Journal of the International Neuropsychological Society (2019), **25**, 800–810

doi:10.1017/S135561771900047X

Differentiating Between Healthy Control Participants and Those with Mild Cognitive Impairment Using Volumetric MRI Data

Renée DeVivo[1,2], Lauren Zajac[1,2], Asim Mian[3], Anna Cervantes-Arslanian[4], Eric Steinberg[5], Michael L. Alosco[5,6], Jesse Mez[5,6], Robert Stern[1,4,5,6] and Ronald Killany[1,2,5,7] for the Alzheimer's Disease Neuroimaging Initiative*

[1]Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, Massachusetts, USA
[2]Center for Biomedical Imaging, Boston University School of Medicine, Boston, Massachusetts, USA
[3]Department of Radiology, Boston Medical Center, Boston, Massachusetts, USA
[4]Department of Neurosurgery, Boston University School of Medicine, Boston, Massachusetts, USA
[5]Boston University Alzheimer's Disease Center, Boston, Massachusetts, USA
[6]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA
[7]Boston University School of Public Health, Boston, Massachusetts, USA

(Received May 25, 2018; Final revision February 20, 2019; Accepted March 28, 2019; First Published Online May 27, 2019)

Abstract

Objective: To determine whether volumetric measures of the hippocampus, entorhinal cortex, and other cortical measures can differentiate between cognitively normal individuals and subjects with mild cognitive impairment (MCI). **Method:** Magnetic resonance imaging (MRI) data from 46 cognitively normal subjects and 50 subjects with MCI as part of the Boston University Alzheimer's Disease Center research registry and the Alzheimer's Disease Neuroimaging Initiative were used in this cross-sectional study. Cortical, subcortical, and hippocampal subfield volumes were generated from each subject's MRI data using FreeSurfer v6.0. Nominal logistic regression models containing these variables were used to identify subjects as control or MCI. **Results:** A model containing regions of interest (superior temporal cortex, caudal anterior cingulate, pars opercularis, subiculum, precentral cortex, caudal middle frontal cortex, rostral middle frontal cortex, pars orbitalis, middle temporal cortex, insula, banks of the superior temporal sulcus, parasubiculum, paracentral lobule) fit the data best ($R^2 = .7310$, whole model test chi-square = 97.16, $p < .0001$). **Conclusions:** MRI data correctly classified most subjects using measures of selected medial temporal lobe structures in combination with those from other cortical areas, yielding an overall classification accuracy of 93.75%. These findings support the notion that, while volumes of medial temporal lobe regions differ between cognitively normal and MCI subjects, differences that can be used to distinguish between these two populations are present elsewhere in the brain.

Keywords: Neuroimaging, Healthy aging, Hippocampus, Entorhinal cortex, Atrophy, Logistic models

INTRODUCTION

The prevalence of Alzheimer's disease (AD) continues to rise as the global population ages. In recent years, researchers have come to view AD as a continuum, rather than a sequence of distinct phases of cognitive and neuropathological changes (Aisen et al., 2017). During the "preclinical" aspects of the continuum, individuals are generally cognitively asymptomatic. However, many "preclinical" individuals go on to develop symptoms, such as episodic memory loss, and this is referred to as mild cognitive impairment (MCI). In recent years, MCI has been clinically characterized by criteria such as self- or informant-reported cognitive complaints, objective cognitive impairment, preserved independence in functional abilities, and the absence of dementia (Petersen et al., 2014). As the disease progresses, cognitive impairment worsens and functional impairment becomes increasingly apparent in everyday life. At this point, a person is considered to have AD dementia (Aisen et al., 2017). With the growing understanding of AD as a continuum, identifying biomarkers of pathophysiological changes has become paramount.

Correspondence and reprint requests to: Ronald Killiany, Boston University School of Medicine, Center for Biomedical Imaging, 700 Albany Street, W701, Boston, MA 02118, USA. E-mail: killiany@bu.edu

*Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Morphometric magnetic resonance imaging (MRI) studies have established that the areas of the brain often first damaged in MCI and AD are the hippocampus and the entorhinal cortex (Du et al., 2001; Killiany et al., 2000, see reviews Pini et al., 2016; Zhou, Zhang, Zhao, Qian, & Dong, 2016). The pathophysiological changes present in AD include the accumulation of amyloid-beta that form plaques, as well as the aggregation of tau-proteins that form neurofibrillary tangles. Such tangles induce neuronal death, resulting in morphometric changes that start in the medial temporal lobe (Gómez-Isla et al., 1996, see review Spires-Jones & Hyman, 2014). In order to obtain sensitive and specific measures of these structures and the various changes that occur, researchers have begun to segment the hippocampus into subfields (de Flores, La Joie, & Chételat, 2015; Pini et al., 2016).

Previous studies reported reductions in the volumes of the whole hippocampus, hippocampal subfields, and entorhinal cortex in the brains of MCI and AD subjects compared with controls (Mueller et al., 2010; Pennanen et al., 2004, see review de Flores et al., 2015). Such studies utilized a cross-sectional approach to affirm subjects as controls, MCI, or AD based upon the characteristics of various regions of interest (ROIs) (Colliot et al., 2008; Du et al., 2001; Hanseeuw et al., 2011; Khan et al., 2015; Mueller et al., 2010; Xu et al., 2000). Many of these studies focused on subjects who already had MCI in order to best predict who with MCI will convert to AD (Khan et al., 2015; Killiany et al., 2000; Plant et al., 2010; Westman et al., 2011) with less emphasis on creating models that accurately classify subjects as control or MCI. These classification studies often use the characteristics of only one ROI as a classifier variable, such as the hippocampus, entorhinal cortex, or a specified hippocampal subfield, and have failed to achieve classification accuracies which exceed 80% (Colliot et al., 2008; Du et al., 2001; Hanseeuw et al., 2011; Mueller et al., 2010; Pennanen et al., 2004; Westman et al., 2011, see review Weiner et al., 2015). Thus, there remains a significant need for models that can correctly classify subjects in the earlier stages of the AD continuum.

The goal of the present study was to utilize morphometric MRI measures to identify a broader set of variables that can classify controls and MCI subjects using logistic regression. The first step was to determine the utility of the whole hippocampus, hippocampal subfields, and entorhinal cortex as classifiers. Next, less studied cortical regions outside the medial temporal lobe were added to the model to determine whether any of these regions could improve the model fit and classification accuracy. While these regions are not as commonly used in classification models, researchers have begun to find consistent patterns of atrophy in MCI and AD in regions beyond the medial temporal lobe, including gyri within the frontal, parietal, and temporal lobes (Hänggi, Streffer, Jäncke, & Hock, 2011; Karas et al., 2004). Knowledge regarding the classification value of these regions in cognitively normal subjects and those with MCI could enhance our understanding of the disorder.

