Discrimination between Alzheimer’s disease, mild cognitive impairment and normal aging by using automated segmentation of the hippocampus

Olivier Colliot, PhD; Gaël Chételat, PhD; Marie Chupin, PhD; Béatrice Desgranges, PhD; Benoît Magnin, PhD; Habib Benali, PhD; Bruno Dubois, MD, PhD; Line Garnero, PhD; Francis Eustache, PhD; and Stéphane Lehéricy, MD, PhD

1Cognitive Neuroscience and Brain Imaging Laboratory, CNRS UPR 640-LENA, Université Pierre et Marie Curie - Paris 6, Hôpital de la Pitié-Salpêtrière, Paris, France
2Inserm-EPHE-Université de Caen Basse-Normandie, Unité E 0218, GIP Cyceron, CHU de Caen, France
3INSERM U678, Université Pierre et Marie Curie - Paris 6, Hôpital de la Pitié-Salpêtrière, Paris, France
4INSERM U610, Université Pierre et Marie Curie - Paris 6, Hôpital de la Pitié-Salpêtrière, Paris, France
5Department of Neuroradiology, Center for Neuroimaging Research-CENIR, Université Pierre et Marie Curie - Paris 6, Hôpital de la Pitié-Salpêtrière, Paris, France

Running title: Automated hippocampal volumetry in Alzheimer’s disease

Key words: Alzheimer’s disease, MCI, hippocampus, magnetic resonance imaging, image processing

Type of manuscript: Original research

Correspondence to:
Olivier Colliot
LENA - CNRS UPR 640
Cognitive Neuroscience and Brain Imaging Laboratory
Hôpital de la Pitié-Salpêtrière
47, boulevard de l’hôpital
75651 Paris Cedex 13
France
Phone: +33 1 42 16 11 86
Fax: +33 1 45 86 25 37
e-mail: olivier.colliot@chups.jussieu.fr
Abstract

Purpose: To prospectively evaluate the accuracy of automated hippocampal volumetry to distinguish between patients with Alzheimer’s disease (AD), patients with mild cognitive impairment (MCI) and elderly controls, using established criteria for patients with AD and MCI as the reference standard.

Materials and Methods: The regional ethics committee approved the study and written informed consent was obtained from all participants. We studied 25 patients with AD (11 males, 14 females, age ± standard-deviation (SD) = 73 ± 6 years, mini-mental score (MMS) = 24.4 ± 2.7), 24 patients with amnestic MCI (10 males, 14 females, age ± SD = 74 ± 8 years, MMS = 27.2 ± 1.4) and 25 elderly healthy controls (13 males, 12 females, age ± SD = 64 ± 8 years). In each participant, the hippocampi were automatically segmented on three-dimensional (3D) T₁-weighted magnetic resonance images (MRI) with high spatial resolution. Segmentations were performed using recently developed software which allows fast segmentation with minimal user input. Group differences in hippocampal volume were assessed using Student’s T-tests. To obtain robust estimates of the p-values, of the correct classification rate, sensitivity and specificity, we used bootstrap methods.

Results: Significant hippocampus volume reductions were detected in all groups of patients (-32% in AD versus controls, p<0.001; -19% in MCI versus controls, p<0.001; -15% in AD versus MCI, p<0.01). Individual classification based on hippocampal volume resulted in 84% correct classification (sensitivity: 84%, specificity: 84%) between AD and controls and 73% (sensitivity: 75%, specificity: 70%) between MCI and controls.
Conclusion: Our automated method can serve as an alternative to manual tracing and may thus prove useful in assisting in the diagnosis of AD.
Introduction

Alzheimer’s disease (AD) is the most common cause of dementia in the elderly (1). Early and accurate diagnosis of AD can be challenging. In recent years, the early clinical signs of AD have been extensively investigated, leading to the concept of amnestic Mild Cognitive Impairment (MCI) (2-4). A challenge for modern neuroimaging is to help in the diagnosis of early AD and particularly in amnestic MCI patients. Early diagnosis of AD patients allows early treatment with cholinesterase inhibitors, which have been shown to delay institutionalization, improve or stabilize cognition and behavioral symptoms (5, 6).

