects on cognitive function (e.g., untreated diabetes, heart or thyroid disease, cancer); no contraindications for MRI scanning and a fasting blood draw; and sufficient visual and hearing acuity to undergo neuropsychological testing. Participants were excluded from this analysis if they had incomplete neuroimaging or neuropsychological data. At the time of enrollment, participants either had 1) no subjective cognitive difficulties or cognitive diagnoses or 2) a clinical diagnosis of mild neurocognitive disorder/MCI. All individuals in the CN group were confirmed as being cognitively unimpaired based on a normed Montreal Cognitive Assessment (20) score of $z > -1.0$ (21). Inclusion in the aMCI group for these analyses required that those with a clinical MCI diagnosis also meet actuarial neuropsychological criteria (22) for aMCI on the testing they completed as part of their participation

in this study. We chose to use these actuarial criteria to define the aMCI group as this approach has been found to minimize false positive MCI diagnoses (23) and yield greater diagnostic specificity (24) and stability over time compared to conventional criteria (25). This required that they had either 1) ≥ 2 impaired scores (i.e., ≤ -1 SD) on memory tests or 2) ≥ 1 impaired score (i.e., ≤ -1 SD) in three separate cognitive domains, at least one being a memory test, using demographically-corrected normed scores from the neuropsychological test battery described below. All participants provided written informed consent, and this study was approved by the Medical University of South Carolina institutional review board.

The final sample included 99 individuals at baseline (aMCI n = 25, CN n = 74) and 63 who returned for follow-up (aMCI n = 11, CN n = 52); median follow-up duration was 1.2 years (range: 0.86 – 2.70 years). At enrollment, the sample was composed of older adults (M_{age} = 70.09 years, SD = 6.78) roughly half of whom were female (54.5%; 54 females), predominantly White (91.9%; 91 White, 8 Black/African American), and had a college education on average (M_{edu} = 16.35 years, SD = 2.37, range: 12 – 20 years). All participants identified as non-Hispanic or Latino and none identified as American Indian/Alaskan Native, Asian, Native Hawaiian/Pacific Islander, or Other. Table 1 provides descriptive statistics for demographic variables in the aMCI and CN groups separately for each timepoint.

Measures

Neuropsychological Test Battery

At baseline and follow-up, participants were administered version 2 of the Alzheimer Disease Centers' Neuropsychological Test Battery in the Uniform Data Set (details of battery published previously; 26). Given that this study focused specifically on individuals with amnestic impairments, we used scores from the following two tests of verbal episodic memory:

Rey Auditory Verbal Learning Test (RAVLT)

The RAVLT is a list-learning test in which participants were presented with 15 words and the test comprises 5 learning trials, an immediate recall trial and a delayed recall trial (20 min) followed by a recognition trial. Raw scores for each condition were adjusted for age using published norms (27).

Logical Memory

The Logical Memory test from the Wechsler Memory Scale-Revised (WMS-R; 28) is a story-learning task. The first story has two immediate recall trials and the second story has one immediate recall trial. Both stories have a single 20-minute delayed recall trial followed by a recognition trial. Raw scores for each condition were adjusted for age, education, and sex using established norms (29).

The episodic memory variable used in the analyses was a Z-score composite calculated by averaging the norm-adjusted scores across the immediate and delayed recall conditions of both the RAVLT and Logical Memory tests.

