International Psychogeriatrics (2014), 26:7, 1067–1081 © International Psychogeriatric Association 2014
doi:10.1017/S1041610213002627

Hippocampal atrophy in people with memory deficits: results from the population-based IPREA study

Luca Ferrarini,[1,2] Baldur van Lew,[1] Johan H. C. Reiber,[1] Claudia Gandin,[3] Lucia Galluzzo,[3] Emanuele Scafato,[3] Giovanni B. Frisoni,[4,5] Julien Milles[1] and Michela Pievani;[4] the IPREA Working Group (Italian PRoject on Epidemiology of Alzheimer's disease)*

[1] *LKEB – Division of Image Processing, Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands*
[2] *Campus IFOM-IEO, Milan, Italy*
[3] *National Center on Epidemiology, Surveillance and Health Promotion, Istituto Superiore di Sanità, Rome, Italy*
[4] *Laboratory of Epidemiology Neuroimaging and Telemedicine, IRCCS Istituto Centro San Giovanni di Dio, Fatebenefratelli, Brescia, Italy*
[5] *Memory Clinic and LANVIE - Laboratory of Neuroimaging of Aging, University Hospitals and University of Geneva, Geneva, Switzerland*

ABSTRACT

Background: Clinical studies have shown that hippocampal atrophy is present before dementia in people with memory deficits and can predict dementia development. The question remains whether this association holds in the general population. This is of interest for the possible use of hippocampal atrophy to screen population for preventive interventions. The aim of this study was to assess hippocampal volume and shape abnormalities in elderly adults with memory deficits in a cross-sectional population-based study.

Methods: We included individuals participating in the Italian Project on the Epidemiology of Alzheimer Disease (IPREA) study: 75 cognitively normal individuals (HC), 31 individuals with memory deficits (MEM), and 31 individuals with memory deficits not otherwise specified (MEMnos). Hippocampal volumes and shape were extracted through manual tracing and the growing and adaptive meshes (GAMEs) shape-modeling algorithm. We investigated between-group differences in hippocampal volume and shape, and correlations with memory deficits.

Results: In MEM participants, hippocampal volumes were significantly smaller than in HC and were mildly associated with worse memory scores. Memory-associated shape changes mapped to the anterior hippocampus. Shape-based analysis detected no significant difference between MEM and HC, while MEMnos showed shape changes in the posterior hippocampus compared with HC and MEM groups.

Conclusions: These findings support the discriminant validity of hippocampal volumetry as a biomarker of memory impairment in the general population. The detection of shape changes in MEMnos but not in MEM participants suggests that shape-based biomarkers might lack sensitivity to detect Alzheimer's-like pathology in the general population.

Key words: hippocampal atrophy, memory, MRI, population-based

Introduction

Alzheimer's disease (AD) is a neurodegenerative disease associated with the progressive accumulation of neurofibrillary tangles and amyloid plaques around neuronal cells, leading to neuronal loss, cognitive deficits, and finally to clinically overt dementia (Selkoe, 2002). Despite considerable efforts by the scientific community to unveil the mechanisms behind AD, there is currently no cure available for patients, and costs for the management and care of the patients are bound to rise considerably with increasing life expectancy worldwide (http://www.alz.org/downloads/facts_figures_2013.pdf). Currently, the greatest hope for AD patients relies on early therapeutic interventions aiming at delaying the progression of the disease. Such measures are likely to be more effective when administered during the very early stages of the disease rather than to patients with full blown AD (Emery, 2011). Indeed, delaying

* For a list of participants in the IPREA working group, see the Appendix 1.
Correspondence should be addressed to: Giovanni B. Frisoni, MD, Laboratory of Epidemiology, Neuroimaging and Telemedicine, IRCCS Istituto Centro San Giovanni di Dio, Fatebenefratelli, via Pilastroni 4, 25125 Brescia, Italy. Phone: +39-030-3501361; Fax: +39-030-3501592. Email: gfrisoni@fatebenefratelli.it. Received 1 Jul 2013; revision requested 6 Sep 2013; revised version received 6 Dec 2013; accepted 12 Dec 2013. First published online 13 February 2014.

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

disease progression by one year in 2015 would reduce the estimated prevalence by about 4 million (8%) in 2050, which in turn would translate into a considerable alleviation of the costs associated to AD (Emery, 2011).

Important steps towards early AD detection have been achieved in the past decade through the development of *in vivo* biomarkers of pathology. Among these, hippocampal atrophy is the most established biomarker of neurodegeneration (Bobinski *et al.*, 2000) and is closely associated with episodic memory deficits (Frisoni *et al.*, 2010). This well-established clinico-pathological correlation motivated the inclusion of hippocampal volume and medial temporal lobe (MTL) atrophy as supportive features for AD diagnosis in the newly published research criteria (Albert *et al.*, 2011). Imaging studies in individuals with mild cognitive impairment (MCI), a population at higher risk to develop AD (Petersen *et al.*, 2001), have shown that hippocampal volumes are already reduced in these people (Jack *et al.*, 2010). Studies using an alternative definition of MCI (e.g. aging-associated cognitive decline (AACD); Levy, 1994) similarly reported significant MTL atrophy associated with cognitive deficits (Pantel *et al.*, 2003). More advanced hippocampal analysis tools, such as shape-based methods, help to distinguish between AD-related changes (affecting the anterior dorsal CA1 field) and age-related changes (affecting the ventral hippocampal head) (Apostolova *et al.*, 2006; Frisoni *et al.*, 2008; Wang *et al.*, 2003). Overall, the discriminant and predictive validity of hippocampal atrophy in clinical populations is good, as demonstrated by studies rooted in a pattern recognition framework (Apostolova *et al.*, 2006; Devanand *et al.*, 2007; Clark *et al.*, 2008; Colliot *et al.*, 2008; Karas *et al.*, 2008; Klöppel *et al.*, 2008; Ferrarini *et al.*, 2009).

Less known is whether hippocampal atrophy can discriminate between normal cognition and cognitive impairment in the general population. This would be of interest for the possible use of hippocampal atrophy to screen populations for preventive interventions. The above-mentioned studies were indeed based on clinical-based populations, which typically include individuals at a more advanced stage of cognitive impairment. Conversely, population-based studies usually include patients in the earliest stages of the disease and have thus a greater chance to detect subtle changes associated with early cognitive impairment. Population-based studies consistently showed that hippocampal atrophy is associated with an increased risk of dementia (see Table 1 for a summary of population-based findings on hippocampal atrophy, and Table S1, available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG, for a complete list of the main population-based studies), supporting the validity of hippocampal atrophy as a biomarker to predict conversion to dementia. Conversely, there is mixed evidence on the discriminant validity of hippocampal atrophy in separating normal aging from people with cognitive impairment (Table 1), thus limiting its potential as a biomarker to screen at-risk populations. Since hippocampal changes are likely to be subtle in the earliest disease stages, finer approaches, such as shape-based analysis methods, which have the potential to detect very small morphological changes, are needed. This study, therefore, aims at assessing hippocampal shape and volume changes in elderly people without dementia, drawn from a population-based study carried out in Italy, namely, the Italian PRoject on the Epidemiology of Alzheimer disease (IPREA; Scafato *et al.*, 2005).

Methods

Cohort

TOTAL IPREA SAMPLE

The participants included in this study were drawn from the epidemiological IPREA study (Scafato *et al.*, 2005). The initial study included 2,985 Italian elderly adults (aged between 65 and 84 years) who underwent a personal and informant interview, a physical and a neurological examination, and comprehensive neuropsychological testing, as described in detail elsewhere (Scafato *et al.*, 2010). A summary of the diagnostic criteria used for participants classification are reported in Appendix 2. Briefly, the AACD criteria were used to classify patients as cognitively normal (HC), patients with cognitive impairment (AACD, AACD-nos, or objective cognitive decline (OCD); see Appendix 2), and patients with dementia (AD, vascular dementia (VaD), Parkinson's disease (PD) and other dementias). Diagnosis of dementia was defined based on internationally accepted criteria for the various forms of dementia. In this population, it has been estimated that the prevalence of cognitive impairment is approximately 45% (9% fulfilling AACD criteria, 19% fulfilling AACD-nos criteria, and 17% fulfilling OCD criteria; Scafato *et al.*, 2010).