Three-dimensional (3D) magnetic resonance imaging (MRI) with high spatial resolution allows visualization of subtle anatomical changes and thus can help in the detection of brain atrophy at the beginning of the disease. The hippocampus is known to be affected in the earliest stages of AD (7, 8). Many studies have thus assessed hippocampal atrophy in AD using manual segmentation on MRI (9-19). These studies have demonstrated that manual MR volumetry of the hippocampus can distinguish patients with AD from elderly controls with a high degree of accuracy (80% to 90%). However, manual segmentation of the hippocampus requires a high degree of anatomical training, is observer-dependent and time consuming (up to more than one hour). Although more suitable in clinical practice, visual evaluation of atrophy on multiplanar MRI is difficult and prone to subjectivity (20).

We have developed an automated method that is able to segment the hippocampus on MRI (21). This method has been compared to manual segmentation in young healthy participants and patients with AD and has proved to be reliable, fast and accurate (about 8% relative volume error when compared to manual segmentation) (21). Thus, the purpose of our study was to prospectively evaluate the accuracy of automated
Colliot et al., *Automated hippocampal volumetry in Alzheimer’s disease*

hippocampal volumetry to distinguish between patients with AD, MCI and elderly controls, using established criteria for patients with AD and MCI as the reference standard.
Materials and Methods

Participants

The regional ethics committee approved the study and written informed consent was obtained from all participants. All patients were capable of giving informed consent. Written informed consent was given by the patients themselves. In our study, 25 patients with AD (11 males, 14 females, age ± standard-deviation (SD) = 73 ± 6 years, range = 62–81 years, mini-mental score (MMS) = 24.4 ± 2.7, range = 19-29) and 24 patients with amnestic MCI (10 males, 14 females, age ± SD = 74 ± 8 years, range = 55–87 years, MMS = 27.2 ± 1.4, range = 24-29) were selected from the database of patients prospectively recruited at the Centre Hospitalo-Universitaire (CHU) of Caen. From this database, we included participants whose scanning parameters followed the protocol described below and whose MRI was free of substantial visible motion artifacts. These samples partly overlap with those of previous publications (22-25). The diagnosis for probable AD was made according to the NINCDS-ADRDA (National Institute of Neurological and Communicative Diseases and Stroke-Alzheimer's Disease and Related Disorders Association) criteria (26). The diagnosis of MCI was based on Petersen et al.’s criteria (27). All MCI patients were evaluated every 6 months over an 18-month period to assess conversion, i.e., whether they met NINCDS-ADRDA criteria of probable AD or not. Patients were declared as converters if they had impaired performances (more than 1.5 SD below the normal means according to age and education when available) in at least one of general intellectual function scales as well as in at least two areas of cognition including memory, leading to impaired daily activities as judged by the clinicians from the consultation interviews. Post hoc exclusion criteria included presence of substantial neurological, psychiatric or any other medical disease that could affect brain functioning or structure, and normal episodic
memory performances at follow-up. At completion of the 18-month follow-up period, eight MCI (8/23=35%) patients were declared as converters (3 males, 5 females, age ± SD = 77 ± 4 years, range = 71–82 years, MMS = 26.5 ± 1.4, range = 24-28), 15 patients still had isolated memory deficits (non-converters, 7 males, 8 females, age ± SD = 72 ± 9 years, range = 55–87 years, MMS = 27.7 ± 1.2, range = 26-29) and one MCI patient (female) refused follow-up and was thus excluded in the analysis between converters and non-converters. The annual conversion rate was thus 23%.

AD and MCI patients were compared to 25 elderly healthy controls (13 males, 12 females, age ± SD = 64 ± 8 years, range = 51–84 years) with normal memory performance, as assessed using tests of episodic, semantic and working memory, and without vascular lesions on MRI. To exclude vascular lesions, all controls were checked to have normal signal intensity on T1-, T2- and/or FLAIR-weighted MRI, and notably no substantial white matter T2-FLAIR-weighted hyperintensities (less than 5 pinpoint hyperintensities, size <4 mm (28)). The controls were screened for the absence of cerebrovascular risk factors, mental disorder, substance abuse, head trauma, substantial MRI or biological abnormality, and incipient dementia using a memory test battery. Control participants were recruited through advertisement in local newspapers. Control participants were required to be over 50 years old. There was no specific sex criterion. The recruitment of participants (controls and patients) began in 1999 and ended in 2004.

