

Published in final edited form as:

Neuroimage. 2010 May 15; 51(1): 488–499. doi:10.1016/j.neuroimage.2009.12.125.

## Automated 3D mapping of baseline and 12-month associations between three verbal memory measures and hippocampal atrophy in 490 ADNI subjects

Liana G. Apostolova<sup>1,2</sup>, Jonathan H. Morra<sup>2</sup>, Amity E. Green<sup>1,2</sup>, Kristy S. Hwang<sup>1,2</sup>, Christina Avedissian<sup>2</sup>, Ellen Woo<sup>1</sup>, Jeffrey L. Cummings<sup>1,3</sup>, Arthur W. Toga<sup>1,2</sup>, Clifford R. Jack Jr.<sup>4</sup>, Michael W. Weiner<sup>5,6</sup>, Paul M. Thompson<sup>1,2</sup>, and the Alzheimer's Disease Neuroimaging Initiative

<sup>1</sup> Department of Neurology, David Geffen School of Medicine, UCLA, Los Angeles, CA

<sup>2</sup> Laboratory of Neuro Imaging, David Geffen School of Medicine, UCLA, Los Angeles, CA

<sup>3</sup> Department of Psychiatry and Biobehavioral Neurosciences, David Geffen School of Medicine, UCLA, Los Angeles, CA

<sup>4</sup> Department of Radiology, Mayo Clinic College of Medicine, Rochester, MN

<sup>5</sup> Departments of Radiology and Medicine, UC San Francisco, San Francisco, CA

<sup>6</sup> Department of Veterans Affairs Medical Center, San Francisco, CA

### Abstract

We used a previously validated automated machine learning algorithm based on adaptive boosting to segment the hippocampi in baseline and 12-month follow-up 3D T1-weighted brain MRIs of 150 cognitively normal elderly (NC), 245 mild cognitive impairment (MCI) and 97 DAT ADNI subjects. Using the radial distance mapping technique, we examined the hippocampal correlates of delayed recall performance on three well-established verbal memory tests – ADAScog delayed recall (ADAScog-DR), the Rey Auditory Verbal Learning Test -DR (AVLT-DR) and Wechsler Logical Memory II-DR (LM II-DR). We observed no significant correlations between delayed recall performance and hippocampal radial distance on any of the three verbal memory measures in NC. All three measures were associated with hippocampal volumes and radial distance in the full sample and in the MCI group at baseline and at follow-up. In DAT we observed stronger left-sided associations between hippocampal radial distance, LM II-DR and ADAScog-DR both at baseline and at follow-up. The strongest linkage between memory performance and hippocampal atrophy in the MCI sample was observed with the most challenging verbal memory test – the AVLT-DR, as opposed to the DAT sample where the least challenging test the ADAScog-DR showed strongest associations with the hippocampal structure. After controlling for baseline hippocampal atrophy, memory performance showed regionally specific associations with hippocampal radial distance in predominantly CA1 but also in subicular distribution.

---

Send correspondence to: Liana G. Apostolova, M.D., Mary S. Easton Center for Alzheimer's Disease Research, 10911 Weyburn Ave, 2<sup>nd</sup> floor, Los Angeles, CA 90095, Phone: +1 310/794-2551, Fax: +1 310/794-3148, lapostolova@mednet.ucla.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Introduction

Alzheimer's disease (AD) is the most common cause of dementia among the elderly. At the time when dementia of Alzheimer's type (DAT) can be clinically diagnosed with current criteria, AD pathology has already spread and irreversibly destroyed the brain parenchyma. Thus recent major efforts in AD research have concentrated on the search for disease-associated biomarkers that can reliably identify patients in prodromal DAT (e.g., MCI and pre-MCI) stages and support the expedited evaluation of novel disease-modifying therapies.

MCI is an intermediate state between normal aging and dementia. Amnesic MCI patients suffer from memory impairment while still enjoying functional lifestyles (Petersen, 2007). Most amnesic MCI patients have the pathological hallmarks of AD - neocortical senile plaques, neurofibrillary tangles, atrophy and neuronal loss in layer II of the entorhinal cortex (Jicha et al., 2006a; Price and Morris, 1999), but have not yet progressed sufficiently to meet criteria for DAT. MCI, and more recently pre-MCI, i.e. cognitively normal elderly who progress and develop MCI and DAT in the future (Apostolova et al., 2008), have become an intense scientific focus.

Historically, DAT clinical trials have relied on cognitive and functional outcome measures alone. In recent years, there has been increased interest in developing laboratory and imaging disease biomarkers in addition to cognitive and functional endpoints or even as substitutes for them, i.e., as surrogate markers (Cummings et al., 2007; Thal et al., 2006). Biomarkers are currently the only feasible approach to quantifying disease-associated changes in the pre-symptomatic AD (pre-DAT) stages (Cummings et al., 2007; Dubois et al., 2007).

One major international AD scientific effort, the Alzheimer's Disease Neuroimaging Initiative (ADNI), was established to collect and evaluate putative clinical, imaging and laboratory AD biomarkers. The ADNI (Principal Investigator: Michael W. Weiner, M.D., VA Medical Center and University of California – San Francisco) is a large multi-site longitudinal MRI and fluorodeoxyglucose positron emission tomography (FDG-PET) study of 200 elderly controls, 400 subjects with amnesic MCI, and 200 patients with DAT (Mueller et al., 2005) (also see <http://www.loni.ucla.edu/ADNI> and [ADNI-info.org](http://ADNI-info.org)).

Hippocampal atrophy remains the best-studied structural AD imaging biomarker to date. Hippocampal atrophy progresses steadily throughout the course of AD (Jack et al., 2000; Jack et al., 1998; Jack et al., 1997) and shows strong correlations with Braak and Braak pathological staging (Bobinski et al., 1997; Bobinski et al., 1995; Schonheit et al., 2004) and verbal memory performance (Apostolova et al., 2006b; de Toledo-Morrell et al., 2000). Hippocampal dysfunction manifests with memory loss (de Toledo-Morrell et al., 2000; Fleischman et al., 2005; Mortimer et al., 2004).

The present study uses ADNI baseline and one-year follow-up cognitive and imaging data from 490 subjects to examine the relationships between three memory tests and hippocampal atrophy. The cognitive portion of the Alzheimer's Disease Assessment Scale (ADAScog) (Welsh et al., 1994) is one of the most commonly used cognitive instruments in clinical trials in DAT (Doody et al., 2008; Mulnard et al., 2000; Rogers et al., 1998; Wilcock et al., 2000) and MCI (Petersen et al., 2005; Salloway et al., 2004). Its verbal memory portion is adopted from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) battery (Welsh et al., 1994) and a delayed recall test has been added. The Rey Auditory Verbal Learning Test (AVLT) and the Wechsler Memory Scale - Logical Memory II test (LM II) are commonly used by neuropsychologists to assess verbal memory and have been extensively validated for use in cognitively normal and demented subjects (Rey, 1964; Wechsler, 1987). LM II also has been utilized as a screening tool in MCI clinical trials (Petersen et al., 2005; Salloway et al., 2004).

## Methods

### Subjects

The ADNI is a 5-year longitudinal study of 800 adults, ages 55–90, including 400 amnesic MCI, 200 DAT and 200 NC subjects. The current analyses used all subjects with available baseline and 1-year follow-up cognitive and imaging data as of September 2008. The sample consisted of 490 subjects, of whom 97 were diagnosed with DAT, 245 with MCI and 148 who were cognitively normal elderly (NC). Diagnosis of DAT was based on the National Institute of Neurological and Communicative Disorders and Stroke and the AD and Related Disorders Association (NINCDS-ADRDA) criteria (McKhann et al., 1984). DAT subjects had Mini-Mental Examination (Folstein et al., 1975) (MMSE) scores between 20–26 and a Clinical Dementia Rating scale (Morris, 1993) (CDR) score of 0.5–1 at baseline; they may be considered mild DAT patients. All MCI subjects had memory complaints but did not meet criteria for dementia. They scored between 24–30 on the MMSE, had a global CDR score of 0.5 and a memory score of 0.5 or greater on the CDR. In addition, they also exhibited objective memory impairment on LM II. NC subjects did not meet criteria for MCI or DAT. Their MMSE scores were between 24–30 and their global CDR was 0. Subjects were excluded if they refused or were unable to undergo magnetic resonance imaging (MRI). Also excluded were those with other neurological disorders, active depression or history of psychiatric diagnosis, including major depression or alcohol or substance dependence, within the past 2 years, and those with less than 6 years of education or were not fluent in English or Spanish. The full list of inclusion/exclusion criteria may be accessed on pages 23–29 of the online ADNI protocol (see [http://www.adni-info.org/images/stories/Documentation/adni\\_protocol\\_03.02.2005\\_ss.pdf](http://www.adni-info.org/images/stories/Documentation/adni_protocol_03.02.2005_ss.pdf)).

### Cognitive measures

We used baseline and follow-up delayed verbal recall scores from three previously validated cognitive tests – the ADAScog delayed recall (ADAScog-DR), AVLT 30 minute delayed recall (AVLT-DR) and LM II delayed recall (LM II-DR). These data are freely distributed to interested researchers (see <http://www.loni.ucla.edu/ADNI> and [ADNI-info.org](http://adni-info.org)). All three measures were used as continuous variables in our analyses. In addition, we also included global cognitive scores from the MMSE and CDR sum of boxes (CDR-SOB) for comparison in some of the analyses. The 3D hippocampal maps showing these correlations in the full sample have been previously published elsewhere (Morra et al., 2008b; Morra et al., 2008c).

