Journal of Alzheimer's Disease 46 (2015) 199–209
DOI 10.3233/JAD-142280
IOS Press

Fully Automated Atlas-Based Hippocampus Volumetry for Clinical Routine: Validation in Subjects with Mild Cognitive Impairment from the ADNI Cohort

Per Suppa[a,b], Harald Hampel[c], Lothar Spies[b], Jochen B. Fiebach[d], Bruno Dubois[c,1], Ralph Buchert[a,*,1] and Alzheimer's Disease Neuroimaging Initiative[2]
[a]*Department of Nuclear Medicine, Charité, Berlin, Germany*
[b]*jung diagnostics GmbH, Hamburg, Germany*
[c]*Université Pierre et Marie Curie, Département de Neurologie, Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A), Hôpital de la Salpêtrière, Paris, France*
[d]*Center for Stroke Research Berlin, Charité, Berlin, Germany*

Handling Associate Editor: Babak Ardekani

Accepted 9 February 2015

Abstract. Hippocampus volumetry based on magnetic resonance imaging (MRI) has not yet been translated into everyday clinical diagnostic patient care, at least in part due to limited availability of appropriate software tools. In the present study, we evaluate a fully-automated and computationally efficient processing pipeline for atlas based hippocampal volumetry using freely available Statistical Parametric Mapping (SPM) software in 198 amnestic mild cognitive impairment (MCI) subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI1). Subjects were grouped into MCI stable and MCI to probable Alzheimer's disease (AD) converters according to follow-up diagnoses at 12, 24, and 36 months. Hippocampal grey matter volume (HGMV) was obtained from baseline T1-weighted MRI and then corrected for total intracranial volume and age. Average processing time per subject was less than 4 minutes on a standard PC. The area under the receiver operator characteristic curve of the corrected HGMV for identification of MCI to probable AD converters within 12, 24, and 36 months was 0.78, 0.72, and 0.71, respectively. Thus, hippocampal volume computed with the fully-automated processing pipeline provides similar power for prediction of MCI to probable AD conversion as computationally more expensive methods. The whole processing pipeline has been made freely available as an SPM8 toolbox. It is easily set up and integrated into everyday clinical patient care.

Keywords: ADNI, Alzheimer's disease, atlas-based segmentation, fully automated, hippocampus volumetry, magnetic resonance imaging, mild cognitive impairment, prediction

*Correspondence to: Ralph Buchert, PhD, Charité – Universitätsmedizin Berlin, Department of Nuclear Medicine, Charitéplatz 1, 10117 Berlin, Germany. Tel.: +49 30 450627059; Fax: +49 30 4507527959; E-mail: Ralph.Buchert@charite.de.

[1]These authors contributed equally as senior authors.

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

ISSN 1387-2877/15/$35.00

INTRODUCTION

The National Institute on Aging–Alzheimer's Association working group [1], the International Working Group [2–4], and the European Federation of the Neurological Societies [5] recommend the use of core feasible biomarkers such as magnetic resonance imaging (MRI)-based hippocampus volumetry to complement clinical criteria with evidence for underlying Alzheimer's disease (AD) pathology in order to improve the prognostic accuracy in amnestic mild cognitive impairment (MCI).

Fully automated tools allow user-independent determination of hippocampal grey matter volume (HGMV) and, therefore, eliminate the need for time consuming manual segmentation of the hippocampus by trained experts. There are various software tools for fully automated hippocampal volumetry available [6–9]. In a systematic comparison of these tools to support the qualification of HGMV as an imaging biomarker for enrichment of clinical trials in predementia stages of AD by the European Medicines Agency (EMA), all these tools provided about the same accuracy for prediction of probable AD in MCI subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [10]: the area under the receiver-operating characteristics curve for prediction of MCI-to-AD conversion within 2 years ranged from 0.73 to 0.76. A short description of the tools evaluated in this EMA qualification process is given in the Discussion section.

However, requirements for clinical routine differ from requirements for clinical trials: for automatic hippocampal volumetry to be useful also in busy everyday clinical routine, the computation needs to be fast, ideally uses only freely available software, is easy to set-up, and still provides similar prognostic value as the more sophisticated methods tested for the EMA qualification.