MATERIALS AND METHODS

Subjects

This study utilized MRI scans from 96 subjects selected from two sources. Forty-two scans were obtained from the Boston University Alzheimer's Disease Center (BU-ADC) Clinical Core Registry: Health Outreach Program for the Elderly (HOPE). The BU-ADC is a center (of 30) funded by the National Institute on Aging that contributes data to the National Alzheimer's Coordinating Center. The BU-ADC registry, including subject recruitment and inclusion/exclusion criteria, has previously been described (Ashendorf et al., 2017; Galetta et al., 2017). All HOPE subjects' diagnoses were made at consensus conferences, following presentation and discussion of all medical history and evaluation results, such as clinical interview, informant input, neuropsychological test scores, and MRI scans viewed for clinical criteria (hippocampal atrophy, white matter signal abnormalities, and evidence of microbleeds). No formal biomarker (amyloid, tau, or genetic) data is presented at such consensus conferences. Subjects were determined to be cognitively normal ($n = 19$) if their objective neuropsychological test scores were within the normal range and they had a Clinical Dementia Rating (Morris, 1993) global score of 0.0. All MCI diagnoses ($n = 23$) were amnestic and followed published criteria (Petersen et al., 2014).

Fifty-four scans were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (for more information, refer to adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership, led by Principle Investigator, Michael W. Weiner, MD. The primary goal of the ADNI has been to elucidate clinical, genetic, imaging, and biochemical biomarkers of AD, and to better understand the progression from normal cognition to MCI to AD (Weiner et al., 2015). Twenty-seven of these subjects were cognitively normal controls and the other 27 subjects were amnestic MCI single domain. All ADNI scans were selected to have come from a Philips 3T scanner to ensure comparable imaging parameters as HOPE scans, and to balance the demographic data of HOPE subjects. All procedures were approved by local IRBs, and subjects gave informed consent at the time of their enrollment in both studies in accordance with the Declaration of Helsinki. In total, there were 96 scans analyzed (46 control and 50 MCI; 54 females and 42 males) in this study. All MCI scans came from subjects that had the amnestic form of MCI. Eight also showed impairment in another domain such as language, executive function, or visuospatial functioning in addition to the impairment of memory.

Both HOPE and ADNI collect demographic data, including age, education, and APOE ε4 status, as well as neuropsychological test scores from the Mini-Mental State Exam (MMSE), Geriatric Depression Scale (GDS), Logical Memory Recall (modified from the Wechsler D. Wechsler Memory Scale-Revised, San Antonio, Texas: Psychological Corporation; 1987), and Part B of the Trailmaking Test

(Table 1). The neuropsychological tests and diagnostic procedures utilized in both studies were based on comparable protocols. Given that these neuropsychological measures were utilized to differentiate and define groups, none of these measures could again be used as outcomes.

Imaging Assessments

The scans we used were 3D magnetization-prepared rapid acquisition of gradient echo sequences. These were acquired on 3T Philips scanners. For the BU-ADC scans, a 32-channel headcoil and sense factor of 2 was used with the following imaging parameters: TR = 6.7 ms, TE = 3.1 ms, flip angle = 9°, reconstructed and acquisition voxel size = .98 × .98 × 1.2 mm, FOV = 250 × 250 × 180 mm, 150 sagittal slices. For the ADNI scans (Jack et al., 2008), an 8-channel headcoil and a sense factor of 1.8 was used with the following imaging parameters: TR = 6.7 ms, TE = 3.1 ms, flip angle = 9°, reconstructed voxel size = 1.05 × 1.05 × 1.20 mm, acquisition voxel size = 1.11 × 1.11 × 1.20 mm, FOV = 270 × 252 × 204 mm, 170 sagittal slices. DICOM scans were downloaded from the ADNI database.

The MRI scans were segmented with FreeSurfer version 6.0 (http://surfer.nmr.mgh.harvard.edu; for additional details, see Desikan et al., 2006; Iglesias et al., 2015). FreeSurfer v6.0 utilizes an improved atlas that can automatically segment hippocampal regions into a greater number of subfields than previous versions have allowed (Iglesias et al., 2015).

Statistical Analysis

Independent samples one-tailed *t*-tests were used to assess differences between the control and MCI groups in terms of neuropsychological outcome and MRI measures. Independent samples two-tailed *t*-tests were used to assess demographic factors, such as age and education, as there was no expected direction to any group differences. A chi-square test was performed to determine if a difference existed in APOE ε4 status between the groups. In terms of these analyses, we had two goals in mind. The first was to establish whether the groups were matched in terms of demographics and APOE ε4 status. To assess the significance of these tests, we used a *p*-value of .01 without correction to increase the stringency of not finding a difference. Our second goal was purely exploratory and involved looking for differences between the groups with the expectation that neuropsychological performance and size of brain regions would be less in the MCI group. The analyses conducted towards this goal were meant to help us know the data – selection into regression models was made independent of these exploratory analyses. These tests were interpreted using an arbitrary *p*-value of .01. Measures generated from FreeSurfer v6.0 included estimated total intracranial volume (eTIV), 68 cortical volumes, 12 subcortical volumes, right and left hippocampal volumes, and 24 hippocampal subfield volumes (for additional segmentation

Table 1. Demographic and neuropsychological test data for control and MCI groups

	Mean for control group ± SD (range from lowest to highest)	Mean for MCI group ± SD (range from lowest to highest)	*p*-Value	Cohen's *d*
Age (in years) ($n = 96$), two-tailed test	75.24 ± 8.50 (54–96)	74.48 ± 6.76 (59–92)	.63	.10
Education (in years) ($n = 96$), two-tailed test	16.02 ± 2.44 (12–20)	16.56 ± 2.57 (12–21)	.3	.22
APOE ε4 status ($n = 94$)	17/45 control participants have at least one ε4 allele. Of the controls, one participant was homozygous for the ε4 allele	16/49 MCI participants have at least one ε4 allele. Of the MCI group, one participant was homozygous for the ε4 allele	.87	.053 (Cohen's *W*)
Logical Memory Immediate($n = 87$)	15.79 ± 2.97 ($n = 43$)	10.82 ± 4.71 ($n = 44$)	.000000039*	1.26
Logical Memory Delayed ($n = 87$)	14.86 ± 3.41 ($n = 43$)	9.25 ± 5.04 ($n = 44$)	.000000170*	1.30
Mini Mental State Examination ($n = 96$)	29.27 ± .96 ($n = 46$)	28.00 ± 1.99 ($n = 50$)	.0000976*	.81
Geriatric Depression Scale ($n = 95$)	.78 ± 1.06 ($n = 45$)	1.56 ± 2.45 ($n = 50$)	.025	.41
Part B of Trailmaking Test ($n = 96$, time to completion in seconds)	70.48 ± 21.05 ($n = 46$)	130.24 ± 82.88 ($n = 50$)	.00000364* ($p < .0001$, Wilcoxon–Mann–Whitney test)	.99

Note: There were 96 subjects total. Smaller sample sizes reflect missing data. All tests were one-tailed *t*-tests unless denoted otherwise. *p*-Value < .01.
MCI, mild cognitive impairment; APOE; apolipoprotein E; ε4, allele 4; SD, standard deviation.
*Denotes significance.