### Imaging data acquisition and preprocessing

All subjects were scanned with a standardized high-resolution MRI protocol (<http://www.loni.ucla.edu/ADNI/Research/Cores/index.shtml>) (Jack et al., 2008; Leow et al., 2006). Images were obtained on scanners developed by one of three manufacturers (General Electric Healthcare, Siemens Medical Solutions, and Philips Medical Systems). ADNI also collects data at 3.0 T from a subset of subjects, but to avoid having to model field strength effects in this study, only 1.5 T images were used. At each visit, two T1-weighted MRI scans were collected using a sagittal 3D MP-RAGE sequence for each subject. The TE/TR/TI (echo, repetition, and inversion time) parameters were optimized for best contrast to noise in a feasible acquisition time. The raw data had an acquisition matrix of 192×192×166 and voxel size 1.25×1.25×1.2 mm<sup>3</sup> in the x-, y-, and z- dimensions (Jack et al., 2008). An in-plane, zero-filled reconstruction (i.e., sinc interpolation) resulted in a 256×256 matrix and a reconstructed voxel size of 0.9375×0.9375×1.2 mm<sup>3</sup> in the x-, y-, and z- dimensions. The image with higher quality (of two that were obtained identically for each subject) was selected by the ADNI MRI quality control center at the Mayo Clinic (in Rochester, MN, USA) (Jack et al., 2008). Phantom-based geometric corrections were applied to ensure that spatial calibration was kept within a specific tolerance level for each scanner involved in the ADNI study (Gunter et al., 2006). Additional image corrections included GradWarp correction for geometric distortion due to gradient non-

linearity (Jovicich et al., 2006), a “B1-correction” for image intensity non-uniformity (Jack et al., 2008) and an “N3” bias field correction, for reducing intensity inhomogeneity (Sled et al., 1998). The B1-correction (Jack et al., 2008) is different from the N3 bias field correction as it adjusts for image intensity inhomogeneity due to the B1 magnetic field non-uniformity using calibration scans. B1 calibration scans are collected to correct the image intensity non-uniformity that results when RF transmission is performed with a more uniform body coil but MRI signal reception is performed with a less uniform head coil. By contrast, the “N3” bias field correction, for reducing intensity inhomogeneity (Sled et al., 1998), is an image post-processing routine that is not dependent on calibration scans derived from the scanner. It essentially adjusts the spatial profile of image intensities using a multiplicative spline function, to make the histogram as sharp as possible. It also aims to adjust for the central bright artifact that can occur due to the dielectric effect. Both the uncorrected and corrected image files are freely available to interested researchers at <http://www.loni.ucla.edu/ADNI>.

All brain scans were linearly registered to the International Consortium for Brain Mapping (ICBM-53) standard brain template (Mazziotta et al., 2001) with a 9-parameter (9P) transformation (3 translations, 3 rotations, 3 scales) using the Minctracc algorithm (Collins et al., 1994). Globally aligned images were resampled in an isotropic space of 220 voxels along each axis ( $x$ ,  $y$ , and  $z$ ), with a final voxel size of  $1 \text{ mm}^3$ .

### Automated segmentation

The hippocampi were segmented with our new automated machine-learning hippocampal segmentation approach (AdaBoost) based on a statistical method called adaptive boosting originally developed by Freund and Shapire (Freund and Shapire, 1997). The technique has been described in detail in several publications by Morra et al. (Morra et al., 2008a; Morra et al., 2008c; Morra et al., 2008d). AdaBoost uses a training set of image data to develop mathematical rules for classifying future data, i.e., for labeling each voxel in a new image as belonging to the hippocampus or not. The training set consists of small number of representative images and their manual segmentations (in this case 21 subjects - 7 NC, 7 MCI and 7 DAT subjects) delineated by an expert (A.E.G., inter-rater reliability: Cronbach’s Alpha=0.97, intra-rater reliability: Cronbach’s Alpha =0.98) using a well-established, detailed anatomical tracing protocol with high inter- and intra-rater reliability. Based on the specific feature information contained in the positive and negative voxels of the training dataset (i.e., those belonging and not belonging to the structure of interest), AdaBoost develops a set of rules and computes the optimal combination of features for accurate segmentation of unknown images. Thousands of local features are taken into account, such as image gradients, local curvature of image interfaces, tissue classification as gray or white matter, and also statistical information on the likely stereotaxic position of the hippocampus. Using established numerical procedures from the fields of machine learning and computer vision (Morra et al., 2009), the training phase estimates the optimal weighting of these features in a mathematical formula that computes the probability of being inside the hippocampus. The algorithm’s performance has been validated in prior reports, and, when labeling new data previously unseen by the algorithm, it has been found to agree with human raters as well as human raters agree with each other (Morra et al., 2008a). Once a successful classification model is created the AdaBoost algorithm is applied to the full study cohort (in this case to all 490 baseline and follow-up scans).

### Radial distance mapping

After converting each hippocampal segmentation into a 3D parametric mesh model, we computed the medial core (a 3D medial curve threading down the center of each structure). The radial distance from each 3D hippocampal surface point to the medial core was computed (Apostolova et al., 2006a; Apostolova et al., 2006b). This provides a measure of the thickness of the structure at each surface point. The cognitive scores were then entered as covariates in

a general linear model predicting the radial distance at each surface point of the mesh models. Associations between cognitive performance and hippocampal radial distance were sought in the full sample (N=490) and separately for the NC, MCI and DAT groups.

## Statistical methods

We used one-way Analyses of Variance (ANOVAs) with a *post hoc* Bonferroni correction for multiple comparisons for continuous variables to examine group differences in baseline and follow-up measures of age, education, MMSE, LM II-DR, AVLT-DR and ADAScog-DR. A chi-squared test for categorical variables was used to determine any group differences in gender. Pearson's correlation analyses were used to investigate possible associations between hippocampal volume and memory performance in the full sample, and separately within each diagnostic group.

Linear regression models were used to map the association between LM II-DR, AVLT-DR, ADAScog-DR and hippocampal radial distance at baseline and at follow-up in all subjects, and separately within each diagnostic group, as well as to map the associations between change in memory performance and hippocampal thinning over the 12 month follow-up period. The 3D statistical maps were further subjected to multiple comparisons correction by permutation analyses (permuting the predictor variable in this case the memory scores) with the stringent threshold of  $p < 0.01$ . Permutation tests on maps have been widely used in the brain mapping literature but there are differences among the approaches. Other than FDR tests there are 3 types of tests commonly used on statistical maps – 1) peak height, 2) cluster size, 3) total supra-threshold volume (total volume of all clusters any size, i.e., set-level inference). These are described in detail in Frakowiak 1997 (Frackowiak et al., 2007). The approach used here, from Thompson et al. (Thompson et al., 2003), differs somewhat from the approach of Nichols and Holmes (Nichols and Holmes, 2002), and the two approaches aim to control different error rates. Some permutation-based approaches, e.g., Nichols and Holmes (Nichols and Holmes, 2002), aim to control the family-wise error rate (chance of one or more false positives in the entire map) based on the permutation distribution of the image-wise maximum statistic, as they build an empirically-based null distribution for the image-wise maximum statistic based on randomizations of the data. The most extreme 5% of the null distribution for the maximal statistic may be used to threshold the raw statistical map. This allows one to reject the null hypothesis at individual voxels while knowing that the chance of family wise error (FWE) is controlled at 5%. Corrected  $p$ -values for each voxel are obtained by evaluating the percentage of the permutation distribution for the maximal statistic that exceeds the voxel statistic. This allows particular points to be declared significant. In contrast, our approach (Thompson et al., 2003), determines a single corrected  $p$ -value for each map (which is reported in Table 3), based on the number of points surviving a particular *a priori* threshold (which we set to 0.01 in our analyses). When this is used, one may argue that an overall-significant map must contain one or more (corrected) significant points, but this does not allow the interpretation of the  $p$ -values at each point as being corrected. Even so, our approach is akin to set-level inference in functional imaging, which is more sensitive to a distributed pattern of weak effects than null distributions based on the maximum statistic. As one moves from peak height to cluster size and set-inference approaches (we use the latter in this and our other papers) there is a trade off of localization ability for statistical power as it is easier for the total supra-threshold volume to catch effects all over the structure even if it does not coalesce into regions that exceed a pre-set number of voxels. We find this to be the best-suited approach for weak distributed effects.

Overall permutation corrected  $p$ -values for our statistical maps are presented in Table 3, but the maps themselves are illustrated with uncorrected  $p$ -values.

We conducted a separate linear regression analysis with hippocampal radial distance at follow-up as the dependent variable and memory test scores in follow-up as the predictor variable while controlling for baseline hippocampal volume in the full dataset and in each diagnostic groups. To avoid issues about in-sync permuting in the settings of two linear regression covariates for these maps we applied a map-wise false discovery rate (FDR) correction (Benjamini and Hochberg, 1995). These methods are widely used for multiple comparisons correction in statistical brain maps, and to derive a statistical threshold (critical value,  $t$ ) where possible, that controls the expected proportion of false positive results (FDR) in the map. If there is such a threshold that controls the FDR at 5%, then the pattern of results in the map is declared, by convention, to be significant overall (Benjamini and Hochberg, 1995). Overall FDR corrected  $p$ -values for our statistical maps are presented in Table 4 while the maps themselves are illustrated with uncorrected  $p$ -values.