Statistical Parametric Mapping (SPM, Wellcome Trust Centre for Neuroimaging, London, UK) [11] is freely available, widely used, well documented open source and, therefore, is an ideal candidate tool kit for adoption in clinical practice and to support standardization [12]. Atlas-based hippocampal volumetry with SPM has been validated against manual tracing by clinical experts [6, 13]. However, a previous study of SPM5-based HGMV in ADNI-MCI subjects found inferior prediction accuracy compared to more sophisticated methods [14]. In the present study, we propose a novel SPM8 processing pipeline for hippocampal volumetry using tissue probability maps from elderly control subjects and evaluate it in ADNI-MCI subjects. This SPM processing pipeline, too, avoids computationally expensive steps in order to guarantee compatibility with the workflow both in busy in-patient memory clinics and primary care or specialist private practices with time restrictions and high patient throughput. Such a fast and easy-to-use method might reduce the barrier for the translation of hippocampal volumetry into everyday clinical routine. In order to support this, the software used in the present study (termed "HV") has been made freely available from the SPM homepage (http://www.fil.ion.ucl.ac.uk/spm/ext/#HV).

MATERIAL AND METHODS

ADNI subjects

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public private partnership. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada.

A total of 4,733 datasets from the ADNI1 database were identified and downloaded in January 2012 using the following search criteria in the ADNI/LONI search mask: Projects: ADNI, Modality: MRI, Series: MP*AGE. Subjects were included if a 1.5 T MRI screening scan was available and if they were diagnosed as MCI at time of inclusion and had received follow-up diagnoses at 12, 24, and 36 months (ADNI-MCIs). ADNI-MCIs who were diagnosed as AD at a time-point during this period but later were reclassified to MCI were excluded. In total, 198 ADNI-MCIs were eligible: 103 ADNI-MCIs who had remained stable over 36 months and 95 ADNI-MCIs who had converted to AD within this time period. Thirty-two MCI-to-AD converters had converted at the 12 months follow-up examination, 43 subjects between the 12 and 24 months examinations, and 20 subjects had converted between the 24 and 36 months examinations.

As control group we included all subjects from ADNI1 for whom a 1.5 T MRI screening scan was available for download and who were documented as normal throughout a period of 36 months after baseline clinical examination. This resulted in the inclusion of 137 subjects. There were no further inclusion or exclusion criteria for the controls. There was no selection according to age or gender.

The MR acquisition protocol used in the ADNI subjects has been described in detail in [15]. In brief, MR imaging was performed on 1.5T GE, Philips, or Siemens systems using acquisition protocols that had been optimized to harmonize image characteristics across different platforms. High resolution T1-weighted MRI scans were collected using a sagittal 3-dimensional magnetization prepared rapid gradient echo (3D-MPRAGE) sequence with an approximate TR = 2400 ms, minimum full TE, approximate TI = 1000 ms, and approximate flip angle of 8° (scan parameters varied between sites, scanner platforms, and software versions). Scans were collected with a 24 cm field-of-view and an acquisition matrix of $192 \times 192 \times 166$ (x, y, z dimensions), to yield a standard voxel size of $1.25 \times 1.25 \times 1.2\,\text{mm}^3$. Images were then reconstructed to give a $256 \times 256 \times 166$ matrix and voxel size of approximately $1 \times 1 \times 1.2\,\text{mm}^3$.

Two 3D-MPRAGE scans had been acquired in the same imaging session (back-to-back scans). We consistently selected the first scan to mimic clinical routine in which there is usually only a single scan available. All images were downloaded as "unpreprocessed" (no gradwarp, B1 non-uniformity or N3-N3 correction, see http://adni.loni.usc.edu/methods/mri-analysis/mri-pre-processing/).

Image preprocessing

3D-MPRAGE images were first segmented (into tissue classes) and stereotactically normalized to a whole brain template using the unified segmentation engine of SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK) which combines both processes in a single step. Correction for signal non-uniformity was also included [16]. The unified segmentation engine is guided by tissue probability maps (TPMs) for grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). TPMs provided by SPM8 are based on scans acquired from healthy young adults. Since the ADNI-MCI population is significantly older, SPM's default TPMs and default whole brain template were replaced by freely available TPMs and a freely available whole brain template which both had been generated from a population of 662 healthy elderly subjects aged between 63 and 75 years [17]. These TPMs feature a voxel resolution of $1 \times 1 \times 1\,\text{mm}^3$ while SPM's default TPMs are $2 \times 2 \times 2\,\text{mm}^3$. Default settings of the segmentation engine were used as described in [18]. Modulation of stereotactically normalized component images for correction of warping-associated volume changes was performed using the determinant of the Jacobian of the transformation field.