MRI Data Acquisition and Analysis

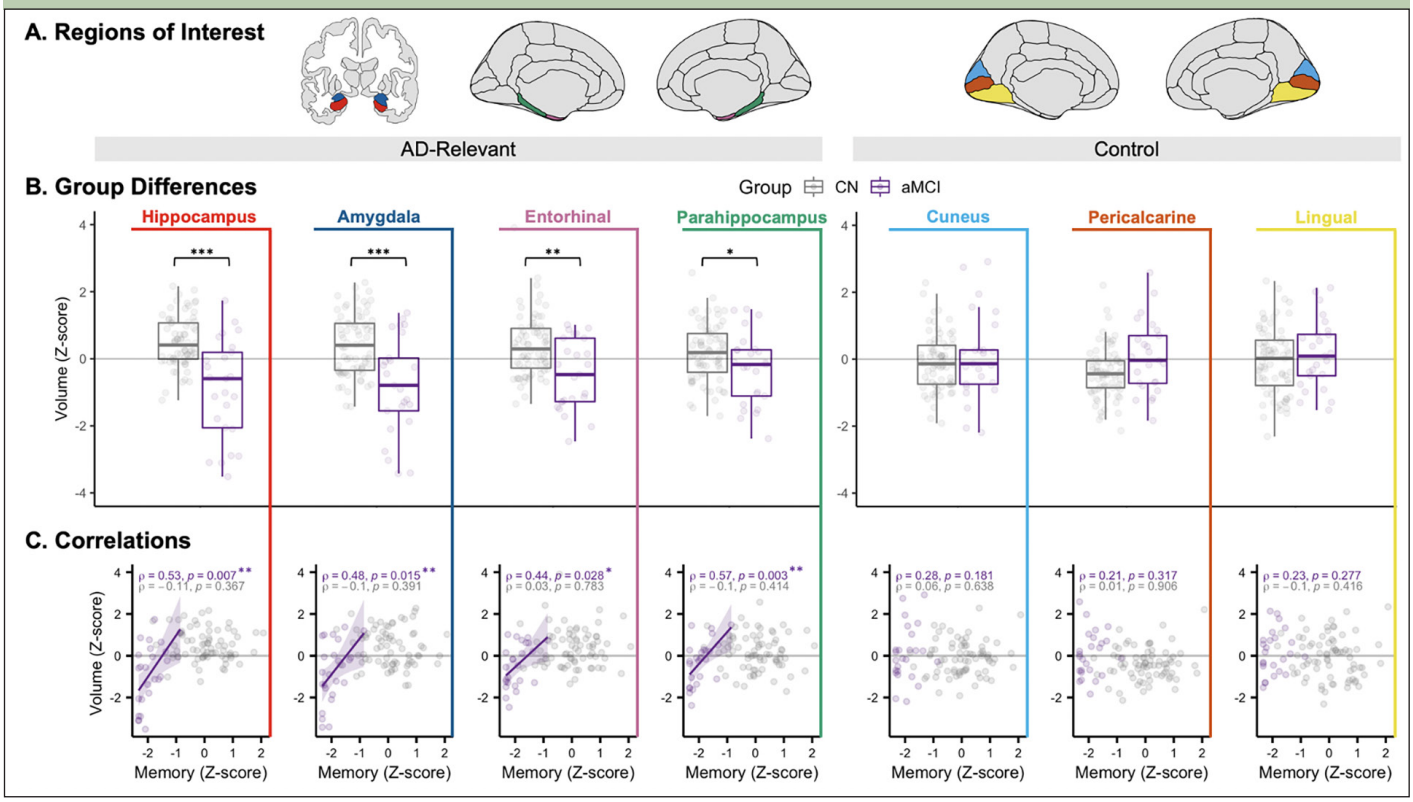
MRI scans were collected at baseline and follow-up on a 3T TIM Trio MR system (Siemens Medical Solutions, Erlangen, Germany; 19) or a 3T Prismafit system (Siemens

Table 1. Baseline Sample Demographics, Characteristics, and Tests of Group Differences (N = 99)

			Between-Group Differences at Baseline	
	CN (n = 74)	aMCI (n = 25)		
	Median (IQR)	Median (IQR)	Stat	p
Age (years)	68.11 (9.02)	74.22 (13.58)	U = 624.0	.016*
Education (years)	16 (4)	16 (2)	U = 918.5	.960
Sex (no. and % female)	44 (0.59)	10 (0.4)	X^2 = 2.12	.145
Race (no. and %)			X^2 = 1.67	.197
White	66 (0.89)	25 (1)	-	-
Black	8 (0.11)	0 (0)	-	-
Episodic Memory (Z-Score)	0.07 (1.15)	-2 (0.5)	U = 1826	<.001***
ROI Volumes (Z-scores)				
Hippocampus	0.41 (1.08)	-0.59 (2.25)	U = 1444	<.001***
Amygdala	0.4 (1.4)	-0.79 (1.57)	U = 1373	<.001***
Entorhinal	0.29 (1.19)	-0.47 (1.89)	U = 1310	.002**
Parahippocampus	0.19 (1.16)	-0.17 (1.38)	U = 1179	.041*
Cuneus	-0.14 (1.16)	-0.14 (1.02)	U = 877	.702
Pericalcarine	-0.44 (0.8)	-0.03 (1.43)	U = 698	.068
Lingual	0.03 (1.35)	0.09 (1.24)	U = 829	.442

Note. CN = cognitively normal control group, aMCI = amnestic mild cognitive impairment group. *p ≤ .05, **p ≤ .01, ***p ≤ .001.

Figure 2. Baseline Group Differences in ROI Volumes and Associations with Episodic Memory



A) Illustrations of bilateral ROIs used in the analyses grouped as AD-relevant (red: hippocampus, dark blue: amygdala, pink: entorhinal cortex, green: parahippocampus) or Control regions (light blue: cuneus, orange: pericalcarine gyrus, yellow: lingual gyrus). B) Boxplots depicting group differences (CN: left, light gray; aMCI: right, purple) in normative volumes (y-axes) for each ROI. C) Scatterplots showing within-group associations (CN: light gray; aMCI: purple) between episodic memory performance (x-axes) and normative volumes (y-axes) for each ROI. Note. CN = cognitively normal control group; aMCI = amnestic mild cognitive impairment group. *p ≤ .05, **p ≤ .01, ***p ≤ .001