**MRI acquisition**

Within an interval of two months at most from inclusion for the controls and a few days for MCI and AD patients, each participant underwent a T1-weighted volume MRI scan, which consisted of a set of 128 adjacent axial cuts parallel to the anterior commissure-posterior commissure (AC-PC) line and with slice thickness 1.5 mm and
pixel size 0.9375x0.9375mm² using the spoiled gradient echo sequence (SPGR) (repetition time (TR)=10.3 ms; echo time (TE)=2.1 ms; field of view (FOV)=24*18cm²; matrix=256*192). All the MRI data sets were acquired on the same scanner (1.5 T Signa Advantage echospeed; General Electric, Milwaukee, WI).

**Automated segmentation of the hippocampus**

The segmentation of the hippocampus was performed using an automated method we previously developed. This approach segments both the hippocampus and the amygdala simultaneously based on competitive region-growing between these two structures. It includes prior knowledge on the relative positions of these structures with respect to anatomical landmarks which are automatically identified. During the iterative segmentation process, eleven sets of landmarks were automatically retrieved at the border of the deforming structures. Two of these landmarks were located at the interface between the hippocampus and the amygdala, based on the alveus and the temporal horn of the lateral ventricle. Three of them were defined for the hippocampus only: two based on the alveus and one on the hippocampal sulcus. One was defined for the amygdala only, based on the isthmus of the temporal lobe. Finally, five were defined for both the hippocampus and the amygdala, based on the parahippocampal gyrus and the temporal horn of the lateral ventricle. More details can be found in a previously published paper (21). It should be noted that all these landmarks were found automatically by the algorithm and that no intervention from the operator was required. It is thus not necessary to be able to locate these landmarks on the MRI to use the automated segmentation software.

Segmentations ([Figures 1 and 2](#)) were processed by a trained operator (O.C.) who was blind to all clinical data and diagnostic categories. The operator had six years of
experience in structural MRI analysis. He was trained to use the segmentation software by processing a set of 20 images.

The method requires the following initialization from the operator. First, a bounding box is manually defined around the amygdalo-hippocampal complex (Figure 1a), by selecting six slices which correspond to the limits of the hippocampus and the amygdala in each direction. The dimension of the bounding box is typically around 30x50x20 voxels. Then, two seeds are placed in the hippocampus and the amygdala, respectively (Figure 1b). These seeds constitute the starting points of the deformation process. They are positioned close to the center of the amygdala and the center of the head of the hippocampus. Lastly, starting from these two seeds, the algorithm automatically aggregates voxels depending on the intensities, regularity of the region, and the detection of anatomical landmarks, and converges to the segmentation of the two structures (Figures 1c and d). Additionally, three parameters of the algorithm can be adjusted depending on the participant at hand: one radiometric parameter and two geometric parameters. The radiometric parameter controls the ratio between the intensity characteristics (mean and SD) of the hippocampus and the amygdala and those of the gray matter. Even though the amygdala and the hippocampus are mainly gray matter structures, their intensity on T1-weighted MRI is slightly different from the intensity of the gray matter. Specifically, we observed that their mean intensity is slightly lower than the mean intensity of the gray matter of the bounding box while their standard-deviation (SD) is higher. The ratio between the mean and SD of the intensity of the amygdala and the hippocampus and those of the gray matter may depend on the image contrast. For this reason, three preset values controlling these intensity ratios can be chosen depending on the visual contrast of the image. One of the geometric parameters controls the degree of shape anisotropy in the head of the hippocampus. The
other geometric parameter controls the anisotropy in the tail of the hippocampus. These two geometric parameters are adjusted when the hippocampus is atrophied. The average total processing time for each amygdalo-hippocampal complex (measured on a subset of ten participants: three controls, two MCI and five AD patients) was 11 minutes for the complete procedure, including bounding box definition, seeds positioning, parameter adjustment and automatic deformation by the algorithm. Intrarater reproducibility was assessed by performing the segmentation twice, after a one week interval, on a randomly selected subsample of 10 participants (three AD, three MCI and four controls). The rater was blind to all previous parameters or visualization adjustments. The mean intra-rater relative volume difference was 7% ± 7 (mean ± SD). The difference between the two measurements was not statistically significant (p=0.4, paired Student’s t-test).