Hippocampal volumetry

The unified segmentation engine of SPM8 provides separate component images for GM, WM, and CSF, all in the template space. HGMV was determined by multiplying the normalized and modulated GM component image with a predefined binary mask for both hippocampi from a freely available toolbox [19], i.e., HGMV was obtained by summing the modulated GM intensity over all voxels within the hippocampus mask. This mask comprises cornus ammonis, fascia dentate, and subiculum substructures as defined by Amunts and coworkers [20]. The binary mask features an isotropic resolution of 1 mm matching the resolution of the component images. Total GM volume (GMV), total WM volume (WMV), and total CSF volume (CSFV) were calculated by summing up the voxel intensities of the normalized and modulated component images of the corresponding tissue class. Total intracranial volume (TIV) was estimated as the sum of GMV, WMV, and CSFV.

We also evaluated the prognostic value of hippocampus white matter volume (HWMV) and hippocampus parenchymal volume (HPV). HWMV was obtained by applying the same hippocampus mask to the patient's modulated WM component image; HPV was computed as HPV = HGMV+HWMV.

Correction for head size and age

HGMV was corrected for TIV and age to reduce inter-subject variability of no interest [21]. For this purpose, first bilinear regression analysis was performed in the group of ADNI-normals (137 subjects) with HGMV as dependent variable and TIV and age as predictors. Then, HGMV of ADNI-MCIs was adjusted to mean age and mean TIV of ADNI-normals according to

$$HGMV_{ad} = HGMV + a_{HGMV}\,(\langle TIV\rangle - TIV) + b_{HGMV}\,(\langle age\rangle - age) \quad (1)$$

where *TIV* and *age* represent TIV and age of the ADNI-MCI subject to be adjusted. a_{HGMV} and b_{HGMV}

are the regression coefficients from the bilinear fit to the ADNI-normals, and $\langle age \rangle$ and $\langle TIV \rangle$ denote mean age (= 75.7 years) and mean TIV (= 1450 ml) of ADNI-normals at baseline.

HWMV and HPV were corrected for TIV and age analogously.

Validation

The fully-automated SPM8 pipeline was validated by head-to-head comparison with a semi-automated method (HV-SNT, Medtronic Surgical Navigation Technologies, Louisville, CO) which has been shown to provide excellent agreement with manual tracing of the hippocampus [22]. HV-SNT values were available for download from the ADNI homepage for 134 of the 198 ADNI-MCIs ($n = 68$ stable MCI). HV-SNT values were corrected for age and TIV using equation 1 with the exact same regression coefficients as for correction of SPM8-HGMV values.

Statistical analysis

Comparison of more than two groups with respect to age, TIV, and Mini-Mental State Examination (MMSE) score was performed by ANOVA. In case of a significant effect, *post-hoc* tests between pairs of groups were performed using the homoscedastic or heteroscedastic unpaired two-sample *t*-test depending on the result of Levene's test for equality of variance.

The power of HGMVad for differentiation between MCI stables and MCI converters was assessed by receiver operating characteristic (ROC) analyses using the open source R package pROC [23]. The area (AUC) under the ROC curve was used as performance measure. Cut-off values on HGMVad for the determination of accuracy, sensitivity, specificity, and prognostic values were determined by maximizing the Youden index J = sensitivity+specificity - 1 [24], which is symmetric in sensitivity and specificity and, therefore, imposes equal penalty on false positive and false negative classifications. The Youden index describes the vertical distance of the ROC curve from the line of chance (diagonal line). Thus, the criterion of maximum Youden index selects the point on the ROC curve with largest vertical distance to the line of chance as operating point.