MRI SUBSTUDY

A subsample of the initial cohort underwent an MRI scan ($n = 567$). The exclusion criterion for MRI was diagnosis of dementia (Scafato *et al.*, 2005). Additionally, for the present study we excluded participants with non-amnestic deficits

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

Table 1. Literature review results: discriminant and predictive validity of hippocampal atrophy as a biomarker of cognitive impairment in population-based studies

REFERENCE[a]	STUDY DESIGN	FOLLOW-UP DURATION	SAMPLE SIZE	HIPPOCAMPAL MEASURE	OUTCOME MEASURES	MAIN RESULTS	DISCRIMINANT VALIDITY	PREDICTIVE VALIDITY
Wolf *et al.* (2001)	Cross-sectional	N.A.	17 controls 12 MCI 10 AD	HV (total, head, body)	Group differences Discrimination accuracy	HV (total): MCI < controls (11%–14% smaller) AD < controls (26%–28% smaller) AD < MCI; HV (body): significantly different over all three groups Accuracy: 77% (HV for MCI vs. controls)	Yes	N.A.
Pantel *et al.* (2003)	Cross-sectional	N.A.	22 controls 21 AACD 12 AD	HV, PHV	Group differences	AACD < controls (PHV) AD < AACD, controls (PHV, HV)	Yes/no	N.A.
Pennanen *et al.* (2004)	Cross-sectional	N.A.	59 controls 65 MCI	HV	Group differences Discrimination accuracy	MCI < controls (8% smaller) Accuracy: 60%	Yes/no	N.A.
Kumar *et al.* (2006)	Cross-sectional	N.A.	522 controls 29 MCI	HV	Group differences Prediction of MCI	No group difference Not a significant predictor	No	N.A.
Scher *et al.* (2007)	Cross-sectional	N.A.	102 controls 24 incident AD	HV, shape analysis	Group differences	HV: incident AD < controls (11% smaller) Shape: incident AD < controls (lateral body: CA1, DG, subiculum)	Yes	N.A.
Reitz *et al.* (2009)	Cross-sectional	N.A.	162 normal 22 naMCI 30 aMCI 17 adults with dementia	HV	Group differences Correlation with memory and language	Individuals with dementia < controls Associated with word total recall in the whole sample but not in normal + MCI or normal alone	No	N.A.
Zhang *et al.* (2011)	Cross-sectional	N.A.	243 controls 146 MCI (81 aMCI, 55 naMCI)	HV	Prediction of MCI, aMCI, naMCI	Significant predictor of: MCI (OR: 1.79) aMCI (OR: 1.81) naMCI (OR: 2.03)	Yes	N.A.
Becker *et al.* (2012)	Cross-sectional	N.A.	29 normal 20 MCI	MTA, HV	Group differences	MCI < controls	Yes	N.A.
Cui *et al.* (2012)	Cross-sectional	N.A.	204 normal 79 aMCI	HV, subcortical GM, CSF, WM	Discriminating features	Not a significant discriminator	No	N.A.
Zhang *et al.* (2012)	Cross-sectional	N.A.	120 controls 135 MCI (74 aMCI, 61 naMCI)	VBM	Group differences	aMCI < controls	Yes	N.A.

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

Table 1. Continued.

REFERENCE[a]	STUDY DESIGN	FOLLOW-UP DURATION	SAMPLE SIZE	HIPPOCAMPAL MEASURE	OUTCOME MEASURES	MAIN RESULTS	DISCRIMINANT VALIDITY	PREDICTIVE VALIDITY
Visser *et al.* (1999)	Longitudinal	Three years (clinical)	18 controls 20 MCI (9 converters, 4 stable) 7 AD	Baseline MTA, HV, PHV	Group differences Correlation with cognitive decline Prediction of clinical outcome	Baseline: AD < controls (MTA, PHV) Follow-up: MCI converters < MCI stable (PHV) PHV correlated with memory change Significant predictors: PHV (OR: 0.26 (0.08–0.86), 77% accuracy) HV (OR: 0.21 (0.05–0.99), 69% accuracy) MTA (OR: 12.2 (1.4–9.5), 77% accuracy)	No	Yes/no
Persson *et al.* (2006)	Longitudinal	Ten years (clinical)	40 elderly: 20 with stable memory 20 with declining memory	Baseline HV	Group differences	Declining < stable	N.A.	Yes
Herruka *et al.* (2008)	Longitudinal	Three to five years (clinical)	21 MCI: 13 stable 8 converters	Baseline HV	Group differences Correlation with memory	Converters < stable Word list delayed recall correlated with HV in the whole sample and in MCI-c	N.A.	Yes
Tapiola *et al.* (2008)	Longitudinal	Three years (clinical)	60 MCI: 47 stable 13 converters to dementia	Baseline HV	Group differences Prediction of conversion to dementia	Converters < stable Significant predictor of conversion to dementia (HR > 0.73)	N.A.	Yes
Godin *et al.* (2010)	Longitudinal	Four years (clinical)	1,032 adults without dementia: 224 with moderate cognitive decline 46 with severe cognitive decline	Baseline HV	Prediction of cognitive decline	HV associated with increased risk of moderate (OR: 0.7 (0.6–0.9)) and severe cognitive decline (OR: 0.5 (0.3–0.7)) HV associated with annual changes in memory and non-memory	N.A.	Yes
den Heijer *et al.* (2010)	Longitudinal	Eight years (clinical and imaging)	518 adults without dementia (50 converters to dementia)	Baseline HV Rate of HV decline	Prediction of conversion to dementia	Baseline HV associated with risk of dementia (HR > 2 (1.5–2.6)) Rate of HV decline associated with: (i) risk of dementia (HR: 1.6 (1.2–2)) (ii) delayed memory decline in dementia-free people (HR > 1.4 (1–2))	N.A.	Yes

Table 1. Continued.

REFERENCE[a]	STUDY DESIGN	FOLLOW-UP DURATION	SAMPLE SIZE	HIPPOCAMPAL MEASURE	OUTCOME MEASURES	MAIN RESULTS	DISCRIMINANT VALIDITY	PREDICTIVE VALIDITY
Sachdev *et al.* (2013)	Longitudinal	Two years (clinical)	223 MCI: 66 reverters to normal cognition 157 non-reverters	Baseline HV	Group differences	Non-reverters < reverters	N.A.	Yes
Zhuang *et al.* (2013)	Longitudinal	Two years (clinical)	155 controls 39 converters to aMCI 27 stable aMCI	Baseline HV	Group differences	Stable MCI < controls Baseline HV correlated with episodic memory	Yes	No

Abbreviations: AD = Alzheimer's disease; MCI = mild cognitive impairment; HV = hippocampal volume; MTA = medial temporal atrophy on visual rating scale; PHV = parahippocampal volume; VBM = voxel-based morphometry; aMCI = amnestic MCI; naMCI = non-amnestic MCI; AACD = aging-associated cognitive decline; DG = dentate gyrus; OR = odds ratio; GM = gray matter; WM = white matter; CSF = cerebrospinal fluid.
YES/NO: questionable or trend for significance.
Inclusion criteria: (i) population-based; (ii) including people with cognitive impairment (MCI or AACD; eventually AD or cognitively impaired); (iii) hippocampal atrophy as the primary outcome. Exclusion criteria: (i) reviews; (ii) primarily methodological/descriptive studies; (iii) correlation studies in cognitively healthy.
[a]Full references are listed in Table S1 (available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG).

and included all who had (i) normal cognition (HC) or (ii) memory deficits (either AACD, AACD-nos, or OCD). Memory deficits were assessed with the Italian version of the Buschke Fuld Selective Reminding Test (Spinnler and Tognoni, 1987), which covers three aspects of episodic memory: short-term recall (STR), long-term recall (LTR), and long-term storage (LTS). Memory scores were defined as abnormal when the score (adjusted for age and education level using normative values for the Italian population; Spinnler and Tognoni, 1987) was 1 SD below the mean for the population, as suggested by Levy (1994).