Normalization with Total Intracranial Volume (TIV)

For subsequent classification of AD patients and healthy controls, hippocampal measurements were normalized to the total intracranial volume (TIV). The TIV was computed using SPM5 (Statistical Parametric Mapping) software (Wellcome Department of Imaging Neuroscience, London, UK) according to the following procedure:

- correction for intensity non-uniformity, spatial normalization to a common stereotaxic space and tissue classification into gray matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) using a unified procedure (29);
- creation of an approximate intra-cranial mask by summing the three unmodulated tissue probability maps and thresholding the result at 0.5 (this removes some non-CSF voxels that were included in the CSF map by the
segmentation process). Modulation is a procedure which allows preserving the amount of tissue during the normalization process by multiplying the tissues by the Jacobian determinants of the deformation (30);
- masking modulated tissue maps with the previous result;
- total intracranial volume (TIV) is the sum of the masked modulated GM, WM and CSF maps.

In the following, NHV refers to the normalized hippocampal volume and is defined as \( \text{NHV} = \frac{TIV_m \times HV}{TIV} \) where \( TIV_m \) is the average TIV computed across all participants which is constant. The role of the constant multiplicative factor \( TIV_m \) is simply to preserve the order of magnitude of NHV similar to that of HV. Lastly, normalized hippocampal volumes were averaged over both hemispheres. In the following, all reported volumes thus represent the average of the left and right hippocampal volumes.

Statistical analysis

The statistical analysis was conducted by O.C. using in-house software developed by B.M. and H.B.

Group analysis. Group differences in normalized hippocampal volume between AD patients, MCI patients, and healthy controls were assessed using Student’s T-tests. To obtain a more robust estimate of the p-value, we used a bootstrap method (31). In brief, this method proceeds as follows. Let us denote \( S_1 \) the group of controls and \( S_2 \) the group of patients and \( S \) the union of these groups. We worked with the null hypothesis that there are no differences between the mean values of the two groups. We resampled the set \( S \) under the null hypothesis, creating resampled sets \( S_1^* \) and \( S_2^* \) by drawing with replacement participants from both groups. For the n-th resampling \( S_1^{*,n} \) and \( S_2^{*,n} \) we
computed the corresponding value $T^{*,n}$ of the T-test. We performed 5000 resampling and calculated the percentile corresponding to the initial value of the T-test in the set of values $\{T^{*,n}, n=1…5000\}$. According to the bootstrap theory, this percentile is a good estimate of the p-value of our test. Differences were considered significant when $p<0.05$.

**Individual analysis.** For the automatic classifications of AD vs controls, MCI vs controls and AD vs MCI, each individual participant was assigned to the closest group. Specifically, if $S_1$ and $S_2$ are two groups of participants with respective means $m_1$ and $m_2$, a new individual with volume $x$ is assigned to $S_1$ if $|x - m_1| < |x - m_2|$ and to $S_2$ otherwise. To obtain a robust estimate of the correct classification rate, sensitivity and specificity, we used a bootstrap procedure for training set selection. To this purpose, we drew without replacement approximately 75% of each group to obtain a training set $\{S_1^*, S_2^*\}$ and to estimate the means $m_1^*$ and $m_2^*$. The remaining 25% were used as a test set. The procedure was repeated 5000 times. We thus obtained the correct classification rates, the sensitivity and the specificity for the 5000 drawings.

The classification was also analyzed using ROC (receiver operating characteristic) curves. The ROC curve indicates the relationship between sensitivity and 1-specificity for each inter-group discrimination. We computed the area under the curve (AUC) which is an index of overall discriminative ability.