MCI subjects with HGMVad less than the cut-off were predicted to convert to clinical AD, MCI subjects with HGMVad equal to or larger than the cut-off were predicted to remain stable. Although the maximum of the Youden index is a rather simple model, it might be affected by statistical noise. Thus, overfitting cannot be ruled out so that estimates of diagnostic accuracy measures are most likely overly optimistic. In order to correct for overfitting, 100 repeats of 20-fold cross-validation were performed. In detail, we randomly split the whole patient sample into 20 mutually exclusive equal-sized subsets, so-called 'folds' (the size can differ between folds by 1 patient). MCI stable subjects and MCI converters were randomized independently into the 20 folds so that the class proportion (stables versus converters) in each fold was about the same as in the whole patient sample. Empirically, this 'stratification' process reduces bias of cross-validation [25]. For the i-th fold, we determined the cut-off according to the Youden criterion from the ROC curve including all patients from the other 19 folds, and then applied this cut-off to the patients in the i-th fold, i.e., each patient in the i-th fold was categorized as true positive (TP), false positive (FP), true negative (TN), or false negative (FN). After this was performed for each of the 20 folds, accuracy, sensitivity, specificity, and predictive values were computed over all patients in the whole sample based on their categorization (as TP, FP, TN, or FN) using the standard formulas. Thus, one run of stratified 20-fold cross-validation provided one estimate for accuracy, sensitivity, specificity and the predictive values. Stratified 20-fold cross-validation was repeated 100 times (with independent randomization into folds), which is a standard approach to approximate 'complete' cross-validation, i.e., the average over all possible folds. Accuracy measures were averaged over the 100 repeats.

Estimating errors of accuracy estimates by variance across repeats of cross-validation is limited by the risk of duplicated training samples. We therefore used Equation (3) in [25] to estimate the 95% confidence interval of the accuracy measures.

RESULTS

There was no difference between stable ADNI-MCIs and the three subgroups of ADNI-MCI converters (within 12, 24, 36 months) with respect to gender-controlled TIV ($p = 0.576$) nor with respect to age ($p = 0.818$) (Table 1). The MMSE score at baseline was higher in the stable ADNI-MCIs than in each of the three subgroups of ADNI-MCI converters ($p \leq 0.039$). The converter subgroups did not differ with respect to MMSE at baseline.

Computation of HGMV and TIV from a MPRAGE data set took less than four minutes per subject on

Table 1
Characteristics of ADNI-MCI subgroups and ADNI-normals at baseline

	Age	MMSE	TIV	HGMV	HGMVad
MCI stable ($n = 103$)	75.20 (7.18)	27.68 (1.66)	1512.5 (171.7)	8.24 (1.08)	8.13 (0.98)
MCI-to-AD converter within 12 months ($n = 32$)	74.66 (6.46)	26.56 (1.83)	1492.4 (173.2)	7.29 (0.99)	7.18 (0.94)
MCI-to-AD converter between 12 and 24 months ($n = 43$)	74.18 (7.87)	26.37 (1.59)	1518.6 (217.2)	7.69 (1.22)	7.52 (1.06)
MCI-to-AD converter between 24 and 36 months ($n = 20$)	75.76 (6.56)	26.50 (1.40)	1385.3 (151.3)	7.58 (0.92)	7.66 (0.92)
Normals ($n = 137$)	75.74 (5.24)	29.18 (0.99)	1450.0 (160.5)	8.69 (1.06)	8.69 (0.99)

Given are mean value (standard deviation). MMSE, Mini-Mental State Examination; TIV, total intracranial volume; HGMV, hippocampal grey matter volume (ml); HGMVad, total hippocampal grey matter volume adjusted to mean age and mean TIV of ADNI-normals (ml).

an Intel Core 2 Duo CPU with 3.33 GHz and 8 GB RAM.

Bilinear regression of HGMV in the group of ADNI-normals with TIV and age as predictors resulted in the following regression coefficients: $a_{HGMV} = 0.0011$ and $b_{HGMV} = -0.0609$ ml/year. These values were used in equation (1) to adjust HGMV for age and TIV (see subsection "Correction for head size and age").

There was a moderate to strong correlation of HGMVad between the fully-automated SPM8 pipeline and the semi-automated HV-SNT method (Pearson correlation coefficient = 0.72, $p = 2.2$ e-16).

ROC curves for the SPM8-based HGMVad are shown in Fig. 1. Maximum AUC of 0.78 was achieved for identification of ADNI-MCIs who converted within 12 months. There was a trend to lower AUCs for detection of ADNI-MCIs who converted within 24 (AUC = 0.72) and 36 months (AUC = 0.71). Cross-validated cut-offs and corresponding accuracies, sensitivities, specificities, and predictive values are listed in Table 2. Hippocampus white matter volume (HWMVad) provided considerably smaller AUC for prediction of MCI-to-AD conversion (0.67, 0.62, and 0.59 for conversion within 12, 24, or 36 months, respectively), whereas the AUC for the hippocampus parenchymal volume (HPVad) was only slightly smaller (0.76, 0.70, and 0.68 for conversion within 12, 24, or 36 months, respectively).