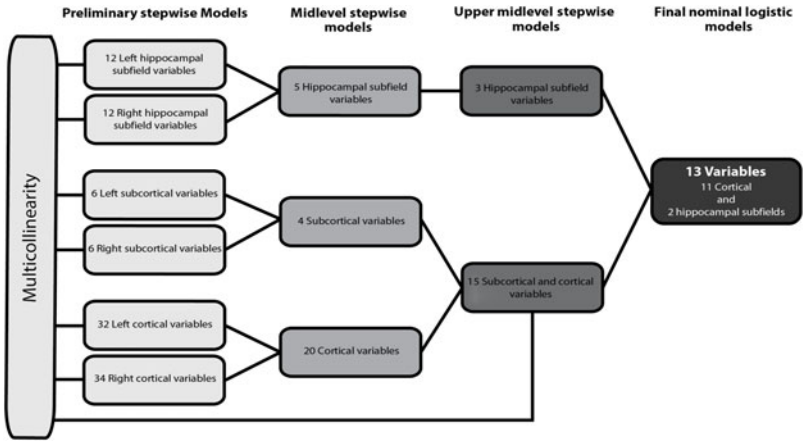

Fig. 1. Outline of stepwise and nominal logistic regression analyses. Overview of the process for determining the "best" classification model. Various preliminary and mid-level analyses were used as a data reduction technique to identify the variables that produce the most accurate classification model.

details, see Iglesias et al., 2015). All volumes generated by FreeSurfer v6.0 were examined and then used in subsequent models. For a complete list of the variables collected from FreeSurfer v6.0, see http://www.bumc.bu.edu/mri/personnel-and-students/renee-devivo/. Data were visually inspected for errors and edited as needed. ANCOVA was performed to determine whether eTIV, age, gender, education, APOE ε4 status, or study (i.e., HOPE or ADNI) had an effect on any morphometric MRI variables. Age and eTIV had an effect on the majority of ROI volumes, while gender, education, APOE ε4, and study had no impact. In order to correct for age and eTIV, residuals were computed based upon the values of all subjects included in this study. Additionally, an overall MANOVA was run using study (HOPE or ADNI) as a between-subjects variable and brain ROIs as a repeated measure to ensure that the use of different headcoils in the two studies did not have any overall effect on the outcome. This analysis was run because different headcoils were used in the ADNI and HOPE (8 vs. 32 channels, respectively) studies.

Using these residuals, targeted nominal logistic regression models were created to determine how well the volumes of individual regions, such as entorhinal cortex, whole hippocampus, and hippocampal subfields, could classify subjects. These regions were utilized first due to their importance in cognitive impairment. To follow up, additional stepwise variable logistical models (mixed, probability to enter/leave $p < .25$) were run using subgroups of ROIs (subcortical, cortical, and hippocampal subfields) to determine which ROIs classified subjects best. To control for multicollinearity, correlations were calculated between the subcortical and cortical regions. Regions that had intercorrelated values >.7 with one or more other regions were excluded from the stepwise models. A series of stepwise logistic models were used as steps for data reduction. The preliminary stepwise logistic models were computed using separate classes of variables as follows: (1) left hippocampal subfield variables, (2) right hippocampal subfield variables, (3) left subcortical variables, (4) right subcortical variables, (5) left cortical variables, and (6) right cortical variables. From the results of these preliminary analyses, mid-level analyses were conducted using (1) hippocampal subfield variables and (2) combined cortical and subcortical group variables. No bilateral regions were included in mid-level stepwise analyses, rather the unilateral region with the most predictive value, or the lowest p value, was used. Following the mid-level stepwise models, an upper mid-level stepwise logistic model was created using the selected cortical and subcortical variables and selected hippocampal subfield variables. Finally, a nominal logistic model was generated to create an optimal classification model of group membership. For a schematic of this data reduction process, see Figure 1. Within each of the logistic models, false discovery rate (FDR) correction was used to establish the significant contributors to each model. Across models, Bonferroni correction (p-value divided by the number of models) was used to determine the appropriate p-value for significance. As a final check, the nominal logistic model was regenerated using a "backward" approach to ensure the order of variable selection was not driving the model.

A leave-one-out prediction of the one-out validation technique was conducted to better estimate how the model would perform on a different dataset (Fan, Batmanghelich, Clark, & Davatzikos, 2008; Misra, Fan, & Davatzikos, 2009). Essentially, we re-ran the upper mid-level stepwise and final nominal logistic regression models leaving one subject out each time until all subjects had been left out once. All statistical analyses were conducted with JMP Pro 13.0 on a Mac Pro running macOS Sierra operating system.

RESULTS

Demographic Data

There were no differences between the groups in terms of age, years of education, or APOE ε4 status (p's > .01) (Table 1).