Healthineers) following a system upgrade in 2017, with acquisition protocols for all sequences matched between the systems. The T1-weighted images used in this study were acquired using the 3D magnetization-prepared rapid acquisition with gradient echo (MPRAGE) sequence with the following parameters, identically acquired from both the TIM Trio and the Prisma: TR/TI/TE = 1900/900/2.26 ms, FOV = 256 × 256 mm^2, a generalized autocalibrating partially parallel acquisition (GRAPPA) factor of 2, voxel size 1.0 × 1.0 × 1.0 mm3. The FreeSurfer v6.0 recon-all processing stream was used to segment each participant's T1-weighted image and obtain regional brain volumes. We selected seven ROIs representing AD-relevant regions (hippocampus, parahippocampus, entorhinal cortex, amygdala; 7,10,30) and control regions (cuneus, lingual gyrus, pericalcarine gyrus; 7,31), shown in Figure 2A. These volumes were submitted to NOMIS (32) which is a free normative volumetric software that generates normative z-scores per ROI per participant, accounting for demographics and scanner characteristics (i.e., age, sex, head size, scanner manufacturer, magnetic field strength), using a normative database of 6,909 cognitively intact individuals ages 18 to 100. We chose to compute bilateral ROIs (by taking the average of z-scores across hemispheres), given meta-analytic evidence of bilateral medial temporal lobe atrophy in MCI and AD (33).

Statistical Analysis

Analyses were conducted in R v1.4.1106 and SPSS v.28. All variables were assessed for normality and nonparametric tests were used as appropriate. First, baseline group differences in demographics, ROI volumes, and episodic memory scores were tested using Mann-Whitney U tests for continuous variables and chi-square tests for categorical variables (i.e., sex and race). Next, we conducted Spearman rank correlations to assess the associations between episodic memory performance and each of the seven ROI volumes in each group separately at baseline. We then tested for group differences in the resulting correlations for each ROI using the 'cocor' R package which computes a Fisher's Z statistic and associated p-value. Given that the aMCI group was significantly older than the CN group, we conducted sensitivity analyses controlling for age in the above tests of group differences and correlations even though both memory and ROI scores were already norm-corrected. These were 1) a set of ANCOVA's with each ROI as the dependent variable, group as a between-subjects factor, and age as a covariate and 2) partial Spearman rank correlations between memory scores and each ROI volume covarying age.

Using the subset of participants who returned for follow-up (Figure 1), we tested for group differences in change in ROI volumes and memory over time by

Table 2. Tests of Within- and Between-Group Differences and Interaction Effects in Participants with Complete Follow-Up Data (N = 63)

	Subsample with Complete Follow-Up Data				A. Between-Group Differences at Baseline		B. Between-Group Differences at Follow-Up		C. Within-Group Differences				D. Group x Time Interactions	
	Baseline		Follow-Up						CN		aMCI			
	CN (n = 52)	aMCI (n = 11)	CN (n = 52)	aMCI (n = 11)										
	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)	F	p	F	p	F	p	F	p	F	p
Episodic Memory (Z-Score)	0.12 (0.96)	-1.91 (0.41)	0.58 (0.95)	-1.87 (0.45)	75.38	<.001***	67.62	<.001***	8.63	.005**	1.55	.219	0.01	.925
ROI Volumes (Z-scores)														
Hippocampus	0.24 (1.02)	-1.13 (2.74)	0.29 (1.14)	-1.36 (2.84)	14.15	<.001***	17.57	<.001***	0.58	.448	7.25	.009**	4.52	.038*
Amygdala	0.24 (1.44)	-0.45 (2.16)	0.10 (1.48)	-1.13 (2.37)	10.56	.002**	10.72	.002**	2.10	.152	0.87	.354	0.06	.808
Entorhinal	0.44 (1.16)	-0.17 (1.8)	0.35 (1.12)	-0.48 (1.17)	4.39	.040*	19.65	<.001***	0.48	.491	7.07	.010**	4.52	.038*
Parahippocampus	0.05 (1.02)	0.15 (0.48)	0.14 (1.1)	-0.34 (0.91)	0.03	.855	0.82	.370	0.58	.451	1.58	.213	2.13	.150
Cuneus	0.02 (1.25)	-0.14 (1.05)	-0.03 (1.19)	-0.16 (1.3)	0.57	.454	0.35	.556	0.06	.803	0.09	.765	0.14	.708
Pericalcarine	-0.4 (0.84)	0.07 (1.24)	-0.26 (1.72)	0.13 (1.54)	2.85	.097	4.44	.039*	0.21	.648	2.73	.104	0.30	.586
Lingual	-0.02 (1.61)	-0.04 (1.61)	-0.42 (1.09)	0.39 (0.74)	1.05	.309	1.52	.222	0.84	.362	0.03	.856	1.72	.195

Note. CN = cognitively normal control group, aMCI = amnestic mild cognitive impairment group. Results presented in Columns A through D are from repeated measures ANCOVAs controlling for follow-up interval (F(1,60)). *p ≤ .05, **p ≤ .01, ***p ≤ .001.

conducting repeated measures ANCOVAs with time as a within-subjects factor, group as a between-subjects factor, and follow-up interval as a covariate. Then, linear regression was used to test whether change in ROI volume was associated with change in episodic memory, and whether this was moderated by group (by testing an interaction with group), while controlling for follow-up interval. Change scores were calculated by subtracting z-scores at baseline from z-scores at follow-up.