In addition we used the map-wise FDR correction method to derive cumulative distribution function (CDF) plots of the  $p$ -values in our main regression maps (Figures 1–3). These cumulative plots of  $p$ -values, or CDFs, allow for direct visual comparisons of the strength of correlations between each memory measure, MMSE and CDR-SOB and hippocampal radial distance. CDF plots can be used to rank statistical maps in terms of their effect sizes. In other words, statistical maps with CDFs that rise more steeply at the origin also to have a higher proportion of voxels with effect sizes exceeding any given fixed threshold. In the CDF plots, the  $x$  axis represents any arbitrary  $p$ -value threshold that is applied to the map (between 0 and 1), the  $y$  axis shows the proportion of the statistical map (i.e., a fraction between 0 and 1) showing effects that are more significant than that chosen  $p$ -value threshold. The  $y=20x$  line denotes the allowed 5% false discovery rate, which is the maximum proportion of false positives that is allowed for a map to be declared significant overall. If a given CDF function curves above the  $y=20x$  line, but then crosses it again at a point other than the origin, then the map shows statistically significant effect. In general, when effect sizes in the maps are greater overall, the CDF crosses the  $y=20x$  line at a higher statistical threshold ( $x$  value), meaning that a broader range of statistical thresholds can be applied to the data – and therefore more voxels reported as significant - while still keeping the false discovery rate below the conventional 5% (see Figure 6). In these graphs, we show the critical value,  $t$ , which is the highest threshold that can be applied to the statistical map while keeping the expected proportion of false positives below 5%. This is termed the “critical” value, e.g. the critical uncorrected  $p$ -value, which can maintain a certain, specified FDR.

## Results

The results from the Bonferroni-corrected ANOVAs and the chi-squared test for demographic and cognitive comparisons are presented in Table 1. There were no group differences in age. The MCI group had significantly more male subjects (65.3%) than the DAT (50.5%,  $p=0.01$ ) and NC (51.3%,  $p=0.004$ ) groups. The DAT group had fewer years of education relative to the MCI and NC groups [mean = 14.9 years for DAT vs. mean = 16 years for MCI ( $p=0.006$ ) and NC ( $p=0.02$ )]. As expected, all cognitive measures at baseline and at follow-up were significantly worse in the DAT and MCI groups relative to the NC group, as were the cognitive differences between the MCI and DAT groups (all  $p<0.001$  after Bonferroni correction for multiple comparisons).

### Volumetric results

The results of the correlation analyses between overall hippocampal volume and the memory tests in the full sample, and within each diagnostic group, are presented in Table 2. All three verbal memory measures showed significant correlations with hippocampal volume in the full sample and in the MCI group. In DAT, LM II-DR was correlated with both the left and right

hippocampus at baseline (left  $r=0.33$ ,  $p=0.001$  and right  $r=0.35$ ,  $p<0.0001$ ) but only with the left hippocampus in follow-up (left  $r=0.21$ ,  $p=0.044$ ), while ADAScog-DR was correlated only with the left hippocampus both at baseline ( $r=-0.28$ ,  $p=0.005$ ) and in follow-up ( $r=0.21$ ,  $p=0.04$ ). There were no significant correlations between hippocampal volume and memory performance in NC.

### 3D mapping results

The uncorrected significance maps for each memory measure are shown in Figures 1–3. Table 3 lists the global permutation-corrected significance of the statistical maps shown in Figures 1–3. In agreement with the volumetric results, no significant associations between hippocampal radial distance and verbal memory performance were detected in NC (see third row images in Figures 1–3), but significant associations were detected in the full sample (see second row images in Figures 1–3). Significant associations between memory and hippocampal radial distance were detected in both DAT and MCI. LM II-DR showed significant associations with hippocampal radial distance bilaterally in MCI at both time points but more significant in follow-up. (see Table 3 and the fourth row images in Figure 1). In DAT, significant associations were seen with LM II-DR bilaterally at baseline but only on the left in follow-up (see Table 3 and the bottom row images in Figure 1). AVLT-DR showed significant associations with hippocampal radial distance bilaterally in MCI both at baseline and at follow-up (see Table 3 and the fourth row images in Figure 2). DAT subjects failed to show significant associations with AVLT-DR at both time points probably as a result of floor effect (see Table 3 and the bottom row images in Figure 2). ADAScog-DR showed significant associations with hippocampal radial distance at baseline on the left hippocampus in MCI and bilaterally at follow-up (see Table 3 and the fourth row images in Figure 3). In DAT significant ADAScog-DR associations were seen only for the left hippocampus at both time points (see Table 3 and the bottom row images in Figure 3). To assess whether naturally occurring hippocampal asymmetry might lead to the findings above (more bilateral associations in MCI and more left sided associations in DAT) we conducted a post hoc left vs. right hippocampal radial distance comparison in each diagnostic group and in the full sample at each time point. Figure 4 demonstrates a very well-conserved asymmetry pattern from normal aging to DAT with larger posterior CA1 radial distances on the left and larger subicular and CA2-3 distances on the right. There were no changes in the asymmetry pattern from baseline to follow-up in any group and the pooled sample. All left vs. right comparisons across the three diagnostic groups and in the pooled sample both at baseline and in follow-up were highly statistically significant ( $p_{\text{corrected}}<0.001$  for all maps depicted in Figure 4).

Our analyses aimed to uncover correlations between hippocampal thinning over 12-months and 12-month change in verbal memory performance did not result in statistically significant results.

The results from the linear regression analyses of the relationship between follow-up memory performance with follow-up hippocampal radial distance while controlling for baseline hippocampal volume are shown in Figure 5 and supplemental figure on the journal's website. Figure 5 shows that after controlling for pre-existing atrophy memory performance shows strong regionally specific correlations in areas corresponding to the CA1 hippocampal subfield and parts of the subiculum. After applying FDR correction the pooled sample maps for all three verbal memory measures as well as the LM II-DR maps for the DAT group remained statistically significant (see Figure 5 and the supplemental figure on the journal's website as well as Table 4 listing the highest p-value threshold (critical value,  $t$ ) that keeps FDR at 5%).

## CDF plots

We used CDF plots to objectively compare and rank the associations between the three memory measures, MMSE and CDR-SOB and hippocampal radial distance, as well as to demonstrate which of the cognitive tests showed the best linkage with hippocampal atrophy within each diagnostic group (Figure 6). In agreement with the results presented so far, Figure 6 showed no significant associations between memory test score and hippocampal radial distance in NC, while the strongest associations were between AVLT-DR and hippocampal radial distance in MCI and between ADAScog-DR and hippocampal radial distance in DAT. MMSE and CDR-SOB were less sensitive than the memory measures in all three diagnostic groups.

## Patterns of correlations between memory measures and hippocampal subregions

As seen in Figures 1–3, the associations between memory performance and hippocampal structure are not uniformly distributed. In MCI and DAT, there is consistently a strong relationship between delayed recall performance and the lateral hippocampal area closely corresponding to the CA1 subfield. The inferior hippocampal surface, which captures most of the subiculum region, is another subregion that shows significant associations with memory performance in MCI and DAT. CA2 and CA3, or the top medial part of the hippocampus, show a correlation in MCI with AVLT-DR at baseline and follow-up, LM II-DR at follow-up and ADAScog-DR at baseline.

## Discussion

DAT is already an epidemic among the elderly in the US and worldwide. To address the pressing need to better understand and treat AD, many researchers are focused on developing and validating AD-related biomarkers - quantitative AD-associated measures that serve as an indirect metric of disease severity. In the present study, we analyzed the relationship between memory loss - the most pervasive AD symptom - and hippocampal atrophy, the most established AD imaging biomarker. We examined how well hippocampal radial distance, a measure of hippocampal thickness, correlates with one task that assesses delayed recall for short stories (LM II) and two tests of delayed recall of a list of unrelated words (AVLT and the verbal memory test from the ADAScog) across the full sample (N=490, consisting of 148 NC, 245 MCI and 97 mild DAT subjects) and separately within each diagnostic group. We used a newly developed high-throughput hippocampal automated segmentation technique that has been previously applied in MCI and DAT (Morra et al., 2008a; Morra et al., 2008c; Morra et al., 2008d). As this method shows promise as a potential analytic tool for clinical trials, we wanted to explore the associations between hippocampal morphology and several of the memory measures that have been repeatedly used as screening tests or as primary and secondary outcomes in MCI and DAT clinical trials.

None of the three verbal memory measures showed significant associations with hippocampal volume/morphology among cognitively normal elderly. In MCI, all three measures showed significant associations with atrophy in both the left and right hippocampi, with AVLT-DR showing the strongest linkages. One explanation for this finding is that subjects with mild cognitive problems find AVLT more challenging than the other two tasks. AVLT consists of a list of 15 words, whereas the ADAScog list consists of only 10 words. In addition, AVLT requires subjects to learn a distractor list following learning of the to-be-remembered words, which can interfere with consolidation of the first list. It is known that memory consolidation is highly dependent upon hippocampal functions. Further, the two word list tests (AVLT and ADAScog-DR) are comprised of unrelated words, in which it would be more difficult for subjects to utilize memory mnemonics (e.g., drawing associations among the words). In contrast, a short story (as used in LM II) provides both a context and built-in associations. Subjects are more readily able to recall the information content of short stories as opposed to



lists comprised of unrelated words as they can more easily utilize memory strategies, such as retaining the theme of the story and using pictorial rehearsal.

The observed stronger associations between LM II-DR and hippocampal atrophy in MCI at follow-up is another interesting observation. It could be due to two separate processes. First for some MCI subjects it could reflect disease progression where the progressive loss of the ability to compensate by means of these memory strategies results in a tighter hippocampal-memory performance association. On the other hand many MCI subjects demonstrated a learning effect (i.e., had improved LM II scores in follow-up, see Table 1). Perhaps the ones benefiting most were the subjects with the least amount of atrophy, which in turn further strengthened the association between memory performance and hippocampal radial atrophy.

The DAT subjects were the only group where differences were detected only on one hippocampus. They showed predominantly left sided hippocampal-memory recall associations. The greatest effect was observed with the cognitive test specifically designed for DAT subjects – the ADAScog. This task is the easiest of the three and showed higher performance levels in DAT subjects, suggesting that the ADAScog would be a better measure for studies hoping to provide above-floor performance levels. DAT subjects showed better performance on LM II-DR at baseline and a significant LM II-DR association with the hippocampal formations. However, this association becomes nonsignificant at follow-up when most DAT subjects can remember only one information unit on average (floor effect, see Table 1). AVLT-DR, the most challenging memory test of the three, showed no associations with hippocampal volume or radial distance in DAT, likely due to a floor effect (DAT subjects recalled an average of less than 1 word, see Table 1).