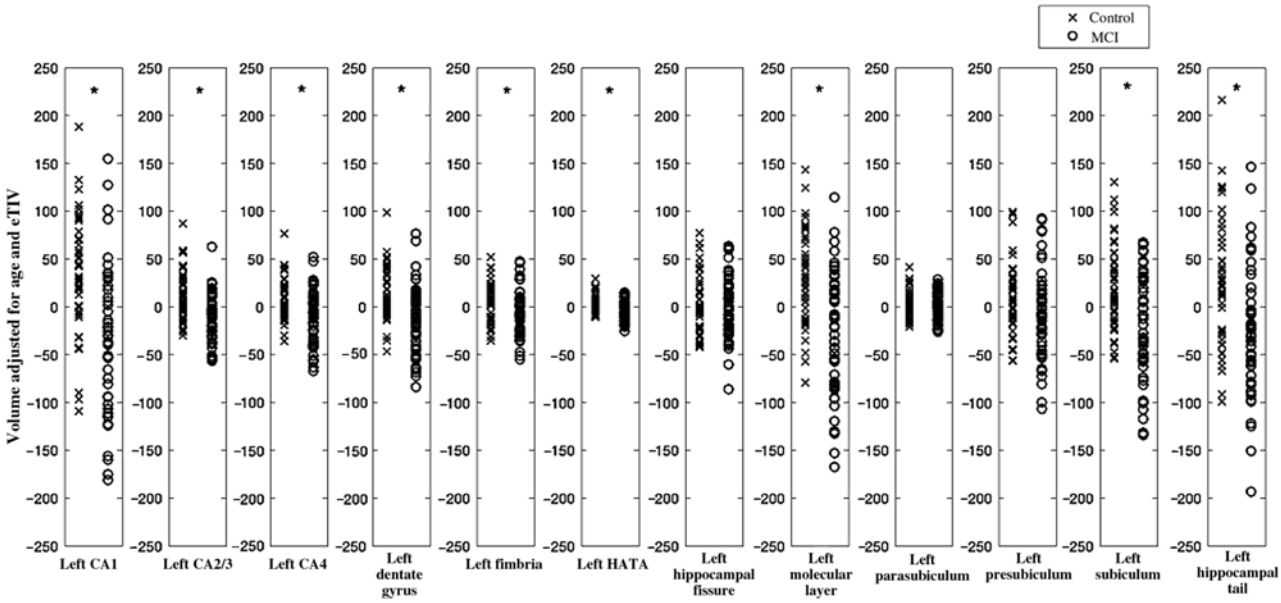

Fig. 2. Volumetric spread of left adjusted hippocampal subfields. The volume (adjusted for age and eTIV) of 12 identified hippocampal subfields was measured in all control and MCI subjects and is plotted. Regions showing significant differences ($p < .01$) are denoted with an asterisk and were found in 9 of the 12 left hippocampal subfields where the MCI subjects had reduced volume compared with the controls. eTIV, estimated total intracranial volume; CA, cornu ammonis; HATA, hippocampal amygdala transition area; MCI, mild cognitive impairment.

As expected, differences were found between the groups in MMSE score, Logical Memory Immediate Raw score, Logical Memory Delayed Raw score, and Part B of the Trailmaking Test (p's $< .01$) (Table 1). There was a trend for MCI subjects to have higher GDS scores ($p = .025$) though neither group expressed clinically relevant scores on the GDS.

MRI Data

When comparing uncorrected volumes, the MCI group had smaller volumes in 15 of the 24 hippocampal subfields, bilateral hippocampal formations, and the right entorhinal cortex, compared with the control group (p's $< .01$). When comparing the residual data (adjusted for age and eTIV) between the control and MCI groups, the same 15 hippocampal subfields, bilateral hippocampal formations, and the right entorhinal cortex remained significantly smaller. In addition, three hippocampal subfields (left fimbria, right hippocampal tail, and right fimbria) and the left entorhinal cortex were significantly smaller in MCI subjects when residual volumes were compared (Figures 2–5, effect sizes reported in Table 2). The results of an overall MANOVA showed that the effect of study (i.e., headcoil) was not significant ($F(1,94) = 2.894$, $p = .09$, Cohen's $d = .36$).

To determine how well the volumes of the entorhinal cortex, hippocampus, and hippocampal subfields identified group membership, these variables were entered into three targeted nominal logistic regression models consisting of both the right and left volumes of each region. Though some were significant, none of these models provided a good model according to the R^2 values obtained: hippocampal subfields ($R^2 = .3629$, whole model test chi-square = 48.2361, $p = .0024$), whole hippocampus ($R^2 = .1817$, whole model test chi-square = 24.1557, $p < .0001$), and entorhinal cortex ($R^2 = .0688$, whole model test chi-square = 9.1421, $p = .0103$). Stepwise analyses using a wider subset of variables were conducted to see if better classification could be obtained. Tests of multicollinearity showed that both the left cuneus and left pericalcarine were correlated with more than one other region with a R value >.7, and these regions were excluded from the following analyses. The six preliminary stepwise models selected five hippocampal subfield variables, four subcortical variables, and twenty cortical variables. Mid-level stepwise analyses conducted on the hippocampal subfield variables and combined subcortical and cortical variables further refined the data selected to three hippocampal subfield variables and fifteen subcortical and cortical variables. All 18 of the selected regions (p's $< .01$, FDR corrected) were entered into an upper mid-level stepwise and subsequent nominal logistic regression model, which was significant ($R^2 = .7120$, whole model test chi-square = 94.63, $p < .0001$), and had a classification rate of .9063 (misclassification rate = .0937) as 40 out of 46 control subjects and 47 out of 50 MCI subjects were classified correctly. Since the two regions (left cuneus and left pericalcarine) that were originally excluded after the tests of multicollinearity were independent of the other regions included in the final models, these regions were forced into the model to see if their addition improved the model. The final model with the left cuneus and left pericalcarine revealed that 13 variables were