**Subgroup analysis.** To ensure that our findings were not biased by age or sex confounding effects, the same group and individual analyses were also performed on smaller groups of age-matched participants. To this purpose, we selected a group of 17 AD patients (6 males, 11 females, age ± standard-deviation (SD) = 70 ± 4 years, range = 62-76 years, mini-mental score (MMS) = 24.1 ± 2.8, range = 19-28), a group of 17 MCI patients (6 males, 11 females, age ± standard-deviation (SD) = 70 ± 6 years, range = 55-
79 years, mini-mental score (MMS) = 27.1 ± 1.3, range = 25-29), and of a group of 17 healthy elderly controls (10 males, 7 females, age ± standard-deviation (SD) = 68 ± 7 years, range = 60-84 years). We also performed the same analysis on smaller groups of sex-matched participants. To this purpose, we selected a group of 22 AD patients (10 males, 12 females, age ± standard-deviation (SD) = 73 ± 5 years, range = 62-81 years, mini-mental score (MMS) = 24.4 ± 2.5, range = 19-28), a group of 22 MCI patients (10 males, 12 females, age ± standard-deviation (SD) = 73 ± 7 years, range = 55-85 years, mini-mental score (MMS) = 27.4 ± 1.4, range = 24-29), and of a group of 22 healthy elderly controls (10 males, 12 females, age ± standard-deviation (SD) = 64 ± 8 years, range = 51-84 years).
Results

Group analysis

Normalized hippocampal volumes were 1.95 cm$^3 \pm 0.46$ (range: 0.98-3.10) for AD patients, 2.30 cm$^3 \pm 0.46$ (1.28-3.10) for MCI patients and 2.86 cm$^3 \pm 0.46$ (1.74-4.05) for control participants. Significant hippocampal volume reductions were found in both AD (-32% (0.91cm$^3$/2.86cm$^3$) volume reduction, p<0.001) and MCI patients (-19% (0.56cm$^3$/2.86cm$^3$) volume reduction, p<0.001), compared to elderly controls (Table 1). AD patients also had significantly smaller hippocampus compared to MCI patients (-15% (0.35cm$^3$/2.30cm$^3$) volume reduction, p<0.01). Among MCI patients, converters were found to have smaller hippocampus at baseline than non-converters (1.97 cm$^3$ vs 2.47 cm$^3$, 20% (0.50cm$^3$/2.47cm$^3$) volume reduction) (Figure 3). The converter subgroup was too small to compute a p-value.

Individual analysis

Correct classification rates were 84% for AD patients and 73% for MCI patients with respect to elderly controls, and 69% for AD patients with respect to MCI patients (Table 2).

Regarding the ROC curves for inter-group discrimination (Figure 4), the area under the curve (AUC) was 0.913 for AD vs controls, 0.808 for MCI vs controls and 0.721 for AD vs MCI, respectively.

Subgroup analysis

For the age-matched subgroups of 17 participants each, normalized hippocampal volumes were 2.06 cm$^3 \pm 0.48$ (range: 0.98-3.10) for AD patients, 2.30 cm$^3 \pm 0.50$ (1.28-3.10) for MCI patients and 2.87 cm$^3 \pm 0.50$ (1.74-4.05) for control participants.
Significant hippocampal volume reductions were found in both AD (-28% (0.81cm$^3$/2.87cm$^3$) volume reduction, p<0.001) and MCI patients (-20% (0.57cm$^3$/2.87cm$^3$) volume reduction, p<0.001), compared to elderly controls. Correct classification rates were 81% for AD patients and 70% for MCI patients with respect to elderly controls.

For the sex-matched subgroups of 22 participants each, normalized hippocampal volumes were 1.93 cm$^3$ ± 0.49 (range: 0.98-3.10) for AD patients, 2.27 cm$^3$ ± 0.46 (1.28-3.10) for MCI patients and 2.84 cm$^3$ ± 0.49 (1.74-4.05) for control participants.