We used two different criteria to assess whether a map was significant after multiple comparisons correction, as both tests are somewhat prevalent in the brain mapping literature, although not always applied to the same data as they are usually considered as alternatives. First, the total supra-threshold surface area (with p-values more extreme than 0.01) was used and a corrected p-value was given for its rank in a null distribution obtained by randomization. And second, we used FDR theory to see if there was a statistical threshold that could be applied to the map that controlled the false discovery rate at the conventional 0.05 level. These are slightly different criteria and they are not always true in the same situations; they agreed for control and DAT groups but permutation gave slightly more powerful results in MCI. These two tests, which have different definitions, are generally true at the same time (i.e., they generally declare the same maps as significant), but for some effects one of the tests may work and the other may not. There is a point of connection between the CDF plots used in the FDR theory and the permutation tests, in that one could look up whether the map, thresholded at  $p=0.01$ , controls the FDR. By contrast, permutation tests find out the null distribution for this suprathreshold area by randomization. For the memory scores in the MCI subjects, the CDF plots show that the suprathreshold area for map thresholded at  $p=0.01$  is around 30 times higher than that which would be expected by chance for AVLT-DR, but only 5–10 times higher than that which would be expected by chance for the other memory scores. This number has to exceed 20 for the FDR to be considered to be controlled at 5% when the data are thresholded at  $p=0.01$ , i.e., 20% of the map has to be significant at  $p=0.01$ . By contrast, the permutation tests establish a non-parametric distribution for the suprathreshold area. As seen in Table 3, it was rare (less than 1 in 20 occurrences) for the suprathreshold area in randomized (null) data to exceed that seen by chance. It has to be concluded that permutation tests at the 0.01 threshold were sometimes more powerful in detecting effects than an FDR test applied at the same threshold. Even so, FDR is adaptive and considers all possible thresholdings of the data, so some effects in MCI, and all effects in DAT, had critical values lower than 0.01. As such they passed FDR as there was some threshold that controlled the false discovery rate.

Our regression analyses assessing associations between the change in memory scores and change in hippocampal radius were not significant. This may be due to the relatively short follow-up interval of only 12 months. Over such a short period of time, the changes in measures of disease progression such as memory scores and hippocampal volumes or thickness can be quite noisy. Table 1 shows that these measures show (1) very small mean changes over 12 month period, and (2) quite large standard deviations. In combination these make it hard to detect significant associations. We plan to repeat these analyses with the 2 or 3-year follow-up data once enough follow-up scans become available.

Once we controlled for baseline hippocampal volume in all subjects we uncovered highly significant regionally specific associations between follow-up memory test performance and hippocampal radial distance predominantly in CA1 distribution.

The CA1 subfield is known to be very susceptible to Alzheimer's type pathology. It is one of the first regions to be affected and throughout the disease course consistently shows the highest pathology burden (Bobinski et al., 1997; Bobinski et al., 1995; Schonheit et al., 2004; Zarow et al., 2005). The observed linkage between CA1 atrophy and verbal memory is in strong agreement with our previous report (Apostolova et al., 2006b) as well as with recent data from another research group. (Fletcher et al., 2008). In addition an association between CA1 and the global ADAScog score tapping more broadly into global cognitive function has also been reported (Csernansky et al., 2005a).

Most studies that have investigated the associations between memory performance and hippocampal atrophy to date have conducted their analyses on either a mixed samples of demented and nondemented subjects or in DAT subjects alone. Only a few studies have separated DAT and NC subjects in their analyses, and among those several reported a lack of or even a negative correlation between memory performance and hippocampal volume in cognitively normal elderly (Chantome et al., 1999; Foster et al., 1999; Kohler et al., 1998; Ylikoski et al., 2000). One recent study extracted hippocampal volumes from postmortem MR images and sought possible correlations between hippocampal volume and verbal memory performance within 1 year of death (Mortimer et al., 2004). The authors reported a weaker association between memory performance and hippocampal volume in nondemented vs. demented individuals. These reports are in line with our findings. Even though the hippocampus plays a crucial role for memory encoding and consolidation in NC, age-associated memory decline depends less on hippocampal and more on extra-hippocampal (e.g., white matter) integrity (Grady, 2008). Therefore, even though we found associations among the hippocampi and memory in the entire sample of controls, MCI, and DAT subjects, our data indicate the importance of analysing specific subgroups separately, as the correlations for the whole sample were driven by the MCI and DAT data.

The observed bilateral hippocampal/memory association in MCI is in agreement with our previous report (Apostolova et al., 2006b). We also observed a pronounced left-sided laterality effect in DAT as already documented by de Toledo-Morrell et al (de Toledo-Morrell et al., 2000). It has also been reported that atrophy of the left but not the right CA1 subfield was predictive of future progression from cognitively normal (CDR=0) to early DAT (CDR=0.5) (Csernansky et al., 2005b). We do not have a good explanation for this bilateral to unilateral shift from MCI to DAT but one could perhaps hypothesize that in the MCI stage, subjects have greater compensatory capabilities and it takes bilateral disruption of the hippocampal networks for verbal memory impairment to manifest itself. As AD pathology increases and the ability to employ complex, higher order compensatory strategies dissipates (i.e., in the DAT stage), the associations between memory and the hippocampus may become detectable on the left side of the brain where the majority of language processing takes place. Another plausible explanation is that the substantially larger MCI group sample size gives us sufficient power to

uncover right-sided verbal memory associations although admittedly our first report of bilateral hippocampal verbal memory associations in MCI came from a much smaller sample size (Apostolova et al., 2006b).

To our knowledge this is the largest study investigating the correlations between verbal memory and hippocampal structural integrity. The strengths of this study are its large size, the detailed subject assessment, the unified MRI protocol across multiple sites and its meticulous data quality control. Additional strengths are the advanced preprocessing and 3D modelling techniques used to map discrete structural-functional correlations from normal aging to dementia. One limitation of the study stems from the limited generalizability. The ADNI study uses rigorous exclusion criteria, as it was designed to closely resemble a clinical trial population. As such, it does not necessarily represent the general elderly population and its findings should be generalized with caution. Another weakness is the etiologic/pathologic uncertainty in the MCI stage. At least 30% of amnesic MCI have been found to harbor non-AD pathology (Jicha et al., 2006b). Such subjects if present in ADNI could be reducing our power to find statistically significant associations. Another limitation is that the link between an inward movement of the hippocampal surface and volumetric atrophy of the underlying subfield is yet to be validated. Such validation is difficult to do with 1.5T magnetic field strength. High-field and ultra-high field structural imaging can potentially provide us with enough resolution for subfield tracing and allow for such validation to take place.

In summary, our findings highlight the importance of the use of hippocampal atrophy as a biomarker by showing significant and robust relationships with the most pervasive symptom of AD, memory loss. Complex measures such as the AVLT-DR seem to be well suited for the MCI population and show strong association with hippocampal radial distance while DAT subjects may require less challenging verbal memory recall measures such as LM II-DR and ADAScog-DR.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Data used in preparing this article were obtained from the Alzheimer's Disease Neuroimaging Initiative database ([www.loni.ucla.edu/ADNI](http://www.loni.ucla.edu/ADNI)). Many ADNI investigators therefore contributed to the design and implementation of ADNI or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators is available at [www.loni.ucla.edu/ADNI/Collaboration/ADNI\\_Citation.shtml](http://www.loni.ucla.edu/ADNI/Collaboration/ADNI_Citation.shtml). All data collection was funded by the following ADNI funding sources (Principal Investigator: Michael Weiner; NIH grant number U01 AG024904): National Institute of Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Foundation for the National Institutes of Health, Pfizer Inc., Wyeth Research, Bristol-Myers Squibb, Eli Lilly and Company, GlaxoSmithKline, Merck & Co. Inc., AstraZeneca AB, Novartis Pharmaceuticals Corporation, the Alzheimer's Association, Eisai Global Clinical Development, Elan Corporation plc, Forest Laboratories, and the Institute for the Study of Aging (ISOA), with participation from the U.S. Food and Drug Administration. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. We thank the members of the ADNI Imaging Core for their contributions to the image pre-processing and the ADNI project.

All study analyses were funded by NIA K23 AG026803 (jointly sponsored by NIA, AFAR, The John A. Hartford Foundation, The Atlantic Philanthropies, The Starr Foundation and an anonymous donor; to LGA), the Turken Foundation (to LGA); NIA P 50 AG16570 (to LGA, JLC and PMT); NIBIB EB01651, NLM LM05639, NCR RR019771 (to PMT); and NIMH R01 MH071940, NCR P41 RR013642 and NIH U54 RR021813 (to AWT).