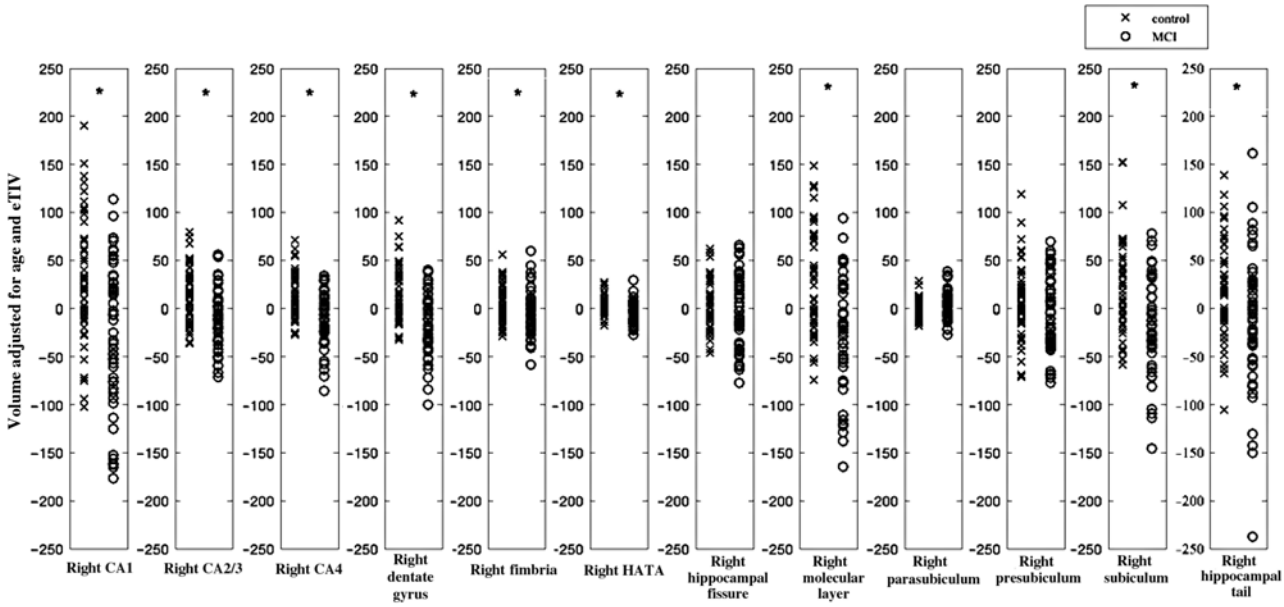

Fig. 3. Volumetric spread of right adjusted hippocampal subfields**.** The volume (adjusted for age and eTIV) of 12 identified hippocampal subfields was measured in all control and MCI subjects and is plotted. Regions showing significant differences ($p < .01$) are denoted with an asterisk and were found in 9 of the 12 right hippocampal subfields where the MCI subjects had reduced volume compared with the controls. eTIV, estimated total intracranial volume; CA, cornu ammonis; HATA, hippocampal amygdala transition area; MCI, mild cognitive impairment.

significant contributors ($R^2 = .7310$, whole model test chi-square = 97.16, $p < .0001$) and had a classification rate of .9375 (misclassification rate = .0625) as 43 out of 46 control subjects and 47 out of 50 MCI subjects were classified correctly. Table 3 outlines the 13 variables that were significant contributors to the final model, and all model fit statistics for the described stepwise and nominal logistic analyses are reported in Table 4. The average classification rate from each leave-out-one analysis was .7708 (misclassification rate = .2292) as 37 out of 46 control subjects and 37 out of 50 MCI subjects were classified correctly.

As an additional exploratory analysis looking at the value of whole hippocampus as a classifier *versus* the value of hippocampal subfields, the final nominal logistic model was re-conducted excluding the hippocampal subfield variables that were chosen, and substituting a unilateral whole hippocampal volume variable instead. When using the whole right hippocampal volume, the model was weakened ($R^2 = .6313$, whole model test chi-square = 83.91, $p < .0001$) and the whole right hippocampus variable was not a significant contributor to this model, which had a classification rate of .875 (misclassification rate = .125). When using the whole left hippocampal volume, the model was essentially similar to when the whole right hippocampal volume was used and also weaker than the original model ($R^2 = .6413$, whole model test chi-square = 85.24, $p < .0001$). Similarly, the whole left hippocampus variable was not a significant contributor in this model. This model had a classification rate of .875 (misclassification rate = .125). As a final check of our analysis using subfields, we re-ran the final nominal logistic model in a "backward" fashion to assure that the order of selection did not bias the analysis. The results of this were the same ($R^2 = .7310$, whole model test chi-square = 97.16, $p < .0001$) with 43 out of 46 control subjects classified correctly and 47 out of 50 MCI subjects classified correctly.

DISCUSSION

Nominal Logistic Model of Group Membership

The present study sought to expand current knowledge regarding what regions of the brain are most influential in classifying subjects who are cognitively normal *versus* those with MCI. Some previous studies have found the entorhinal cortex to be the most effective at discriminating between controls and subjects with cognitive impairment (Killiany et al., 2000; Pennanen et al., 2004), while others have found the entire hippocampal formation or various hippocampal subfield volumes to perform best (Du et al., 2001; Hanseeuw et al., 2011; Mueller et al., 2010; Xu et al., 2000). Regardless, a majority of such cross-sectional studies built classification models utilizing only one region and reported 60–81% accuracy (Hanseeuw et al., 2011; Mueller et al., 2010; Pennanen et al., 2004). Studies including multiple ROIs in their models often achieve greater accuracy as was the goal of the present study (Convit et al., 2000; Convit et al., 1997; Hänggi et al., 2011; Killiany et al., 2000).

Initially, we utilized the volumes of the whole hippocampal formations, hippocampal subfields, and entorhinal cortices to build three targeted classification models, and individually, these subgroups performed poorly at classification. However, when ROIs outside the medial temporal regions were added to our classification model, we were able

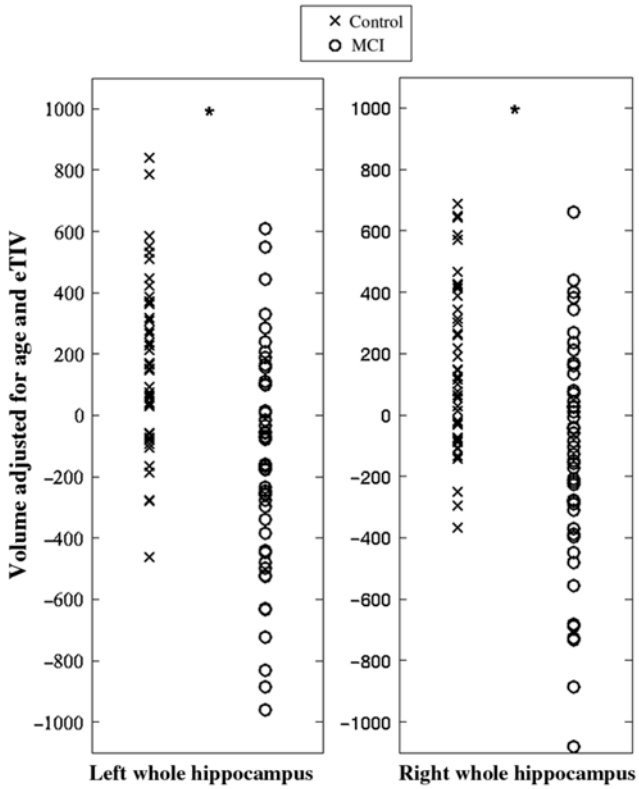

Fig. 4. Volumetric spread of adjusted whole hippocampal formations. The volume (adjusted for age and eTIV) of both whole hippocampal formations was measured in control and MCI subjects and is plotted. Both right and left hippocampal formations in MCI subjects showed a significant reduction in volume compared with controls and is indicated by an asterisk ($p < .01$). eTIV, estimated total intracranial volume; MCI, mild cognitive impairment.