Two hundred and six participants were excluded due to incomplete or missing data ($n = 170$ due to missing or incomplete clinical/neuropsychological data; and $n = 36$ due to MRI artifacts precluding hippocampal measurement), and 180 participants were excluded because they did not fulfill the clinical criteria ($n = 179$ showed cognitive deficits in non-memory domains only; and $n = 1$ had pseudo-dementia). A total of $n = 181$ participants were eligible for the study, of whom 119 were HC and 62 showed memory deficits. Of these, 19 participants had OCD, 12 fulfilled the criteria for AACD, and the remaining 31 fulfilled the criteria for AACD-nos (MEMnos). The OCD and AACD groups were pooled together (MEM) since there were no differences between the two groups in terms of memory deficits. The MEM-nos group was analyzed separately to distinguish pure memory deficits (MEM) from deficits which might be related to other not otherwise specified conditions (AACD-nos criteria). Finally, since the initial HC sample ($n = 119$) was slightly younger (age: 72 ± 4) than the MEM and MEMnos groups ($p < 0.001$ on analysis of variance (ANOVA)), a subgroup of HC was age-matched to the MEM and MEMnos groups by selecting older controls (age > 70; $n = 75$). Demographic data of the three groups are shown in Table 2. Comparisons between the total IPREA sample ($n = 2{,}985$) and the subsample used for the present study ($n = 137$) are shown in Table S2 (available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG).

The design and methodology of the IPREA study have been approved by an international group of experts (Scafato *et al.*, 2005). The project has been approved by the Ethics Committee of the Istituto Superiore di Sanità (ISS). Written informed consent was obtained from all the participants.

MRI acquisition and pre-processing

All the MRI scans were collected on the same mobile Siemens Symphony 1 Tesla system between

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

Table 2. Descriptive statistics for demographic data, memory scores, and hippocampal volumes in the study cohort

	HC	MEM	MEMnos	*p*
N	75	31	31	
Age	75 (3)	74 (5)	76 (5)	0.22
Gender (female)	35 (47%)	10 (32%)	13 (43%)	0.39
Education (years)	7 (3)	6 (2)	7 (4)	0.39
MMSE	28.1 (1.6)	26.9 (2.4)†	26.7 (3.1)∗	0.006
Memory				
STR	123 (26)	49 (12)°	50 (19)∗	<0.001
LTR	87 (44)	26 (18)°	24 (13)∗	<0.001
LTS	9 (2)	5 (1)°	5 (2)∗	<0.001
Volumetry				
Left hippocampus (mm3)	3,740 (477)	3,449 (630)°	3,480 (530)†	0.011
Right hippocampus (mm3)	3,974 (449)	3,706 (638)°	3,730 (561)†	0.019

Values are mean (SD) or number (%); *p* denotes significance on the Analysis of Variance (ANOVA) for continuous variables or the χ^2 test for dichotomous variables.
°$p < 0.05$ in MEM versus HC (Bonferroni *post hoc*); ∗$p < 0.05$ in MEMnos versus HC (Bonferroni *post hoc*); †$p < 0.10$ (trend) in MEM versus HC or MEMnos versus HC (Bonferroni *post hoc*).
MEM = participants with memory deficits according to the AACD criteria; MEMnos = participants with memory deficits not otherwise specified; HC = healthy controls; STR = short-term recall; LTR = long-term recall; LTS = long-term storage.

March 2004 and May 2004. Quality control was performed at each stopover using daily routine procedures. High-resolution sagittal T1-weighted gradient echo sequences were acquired using the following parameters: TR = 11.4 ms, TE = 4.4 ms, field of view = 250 mm, acquisition matrix = 256 × 256, slice thickness = 1.3 mm, and flip angle = 80°.

Hippocampal volumetry

The three-dimensional (3D) images were processed using a combination of scripts written in Perl (http://www.perl.com), based on the Minc toolkit developed at the McConnel Brain Imaging Centre (Montreal Neurological Institute, McGill University, Montreal, Canada). Processing included correction for magnetic field non-uniformities, intensity normalization, and brain-to-brain linear registration (9 degrees of freedom (dof)) to a standard template in the stereotaxic space (ICBM152) and re-sampling to an isotropic 1.5 mm voxel size. Each registered image was visually compared to the template using Register (part of Minc toolkit) and, when the automatic registration failed (mainly due to high scalp brightness), a manual registration was performed based on eleven anatomical landmark points distributed over the cerebrum and brainstem (the most anterior point of the temporal poles, the most posterior aspect on the occipital lobe, the most anterior point on the frontal lobe, the central sulcus, the inferior ventral aspect of the pons-midbrain cleft, the genu and splenium of the corpus callosum, the inter-thalamus adhesion, and the eyes). The hippocampi were manually traced by a single expert tracer with Display (part of Minc toolkit) on contiguous coronal slices, simultaneously checking tracing accuracy on the sagittal and axial planes. Tracings included the hippocampus proper, dentate gyrus, subiculum (subiculum proper and presubiculum), alveus, and fimbria (Pruessner *et al.*, 2000). Test-retest reliability was assessed on 20 participants, intraclass correlation coefficient being 0.93 for the right and 0.94 for the left hippocampus.

To obtain the original hippocampal volumes (i.e. in the native space), the brain with the traced region of interest was back-transformed from the stereotaxic to the native space. The accuracy of the back-transformation was confirmed by visual inspection. Hippocampal volumes were normalized to intracranial volume to control for brain size differences across the participants. Intracranial volume was obtained with SIENAX (part of FMRIB's Software Library – FSL) by computing the scaling factor mapping individual images to the standard MNI152 template (Smith *et al.*, 2002). The reciprocal of this value was then computed to obtain the intracranial volume.

Modeling of the hippocampal shape

The modeling of all hippocampal shapes was performed using GAMEs (Ferrarini *et al.*, 2007; https://darwinnandoe.lumc.nl/drupal6/), a method based on growing and adaptive meshes. GAMEs has been successfully applied in several prior studies, including investigation of shape changes

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

in brain ventricles (Ferrarini *et al.*, 2006; 2008), basal ganglia structures (de Jong *et al.*, 2011), and hippocampi (Ferrarini *et al.*, 2009). A detailed description of the method can be found in Ferrarini *et al.* (2007): briefly, a mesh is first grown (adding nodes and edges) until convergence to a pre-defined set of surface points (chosen as the average volume of the HC group). Subsequently, the mesh topology is frozen (number of nodes and edges), and only the locations of nodes are allowed to adapt to each individual set of hippocampal surface points. The adaptation is performed applying the Kohonen self-organizing map algorithm (Kohonen, 1990), thus preserving topology. The set of final meshes represents a point distribution model (PDM), in which each node in a mesh is uniquely associated to the anatomically equivalent nodes in all other meshes. We applied GAMEs to the modeling of both left and right hippocampi in the stereotaxic ICBM152 space (see Figure S1(a), available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG). The modeling resulted in a total of 137 meshes, one per participant, for each hippocampus. The optimal number of nodes per mesh was found to be 313, with each node locally representing a surface area of approximately 2 mm^2. This was chosen considering the resolution of the data used to create the model (approximately 1.5 $\times$ 1.5 $\times$ 1.5 mm in standard space). Choosing a higher accuracy for the mesh (i.e. more nodes) would be meaningless, since it would approach the limiting resolution of the original data.