## References

- Apostolova LG, Dinov ID, Dutton RA, Hayashi KM, Toga AW, Cummings JL, Thompson PM. 3D comparison of hippocampal atrophy in amnesic mild cognitive impairment and Alzheimer's disease. *Brain* 2006a;129:2867–2873. [PubMed: 17018552]
- Apostolova LG, Dutton RA, Dinov ID, Hayashi KM, Toga AW, Cummings JL, Thompson PM. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. *Arch Neurol* 2006b;63:693–699. [PubMed: 16682538]
- Apostolova LG, Mosconi L, Thompson PM, Green AE, Mistur R, Tsui WH, de Leon MJ. Subregional hippocampal atrophy predicts future decline to Alzheimer's dementia in cognitively normal subjects. *Neurobiology of Aging*. 2008 in press. 10.1016/j.neurobiolaging.2008.08.008
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Series B* 1995;57:289–300.
- Bobinski M, Wegiel J, Tarnawski M, Bobinski M, Reisberg B, de Leon MJ, Miller DC, Wisniewski HM. Relationships between regional neuronal loss and neurofibrillary changes in the hippocampal formation and duration and severity of Alzheimer disease. *Journal of Neuropathology & Experimental Neurology* 1997;56:414–420. [PubMed: 9100672]
- Bobinski M, Wegiel J, Wisniewski HM, Tarnawski M, Reisberg B, Mlodzik B, de Leon MJ, Miller DC. Atrophy of hippocampal formation subdivisions correlates with stage and duration of Alzheimer disease. *Dementia* 1995;6:205–210. [PubMed: 7550600]
- Chantome M, Perruchet P, Hasboun D, Dormont D, Sahel M, Sourour N, Zouaoui A, Marsault C, Duyme M. Is there a negative correlation between explicit memory and hippocampal volume? *Neuroimage* 1999;10:589–595. [PubMed: 10547336]
- Collins DL, Neelin P, Peters TM, Evans AC. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* 1994;18:192–205. [PubMed: 8126267]
- Csernansky JG, Wang L, Miller JP, Galvin JE, Morris JC. Neuroanatomical predictors of response to donepezil therapy in patients with dementia. *Arch Neurol* 2005a;62:1718–1722. [PubMed: 16286546]
- Csernansky JG, Wang L, Swank J, Miller JP, Gado M, McKeel D, Miller MI, Morris JC. Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. *Neuroimage* 2005b;25:783–792. [PubMed: 15808979]
- Cummings JL, Doody R, Clark C. Disease-modifying therapies for Alzheimer disease: challenges to early intervention. *Neurology* 2007;69:1622–1634. [PubMed: 17938373]
- de Toledo-Morrell L, Dickerson B, Sullivan MP, Spanovic C, Wilson R, Bennett DA. Hemispheric differences in hippocampal volume predict verbal and spatial memory performance in patients with Alzheimer's disease. *Hippocampus* 2000;10:136–142. [PubMed: 10791835]
- Doody RS, Gavrilova SI, Sano M, Thomas RG, Aisen PS, Bachurin SO, Seely L, Hung D. Effect of dimebon on cognition, activities of daily living, behaviour, and global function in patients with mild-to-moderate Alzheimer's disease: a randomised, double-blind, placebo-controlled study. *Lancet* 2008;372:207–215. [PubMed: 18640457]
- Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology* 2007;6:734–746. [PubMed: 17616482]
- Fleischman DA, Wilson RS, Gabrieli JD, Schneider JA, Bienias JL, Bennett DA. Implicit memory and Alzheimer's disease neuropathology. *Brain* 2005;128:2006–2015. [PubMed: 15975947]
- Fletcher, E.; Persianinova, M.; Carmichael, OT.; Beiser, A.; Seshadri, S.; Au, R.; Wolf, PA.; DeCarli, C. Hippocampal shape change correlates with delayed memory performance in middle aged adults of the Framingham heart study. *International Conference on Alzheimer's Disease and Related Disorders*; Chicago, IL. 2008.
- Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatric Res* 1975;12:189–198.
- Foster JK, Meikle A, Goodson G, Mayes AR, Howard M, Sunram SI, Cezayirli E, Roberts N. The hippocampus and delayed recall: bigger is not necessarily better? *Memory* 1999;7:715–732. [PubMed: 10659094]

- Frackowiak, RSJ.; Friston, KJ.; Frith, C.; Dolan, R.; Mazziotta, JC., editors. *Human Brain Function*. Academic Press; USA: 2007.
- Freund Y, Shapire R. A decision-theoretic generalization of online learning and an application to boosting. *J Comp Sys Sci* 1997;55:119–139.
- Grady CL. Cognitive neuroscience of aging. *Ann N Y Acad Sci* 2008;1124:127–144. [PubMed: 18400928]
- Gunter, J.; Bernstein, M.; Borowski, B.; Felmlee, J.; Blezek, D.; Mallozzi, R. Validation testing of the MRI calibration phantom for the Alzheimer's Disease Neuroimaging Initiative Study. ISMRM 14th Scientific Meeting and Exhibition; 2006.
- Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, JLW, Ward C, Dale AM, Felmlee JP, Gunter JL, Hill DL, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, DeCarli CS, Krueger G, Ward HA, Metzger GJ, Scott KT, Mallozzi R, Blezek D, Levy J, Debbins JP, Fleisher AS, Albert M, Green R, Bartzokis G, Glover G, Mugler J, Weiner MW. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* 2008;27:685–691. [PubMed: 18302232]
- Jack CR Jr, Petersen RC, Xu Y, O'Brien PC, Smith GE, Ivnik RJ, Boeve BF, Tangalos EG, Kokmen E. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. *Neurology* 2000;55:484–489. [PubMed: 10953178]
- Jack CR Jr, Petersen RC, Xu Y, O'Brien PC, Smith GE, Ivnik RJ, Tangalos EG, Kokmen E. Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology* 1998;51:993–999. [PubMed: 9781519]
- Jack CR Jr, Petersen RC, Xu YC, Waring SC, O'Brien PC, Tangalos EG, Smith GE, Ivnik RJ, Kokmen E. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. [see comment]. *Neurology* 1997;49:786–794. [PubMed: 9305341]
- Jicha GA, Parisi JE, Dickson DW, Johnson K, Cha R, Ivnik RJ, Tangalos EG, Boeve BF, Knopman DS, Braak H, Petersen RC. Neuropathologic Outcome of Mild Cognitive Impairment Following Progression to Clinical Dementia. *Archives of Neurology* 2006a;63:674–681. [PubMed: 16682537]
- Jicha GA, Parisi JE, Dickson DW, Johnson K, Cha R, Ivnik RJ, Tangalos EG, Boeve BF, Knopman DS, Braak H, Petersen RC. Neuropathologic outcome of mild cognitive impairment following progression to clinical dementia. *Arch Neurol* 2006b;63:674–681. [PubMed: 16682537]
- Jovicich J, Czanner S, Greve D, Haley E, van der Kouwe A, Gollub R, Kennedy D, Schmitt F, Brown G, Macfall J, Fischl B, Dale A. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *Neuroimage* 2006;30:436–443. [PubMed: 16300968]
- Kohler S, Black SE, Sinden M, Szekely C, Kidron D, Parker JL, Foster JK, Moscovitch M, Winocour G, Szalai JP, Bronskill MJ. Memory impairments associated with hippocampal versus parahippocampal-gyrus atrophy: an MR volumetry study in Alzheimer's disease. *Neuropsychologia* 1998;36:901–914. [PubMed: 9740363]
- Leow AD, Klunder AD, Jack CR Jr, Toga AW, Dale AM, Bernstein MA, Britson PJ, Gunter JL, Ward CP, Whitwell JL, Borowski BJ, Fleisher AS, Fox NC, Harvey D, Kornak J, Schuff N, Studholme C, Alexander GE, Weiner MW, Thompson PM. Longitudinal stability of MRI for mapping brain change using tensor-based morphometry. *Neuroimage* 2006;31:627–640. [PubMed: 16480900]
- Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci* 2001;356:1293–1322. [PubMed: 11545704]
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34:939–944. [PubMed: 6610841]
- Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR Jr, Weiner MW, Thompson PM. Validation of a fully automated 3D hippocampal

segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *Neuroimage* 2008a;43:59–68. [PubMed: 18675918]

Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR, Schuff N, Weiner MW, Thompson PM. Mapping hippocampal degeneration in 400 subjects with a novel automated segmentation approach. ISBI. 2008b in press.

Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Toga AW, Jack CR Jr, Schuff N, Weiner MW, Thompson PM. Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. *Neuroimage*. 2008c

Morra JH, Tu Z, Apostolova LG, Green AE, Toga AW, Thompson PM. Automatic subcortical segmentation using a contextual model. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv* 2008d;11:194–201.

Morra, JH.; Tu, Z.; Toga, AW.; Thompson, PM. Machine learning for brain image segmentation. Gonzalez, F.; Romero, E., editors. *Biomedical Image Analysis and Machine Learning Technologies*; 2009.

Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 1993;43:2412–2414. [PubMed: 8232972]

Mortimer JA, Gosche KM, Riley KP, Markesbery WR, Snowdon DA. Delayed recall, hippocampal volume and Alzheimer neuropathology: findings from the Nun Study. *Neurology* 2004;62:428–432. [PubMed: 14872025]

Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L. Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement* 2005;1:55–66. [PubMed: 17476317]

Mulnard RA, Cotman CW, Kawas C, van Dyck CH, Sano M, Doody R, Koss E, Pfeiffer E, Jin S, Gamst A, Grundman M, Thomas R, Thal LJ. Estrogen replacement therapy for treatment of mild to moderate Alzheimer disease: a randomized controlled trial. *Alzheimer's Disease Cooperative Study*. *JAMA* 2000;283:1007–1015. [PubMed: 10697060]

Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* 2002;15:1–25. [PubMed: 11747097]

Petersen RC. Mild cognitive impairment. *Continuum Lifelong Learning Neurol* 2007;13:13–36.

Petersen RC, Thomas RG, Grundman M, Bennett D, Doody R, Ferris S, Galasko D, Jin S, Kaye J, Levey A, Pfeiffer E, Sano M, van Dyck CH, Thal LJ. Vitamin E and donepezil for the treatment of mild cognitive impairment. *N Engl J Med* 2005;352:2379–2388. [PubMed: 15829527]

Price JL, Morris JC. Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. *Annals of Neurology* 1999;45:358–368. [PubMed: 10072051]

Rey, A. *L'examen Clinique En Psychologie*. Presses Universitaires de France; Paris, France: 1964.