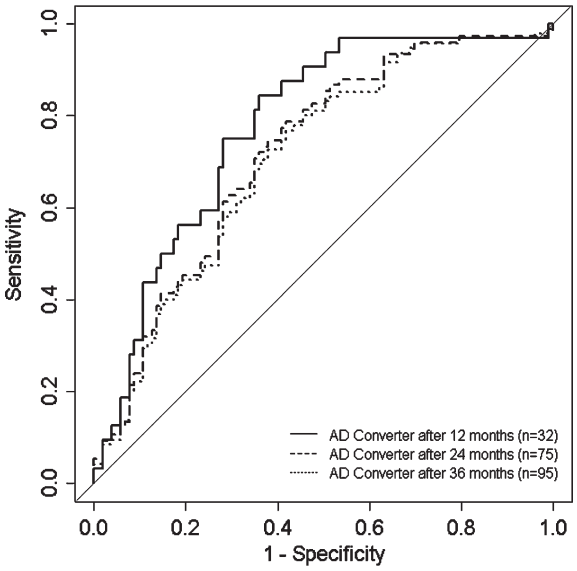

Fig. 1. Receiver operating characteristic curve for the identification of ADNI-MCI subjects who converted to probable AD within 12 (continuous line), 24 (broken line) and 36 months (dotted line) by total hippocampal volume corrected for total intracranial volume and age (HGMVad). Area under the curve is 0.78, 0.72, and 0.71, respectively.

The semi-automatic HV-SNT method (with the same correction for TIV and age) provided only slightly higher AUCs (0.80, 0.72, and 0.73 for 12,

Table 2
Area (AUC) under the ROC curve, cut-off value on HGMVad determined by the maximum Youden index, and accuracy measures for prediction of ADNI-MCI to probable AD conversion within 12, 24 or 36 months by hippocampal grey matter volume (HGMVad) adjusted to mean age and mean TIV of ADNI-normals. All accuracy measures were cross-validated by 100 repeats of 20-fold cross-validation. 95% confidence intervals (CI) are given in brackets. The 95%-CI for the AUC was obtained as described in [43], the 95%-CIs for the accuracy measures were estimated according to [25]. The standard deviation of the cut-off is given in round brackets

			Cross validated				
Interval [months]	AUC	Cut-off [ml]	Accuracy	Sensitivity	Specificity	PPV	NPV
12	0.78	7.85	0.66	0.75	0.63	0.39	0.89
	[0.69–0.86]	(0.1)	[0.58–0.73]	[0.67–0.82]	[0.55–0.71]	[0.31–0.47]	[0.83–0.93]
24	0.72	8.09	0.63	0.72	0.56	0.54	0.73
	[0.64–0.79]	(0.02)	[0.56–0.70]	[0.65–0.78]	[0.49–0.63]	[0.47–0.61]	[0.66–0.79]
36	0.71	8.12	0.63	0.73	0.55	0.6	0.69
	[0.63–0.78]	(0.01)	[0.56–0.69]	[0.66–0.79]	[0.48–0.62]	[0.53–0.67]	[0.62–0.75]

ml, milliliter; PPV, positive predictive value; NPV, negative predictive value.

24, and 36 months prediction). The difference in AUC between the HV-SNT method and the SPM8 processing pipeline was not significant statistically (two-tailed p according to [26] was 0.799, 1.000, and 0.726 for conversion within 12, 24, or 36 months, respectively).

DISCUSSION

The aim of the present study was to make a software tool for fully automated hippocampus volumetry freely available that (i) provides high processing speed by avoiding computationally expensive steps in order to guarantee compatibility with the workflow in busy clinical routine patient care, and (ii) results in similar performance for prediction of MCI-to-AD conversion as the tools evaluated in the EMA qualification process of HGMV as an imaging biomarker for enrichment of clinical trials in predementia stages of AD [10], namely FreeSurfer [27], NeuroQuant [28, 29], Learning Embeddings for Atlas Propagation [9], and Hippocampus Multi-Atlas Propagation and Segmentation [8, 30].