Significant hippocampal volume reductions were found in both AD (-32% (0.91cm$^3$/2.84cm$^3$) volume reduction, p<0.001) and MCI patients (-20% volume (0.57cm$^3$/2.84cm$^3$) reduction, p<0.001), compared to elderly controls. Correct classification rates were 82% for AD patients and 71% for MCI patients with respect to elderly controls.
Discussion

Our study showed that automated segmentation was able to detect significant volume differences in both AD patients and patients with amnestic MCI. The results are in concordance with a vast number of studies based on manual hippocampal segmentation which have shown hippocampal atrophy in patients with Alzheimer’s disease (e.g., (10-15, 18, 19) and in patients with MCI (32-37). Compared to elderly controls, we found an average 32% hippocampal volume loss in patients with AD. This value is in the range of those reported in studies that used manual volumetry, with volume loss in AD comprised between 23% and 34% (10, 12, 17, 18, 33). In patients with MCI, values ranged from 8% to 15% (32-35, 37). We found a slightly higher relative volume loss in these patients (19%). However, it is important to note that MCI criteria are highly variable from one study (or laboratory) to another, as are the annual rates of conversion to AD. As a result, the degree of hippocampal atrophy, which appropriately depends on the underlying aetiology, the severity of symptoms, and the proportion of converters, similarly shows inconsistencies. In our opinion, the slightly higher degree of atrophy found in our study should be related to the relatively high conversion rate of our MCI sample (23% each year), which is higher than that usually reported (about 15%). This rate should in turn reflect the use of strict criteria, including objective memory deficits, normal global cognition (as attested using objective tests), and excluding impairment in other areas of cognition (also objectively assessed) or depression.

Using normalized hippocampal volumes, 84% of AD patients were correctly classified with respect to elderly controls. This value falls within the range of classification results based on manual segmentation of the hippocampus, which was comprised between 82% and 90% for AD (10, 12, 17, 18, 33). For MCI patients,
classification rates ranged from 60% to 74% (32-35, 37). We found a discrimination rate of 73%. As for the degree of atrophy, this relatively high value should be related to the high conversion rate and thus to the use of strict criteria.

We found that MCI patients who later converted to AD had a 20% smaller hippocampal volume at baseline than non-converters. This result should be interpreted with caution due to the small number of converters. Nevertheless, this is in agreement with several studies based on manual segmentation which have reported that baseline hippocampal volume is an indicator of future progression to AD (38-42). This is also in concordance with studies based on visual rating which demonstrated medial temporal atrophy in patients who subsequently converted to AD (43-45).

While several automated hippocampal segmentation methods have been proposed (46-51), few of them have been applied in patients with AD and/or MCI and very rare studies reported the accuracy of their technique to classify MCI or AD and controls. Carmichael et al. have assessed the performance of automated atlas-based segmentation using several freely-available registration methods (AIR, SPM, FLIRT and a fully deformable approach) in AD and MCI patients (52). They conclude that these approaches are less precise when applied to AD patients than controls but this should be tempered by the fact that these techniques were not specifically designed for this task. Fischl et al. proposed a general method, based on a probabilistic atlas, to automatically label different noncortical structures, including the hippocampus, and applied this technique to patients with mild and questionable AD (53). The method detected significant group differences in terms of hippocampal volume but the authors did not investigate the classification of individual participants. Csernansky et al. used
the high-dimensional brain mapping (HDBM) approach, based on fluid registration with a template, to obtain hippocampal volumes and hippocampal shape differences between patients with very mild AD and controls (54). Using a classification based on both volume and shape features, they achieved a sensitivity of 83% and a specificity of 78%. However, they did not assess the classification performance based on volume alone. Using a similar HDBM approach, Hsu et al. compared automated and manual segmentations in AD and cognitively impaired patients (55). They reported good correlations between manual and automated measurements. However, they did not investigate the accuracy of this technique for the classification of individual patients.

The results of our study require confirmation in larger groups of participants. However, the use of bootstrap resampling techniques allows computing robust estimates for relatively small groups of patients. In order to keep the control group as large as possible, we decided not to exclude control participants based on age or gender. As a consequence, the mean age of the healthy controls was lower than those of the AD and MCI patients. However, to ensure that our findings were not biased by age or sex confounding effects, we also performed the same analysis on smaller groups of age-matched and sex-matched participants and obtained similar results. Nevertheless, future studies on larger age- and sex-matched groups of participants are required to confirm our results. Finally, in our study, the automated segmentations were performed by only one rater, excluding the ability to evaluate inter-operator repeatability. In a previous paper (21) we have assessed the inter-rater reproducibility of the automated segmentation in healthy controls and patients with AD and reported a high reproducibility (the average volume difference between operators was 4% for healthy controls and 7% for AD patients).
Using automated segmentation of the hippocampus, we were able to individually classify Alzheimer’s disease, mild cognitive impairment and control participants with a high degree of accuracy. This method can serve as an alternative to manual tracing and may become a useful tool to assist in the diagnosis of Alzheimer’s disease.
References