Results

Table 1 summarizes the sample demographics, ROI volumes, episodic memory performance, and results of tests of group differences at baseline. The aMCI group was significantly older than the CN group at baseline (p = .016), but there were no differences between groups in sex, race, or years of education.

ROI Volumes Differ Between Groups and Relate to Memory

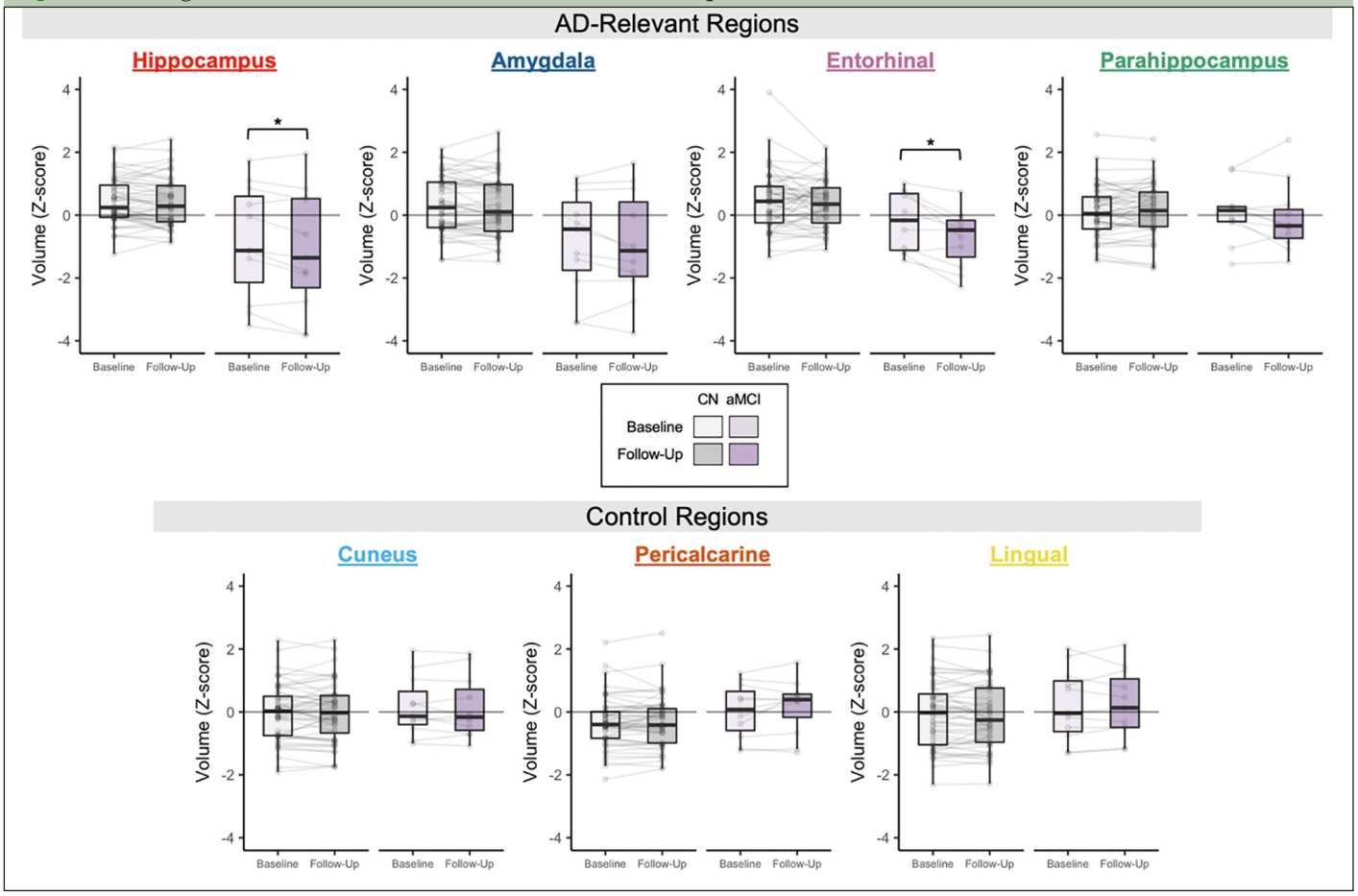
As expected, the aMCI group performed roughly 2 SD below the CN group on tests of episodic memory (p < .001) and had lower volumes than the CN group in all four AD-relevant ROIs (p's < .041; Figure 2B). In contrast, the groups did not differ significantly in any of the three control ROIs (p's > .068). These findings remained in sensitivity analyses controlling for age, aside from the group difference in parahippocampal volume becoming marginal (p = .075). Next, we assessed associations between ROI volumes and episodic memory within groups (Figure 2C). For all four AD-relevant regions, lower volumes were associated with poorer episodic memory performance in the aMCI group with large effect sizes (average Spearman's rho correlation = .51), but they were not associated with memory in the CN group (average Spearman's rho correlation = -.07). The magnitude of these associations was significantly greater in the aMCI group than the CN group (Hippocampus: Z = 2.84, p = .005; Parahippocampus: Z = 3.07, p = .002; Entorhinal Cortex: Z = 1.80, p = .072; Amygdala: Z = 2.56, p = .010). Volumes of the three control regions were not significantly associated with episodic memory performance in either group. All correlational results remained consistent in sensitivity analyses controlling for age.