In FreeSurfer the segmentation of brain structures including the hippocampus is based on a probabilistic atlas (created from a manually labelled training set) providing means and variances of the MRI intensities of the considered brain structures separately for each location and each tissue class [6]. In addition, Freesurfer uses information on local spatial relationships between brain structures (e.g., 'the hippocampus is posterior and inferior of the amygdala') which means that the probability of a voxel to belong to a given brain structure is computed not only from its MRI intensity and prior probabilities (specified in the probabilistic atlas) but also as a function of the structure label of neighboring voxels. This additional prior information is also generated from the manually labeled training set (using spatially varying and non-isotropic Markov random fields). Registration of the MRI of the individual subject with the atlas space is performed by a low-dimensional affine transformation (12 degrees of freedom).

NeuroQuant uses discrete cosine transform nonlinear registration to transform the MR image of an individual brain to a probabilistic atlas customized for labeling anatomy in elderly subjects [29]. The automated segmentation of brain structures is performed with a similar method as used by FreeSurfer [6] (see above).

Learning Embeddings for Atlas Propagation provides a set of 30 MR brain images of young healthy subjects (median 30.5 y) in each of which a set of 83 anatomical structures including the hippocampi has been labelled manually [9]. These 30 atlases together with the MR image on an individual subject to be analyzed are embedded in a coordinate system in which the distance between two images is determined by their similarity (based on normalized mutual information) so that similar images are close to each other. Each one of the 30 atlases (or of a subset) is then propagated through the manifold defined by this coordinate system towards the individual MRI. This propagation is performed in several steps ('from neighbor to neighbor to neighbor... ') in order to avoid the need to estimate large deformations which most likely would be required for direct transformation of an atlas to the individual MRI. The rationale for this is that estimation of large deformations is particularly prone to errors. After all atlases have been propagated to the individual MRI, a probabilistic atlas is generated from the structure labels of the transformed atlases. This probabilistic segmentation is refined by using intensity information from the individual image [9].

Hippocampus Multi-Atlas Propagation and Segmentation starts with a hippocampal template library of manually segmented regions from 55 subjects, 36 clinically diagnosed probable AD patients, and 19 age-matched healthy controls [8, 30]. The individual subject's MR image is registered to the healthy control subject to which all of the template library scans had been registered. The registration is performed in two steps: non-linear brain-to-brain registration based on free form deformation [31] followed by linear hippocampus-to-hippocampus registration (6 degrees of freedom) to improve alignment of the hippocampi. After registration, an intensity threshold is applied to exclude white matter and CSF voxels. The resulting hippocampus image of the individual subject is compared to all hippocampi in the template library using cross-correlation. Finally, the hippocampus segmentations from the best 8 matches are combined using 'simultaneous truth and performance level estimation' with constraints on spatial smoothness [32].

The processing pipeline proposed in the present study is based on combined segmentation and nonlinear brain-to-brain registration of the MR image of the individual subject to a whole brain template using the unified segmentation engine of SPM8. Unified segmentation is guided by *a priori* tissue probability maps generated from a population of healthy elderly subjects [17]. After combined segmentation/registration, the grey matter volume of the hippocampus is obtained by applying a binary hippocampus mask predefined

in template space, i.e., by summing up all voxel intensities of the registered and modulated grey matter component image within the mask. The latter is clearly less sophisticated than the elaborated delineation methods implemented in the software tools evaluated by the EMA. However, the results of the present study demonstrate that the SPM8 processing pipeline provides similar performance for prediction of MCI-to-AD conversion as these methods. The pipeline avoids computationally expensive steps and, therefore, allows hippocampal volumetry to be performed within a few minutes. In addition, the pipeline is entirely built upon freely available software. Furthermore, the pipeline is easily adapted, for example by changing the template and the tissue probability maps (just a few files to be replaced) to optimize the processing for a specific patient group. It is also easily extended to estimate the volume of additional brain structures, simply by adding corresponding masks.

Clerx and coworkers, using Learning Embeddings for Atlas Propagation in MCI subjects from another multi-center trial than ADNI, found an AUC of 0.71 for prediction of conversion within 2 years [33]. The AUC for 2 years prediction provided by the novel SPM8-based processing pipeline was 0.72 in the present study (Table 2), and thus in very good agreement.