Rogers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Donepezil Study Group. *Neurology* 1998;50:136–145. [PubMed: 9443470]

Salloway S, Ferris S, Kluger A, Goldman R, Griesing T, Kumar D, Richardson S. Efficacy of donepezil in mild cognitive impairment: a randomized placebo-controlled trial. *Neurology* 2004;63:651–657. [PubMed: 15326237]

Schonheit B, Zarski R, Ohm TG. Spatial and temporal relationships between plaques and tangles in Alzheimer-pathology. *Neurobiology of Aging* 2004;25:697–711. [PubMed: 15165691]

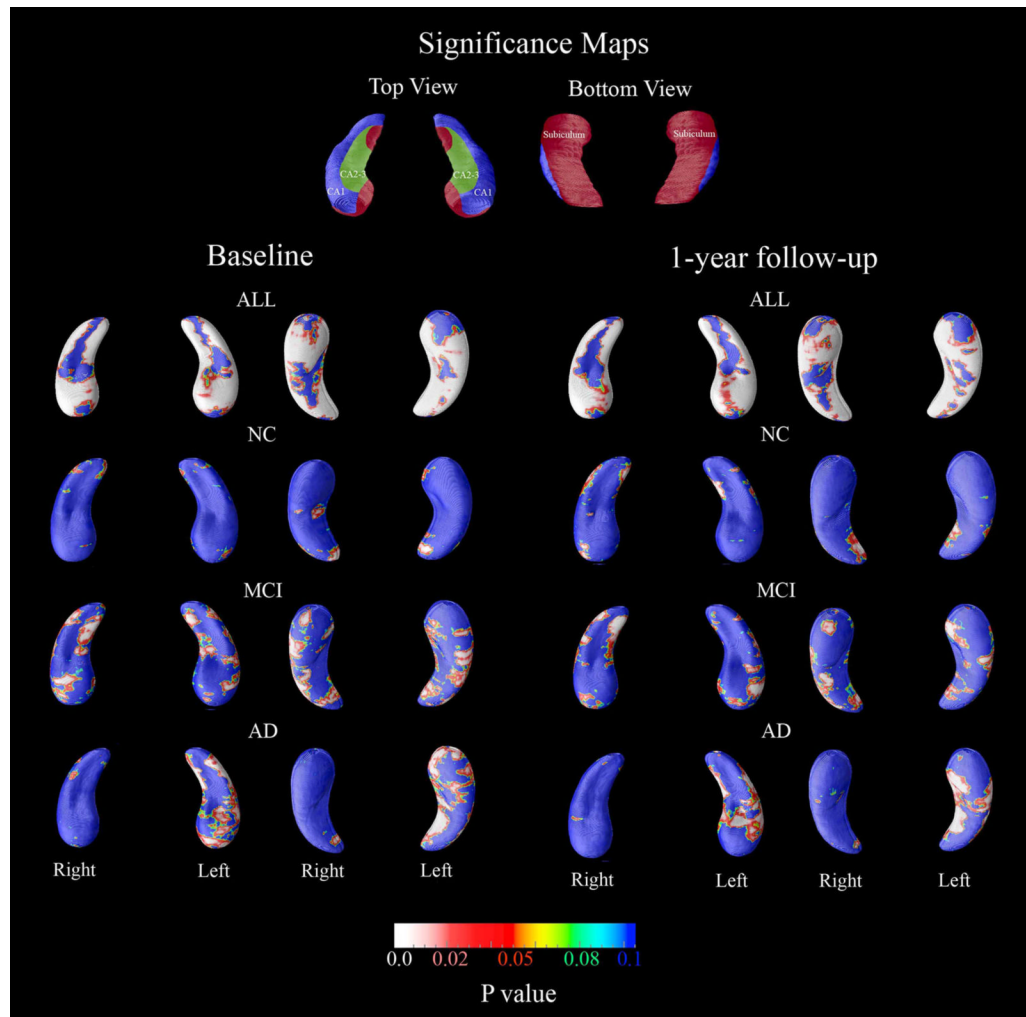
Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 1998;17:87–97. [PubMed: 9617910]

Thal LJ, Kantarci K, Reiman EM, Klunk WE, Weiner MW, Zetterberg H, Galasko D, Pratico D, Griffin S, Schenk D, Siemers E. The role of biomarkers in clinical trials for Alzheimer disease. *Alzheimer Disease & Associated Disorders* 2006;20:6–15. [PubMed: 16493230]

Thompson PM, Hayashi KM, de Zubicaray G, Janke AL, Rose SE, Semple J, Herman D, Hong MS, Dittmer SS, Doddrell DM, Toga AW. Dynamics of gray matter loss in Alzheimer's disease. *Journal of Neuroscience* 2003;23:994–1005. [PubMed: 12574429]

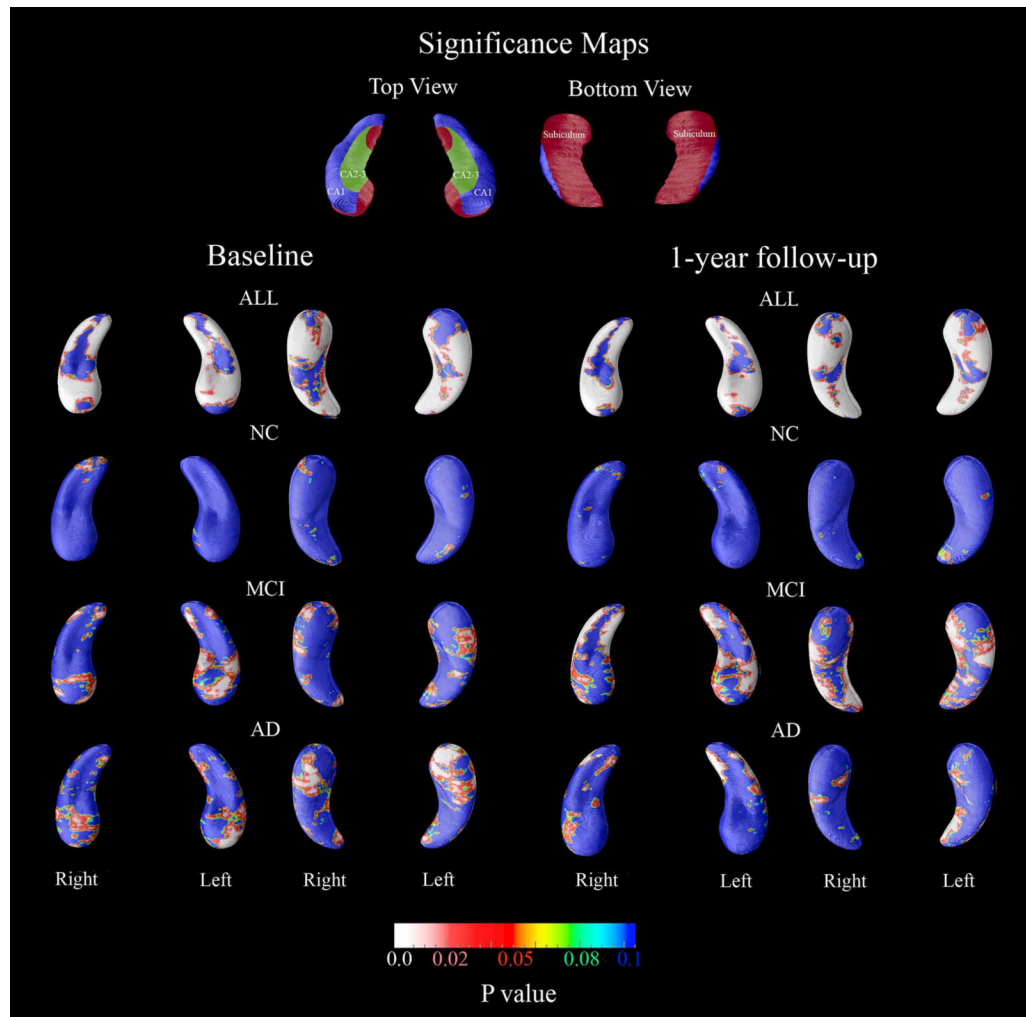
Wechsler, D. *Wechsler Memory Scale - Revised*. The Psychological Corporation; San Antonio, TX: 1987.

- Welsh KA, Butters N, Mohs RC, Beekly D, Edland S, Fillenbaum G, Heyman A. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology* 1994;44:609–614. [PubMed: 8164812]
- Wilcock GK, Lilienfeld S, Gaens E. Efficacy and safety of galantamine in patients with mild to moderate Alzheimer's disease: multicentre randomised controlled trial. Galantamine International-1 Study Group. *BMJ* 2000;321:1445–1449. [PubMed: 11110737]
- Ylikoski R, Salonen O, Mantyla R, Ylikoski A, Keskivaara P, Leskela M, Erkinjuntti T. Hippocampal and temporal lobe atrophy and age-related decline in memory. *Acta Neurol Scand* 2000;101:273–278. [PubMed: 10770526]
- Zarow C, Vinters HV, Ellis WG, Weiner MW, Mungas D, White L, Chui HC. Correlates of hippocampal neuron number in Alzheimer's disease and ischemic vascular dementia. *Annals of Neurology* 2005;57:896–903. [PubMed: 15929035]