Expected Longitudinal Changes in ROI Volumes

The subsets of participants who did (n = 63) and did not return for follow-up (n = 36) did not differ significantly in age (U = 1261, p = .358), years of education (U = 2030, p = .057), sex (X^2 = 2.32, p = .130), or race (X^2 = 0.00, p = 1.00). Similarly, in those with complete follow-up data, there were no differences between the CN (n = 52) and aMCI (n = 11) groups in baseline age (U = 245, p = .463), years of education (U = 231, p = .308), sex (X^2 = 1.75, p = .186), or race (X^2 = 0.21, p = .647). The median follow-up interval did not differ significantly between groups (U = 257, p = .606), and was approximately 1.2 years for both (CN: 1.18 years; aMCI: 1.15 years).

Table 2 presents baseline and follow-up data for those with complete data at both timepoints as well as results from the repeated measures ANCOVAs controlling for follow-up interval. There was a main effect of group on memory performance, such that the aMCI group performed significantly worse than the CN group on aggregate across timepoints (F(1,60) = 82.47, p < .001, η^2_p = .58). Although median episodic memory performance in the aMCI group remained 2 SD below that of the CN group at follow-up, change in memory performance over time did not differ between groups when controlling for follow-up interval (interaction p = .925). The main effects of Group on ROI volumes again localized to AD-relevant regions, with the aMCI group exhibiting lower volumes across timepoints in the hippocampus (F(1,60) = 16.26, p < .001, η^2_p = .21), entorhinal cortex (F(1,60) = 11.59, p = .001, η^2_p = .16), and amygdala (F(1,60) = 10.95, p = .002, η^2_p = .15), but not the parahippocampus (F(1,60)

Figure 3. Changes in ROI Volumes from Baseline to Follow-Up



Boxplots depict normative volumes (y-axes) for AD-relevant (top row) and Control (bottom row) ROIs. Each graph contains a pair of boxplots for each group (CN: left pair, gray; aMCI: right pair, purple), showing volumes at each timepoint (Baseline: left, light colors; Follow-up: right; dark colors). Significant Group X Time interactions indicated with asterisks. Note. CN = cognitively normal control group; aMCI = amnestic mild cognitive impairment group. *p ≤ .05

= 0.33, p = .568, η^2_p = .005) or any of the control regions (p's > .06). When controlling for follow-up interval, there were two Group X Time interactions (Table 2D; Figure 3). The first was for hippocampal volume (interaction p = .038, η^2_p = .07) such that the aMCI group exhibited significant hippocampal volume decline over time (pairwise comparison p = .009), but the CN group did not (pairwise comparison p = .448); this related to a median percent annual decline of 17% for aMCI vs. 8% in CN. The second was for entorhinal volume (interaction p = .038, η^2_p = .07) such that there were significant declines in the aMCI group over time (pairwise comparison p = .010) but not the CN group (pairwise comparison p = .491); median percent annual decline was 54% for aMCI vs. 5% for CN.

Change in memory scores was not significantly associated with change in any of the ROIs across groups when controlling for follow-up interval. However, there was one trending moderating effect of group (interaction β = 0.45, SE = 0.28, p = .110, η^2_p = .04) such that decline in memory was marginally associated with decreased hippocampal volume in the aMCI group (β = -0.47, SE = 0.26, p = .080) but not the CN group (β = -0.02, SE = 0.08, p = .787).

Discussion

This study sought to validate normative morphometric estimates produced by the open-source FreeSurfer and NOMIS software in the context of AD-related neurodegeneration. Comparing patients with aMCI and CN, we demonstrate that these metrics successfully replicate established patterns of atrophy in several AD-relevant regions and expected associations with episodic memory deficits in the aMCI group. Further, we provide preliminary evidence of sensitivity of these metrics to disease progression over time. These findings support the utility of this free academic software in quantifying neurodegeneration in aMCI, which can be used in lieu of proprietary alternatives in research settings.