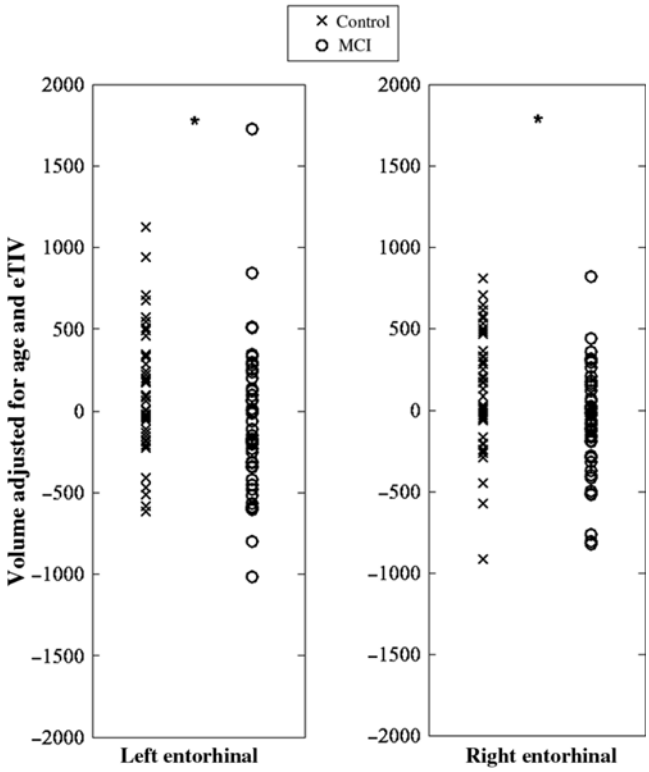

Fig. 5. Volumetric spread of adjusted entorhinal cortices. The volume (adjusted for age and eTIV) of both entorhinal cortices was measured in control and MCI subjects and is plotted. Both right and left entorhinal cortices in MCI subjects showed a significant reduction in volume compared with controls and is indicated by an asterisk ($p <$ .01). eTIV, estimated total intracranial volume; MCI, mild cognitive impairment.

to correctly identify 43 out of 46 control subjects and 47 out of 50 MCI subjects for an overall classification accuracy of 93.75% with 93.48% sensitivity and 94% specificity. Table 3 shows the 13 volumes that contributed most to this model. Notably, although subcortical, cortical, and hippocampal subfield variables were all taken into account in the data reduction stepwise models, only cortical regions and hippocampal subfields were chosen by the data reduction models to go into the final logistic model.

Cross-sectional studies exploring the classification value of hippocampal subfield volumes *versus* whole hippocampal volumes have found disparate results (Hanseeuw et al., 2011; Pluta et al., 2012). Some have found the CA1 volume to be a stronger classifier than whole hippocampal volume (La Joie et al., 2013; Pluta et al., 2012), while other studies have found the subiculum volume, but not the CA1 volume, to be a more effective classifier than whole hippocampal volume (Hanseeuw et al., 2011). Our findings are in line with those promoting the classification value of the subiculum volume. When we created a model including either the right or the left whole hippocampus volume, it did not perform as well as our model that included the parasubiculum and subiculum (misclassification rates were worse and R^2 values were weaker). Such findings have been highlighted previously (Mueller et al., 2010, see review Pini et al., 2016).

Thus far, relatively less research has been done analyzing the influence of cortical structures in classification models for control and MCI populations. Killiany et al. found the volume of the caudal portion of the anterior cingulate and the banks of the superior temporal sulcus to be important for discriminating controls, stable MCI, and MCI who decline (Killiany et al., 2000). We also found these two regions to be important classifiers in this study.

One of the first cross-sectional studies that used classification models to identify controls and MCI subjects based on hippocampal volume obtained a classification accuracy of 73.4% (Convit et al., 1997). Three years later, this group found that adding the fusiform gyrus and combined middle and inferior temporal gyri helped to identify who with MCI would further decline to AD (Convit et al., 2000). We also found the middle temporal gyrus to be important in our classification model, while the fusiform gyrus was less meaningful. Though these studies had subjects of different cognitive capacities, the overlapping ROIs in both studies suggest that volume loss in the middle temporal gyrus and the fusiform gyrus may contribute to cognitive impairment.

Table 2. Between-group measures of effect size

Region (age and eTIV-corrected)	Cohen's *d*
Left whole hippocampus	1.045175658
Left molecular layer	1.001047715
Left CA1	.97950074
Right molecular layer	.956542325
Left CA4	.925790159
Right dentate gyrus	.925375322
Right whole hippocampus	.925312224
Left subiculum	.90302485
Left CA2/3	.902831378
Left dentate gyrus	.902202961
Right CA4	.885329746
Right CA2/3	.852106001
Right CA1	.825208486
Left hippocampal tail	.810886649
Left HATA	.778884444
Right subiculum	.763229419
Right HATA	.641063049
Right hippocampal tail	.582298616
Right entorhinal cortex	.576953066
Left fimbria	.572899506
Left entorhinal cortex	.520818628
Right fimbria	.497658825
Left presubiculum	.469300487
Right presubiculum	.363948262
Left hippocampal fissure	.264820413
Right hippocampal fissure	.199064252
Left parasubiculum	−.036346266
Right parasubiculum	−.116042635

Note: CA, cornu ammonis; eTIV, estimated intracranial volume; HATA, hippocampal amygdala transition area.

Table 3. ROI classifiers of group membership

Region (age and eTIV-corrected)	FDR *p*-value (effect likelihood ratio test)
Left superior temporal cortex	.00001*
Right caudal anterior cingulate	.00001*
Right pars opercularis	.00001*
Left subiculum	.00002*
Right precentral cortex	.00011*
Left caudal middle frontal cortex	.00011*
Left rostral middle frontal cortex	.00015*
Left pars orbitalis	.00015*
Left middle temporal cortex	.00090*
Right insula	.00116*
Left banks of the superior temporal sulcus	.00117*
Right parasubiculum	.00287*
Right paracentral lobule	.00287*
Left fusiform cortex	.01654
Left transverse temporal cortex	.03894
Left cuneus cortex	.12556
Left entorhinal cortex	.12738
Left pericalcarine	.28605

Note: *p*-Value < .01.
eTIV, estimated intracranial volume; FDR, false discovery rate.
*Denotes significance.

Instead of creating classification models, other studies identify regions that show the greatest amount of volume loss in MCI subjects compared with controls. Such studies examining volumetric differences often report regions within the parietal and lateral temporal lobes to be implicated in the early stages of the AD continuum following the medial temporal lobe (Desikan et al., 2009; Fan et al., 2008; Hängii et al., 2011). These studies report volume reductions in many of the regions that were important in our classification model, such as the insula (Fan et al., 2008; Karas et al., 2004), the superior temporal cortex (Fan et al., 2008; Karas et al., 2004), and the middle temporal cortex (Fan et al., 2008). Therefore, while these studies did not make classification models like the present study, the re-occurring significance of regions such as the insula and middle temporal cortex suggests that such regions have a role in cognitive impairment and disease.