After having modeled each hippocampus, local normal versors (i.e. vectors of length 1) were identified along the surface model (see Figure S1(b), available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG). Subsequently, for each given node and normal versor, all participants' local node positions were projected along the normal directions (see Figure S1(c), available as supplementary material attached to the electronic version of this paper at www.journals.cambridge.org/jid_IPG): this resulted in a reduction of dimensionality for the coding of each node across the participants, from its original 3D coordinates to a one-dimensional (1D) coordinate (i.e. its projection along the local normal versor). Both the 3D and 1D representations of nodes were considered for subsequent analyses: specifically, the 3D representation was used for the statistical analysis of local shape differences between groups, while the 1D representation was used for correlation analysis of focal changes with memory scores and age.

Statistical analysis

Statistical differences between groups in demographic, clinical, and volumetric features were assessed with the ANOVA test (*post hoc*: Bonferroni correction).

Second, focal morphological differences between groups were analyzed. To this goal, a non-parametric multidimensional test (Hotelling's T2 test) was used. The statistical background has already been provided in detail in previous works (Ferrarini *et al.*, 2006; 2007). In brief consider a specific surface location (i.e. to specific node in the mesh model): such a node is uniquely associated to anatomically equivalent locations across all participants and groups. Hence, focusing on a given comparison (e.g. HC vs. MEM), two clouds of 3D space locations can be identified. As a first step, the Hotelling's T2 statistic is evaluated. Next, the association between node locations and groups is permuted randomly for a given number of times (i.e. 10,000). At each permutation, the Hotelling's T2 statistic is evaluated. Due to random permutation, the set of statistics obtained in this second phase follows the null hypothesis of no difference between groups. Eventually, a p-value for the original comparison (e.g. HC vs. MEM) is obtained by considering the proportion of times in which a higher statistics was obtained by chance under the null-hypothesis (for more details about non-parametric permutation tests, the reader is referred to Appendix 3).

A correlation analysis was performed to investigate the relationship between hippocampal features (volumes and shape changes) and socio-demographic and cognitive variables (age and memory deficits) using the Pearson correlation test. Correlations were performed separately for the right and left hippocampi. For the morphological features, the local 3D spatial information (node locations across the participants) was first turned into a 1D representation. The cognitive variables were the STR, LTR, and LTS corrected scores. The correlations were performed both across the entire cohort (HC, MEM, and MEMnos) and separately within each patient group (MEM and MEMnos). The threshold for significance was set at $p < 0.05$ for all the correlation analyses. Multiple comparisons' correction was performed for volumetric data with the False Discovery Rate procedure (Benjamini and Yekutieli, 2001) at $q = 0.10$ and $q = 0.05$ levels. For shape-based analysis, p-values were corrected locally with Hotelling's T2 statistic and permutation test (see previous paragraph).

Finally, the accuracy of hippocampal volumetry in separating MEM and MEMnos groups from HC, and participants with borderline cognition from

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

controls, was estimated using receiver operating characteristic (ROC) curves and computing the corresponding area under the curve (AUC).

Results

Demographic and clinical data

No significant difference was detected across groups for age, gender, and years of education ($p > 0.22$ on ANOVA; Table 2), which was in line with the matching procedure. Memory scores were significantly lower in MEM ($p < 0.001$ for STR, LTR, and LTS; Bonferroni *post hoc*) and MEMnos ($p < 0.001$) groups compared with HC, in line with the selection criteria. Mini-Mental State Examination (MMSE) was significantly lower in MEMnos ($p = 0.02$ on Bonferroni) and showed a trend for lower scores in MEM ($p = 0.06$ on Bonferroni) participants compared with HC. No significant differences were detected between the MEM and the MEMnos groups in any of the demographic and clinical variables ($p > 0.05$ on Bonferroni *post hoc*).

Volumetric and morphological group comparisons

Hippocampal volumes normalized to intracranial volume differed significantly between groups ($p < 0.02$ on ANOVA; Table 2). *Post hoc* Bonferroni showed that the left and right normalized hippocampi were smaller in MEM compared with HC (8% and 7% smaller for the left and right, respectively; $p = 0.03$ for the left and $p = 0.05$ for the right) and were marginally lower in MEMnos compared with HC ($p = 0.07$ and $p = 0.09$).

Shape analysis showed no significant difference between MEM and HC ($p > 0.05$ over the whole hippocampal surface, data not shown). Conversely, hippocampal shape differed significantly between MEMnos and HC, significant differences mapping to the posterior hippocampus (primarily the right CA1 tail and the ventral subiculum bilaterally; Figure 1, left panel). Direct comparisons between MEM and MEMnos detected hippocampal shrinkage in the latter group in the right CA1 tail (Figure 1, right panel).

Volumetric correlation analysis

In the whole cohort, normalized hippocampal volumes correlated significantly with age ($r = -0.21$ and $p = 0.01$ for the left, and $r = -0.25$ and $p = 0.003$ for the right hippocampus; Table 3) and memory scores ($p < 0.001$ for all memory variables; Table 3) and survived the false discovery rate (FDR) correction ($p < 0.033$ at $q = 0.10$ and $p < 0.017$ at $q = 0.05$). When the analysis was restricted to patients with memory deficits, a positive correlation was found in the MEM group between the STR scores and the right normalized hippocampal volume ($r = 0.38$, $p = 0.04$; Table 3) and the left normalized hippocampus and LTS scores ($r = 0.43$, $p = 0.04$; Table 3). These correlations survived FDR-correction at $q = 0.10$ ($p < 0.042$) but not at $q = 0.05$ ($p < 0.021$). No significant correlation was found in the MEMnos group ($p > 0.05$; Table 3).

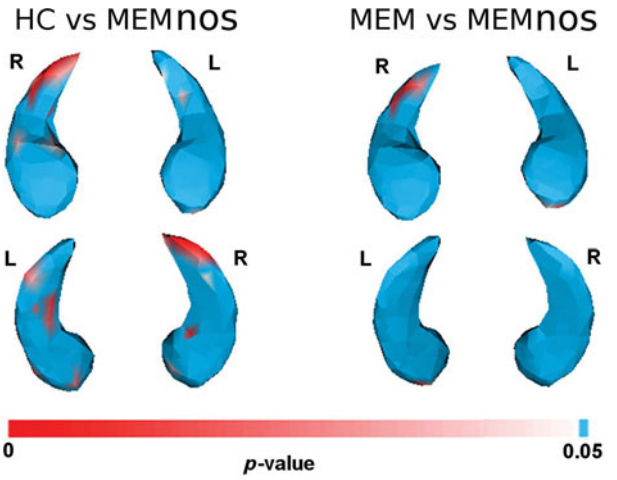


Figure 1. (Colour online) Significant hippocampal shape differences between groups. The maps show significant *p*-values ($p < 0.05$, red to white colors) for the comparisons between HC and MEMnos, and between MEM and MEMnos. No significant difference was detected between the MEM and HC groups ($p > 0.05$).

Morphological correlation analysis

The results of morphological correlation analysis are shown in color-coded hippocampal maps representing the significance of the correlation (*p*-values) and the corresponding signed R^2 values (Figures 2–4). These findings are described in detail in the following sections.

Whole cohort

Correlations between hippocampal shape changes and STR and LTR memory deficits were localized in the right CA1 tail and in the ventral subiculum bilaterally (Figure 2(a)–2(b)). Long-term storage memory scores were only weakly associated with shape changes, mainly in the anterior ventral hippocampus (subiculum, Figure 2(c)). Memory scores were negatively correlated with atrophy (i.e. lower memory scores corresponding to greater atrophy). Age was positively correlated with atrophy (i.e. lower ages corresponding to lower atrophy), significant associations mapping to the anterior hippocampus (head of the CA1 field and anterior

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

Table 3. Correlations between hippocampal volumes, age, and memory deficits in the whole study sample, in the MEM sub-group, and in the MEMnos sub-group

	LEFT HIPPOCAMPUS		RIGHT HIPPOCAMPUS	
	r	*p*	*r*	*p*
Whole sample				
Age	− 0.21	0.01*	− 0.25	0.003*
STR	0.29	0.001*	0.33	<0.001*
LTR	0.29	0.001*	0.32	<0.001*
LTS	0.33	<0.001*	0.33	<0.001*
MEM				
Age	− 0.27	0.15	− 0.29	0.11
STR	0.34	0.06	0.38	0.04*
LTR	0.20	0.27	0.21	0.26
LTS	0.43	0.04*	0.35	0.09
MEMnos				
Age	− 0.22	0.24	− 0.34	0.06
STR	0.13	0.50	0.14	0.44
LTR	0.10	0.59	0.08	0.67
LTS	0.19	0.39	− 0.05	0.82

r denotes Pearson's correlation coefficient; *p* denotes Pearson's significance.
*$p < 0.05$.
MEM = participants with memory deficits according to the AACD criteria; MEMnos = participants with memory deficits not otherwise specified; STR = short-term recall; LTR = long-term recall; LTS = long-term storage.