Our primary aim was to evaluate whether normative volumetric estimates generated from FreeSurfer and NOMIS reproduced well-established findings from the AD literature in a cross-sectional sample of individuals with aMCI and CN controls. As expected, we found that the aMCI group had lower AD-relevant regional volumes and poorer episodic memory performance. Lower volumes of AD-relevant regions were associated

with poorer episodic memory performance in the aMCI group but not the CN group, and there was no relationship between the control regional volumes and memory, highlighting the specificity of these metrics to AD-related atrophic and cognitive changes. These results are consistent with those previously demonstrated by the proprietary NeuroQuant® (34–36) and NeuroReader® software packages (7, 37) and are directly in line with the field's current understanding of medial temporal lobe atrophy and memory decline along the AD continuum (10, 38).

We also conducted an exploratory analysis of how these volumetric estimates and their relationship to memory performance change over time. Although the aMCI group performed worse than the CN group at both timepoints, change in memory performance did not differ between groups. The observation that memory scores in both groups increased over time (reaching statistical significance for the CN group) most likely reflects practice effects (i.e., improvements as a function of repeat testing). Additionally, the relatively short follow-up interval (median ~1 year) was likely too short to see appreciable declines in the aMCI group. However, we did observe a trending association between decreasing hippocampal volume and declining memory in the aMCI group. This interaction may not have reached statistical significance given the small number of aMCI patients seen for follow-up (n = 11) and the restricted range of observed change in memory scores (range of Z-score change: -0.25 – 1.31) and hippocampal volume (range of Z-score change: -0.69 – 0.22) in this group, which is expected given the relatively short follow-up period. Nonetheless, this effect was in the expected direction and small-medium in magnitude (η^2_p = .04), suggesting that these volumetric estimates have sensitivity to detect AD-related changes in volume that may correspond with cognitive decline.

Given that the aims of this study were to specifically evaluate the validity of these volumetric estimates in aMCI, we explicitly focused on regions of the medial temporal lobe, as they demonstrate early and prominent atrophy in the course of AD (7, 10, 39), and episodic memory, as it is the cognitive domain most significantly affected in aMCI (40). However, we acknowledge that MCI is a heterogeneous syndrome in terms of both neuropathology and cognitive presentation. Thus, future work is needed to expand upon the current findings to consider additional brain regions in a more heterogenous MCI sample with more variegated cognitive presentations and etiologies, with other measures such as a global set of regions comprising the AD cortical signature (41) and other cognitive domains affected in AD (e.g., executive function; (42). Other limitations of this study include the demographically homogeneous sample (thereby diminishing the generalizability of these results) and the limited availability of follow-up data relative to baseline data that may bias these results. Nonetheless, these preliminary findings motivate future efforts to test their reproducibility using data from large repositories (such as the Alzheimer's Disease Neuroimaging Initiative and the National Alzheimer's Coordinating Center), given the availability of the MRI and neuropsychological approaches used here.

This study supports the validity of the open-source FreeSurfer and NOMIS software packages as alternatives to costly proprietary software for objectively quantifying neurodegeneration in patients with aMCI in research studies. It is important to note that our findings do not apply to clinical settings as neither FreeSurfer nor NOMIS were created for use as medical devices. These software packages cannot legally be used in clinical practice unless they undergo an extensive vetting process with oversight from conception and design/development to clinical trials, production, and commercial use. This involves meeting international standards for risk management and quality assurance, conducting clinical trials, obtaining approval from relevant regulatory bodies (e.g., the FDA), and establishing quality control and monitoring systems. Nevertheless, the availability of these free tools drastically improves accessibility for researchers who seek to generate regional brain volume estimates that are normatively adjusted for important confounding factors, such as patient demographics. The current findings suggest that these normative morphometric values accurately detect AD-related atrophy and exhibit expected associations with memory dysfunction in aMCI. Future work can establish the prognostic utility of these metrics by testing their sensitivity to disease progression and identification of older adults at risk of developing AD.

Funding: Research reported in this publication was supported by the NIH National Institute On Aging under Award Numbers K23AG044434 and R01AG054159 (to A.B.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The sponsor had no role in the design and conduct of the study; in the collection, analysis, and interpretation of data; in the preparation of the manuscript; or in the review or approval of the manuscript. This work was also supported by an Alzheimer's Association Grant (AARF-21-850073) to S. F-Z.

Conflicts of interest: The authors have no conflicts of interest to report.

Acknowledgments: We thank Dr. D. Szeles, Dr. E. Scott, and A. Sorrell for their help with participant recruitment, data collection, and data entry in support of this project. Additionally, we are grateful to our participants for their commitment to research and willingness to give their time and effort to this study.