A similar, SPM5-based processing pipeline has previously been evaluated by Risacher and colleagues in ADNI-MCI subjects [14]. Using Cohen's effect size of the HGMV difference between MCI converters and MCI stables as performance measure, these authors found an effect size of about 0.4 (derived from Fig. 7 in [14]) for one-year prediction of MCI-to-AD conversion. For comparison, Cohen's effect size of the SPM8-based HGMV in the present study is 0.72 for one-year prediction, even larger than the effect size Risacher et al. reported for FreeSurfer-HGMV [14]. Improved performance of the novel SPM8-pipeline compared to the previously described SPM5-pipeline is mainly due to the use of the more appropriate *a priori* TPMs provided by Lemaitre and co-workers [17], which (i) have been generated from elderly, i.e., more age-matched subjects and (ii) feature $1 \times 1 \times 1$ rather than $2 \times 2 \times 2$ mm^3 voxel resolution. This was confirmed by the following experiments. For the first experiment we down-sampled brain template and tissue probability maps from the elderly subjects to $2 \times 2 \times 2$ mm^3 voxel size. Apart from this, preprocessing of individual MRIs (both MCI subjects and ADNI-normals), hippocampal volumetry, correction for head size and age as well as ROC analysis were repeated as described in the methods section. The AUC showed a slight degradation (0.76, 0.70, and 0.69 for conversion within 12, 24, or 36 months, respectively). In the second experiment we used the default SPM template and the default SPM TPMs. This resulted in substantial degradation of the AUCs (0.69, 0.67, and 0.67 for conversion within 12, 24, or 36 months, respectively). These results indicate that improved performance of the novel SPM8 processing pipeline compared to the SPM5 pipeline evaluated by Risacher and co-workers [14] is primarily due to the difference in the subject sample from which template and tissue probability maps have been generated: elderly controls appears more appropriate than young controls, most likely because they are better age-matched. The unified segmentation engine is the same in SPM5 and SPM8 [11].

In a previous study [34] we tested scaling of HGMV to the individual total GMV, i.e., the HGMV/GMV ratio, to reduce variability of HGMV associated with inter-subject variability of the head size. In this previous study the simple scaling approach had resulted in a slightly better diagnostic accuracy for the differentiation between an AD- and a non-AD-group than adjustment for age and TIV based on a bilinear fit in a control group (AUC = 0.88 versus 0.86; see Table 2 in [34]). However, scaling to total GMV is limited in case of spatially more extended atrophy with substantial loss of GM outside of the hippocampus, since then scaling to total GMV counters the effect of AD on hippocampal volume. Thus, correction for TIV and age based on a bilinear fit in healthy controls might be more robust and, therefore, might be preferred in everyday patient care. This was the rationale for using this method in the present study. Nevertheless, we also performed scaling of the HGMV to the individual total GMV and found AUC and accuracy measures of the resulting HGMV/GMV ratio for the prognosis of MCI-to-AD conversion to be lower than the corresponding values for the HGMV adjusted for age and TIV (results not shown). This might be explained by the fact that the MRI data of the multi-centric ADNI have been acquired with a variety of different MR scanner types, in contrast to the monocentric MRI data analyzed in [34]. Multi-centric MRI acquisition causes additional variability in all volumetric measures including both GMV and TIV. However, Huppertz and coworkers, using an SPM5 processing pipeline, found inter-scanner variability (as measured by the coefficient of variation across multiple scanners) to be about 20% larger for GMV than for TIV [35]. The results of the present study now suggest that GMV and HGMV do not vary by the same amount in the same direction

between different scanners so that their inter-scanner variabilities do not cancel in the HGMV/GMV ratio. Thus, TIV might be more appropriate than GMV to correct hippocampus volume for head size in multicenter studies.

When comparing grey, white, and parenchymal hippocampus volume for prediction of MCI-to-AD conversion we found both hippocampus white matter volume (HWMVad) and hippocampus parenchymal volume (HPVad) to provide smaller AUC than hippocampus grey matter volume (HGMVad). This suggests that hippocampal volumetry in MCI subjects should be restricted to grey matter. However, the fact that the SPM8 pipeline easily allows determination of both grey and white matter volume of the hippocampus separately might be useful in other indications, for example in depression [36].

Observer input required to supervise the processing quality of the fully-automated SPM8 processing pipeline is restricted to visual inspection of coronal slices of the patient's grey matter probability map with the hippocampus masks overlaid as contours in order to detect failures of stereotactical normalization and/or grey matter segmentation. This display is automatically generated by the pipeline. However, there was no failure in any of the subjects (198 ADNI MCI subjects and 137 ADNI normals), although no subject of the ADNI cohort was excluded based on technical constraints such as poor MR image quality. This demonstrates the robustness of the method, which is an important prerequisite for use in everyday clinical routine.