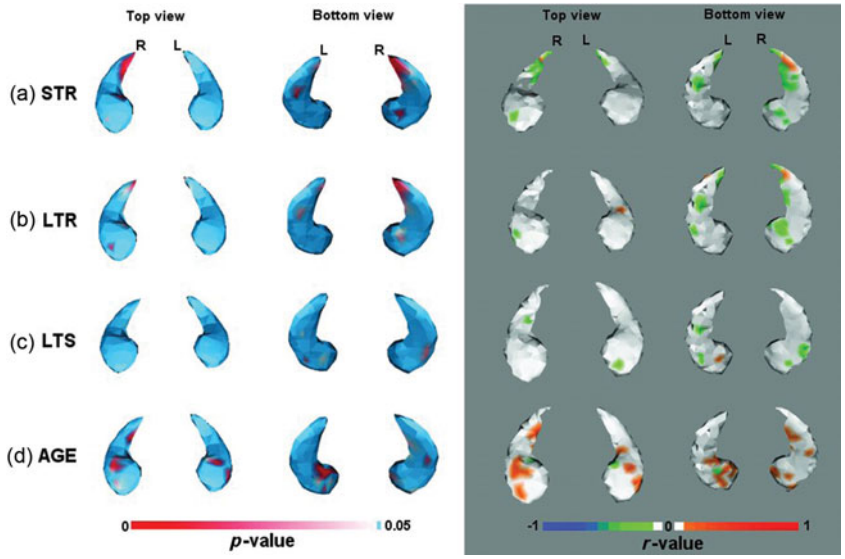


Figure 2. (Colour online) Correlations between local hippocampal changes and memory scores and age in the entire cohort. Maps show significant *p*-values ($p < 0.05$; red to white colors; left panel) and the corresponding R^2 correlation values (positive correlations: orange-to-red, negative correlations: green-to-blue; right panel).

ventral subiculum) bilaterally, and to the right medial subiculum (Figure 2(d)).

MEM GROUP

When the analysis was restricted to the MEM group, no correlation between atrophy and STR and LTR scores was detected (Figure 3(a)–3(b)). Conversely, LTS scores were significantly correlated with hippocampal changes in the head of the CA1 field (Figure 3(c)). The association between age and shape changes in the anterior hippocampus (head of the CA1 field and anterior ventral subiculum) was still significant in this subgroup (Figure 3(d)).

MEMNOS GROUP

Within the MEMnos group, no clear pattern of correlation was observed between hippocampal

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

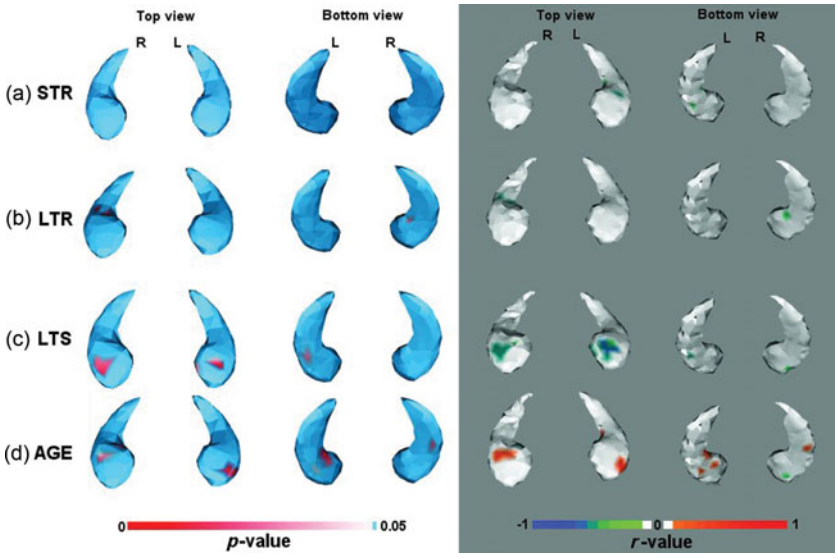


Figure 3. (Colour online) Correlations between local hippocampal changes and memory scores and age in the MEM group. Maps show significant *p*-values ($p < 0.05$; red to white colors; left panel) and the corresponding R^2 correlation values (positive correlations: orange to red, negative correlations: green to blue; right panel).

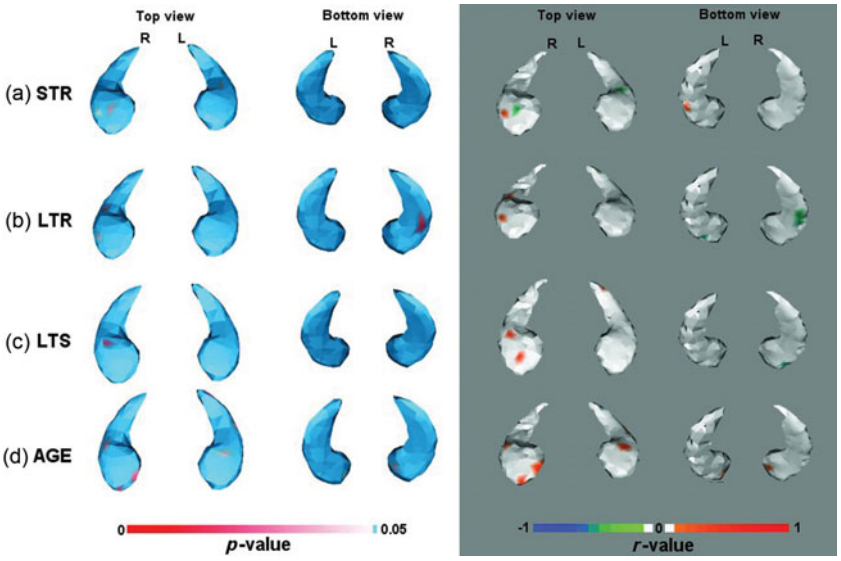


Figure 4. (Colour online) Correlations between local hippocampal changes and memory scores and age in the MEMnos group. Maps show significant *p*-values ($p < 0.05$; red to white colors; left panel) and the corresponding R^2 correlation values (positive correlations: orange to red, negative correlations: green to blue; right panel).

shape and memory scores and age (Figure 4). Only very small areas were detected which approached non-significance ($p = 0.05$).

ROC analysis

Hippocampal volumes discriminated MEM from HC with an accuracy of 0.63 and 0.66 for the left and right hippocampi respectively. Similar values were obtained for the comparison between MEMnos and HC (0.64 and 0.61). When considering participants with borderline deficits, hippocampal volumes showed chance-level accuracy (0.54 and 0.60).

Discussion

The aim of this work was to investigate the association between memory deficits and hippocampal changes in a large population-based elderly cohort. Participants were divided in three groups: normal controls, MEM, and MEMnos. We found that: (i) hippocampal volumes were smaller in the MEM group compared with controls; (ii)

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

hippocampal volumes correlated with long-term memory deficits within the MEM group, significant associations mapping to the anterior dorsal CA1 hippocampus.