References

1. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia. 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018
2. Farias ST, Mungas D, Reed BR, Harvey D, DeCarli C. Progression of Mild Cognitive Impairment to Dementia in Clinic- vs Community-Based Cohorts. Arch Neurol. 2009;66(9):1151-1157. doi:10.1001/archneurol.2009.106
3. Mauri M, Sinforiani E, Zucchella C, Cuzzoni MG, Bono G. Progression to dementia in a population with amnestic mild cognitive impairment: clinical variables associated with conversion. Funct Neurol. 2012;27(1):49-54.
4. Jungwirth S, Zehetmayer S, Hinterberger M, Tragl KH, Fischer P. The validity of amnestic MCI and non-amnestic MCI at age 75 in the prediction of Alzheimer's dementia and vascular dementia. International Psychogeriatrics. 2012;24(6):959-966. doi:10.1017/S1041610211002870
5. Leube DT, Weis S, Freymann K, et al. Neural correlates of verbal episodic memory in patients with MCI and Alzheimer's disease—a VBM study.

International Journal of Geriatric Psychiatry. 2008;23(11):1114-1118. doi:10.1002/gps.2036
6. Sexton CE, Mackay CE, Lonie JA, et al. MRI correlates of episodic memory in Alzheimer's disease, mild cognitive impairment, and healthy aging. Psychiatry Research: Neuroimaging. 2010;184(1):57-62. doi:10.1016/j.pscychresns.2010.07.005
7. Umfleet LG, Butts AM, Janecek JK, et al. Memory Performance and Quantitative Neuroimaging Software in Mild Cognitive Impairment: A Concurrent Validity Study. Journal of the International Neuropsychological Society. 2020;26(10):954-962. doi:10.1017/S1355617720000454
8. Leow AD, Yanovsky I, Parikshak N, et al. Alzheimer's Disease Neuroimaging Initiative: A one-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition. Neuroimage. 2009;45(3):645-655. doi:10.1016/j.neuroimage.2009.01.004
9. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI). Neurology. 2010;74(3):201-209. doi:10.1212/WNL.0b013e3181cb3e25
10. Tabatabaei-Jafari H, Shaw ME, Cherbuin N. Cerebral atrophy in mild cognitive impairment: A systematic review with meta-analysis. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2015;1(4):487-504. doi:10.1016/j.dadm.2015.11.002
11. Wang PN, Liu HC, Lirng JF, Lin KN, Wu ZA. Accelerated hippocampal atrophy rates in stable and progressive amnestic mild cognitive impairment. Psychiatry Research: Neuroimaging. 2009;171(3):221-231. doi:10.1016/j.pscychresns.2008.05.002
12. Yao Z, Hu B, Liang C, Zhao L, Jackson M, Initiative the ADN. A Longitudinal Study of Atrophy in Amnestic Mild Cognitive Impairment and Normal Aging Revealed by Cortical Thickness. PLOS ONE. 2012;7(11):e48973. doi:10.1371/journal.pone.0048973
13. Fischl B. FreeSurfer. NeuroImage. 2012;62(2):774-781. doi:10.1016/j.neuroimage.2012.01.021
14. Kang KM, Sohn CH, Byun MS, et al. Prediction of Amyloid Positivity in Mild Cognitive Impairment Using Fully Automated Brain Segmentation Software. Neuropsychiatr Dis Treat. 2020;16:1745-1754. doi:10.2147/NDT.S252293
15. Ochs AL, Ross DE, Zannoni MD, Abildskov TJ, Bigler ED. Comparison of Automated Brain Volume Measures obtained with NeuroQuant® and FreeSurfer. Journal of Neuroimaging. 2015;25(5):721-727. doi:10.1111/jon.12229
16. Ross DE, Ochs AL, Tate DF, et al. High correlations between MRI brain volume measurements based on NeuroQuant® and FreeSurfer. Psychiatry Research: Neuroimaging. 2018;278:69-76. doi:10.1016/j.pscychresns.2018.05.007
17. Yim Y, Lee JY, Oh SW, et al. Comparison of Automated Brain Volume Measures by NeuroQuant vs. Freesurfer in Patients with Mild Cognitive Impairment: Effect of Slice Thickness. Yonsei Med J. 2021;62(3):255-261. doi:10.3349/ymj.2021.62.3.255
18. Potvin O, Dieumegarde L, Duchesne S, Initiative ADN. NOMIS: Quantifying morphometric deviation from normality over the lifetime in the adult human brain. bioRxiv. Published online 2021. doi:10.1101/2021.01.25.428063
19. Benitez A, Jensen JH, Falangola MF, Nietert PJ, Helpern JA. Modeling white matter tract integrity in aging with diffusional kurtosis imaging. Neurobiology of Aging. 2018;70:265-275. doi:10.1016/j.neurobiolaging.2018.07.006
20. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. Journal of the American Geriatrics Society. 2005;53(4):695-699. doi:10.1111/j.1532-5415.2005.53221.x
21. Rossetti HC, Lacritz LH, Cullum CM, Weiner MF. Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. Neurology. 2011;77(13):1272-1275. doi:10.1212/WNL.0b013e318230208a