Demographics

In this study, the groups were well matched in terms of their demographics (i.e., age, education, APOE ε4 status, and gender) (Table 1). It is noteworthy that there was a slightly greater percentage of individuals in the control group who had APOE ε4 than in the MCI group. Since APOE ε4 status was not used as a selection criterion, this merely reflects the sample to which we had access. Neuropsychological measures revealed expected differences between the control and MCI groups. The control group performed better than the MCI group on four of the five neuropsychological tasks (MMSE, Logical Memory Immediate Recall, Logical Memory Delayed Recall, and Part B of the Trailmaking Test) (Table 1). A difference in GDS score approached significance ($p = .025$) with the MCI group endorsing, on average, .78 points higher on this scale. It is feasible to assume that this difference was a result of the GDS specifically asking about a decrease in memory, rather than a reflection of true depression symptoms.

MRI

We found smaller residual volumes in the MCI group in 18 of the 24 hippocampal subfields, bilateral hippocampal formations, and bilateral entorhinal cortices, compared with the controls. These findings are consistent with previous reports (Du et al., 2001; Hanseeuw et al., 2011; Killiany et al., 2000; La Joie et al., 2013; Pennanen et al., 2004, see review Zhou et al., 2016).

LIMITATIONS

While the best fit model created in this study is promising, there are limitations to consider. This study utilized sufficient data to meet the intended goals, but the sample size remains

Table 4. Model statistics of all stepwise and nominal logistic regressions

Stage	Variables	χ^2	p-Value	R^2	AIC	BIC	Classification accuracy (out of 96) Correct	Miss
Targeted nominal logistic	24 hippocampal subfields	48.23612	.0024*	.3629	153.253	198.79	72	24
Targeted nominal logistic	2 hippocampal formations	24.1557	<.0001*	.1817	115.023	122.455	68	28
Targeted nominal logistic	2 entorhinal cortices	9.1421	.0103	.0688	130.036	137.468	60	36
Preliminary stepwise	12 right hippocampal subfields	27.04459	<.0001*	.2035	114.313	124.13	64	32
Preliminary stepwise	12 left hippocampal subfields	30.81566	<.0001*	.2318	112.769	124.924	67	29
Preliminary stepwise	6 right subcortical	15.5924	.0036*	.1173	127.992	140.147	63	33
Preliminary stepwise	6 left subcortical	21.52908	<.0001*	.162	119.828	129.646	66	30
Preliminary stepwise	34 right cortical	58.2929	<.0001*	.4386	110.625	143.09	78	18
Preliminary stepwise	32 left cortical	49.2495	<.0001*	.3705	116.853	147.569	77	19
Mid-level stepwise	5 hippocampal subfields	29.79788	<.0001*	.2242	111.559	121.377	67	29
Mid-level stepwise	20 cortical and 4 subcortical	79.07338	<.0001*	.5949	92.7302	126.874	85	11
Upper mid-level stepwise	15 subcortical/cortical and 3 hippocampal subfields	94.6322	<.0001*	.712	80.1315	115.879	87	9
Final nominal logistic	11 cortical and 2 hippocampal subfields	97.16355	<.0001*	.731	83.754	122.477	90	6

Note: p-Value < .01.
AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion.
*Denotes significance.

modest. This was driven by a desire to make optimal use of control and MCI subjects from our local ADC population as it more closely resembles a clinical population. We supplemented the subject number using subjects from the ADNI study, though we did not want to overwhelm the study with the "clinical trials population" found in ADNI (Petersen et al., 2013). Furthermore, when working with classification models such as nominal logistic regression, the ultimate goal is to build a model using one dataset and then apply it to a parallel dataset. The very large number of comparisons and variables included in the models limited the strength of our findings. Since we aimed to investigate regions that are less studied in MCI, we used all available subjects to build this model. As such, we realize the potential to be able to refine our findings in a future study as more local subjects become available.

Further research creating classification models of control and MCI subjects utilizing medial temporal regions as well as less studied cortical regions will likely help clarify whether the 13 regions in our model are truly classifiers of cognitive impairment *versus* characteristics of our sample.

CONCLUSION

The results of our study highlight the utility of using measures of the hippocampus, entorhinal cortex, and hippocampal subfields in classification models of controls and MCI subjects. Moreover, our findings provide evidence of the value of less studied regions beyond the medial temporal lobe and their ability to aid in differentiating between those who exhibit normal aging and those who do not.

ACKNOWLEDGMENTS

The authors wish to thank the BU-ADC and ADNI subjects for dedicating their time to create important databases that promise to advance our knowledge of aging through studies such as the present one. No authors involved in the writing of this manuscript had any conflicts of interest with this publication, and there has been no financial support for this work that could have influenced its outcome.

Data collection and sharing for this project was funded by the BU-ADC (NIA P30 AG013846), the ADNI (National Institutes of Health grant U01 AG024904), and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization

is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California (USC). ADNI data are disseminated by the Laboratory of NeuroImaging at USC.

REFERENCES

Aisen, P.S., Cummings, J., Jack, C.R., Morris, J.C., Sperling, R., Frölich, L., Jones, R.W., Dowsett, S.A., Matthews, B.R., Raskin, J., & Scheltens, P. (2017). On the path to 2025: understanding the Alzheimer's disease continuum. *Alzheimer's Research & Therapy*, *9*, 60.

Ashendorf, L., Alosco, M.L., Bing-Canar, H., Chapman, K.R., Martin, B., Chaisson, C.E., Dixon, D., Steinberg, E.G., Tripodis, Y., Kowall, N.W., & Stern, R.A., (2017). Clinical utility of select neuropsychological assessment battery tests in predicting functional abilities in dementia. *Archives Clinical Neuropsychology*, *33*, 530–540. doi: 10.1093/arclin/acx100.

Colliot, O., Chételat, G., Chupin, M., Desgranges, B., Magnin, B., Benali, H., Dubois, B., Garnero, L., Eustache, F., & Lehéricy, S. (2008). Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the Hippocampus. *Radiology*, *248*, 194–201.

Convit, A., de Asis, J., de Leon, M., Tarshish, C., de Santi, S., & Rusinek, H. (2000). Atrophy of the medial occipitotemporal, inferior, and middle temporal Gyri in non-demented elderly predict decline to Alzheimer's disease. *Neurobiology of Aging*, *21*, 19–26.