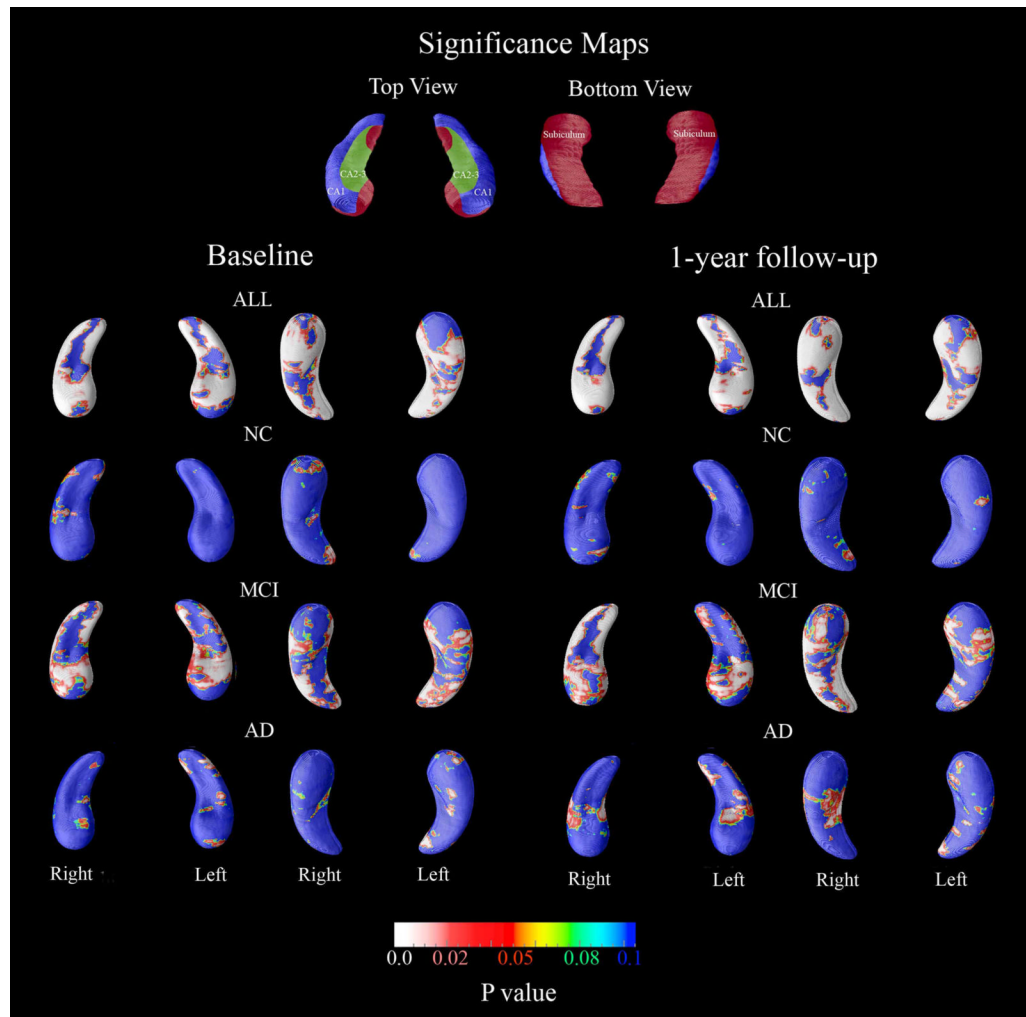


**Figure 1.** 3D significance maps showing the associations between LM II-DR and hippocampal radial distance. In the significance maps red and white colors denote  $p_{\text{uncorrected}} < 0.05$ . The final permutation corrected global p-values are listed in Table 3.

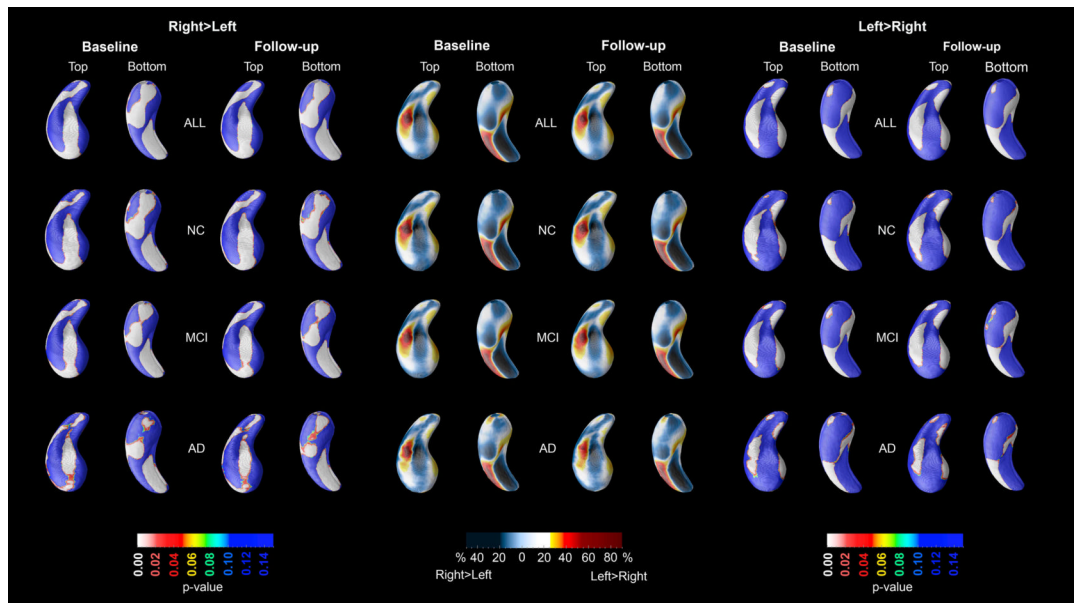




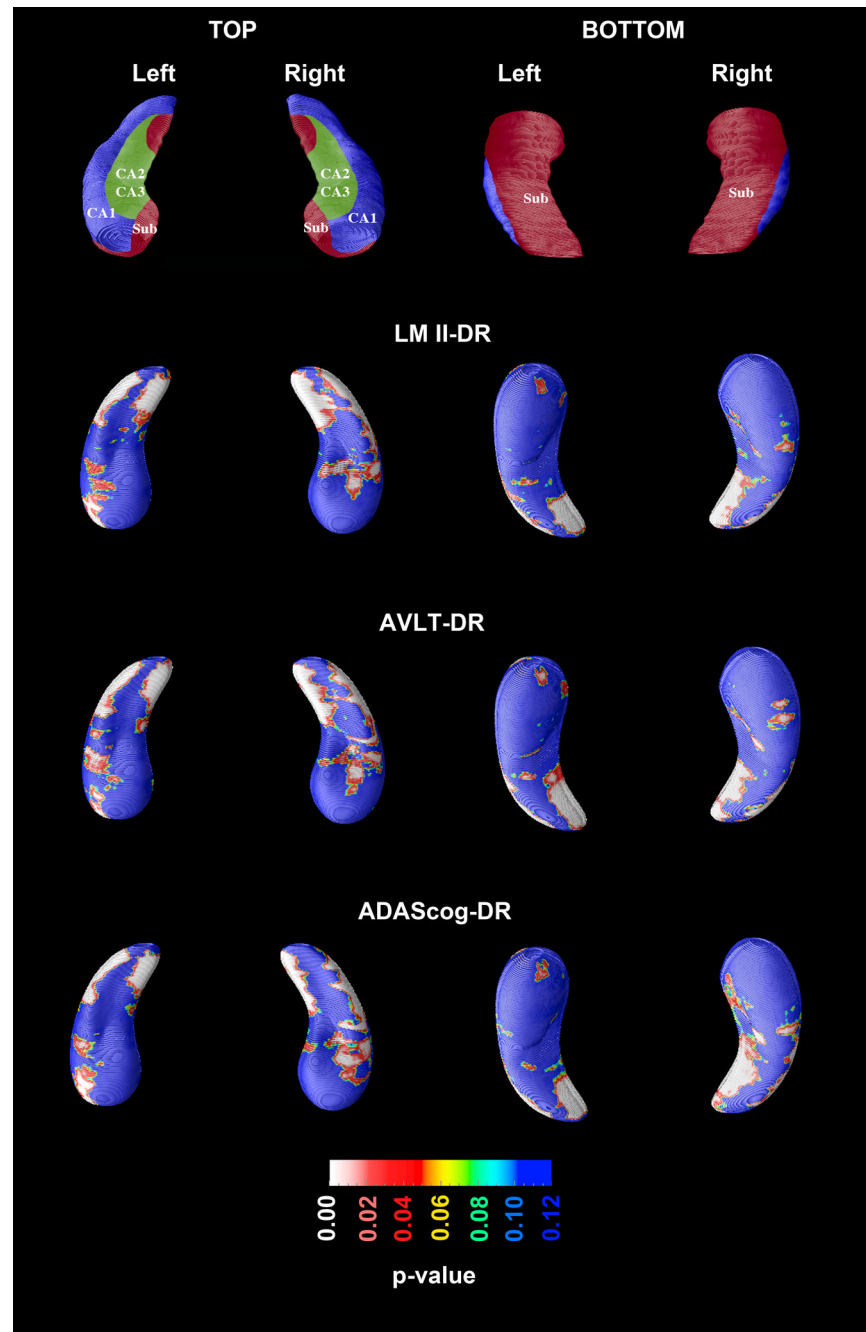
**Figure 2.** 3D significance maps showing the associations between AVLT-DR and hippocampal radial distance. In the significance maps, red and white colors denote  $p_{\text{uncorrected}} < 0.05$ . The final permutation corrected global p-values are listed in Table 3.