22. Jak AJ. Quantification of Five Neuropsychological Approaches to Defining Mild Cognitive Impairment. The American Journal of Geriatric Psychiatry. 2009;17:368-375. doi: 10.1097/JGP.0b013e31819431d5
23. Edmonds EC, Delano-Wood L, Clark LR, et al. Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. Alzheimer's & Dementia. 2015;11(4):415-424. doi:10.1016/j.jalz.2014.03.005
24. Fountain-Zaragoza S, Braun SE, Horner MD, Benitez A. Comparison of conventional and actuarial neuropsychological criteria for mild cognitive impairment in a clinical setting. Journal of Clinical and Experimental Neuropsychology. 2021;43(7):753-765. doi:10.1080/13803395.2021.2007857
25. Bondi MW. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. Journal of Alzheimer's Disease. 2014;42:275-289. doi:10.3233/JAD-140276
26. Weintraub S, Salmon D, Mercaldo N, et al. The Alzheimer's Disease Centers' Uniform Data Set (UDS): The Neuropsychological Test Battery. Alzheimer Dis Assoc Disord. 2009;23(2):91-101. doi:10.1097/WAD.0b013e318191c7dd
27. Schmidt M. Rey Auditory Verbal Learning Test: A Handbook. Western Psychological Services; 1996.
28. Wechsler D. Wechsler memory scale revised. The psychological corporation. New York. Published online 1987.
29. Shirk SD, Mitchell MB, Shaughnessy LW, et al. A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. Alzheimer's Research and Therapy. 2011;3(6):32. doi:10.1186/alzrt94
30. Yue L, Wang T, Wang J, et al. Asymmetry of Hippocampus and Amygdala Defect in Subjective Cognitive Decline Among the Community Dwelling Chinese. Front Psychiatry. 2018;9:226. doi:10.3389/fpsyt.2018.00226
31. Fjell AM, Walhovd KB, Fennema-Notestine C, et al. One-Year Brain Atrophy Evident in Healthy Aging. J Neurosci. 2009;29(48):15223-15231. doi:10.1523/JNEUROSCI.3252-09.2009
32. Potvin O, Dieumegarde L, Duchesne S, for the Alzheimer's Disease Neuroimaging Initiative, CIMA-Q. NOMIS: Quantifying Morphometric Deviation from Normality over the Lifetime in the Adult Human Brain. Neuroscience; 2021. doi:10.1101/2021.01.25.428063
33. Yang J, Pan P, Song W, et al. Voxelwise meta-analysis of gray matter anomalies in Alzheimer's disease and mild cognitive impairment using anatomic likelihood estimation. Journal of the Neurological Sciences. 2012;316(1):21-29. doi:10.1016/j.jns.2012.02.010
34. England HB, Gillis MM, Hampstead BM. RBANS Memory Indices Are Related to Medial Temporal Lobe Volumetrics in Healthy Older Adults and Those with Mild Cognitive Impairment. Arch Clin Neuropsychol. 2014;29(4):322-328. doi:10.1093/arclin/acu012
35. Fyock CA, Hampstead BM. Comparing the relationship between subjective memory complaints, objective memory performance, and medial temporal lobe volumes in patients with mild cognitive impairment. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2015;1(2):242-248. doi:10.1016/j.dadm.2015.03.002
36. Stelmokas J, Yassay L, Giordani B, et al. Translational MRI Volumetry with NeuroQuant: Effects of Version and Normative Data on Relationships with Memory Performance in Healthy Older Adults and Patients with Mild Cognitive Impairment. Journal of Alzheimer's Disease. 2017;60(4):1499-1510. doi:10.3233/JAD-170306
37. Emmert NA, Reiter KE, Butts A, et al. Hippocampal Volumes in Amnestic and Non-Amnestic Mild Cognitive Impairment Types Using Two Common Methods of MCI Classification. Journal of the International Neuropsychological Society. Published online June 16, 2021:1-10. doi:10.1017/S1355617721000564
38. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018
39. Driscoll I, Davatzikos C, An Y, et al. Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. Neurology. 2009;72(22):1906-1913. doi:10.1212/WNL.0b013e3181a82634
40. Leyhe T, Müller S, Milian M, Eschweiler GW, Saur R. Impairment of episodic and semantic autobiographical memory in patients with mild cognitive impairment and early Alzheimer's disease. Neuropsychologia. 2009;47(12):2464-2469. doi:10.1016/j.neuropsychologia.2009.04.018
41. Wang M, Roussos P, McKenzie A, et al. Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. Genome Medicine. 2016;8(1):104. doi:10.1186/s13073-016-0355-3
42. Zheng D, Sun H, Dong X, et al. Executive dysfunction and gray matter atrophy in amnestic mild cognitive impairment. Neurobiology of Aging. 2014;35(3):548-555. doi:10.1016/j.neurobiolaging.2013.09.007

How to cite this article: S. Fountain-Zaragoza, O. Horn, K.E. Thorn, et al. Validity of Normative Volumetric Estimates from Open Access Software in Amnestic Mild Cognitive Impairment. J Prev Alz Dis 2023;2(10):236-243; http://dx.doi.org/10.14283/jpad.2023.19