Convit, A., de Leon, M., Tarshish, C., de Santi, S., Tsui, W., Rusinek, H., & George, H. (1997). Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. *Neurobiology of Aging*, *18*(2), 131–138.

de Flores, R., La Joie, R., & Chételat, G. (2015). Structural imaging of hippocampal subfields in healthy aging and Alzheimer's disease. *Neuroscience*, *309*, 29–50.

Desikan, R.S., Cabral, H.J., Fischl, B., Guttmann, C.R., Blacker, D., Hyman, B.T., Albert, M.S., & Killiany, R.J. (2009). Temporoparietal MR imaging measures of atrophy in subjects with mild cognitive impairment that predict subsequent diagnosis of Alzheimer disease. *American Journal of Neuroradiology*, *30*(3), 532–538.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., & Albert, M.S. (2006). An automated labeling system for Subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980.

Du, A.T., Schuff, N., Amend, D., Laakso, M.P., Hsu, Y.Y., Jagust, W.J., Yaffe, K., Kramer, J.H., Reed, B., Norman, D., & Chui, H.C. (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*(4), 441–447.

Fan, Y., Batmanghelich, N., Clark, C., & Davatzikos, C. (2008). Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *NeuroImage*, *39*, 1731–1743.

Galetta, K.M., Chapman, K.R., Essis, M.D., Alosco, M.L., Gillard, D., Steinberg, E., Dixon, D., Martin, B., Chaisson, C., Kowall, N.W., & Tripodis, Y. (2017). Screening utility of the King–Devick test in mild cognitive impairment and Alzheimer's disease dementia. *Alzheimer Disease Associated Disorders*, *31*(2), 152–158.

Gómez-Isla, T., Price, J., McKeel Jr., D. Morris, J. Growdon, J., & Hyman, B. (1996). Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *Journal of Neuroscience*, *16*(14), 4491–4500.

Hängii, J., Streffer, J., Jäncke, L., & Hock, C. (2011). Volumes of lateral temporal and parietal structures distinguish between healthy aging, mild cognitive impairment, and Alzheimer's disease. *Journal of Alzheimer's Disease*, *26*, 719–734.

Hanseeuw, B., Van Leemput, K., Kavec, M., Grandin, C., Seron, X., & Ivanoiu, A. (2011). Mild cognitive impairment: differential atrophy in the hippocampal subfields. *American Journal of Neuroradiology*, *32*(9), 1658–1661.

Iglesias, J.E., Augustinack, J.C., Nguyen, K., Player, C.M., Player, A., Wright, M., Roy, N., Frosch, M.P., McKee, A.C., Wald, L.L., & Fischl, B. (2015). A computational atlas of the Hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. *NeuroImage*, *115*, 117–137.

Jack Jr, C.R. Bernstein, M.A. Fox, N.C. Thompson, P. Alexander, G. Harvey, D. Borowski, B. Britson, P.J. Whitwell, L.J. Ward, C., & Dale, A.M. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, *27*(4), 685–691.

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

Khan, W., Westman, E., Jones, N., Wahlund, L.O., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Soininen, H., Spenger, C., & Lovestone, S., (2015). Automated hippocampal subfield measures as predictors of conversion from mild cognitive impairment to Alzheimer's disease in two independent cohorts. *Brain Topography*, *28*(5), 746–759.

Killiany, R.J., Gomez-Isla, T., Moss, M., Kikinis, R., Sandor, T., Jolesz, F., Tanzi, R., Jones, K., Hyman, B.T., & Albert, M.S., (2000). Use of structural magnetic resonance imaging to predict who will get Alzheimer's disease. *Annals of Neurology*, *47*(4), 430–439.

La Joie, R., Perrotin, A., De La Sayette, V., Egret, S., Doeuvre, L., Belliard, S., Eustache, F., Desgranges, B., & Chételat, G. (2013). Hippocampal subfield volumetry in mild cognitive impairment, Alzheimer's disease and semantic dementia. *NeuroImage: Clinical*, *3*, 155–162.

Misra, C., Fan, Y., & Davatzikos, C. (2009). Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI. *NeuroImage*, *44*, 1415–1422.

Morris, J. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, *43*, 2412–2414.

Mueller, S., Schuff, N., Yaffe, K., Madison, C., Miller, B., & Weiner, M. (2010). Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's Disease. *Human Brain Mapping*, *31*(9), 1339–1347.

Pennanen, C., Kivipelto, M., Tuomainen, S., Hartikainen, P., Hänninen, T., Laakso, M.P., Hallikainen, M., Vanhanen, M., Nissinen, A., Helkala, E.L., & Vainio, P. (2004). Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiology of Aging*, *25*(3), 303–310.

Petersen, R., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., & Fratiglioni, L. (2014). Mild cognitive impairment: a concept in evolution. *Journal of Internal Medicine*, *275*, 214–228.

Petersen, R.C., Aisen, P., Boeve, B.F., Geda, Y.E., Ivnik, R.J., Knopman, D.S., Mielke, M., Pankratz, V.S., Roberts, R., Rocca, W.A., & Weigand, S. (2013). Criteria for mild cognitive impairment due to Alzheimer's disease in the community. *Annals of Neurology*, *74*(2), 199–208.

Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marizzoni, M., & Frisoni, G. (2016). Brain atrophy in Alzheimer's disease and aging. *Ageing Research Reviews*, *30*, 25–48.

Plant, C., Teipel, S.J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., Bokde, A.W., Hampel, H., & Ewers, M., (2010). Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *NeuroImage*, *50*(1), 162–174.

Pluta, J., Yushkevich, P., Das, S., & Wolk, D. (2012). In vivo analysis of hippocampal subfield atrophy in mild cognitive impairment via semi-automatic segmentation of T2-weighted MRI. *Journal of Alzheimer's Disease*, *31*(1), 85–99.

Spires-Jones, T. & Hyman, B. (2014). The intersection of amyloid beta and tau at synapses in Alzheimer's disease. *Neuron*, *82*(4), 756–771.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Cedarbaum, J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., & Luthman, J. (2015). Update of the Alzheimer's disease neuroimaging initiative: A review of papers published since its inception. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *11*(6), e1–e120.

Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J.S., Tunnard, C., Liu, Y., Collins, L., Evans, A., Mecocci, P., Vellas, B., & Tsolaki, M., (2011). Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *NeuroImage*, *54*(2), 1178–1187.

Xu, Y., Jack, C.R., O'brien, P.C., Kokmen, E., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, R.G., & Petersen, R.C. (2000). Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. *Neurology*, *54*(9), 1760–1767.

Zhou, M., Zhang, F., Zhao, L., Qian, J., & Dong, C. (2016). Entorhinal cortex: A good biomarker of mild cognitive impairment and mild Alzheimer's disease. *Reviews in the Neurosciences*, *27*(2), 185–195.