Table 1. Pairwise intergroup comparisons of hippocampal volumes

<table>
<thead>
<tr>
<th></th>
<th>Normalized</th>
<th>Volume reduction</th>
<th>Statistical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hippocampal volume (cm$^3$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD vs controls</td>
<td>1.95 vs 2.86</td>
<td>-32%</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>MCI vs controls</td>
<td>2.30 vs 2.86</td>
<td>-19%</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>AD vs MCI</td>
<td>1.95 vs 2.30</td>
<td>-15%</td>
<td>p&lt;0.01</td>
</tr>
</tbody>
</table>

Note. – Group differences were assessed using Student’s T-tests and the p-values were estimated using the bootstrap method. AD: Alzheimer’s disease. MCI: mild cognitive impairment.
Table 2. Classification between AD patients, MCI patients and elderly controls, based on normalized hippocampal volume.

<table>
<thead>
<tr>
<th></th>
<th>AD vs controls</th>
<th>MCI vs controls</th>
<th>AD vs MCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification rate</td>
<td>84%</td>
<td>73%</td>
<td>69%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>84%</td>
<td>75%</td>
<td>67%</td>
</tr>
<tr>
<td>Specificity</td>
<td>84%</td>
<td>70%</td>
<td>71%</td>
</tr>
</tbody>
</table>

Note. – For each inter-group discrimination, robust estimates of the correct classification rate, sensitivity and specificity were computed using the bootstrap method (see text for details). AD: Alzheimer’s disease. MCI: mild cognitive impairment.
List of figures

Figure 1. The different steps of the hippocampal segmentation. (a) First, the user defines a bounding box around the amygdalo-hippocampal complex. This panel illustrates four of the limits of the bounding box on a sagittal reconstruction. (b) Two seeds are placed in the hippocampus head and the amygdala, respectively. The panel illustrates two seed voxels in the hippocampus (in red) and the amygdala (in green), on a sagittal reconstruction. The two seeds do not necessarily belong to the same slice. (c) Starting from these two seeds, the algorithm automatically segments the two structures. The panel illustrates the final result on a sagittal reconstruction. (d) 3D surface renderings corresponding to the automated segmentations of the hippocampus and the amygdala. In this example, the T1-weighted MRI was acquired using the inversion recovery fast spoiled gradient echo sequence (IR-FSPGR) (124 adjacent axial slices; slice thickness 1.3 mm; pixel size 0.9375x0.9375mm$^2$; repetition time (TR)=14.3 ms; echo time (TE)=6.3 ms; inversion time (TI)=600 ms; field of view (FOV)=24*18cm$^2$; matrix=256*192). It should be noted that this image is presented to illustrate the segmentation method and is not part of our study.

Figure 2. Automated segmentation of the hippocampus. Left panels: patient with Alzheimer’s disease (coronal and sagittal reconstructions, from left to right). Right panels: healthy elderly control (coronal and sagittal reconstructions, from left to right). T1-weighted MRI were acquired using the spoiled gradient echo sequence (SPGR) (128 adjacent axial slices parallel to the anterior commissure - posterior commissure (AC-PC) line; slice thickness 1.5 mm; pixel size 0.9375x0.9375mm$^2$; repetition time (TR)=10.3 ms; echo time (TE)=2.1 ms; field of view (FOV)=24*18cm$^2$;
matrix=256*192).

**Figure 3.** Normalized hippocampal volumes of converters and non-converters MCI. The circles indicate the volume of each individual participant while the horizontal bar indicates the mean of each group.

**Figure 4.** Receiver operating characteristic (ROC) curves for inter-group classification. (a) AD vs controls. (b) MCI vs controls. (c) MCI vs AD.
Figure 1
Figure 2
Figure 3
Figure 4