These findings are consistent with previous studies showing that hippocampal atrophy can discriminate between normal aging and memory impairment in the general population. A number of previous studies failed to detect any difference in hippocampal volumes between patients and controls (see Table 1: Pantel *et al.*, 2003; Kumar *et al.*, 2006; Reitz *et al.*, 2009; Cui *et al.*, 2012). Others, conversely, reported significant volume reductions, ranging between 7%–14% along the spectrum from MCI to incident AD (see Table 1; Wolf *et al.*, 2001; Pennanen *et al.*, 2004; Scher *et al.*, 2007; Zhang *et al.*, 2011). Our study is in agreement with the latter studies, as we observed hippocampal volume reductions of 7%–8% in participants with memory deficits, and overall, supports the discriminant validity of hippocampal volumetry in the general population. Some caution is, however, warranted when generalizing these results since hippocampal volumetry showed a low accuracy in discriminating participants at the individual level.

The observed correlations between hippocampal atrophy and memory deficits are in line with the known pathophysiology of memory decline in patients with AD pathology. In the whole sample, memory deficits were associated with hippocampal reductions in regions that have been reported to be specific to AD (dorsal head of the CA1 field; Apostolova *et al.*, 2006; Frisoni *et al.*, 2008) but also in regions non-uniquely associated to AD (CA1 tail and anterior ventral subiculum; Apostolova *et al.*, 2006; Frisoni *et al.*, 2008). When the analysis was restricted to the MEM group, the only significant associations was in the CA1 dorsal head, suggesting that the observed effects are specific to patients with an AD-like profile. Although some degree of overlap exists in the hippocampal head between age- and memory-related patterns (Frisoni *et al.*, 2008), overall we observed only a marginal overlap, mainly in the right dorsolateral CA1 field. The hypothesis of an AD-like pattern in the MEM group is further supported by the observation that congruent memory-hippocampus associations were observed in MEM but not in MEMnos participants. Moreover, MEMnos showed a hippocampal reduction of about 6%–7% compared with controls, similarly to the MEM participants, however, this reduction fell short of statistical significance.

Shape-based analysis detected no significant difference between MEM and HC, while MEMnos showed shape changes in the posterior hippocampus compared with HC and MEM groups. The lack of significant shape changes in the MEM group is unexpected, since shape-based analysis is expected to detect very subtle and focal changes otherwise not measurable with volumetry. A possible explanation for this negative finding might be that the hippocampus of our participants was affected by small and generalized (rather than focal) reductions. However, this explanation is unlikely since a large amount of literature has previously shown that atrophy affects specific hippocampal regions and does not involve the whole hippocampus even in full-blown AD (Wang *et al.*, 2003; Apostolova *et al.*, 2006; Frisoni *et al.*, 2008; Ferrarini *et al.*, 2009). An alternative explanation might be that shape-based biomarkers are influenced to a greater extent than volumetric measures by population variability, a factor that might have reduced sensitivity to detect group-differences in our sample. Since no previous study has assessed shape changes in a population-based sample, except for the study by Sher *et al.* (2007), which, however, analyzed patients at a more advanced disease stage (i.e. incident AD), future studies are needed to clarify this issue.

The pattern of hippocampal changes in the MEMnos group showed a good correspondence with regions typically (albeit not specifically) involved in AD, such as the tail of the CA1 field and the posterior subiculum. On the one hand, this result might indicate that hippocampal changes in this subgroup are related to a neurodegenerative process as well. On the other hand, the pathophysiological mechanism underlying these changes can hardly be ascribed to AD pathology. Indeed, as we observed above, the lack of an association between memory deficits and morphological/volume changes in this group does not support the view of an AD aetiology. Since these participants represent a clinically heterogeneous group, who did not fulfill the criteria for AACD (due to the lack of evidence of a gradual onset of cognitive deficits from at least six months, or the past or current presence of medical/psychiatric conditions or use of psychoactive substances), other factors are likely responsible for the observed effect in hippocampal morphology. Although we cannot speculate on the possible mechanisms underlying these abnormalities, it is likely that other factors are involved, e.g. environmental or genetic factors, as well as other pathologies. Since this pattern was also observed in the comparison between MEMnos and MEM, an alternative explanation might be that these morphological changes simply represent a specific feature of this heterogeneous sample. Notwithstanding this uncertainty, we observe that volumetric and shape-based analysis was able to detect significant differences between the two study groups

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

in morphology and in cognitive-morphological associations. This is encouraging in the perspective of the use of hippocampal biomarkers for the differential diagnosis and prediction of dementia.

Some limitations of the study ought to be recognized. The relatively small size of the samples might have introduced some bias. However, although smaller than the original IPREA dataset, our cohort was drawn from a representative sample of the general elderly Italian population. Another possible limitation is that we did not collect biological markers of AD pathology, such as amyloid levels on cerebrospinal fluid (CSF) or amyloid positron emission tomography (PET), and hypometabolism on fluorodeoxyglucose (FDG)-PET (Albert *et al.*, 2011). These examinations indeed were not part of the original study protocol. The IPREA study, however, is an ongoing project and clinical follow-up will be collected for all the participants. These data will indeed be used in future studies to ascertain stability or conversion of participants with memory deficits. Finally, no phantom data were collected to compare MRI signal across sites. Notwithstanding this limitation, the use of mobile MRI can offer some advantages in the context of epidemiological studies. Mobile MRI systems indeed enable: (i) to collect a large number of MRI scans according to a tight schedule, otherwise challenging for standard neuroradiology units, and (ii) to maximize comparability across sites thanks to the use of fully harmonized machines and protocol parameters.

In conclusion, shape and volume hippocampal analysis showed that cognitive deficits are associated with regionally specific hippocampal changes in a population-based elderly cohort. Direct group comparisons showed smaller hippocampal volumes but failed to detect significant shape differences between patients with pure memory deficits and controls. Overall, these results are consistent with the view that hippocampal changes occur early in patients suspected to be at greater risk for AD.

Conflict of interest

None.

Description of author's role

L. Ferrarini, J. Milles, G.B. Frisoni, and M. Pievani designed the study. L. Ferrarini and M. Pievani wrote the paper. L. Ferrarini and M. Pievani were responsible for the statistical design of the study and for carrying out the statistical analysis. B. van Lew, J.H.C. Reiber, G.B. Frisoni, and J. Milles assisted with the writing of the paper. E. Scafato is the principal investigator of the IPREA. L. Galluzzo, C. Gandin, and E. Scafato collected the data and assisted with the writing of the paper.

Acknowledgments

This project was supported by a grant from the Italian Ministry of Health ("Progetti strategici di ricerca finalizzata sulla malattia di Alzheimer", Art. 12 D. Lgs. 502/92 and Art. 12bis D. Lgs. 229/99), Italy. The project was further supported by the Gisela Thier Foundation.

References

Albert, M. S. *et al.* (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7, 270–279. doi:10.1016/j.jalz.2011.03.008.

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. Washington, DC: American Psychiatric Association.

Amodio, P. *et al.* (2002). Variability of trail making test, symbol digit test and line trait test in normal people. A normative study taking into account age-dependent decline and sociobiological variables. *Aging Clinical and Experimental Research*, 14, 117–131.

Apostolova, L. G. *et al.* (2006). Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. *Archives of Neurology*, 63, 693–699.

Basso, A., Capitani, E. and Laiacona, M. (1987). Raven's coloured progressive matrices: normative values on 305 adult normal controls. *Functional Neurology*, 2, 189–194.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann Statist*, 29, 1165–1188.

Bobinski, M. *et al.* (2000). The histological validation of post mortem magnetic resonance imaging-determined hippocampal volume in Alzheimer's disease. *Neuroscience*, 95, 721–725.

Clark, C. M. *et al.* (2008). Biomarkers for early detection of Alzheimer pathology. *Neurosignals*, 16, 11–18.

Colliot, O. *et al.* (2008). Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus. *Radiology*, 248, 194–201. doi:10.1148/radiol.2481070876.

Cui, Y. *et al.* (2012). Automated detection of amnestic mild cognitive impairment in community-dwelling elderly adults: a combined spatial atrophy and white matter alteration approach. *Neuroimage*, 59, 1209–1217.

de Jong, L. W. *et al.* (2011). Shape abnormalities of the striatum in Alzheimer's disease. *Journal of Alzheimer's Disease*, 23, 49–59. doi:10.3233/JAD-2010-101026.