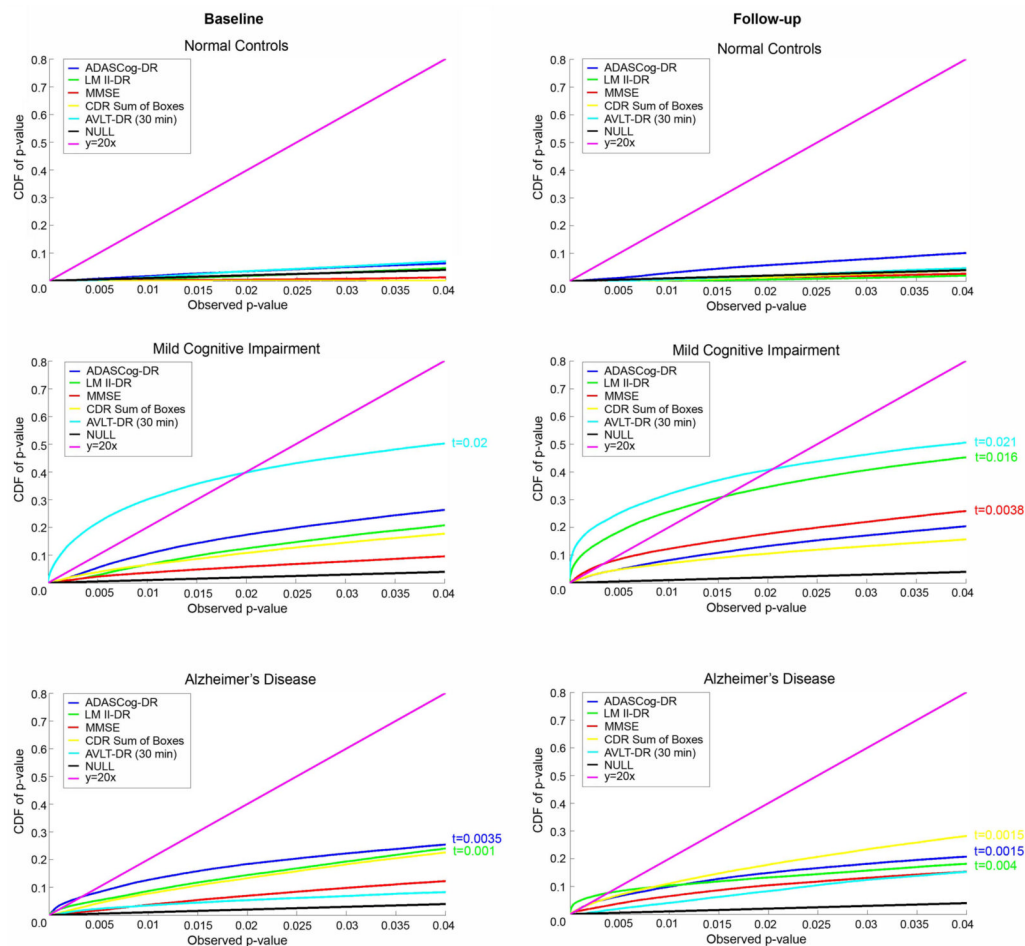
**Figure 3.** 3D significance maps showing the associations between ADAScog-DR and hippocampal radial distance. In the significance maps, red and white colors denote  $p_{\text{uncorrected}} < 0.05$ . The final permutation corrected global p-values are listed in Table 3.



**Figure 4.**  
Left vs. right asymmetry maps for each diagnostic group and the pooled sample.



**Figure 5.** 3D significance maps showing the associations between follow-up memory test scores and 12-month hippocampal radial distance while controlling for baseline hippocampal volume in the full sample. The maps of the associations in each diagnostic group can be seen in the Supplementary material section on the journal's website. In the significance maps, red and white colors denote  $p_{\text{uncorrected}} < 0.05$ . The final FDR corrected global p-values are listed in Table 4.



**Figure 6.** Cumulative distribution function (CDF) plots. These plots based on the FDR method objectively compare and rank the associations between each memory measure and hippocampal radial distance; the higher a function crosses the purple diagonal line (the  $y=20x$  line), the greater its significance. If a function does not cross the  $y=20x$  line at any point other than the origin then pattern of effects in the statistical map used to generate the plot was not greater than what is expected by chance alone under the null hypothesis (i.e.,  $<5\%$ ). The intersection of the CDF and the diagonal line is here called the critical value, denoted here by  $t$ , i.e., the highest threshold that can be applied to the statistical map while keeping the expected proportion of false positives below 5%.

**Table 1**

Between-group statistical comparisons of demographic and clinical variables conducted with one-way ANOVA for continuous variables and Pearson's chi-square analysis for categorical variables.

Variable, mean (SD)	NC N=150	MCI N=245	DAT N=97	P value
Age at baseline, years	75.9 (4.9)	74.9 (7.2)	75.8 (7.3)	0.3
Gender, % male	51.3%	65.3%	50.5%	<b>0.004</b>
Education, years	15.98 (2.8)	15.98 (3.0)	14.88 (3.1)	<b>0.005</b>
MMSE at baseline	29.1 (.9)	26.9 (1.8)	23.39 (1.9)	<b>&lt;0.001</b>
MMSE at follow-up	29.1 (1.1)	26.3 (3.0)	21.52 (4.4)	<b>&lt;0.001</b>
ADAScog-DR at baseline	2.8 (1.7)	6.3 (2.3)	8.5 (1.6)	<b>&lt;0.001</b>
ADAScog-DR at follow-up	2.7 (2.0)	6.2 (3.1)	8.2 (3.0)	<b>&lt;0.001</b>
<b>ADAScog-DR change</b>	<b>-0.1 (1.7)</b>	<b>-0.1 (2.6)</b>	<b>-0.3 (2.8)</b>	0.8
AVLT-DR at baseline	7.24 (3.6)	2.7 (3.1)	0.9 (1.9)	<b>&lt;0.001</b>
AVLT-DR at follow-up	7.95 (3.6)	2.3 (3.2)	0.3 (1.2)	<b>&lt;0.001</b>
<b>AVLT-DR change</b>	<b>-0.7 (3.3)</b>	<b>0.4 (2.3)</b>	<b>0.5 (1.7)</b>	<b>&lt;0.001</b>
LM II-DR at baseline	12.9 (3.4)	3.8 (2.8)	1.3 (1.9)	<b>&lt;0.001</b>
LM II-DR at follow-up	13.5 (4.3)	4.3 (4.5)	1.0 (3.0)	<b>&lt;0.001</b>
<b>LM II-DR change</b>	<b>-0.6 (3.6)</b>	<b>-0.5 (3.4)</b>	<b>0.2 (2.8)</b>	0.14
<b>Left hippocampal volume change</b>	<b>65 (251)</b>	<b>92 (232)</b>	<b>200 (281)</b>	<b>&lt;0.0001</b>
<b>Right hippocampal volume change</b>	<b>0.5 (250)</b>	<b>123 (251)</b>	<b>167 (269)</b>	<b>&lt;0.0001</b>

**Pearson** correlations (*r*) between hippocampal volume and cognitive measures after correcting for age, gender and education. Significant correlations appear in bold.

**Table 2**

Time point	Diagnostic group	LMII-DR		AVLT-DR		ADAScog-DR	
		Left	Right	Left	Right	Left	Right
Baseline	All subjects (N=490)	<b>0.3</b> p<0.0001	<b>0.3</b> p<0.0001	<b>0.26</b> p<0.0001	<b>0.31</b> p<0.0001	<b>-0.32</b> p<0.0001	<b>-0.32</b> p<0.0001
	NC (N=198)	0.02	0.01	-0.12	0.02	-0.02	-0.02
	MCI (N=245)	<b>0.19</b> p=0.003	<b>0.13</b> p=0.045	<b>0.27</b> p<0.0001	<b>0.3</b> p<0.0001	<b>-0.2</b> p=0.002	<b>-0.23</b> p<0.0001
Follow-up	DAT (N=98)	<b>0.33</b> p=0.001	<b>0.35</b> p<0.0001	0.13	0.11	<b>-0.28</b> p=0.005	-0.13
	All subjects (N=490)	<b>0.33</b> p<0.0001	<b>0.36</b> p<0.0001	<b>0.3</b> p<0.0001	<b>0.38</b> p<0.0001	<b>-0.32</b> p<0.0001	<b>-0.3</b> p<0.0001
	NC (N=198)	-0.05	0.02	-0.04	0.04	-0.04	-0.05
Follow-up	MCI (N=245)	<b>0.27</b> p<0.0001	<b>0.25</b> p<0.0001	<b>0.24</b> p<0.0001	<b>0.31</b> p<0.0001	<b>-0.2</b> p=0.002	<b>-0.17</b> p=0.008
	DAT (N=98)	<b>0.21</b> p=0.044	0.08	0.17	0.11	<b>-0.21</b> p=0.04	-0.08

**Table 3**

Permutation-corrected global significance of the maps shown in Figures 1–3.

Time point	Diagnostic group	LMII-DR		AVLT-DR		ADAS <sub>cog</sub> -DR	
		Left	Right	Left	Right	Left	Right
Baseline	All subjects (N=490)	<b>0.0001</b>	<b>0.0001</b>	<b>0.0001</b>	<b>0.0001</b>	<b>0.0001</b>	<b>0.0001</b>
	NC (N=198)	0.7	0.4	0.9	0.2	0.5	0.9
	MCI (N=245)	<b>0.05</b>	<b>0.03</b>	<b>0.0003</b>	<b>0.0001</b>	<b>0.02</b>	<b>0.0017</b>
Follow-up	DAT (N=98)	<b>0.005</b>	0.08	0.1	0.2	<b>0.0015</b>	0.2
	All subjects (N=490)	<b>0.0001</b>	<b>0.0001</b>	<b>0.0001</b>	<b>0.0001</b>	<b>0.0001</b>	<b>0.0001</b>
	NC (N=198)	0.6	0.5	0.5	0.7	0.7	0.5
Follow-up controlling for baseline hippocampal volume	MCI (N=245)	<b>0.0004</b>	<b>0.0002</b>	<b>0.002</b>	<b>0.0001</b>	<b>0.009</b>	<b>0.03</b>
	DAT (N=98)	<b>0.03</b>	0.09	0.1	0.1	<b>0.04</b>	0.4
	All subjects (N=490)	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>



Highest p-value threshold (critical value,  $t$ ) that keeps FDR at 5% in the linear regression analyses of the relationship between follow-up memory performance with follow-up hippocampal radial distance while controlling for baseline hippocampal volume (NS- not significant).

**Table 4**

Verbal Measure	ALL		NC		MCI		AD	
	Left	Right	Left	Right	Left	Right	Left	Right
LM II-DR	0.01	0.01	NS	NS	NS	NS	0.001	0.0008
AVLT-DR	0.01	0.01	NS	NS	NS	NS	NS	NS
ADAScog-DR	0.009	0.01	NS	NS	NS	NS	NS	NS