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

den Heijer, T. *et al.* (2010). A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline. *Brain*, 133, 1163–1172.

Devanand, D. P. *et al.* (2007). Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology*, 68, 828–836.

Emery, V. O. (2011). Alzheimer disease: are we intervening too late? Pro. *Journal of Neural Transmission*, 118, 1361–1378. doi:10.1007/s00702-011-0663-0.

Erkinjuntti, T. *et al.* (2000). Research criteria for subcortical vascular dementia in clinical trials. *Journal of Neural Transmission, Supplementum*, 59, 23–30.

Ferrarini, L., Palm, W. M., Olofsen, H., van Buchem, M. A., Reiber, J. H. and Admiraal-Behloul, F. (2006). Shape differences of the brain ventricles in Alzheimer's disease. *Neuroimage*, 32, 1060–1069.

Ferrarini, L., Olofsen, H., Palm, W. M., van Buchem, M. A., Reiber, J. H. and Admiraal-Behloul, F. (2007). GAMEs: growing and adaptive meshes for fully automatic shape modeling and analysis. *Medical Image Analysis*, 11, 302–314.

Ferrarini, L. *et al.* (2008). Ventricular shape biomarkers for Alzheimer's disease in clinical MR images. *Magnetic Resonance in Medicine*, 59, 260–267. doi:10.1002/mrm.21471.

Ferrarini, L., Frisoni, G. B., Pievani, M., Reiber, J. H., Ganzola, R. and Milles, J. (2009). Morphological hippocampal markers for automated detection of Alzheimer's disease and mild cognitive impairment converters in magnetic resonance images. *Journal of Alzheimer's Disease*, 17, 643–659. doi:10.3233/JAD-2009-1082.

Frisoni, G. B. *et al.* (2008). Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. *Brain*, 131, 3266–3276. doi:10.1093/brain/awn280.

Frisoni, G. B., Fox, N., Jack, C. R., Scheltens, P. and Thompson, P. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*, 6, 67–77. doi:10.1038/nrneurol.2009.215.

Godin, O. *et al.* (2010). Joint effect of white matter lesions and hippocampal volumes on severity of cognitive decline: the 3C-Dijon MRI study. *Journal of Alzheimer's Disease*, 20, 453–463.

Grigoletto, F., Zappalà, G., Anderson, D. W. and Lebowitz, B. D. (1999). Norms for the Mini-Mental State Examination in a healthy population. *Neurology*, 53, 315–320.

Herukka, S. K., Pennanen, C., Soininen, H. and Pirttilä, T. (2008). CSF Abeta42, tau and phosphorylated tau correlate with medial temporal lobe atrophy. *Journal of Alzheimer's Disease*, 14, 51–57.

Jack, C. R., Jr. *et al.* (2010). Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain*, 133, 3336–3348. doi:10.1093/brain/awq277.

Karas, G. *et al.* (2008). Amnestic mild cognitive impairment: structural MR imaging findings predictive of conversion to Alzheimer disease. *American Journal of Neuroradiology*, 29, 944–949. doi:10.3174/ajnr.A0949.

Katz, S., Downs, T. D., Cash, H. R. and Grotz, R. C. (1970). Progress in development of the index of ADL. *Gerontologist*, 10, 20–30.

Klöppel, S. *et al.* (2008). Automatic classification of MR scans in Alzheimer's disease. *Brain*, 131, 681–689. doi:10.1093/brain/awm319.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78, 1464–1480.

Kumar, R. *et al.* (2006). Clinical and neuroimaging correlates of mild cognitive impairment in a middle-aged community sample: the personality and total health through life 60+ study. *Dementia and Geriatric Cognitive Disorders*, 21, 44–50.

Lawton, M. P. and Brody, E. M. (1969). Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*, 9, 179–186.

Levy, R. (1994). Aging-associated cognitive decline. Working Party of the International Psychogeriatric Association in collaboration with the World Health Organization. *International Psychogeriatrics*, 6, 63–68.

McKeith, I. G. *et al.* (1996). Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. *Neurology*, 47, 1113–1124.

McKhann, G. M., Albert, M. S., Grossman, M., Miller, B., Dickson, D. and Trojanowski, J. Q. (2001). Clinical and pathological diagnosis of frontotemporal dementia: report of the work group on frontotemporal dementia and Pick's disease. *Archives of Neurology*, 58, 1803–1809.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D. and Stadlan, E. M. (1984). 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*, 34, 939–944.

Measso, G., Cavarzeran, F., Zappalà, C., Lebowitz, B. D., Crook, T. H. and Pirozzolo, F. J. (1993). The Mini-Mental State Examination: normative study of an Italian random sample. *Dev Neuropsychology*, 9, 77–85.

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

Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15, 1–25.

Novelli, G., Papagno, C., Capitani, E., Laiacona, M., Vallar, G. and Cappa, S. F. (1986). Tre test di ricerca e produzione lessicale. *Archivio di Psicologia, Neurologia e Psichiatria*, 47, 477–506.

Pantel, J., Kratz, B., Essig, M. and Schröder, J. (2003). Parahippocampal volume deficits in subjects with aging-associated cognitive decline. *American Journal of Psychiatry*, 160, 379–382.

Pennanen, C. *et al.* (2004). Hippocampus and entorhinal cortex in mild cognitive impairment and early AD. *Neurobiology of Aging*, 25, 303–310.

Persson, J. *et al.* (2006). Structure-function correlates of cognitive decline in aging. *Cerebral Cortex*, 16, 907–915.

Petersen, R. C. *et al.* (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, 58, 1985–1992.

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

Pruessner, J. C. *et al.* (2000). Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. *Cerebral Cortex*, 10, 433–442.

Reitz, C. *et al.* (2009). Linking hippocampal structure and function to memory performance in an aging population. *JAMA Neurology*, 66, 1385–1392.

Roman, G. C. *et al.* (1993). Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN international workshop. *Neurology*, 43, 250–260.

Sachdev, P. S. *et al.* (2013). Factors predicting reversion from mild cognitive impairment to normal cognitive functioning: a population-based study. *PLoS One*, 8, e59649. doi:10.1371/journal.pone.0059649.

Scafato, E. *et al.* (2005). Italian Project on Epidemiology of Alzheimer's disease (I.PR.E.A.): study design and methodology of cross-sectional survey. *Aging Clinical and Experimental Research*, 17, 29–34.

Scafato, E. *et al.* (2010). Prevalence of aging-associated cognitive decline in an Italian elderly population: results from cross-sectional phase of Italian PRoject on Epidemiology of Alzheimer's disease (IPREA). *Aging Clinical and Experimental Research*, 22, 440–449. doi:10.3275/6970.

Scher, A. I. *et al.* (2007). Hippocampal shape analysis in Alzheimer's disease: a population-based study. *Neuroimage*, 36, 8–18.

Selkoe, D. J. (2002). Alzheimer's disease is a synaptic failure. *Science*, 298, 789–791.

Smith, S. M. *et al.* (2002). Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage*, 17, 479–489.

Spinnler, H. and Tognoni, G. (1987). Standardizzazione e taratura italiana di test neurologici. *Italian Journal of Neurological Sciences*, 8, 1–120.

Tapiola, T. *et al.* (2008). MRI of hippocampus and entorhinal cortex in mild cognitive impairment: a follow-up study. *Neurobiology of Aging*, 29, 31–38.

Thompson, P. M. *et al.* (2004). Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage*, 22, 1754–1766.

Visser, P. J. *et al.* (1999). Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment. *Journal of Neurology*, 246, 477–485.

Wang, L. *et al.* (2003). Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. *Neuroimage*, 20, 667–682.

Wolf, H. *et al.* (2001). Hippocampal volume discriminates between normal cognition; questionable and mild dementia in the elderly. *Neurobiology of Aging*, 22, 177–186.

Zhang, H. *et al.* (2011). Grey matter atrophy of basal forebrain and hippocampus in mild cognitive impairment. *Journal of Neurology, Neurosurgery and Psychiatry*, 82, 487–493.

Zhang, H. *et al.* (2012). Gray matter atrophy patterns of mild cognitive impairment subtypes. *Journal of the Neurological Sciences*, 315, 26–32.

Zhuang, L. *et al.* (2013). Microstructural white matter changes, not hippocampal atrophy, detect early amnestic mild cognitive impairment. *PLoS One*, 8, e58887. doi:10.1371/journal.pone.0058887.

Appendix 1

IPREA working group

E. Scafato, G. Farchi, S. Giampaoli, S. Mariotti, C. Gandin, L. Galluzzo, S. Ghirini, S. Martire, and L. Di Pasquale, Istituto Superiore di Sanità (ISS), Rome; S. Maggi, G. Crepaldi, G. Enzi, and P. Gallina, Italian Research Council (CNR), Rome and University of Padova, Padova; D. Inzitari, M. Baldereschi, and A. Di Carlo, University of Florence and CNR, Florence; G.B. Frisoni and S. Galluzzi, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Fatebenefratelli, Brescia; C. Gandolfo and M. Conti, University of Genova, Genoa; D. Postacchini, G. Cruciani and C. Giuli, INRCA, Fermo (AP); A. Capurso, V. Solfrizzi and F. Panza, University of Bari and IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia; F. Rengo and P. Abete, University of Naples, Naples; M. Motta, University of Catania, Catania; R. Negrini, P. Forti and P. Tabanelli, USL Bologna Nord, Bologna; A. Cocchi and G. Zuccal, University of Rome "Sacro Cuore", Rome; F. Cacciatore, C. Calabrese, G. Sica, and A. Estraneo, Salvatore Maugeri Foundation, IRCCS, Pavia; D. Consoli, F. Naso, G. Torcasio, ASL 8, Vibo Valentia; P. Mecocci, P. Rinaldi, V. Serafini, U. Senin, University of Perugia, Perugia.

Appendix 2

Diagnostic and AACD criteria

Aging-associated cognitive decline criteria were developed in 1994 to describe the transitional phase between normal aging and dementia by the International Psychogeriatric Association (IPA) in collaboration with the World Health Organization (Levy, 1994). According to these criteria, AACD definition is based on the results of a complete clinical and neuropsychological assessment including (i) ADL and IADL assessment (Lawton and Brody, 1969; Katz *et al.*, 1970), (ii) clinical dementia rating (CDR; Morris, 1993), (iii) MMSE battery for global cognition (Measso *et al.*, 1993; Grigoletto *et al.*, 1999), and (iv) neuropsychological tests for the assessment of the major cognitive domains: (1) memory and learning with the Buschke Fuld Selective Reminding Test (Spinnler and Tognoni, 1987); (2) attention with the Trail Making Test, A and B (Amodio *et al.*, 2002); (3) verbal ability with the Verbal Fluency Test for semantic categories (Novelli *et al.*, 1986); (4) visuoconstructive function with the Constructional Praxis test (Spinnler and Tognoni, 1987); and (5) problem-solving with the Raven colored progressive matrices (Basso *et al.*, 1987).

Based on this evaluation, the following diagnostic categories were defined:

1. Participants were classified as cognitively normal if they showed:

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press

(a) normal cognition on MMSE and neuropsychological tests
(b) normal activities of daily living (ADL and IADL)
(c) CDR = 0

2. Participants were classified as AACD if they showed:
(a) objective cognitive deficits in at least one cognitive domain
(b) subjective gradual cognitive decline (reported by the participant or an informant) present for at least six months
(c) absence of any present or past medical/psychiatric conditions or any psychoactive substances use that could explain the cognitive deficits
(d) normal activities of daily living
(e) CDR = 0.5

3. Participants were classified as those with dementia if they showed:
(a) abnormal neuropsychological tests
(b) abnormal activities of daily living
(c) CDR > 0.5

Additionally, the following three groups who did not fulfill the above criteria were identified (Scafato *et al.*, 2010):

4. People with cognitive complaints (CC):
(a) normal cognition on MMSE and neuropsychological tests
(b) cognitive complains by the subject and/or the informant
(c) normal activities of daily living
(d) CDR = 0

5. People with objective evidence of cognitive decline without cognitive complains (OCD):
(a) objective cognitive deficits in at least one cognitive domain
(b) absence of cognitive complains (nor by the subject nor by the informant)
(c) normal activities of daily living
(d) CDR = 0

6. People with AACD not otherwise specified (AACD-nos, corresponding to the AACD-3 group in the paper by Scafato *et al.*, 2010). These participants fulfilled three out of five criteria for AACD (category 2). The criteria not met were the second (gradual onset of symptoms) and the third (absence of comorbidities).

For category 3 (people with dementia), a diagnosis of dementia was defined based on international criteria for the various forms of dementia: National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS–ADRDA) criteria for AD (McKhann *et al.*, 1984); National Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherché et l'Enseignement en Neurosciences (NINDS–AIREN) criteria for VaD (Roman *et al.*, 1993) and Erkinjuntti criteria for subcortical VaD (Erkinjuntti *et al.*, 2000); the McKhann criteria for frontotemporal dementia (FTD) (McKhann *et al.*, 2001); the McKeith criteria for dementia with Lewy bodies (DLB) (McKeith *et al.*, 1996); and DSM-IV-TR criteria for depressive pseudo-dementia, dementia associated with Parkinson's disease, medical conditions, and drug abuse (American Psychiatric Association, 1994).

Appendix 3

Permutation test

Permutation tests have been successfully used for the analysis of brain images (Nichols and Holmes, 2002; Thompson *et al.*, 2004): they require limited assumptions and are corrected for multiple comparisons.

Given two populations G1 and G2, can we localize statistically significant differences on the average surface? The outcome of the permutation tests is, in the first place, a *p*-value for the omnibus hypothesis the two groups G1 and G2 are drawn from the same population. Moreover, we obtain the *p*-values for each node in the model, telling us whether the distribution of that node in space is the same in G1 and G2 or not. Since all meshes are co-registered to a standard space, significant differences in space distribution of a certain location indicates either a significant enlargement or shrinking of one population with respect to the other (hence pointing to atrophy).

Permutation tests can be summarized as follows:

1. considering two groups G1 and G2:

(a) for each node in the model, build up two clouds of points, C1 and C2, considering the positions the node assumes through all the shapes in G1 and G2;
(b) C1 and C2 are compared via a Hotelling's T2 statistic test: outcome of the test is the *t*-value for the node comparison (is the node distributed in space in significantly different ways?); such a test tests both the average positions in space for the two clouds, and their variances.

2. for $N_{\text{perm}} = 10{,}000$ times, two groups of shapes A and B are built up by randomly mixing G1 and G2, and point 1 is performed on them. Only the highest *t*-value is stored for each iteration;
3. a critical *t*-value t_c is evaluated as the *k*th highest value of all the N_{perm} *t*-values previously stored (plus the t_{Max} for the original division in G1–G2), where

$$k = \alpha \times N_{\text{perm}}\, c + 1,\ \alpha = 0.05 \tag{1}$$

4. the *p*-value for the omnibus hypothesis "G1 and G2 are the same" is evaluated as

$$p_{\text{value}} = N/N_{\text{tests}},\ \text{where} \tag{2}$$

$$N = \#\{\text{stored}\, t_{\text{values}}\, |t_{\text{values}} > t_c\}, \tag{3}$$

$$N_{\text{tests}} = N_{\text{perm}} + 1; \tag{4}$$

5. finally, point 4 is applied to each single node, counting how many *t*-values are higher than the *t*-value associated with a particular node in the original G_1–G_2 grouping of shapes, and dividing the number for $N_{\text{perm}} + 1$; this leads to a *p*-value (corrected for multi-tests) for each node in the model.

https://doi.org/10.1017/S1041610213002627 Published online by Cambridge University Press