As a limitation of the present study, it should be noted that hippocampus atrophy is by no means specific for AD, since many other diseases present with hippocampal atrophy, particularly when associated with episodic memory problems. For example, Möller and co-workers recently reported even more pronounced atrophy of the right hippocampus in patients with behavioral variant frontotemporal dementia (bvFTD) compared to patients with Alzheimer's disease matched for age, gender, and education, although MMSE was significantly higher in the bvFTD than in the AD group (mean MMSE = 24 versus 21) [37]. Abdulla and co-workers recently reported hippocampal volume loss and correlation with verbal memory performance in patients with amyotrophic lateral sclerosis [38]. Particularly relevant in this context might be 'hippocampal sclerosis of aging' which is increasingly recognized in recent years and probably indicates a separate disease process. It is a key AD mimic, i.e., the majority of patients with relatively pure 'hippocampal sclerosis of aging' is diagnosed clinically as having probable AD [39]. 'Hippocampal sclerosis of aging' is characterized by cell loss, gliosis, and atrophy in the hippocampus that is out of proportion to AD-specific neuropathology [39].

It should also be noted that the positive predictive values of the HGMV found in the present study were quite low: 0.39, 0.54, and 0.60 for MCI-to-AD conversion within 12, 24, and 36 months, respectively (Table 2). These values are clearly too low to encourage the use of hippocampus volumetry for prediction of MCI-to-AD conversion in individual patients in clinical patient care. In contrast, the negative predictive values were considerably higher: 0.89, 0.73, and 0.69 for conversion within 12, 24, and 36 months, respectively (Table 2). Thus, the primary benefit from hippocampal volumetry in clinical routine is expected to be the 'exclusion' of MCI-to-AD conversion in a significant fraction of patients, rather than its prediction. Considering the enormous mental stress caused by the suspicion of Alzheimer's disease to patients and their families, this is of high relevance.

Finally, it should be noted that the novel SPM8 processing pipeline overestimates the HGMV. Bilateral HGMV in the ADNI-normals was 8.69 ml (Table 1), which is considerably larger than actual HGMV in healthy elderly subjects [36, 40]. The reason for this overestimation is the following. There is some residual anatomical inter-subject variability after unified segmentation, more pronounced in case of strongly atrophic brains. This residual anatomical inter-subject variability cannot be taken into account when using a fixed predefined hippocampus mask for volumetry, but results in additional inter-subject variability of the HGMV estimates. The relative magnitude of the additional variability increases with decreasing size of the hippocampus mask. We therefore selected the rather large hippocampus mask provided by Eickhoff and coworkers [19]. This mask has been obtained by binarization of a probabilistic hippocampus template and, therefore, includes all voxels with non-vanishing probability of belonging to the hippocampus. This guarantees that the hippocampus is completely included within the mask in each individual subject despite the residual anatomical inter-subject variability after unified segmentation. As a consequence, the mask also includes neighboring non-hippocampal voxels of the medial temporal lobe and, therefore, results in overestimation of the hippocampus volume. However, despite this overestimation, the SPM8-HGMV values showed moderate to strong correlation with the results of the semi-automated HV-SNT

(Pearson correlation coefficient = 0.72), in good agreement with previous reports on the correlation between fully-automated versus manual segmentation of the hippocampus in ADNI subjects [41]. Furthermore, the AUC of the SPM8-HGMV for MCI-to-AD conversion within 2 years of 0.72 is in good agreement with the AUC values reported from the EMA qualification process which range between 0.60 and 0.77 (including results from both de novo analysis and literature review) [10]. This suggests that the method for the estimation of the hippocampus volume does not have a large impact on its prognostic value in MCI subjects. It also suggests that most accurate anatomical delineation of the hippocampus volume is not critical for its performance in this application scenario. This might be related to intrinsic limitations of the hippocampus volume as prognostic marker in MCI subjects, i.e., the existence of an upper threshold for its prognostic accuracy considerably below 100%, which also the most accurate volumetry cannot surpass. Improved prognostic value might be achieved by volumetry of hippocampal subfields [42]. However, whether the novel SPM8 processing pipeline might be extended to estimate grey matter volume in hippocampal subfields remains to be tested.

CONCLUSION

The SPM8-based processing pipeline for fully-automated hippocampal volumetry proposed in the present study provides similar performance for prediction of MCI-to-AD conversion as more complex tools. The processing pipeline builds entirely upon freely available software and operates in near real time (a few minutes per subject). This may facilitate integration of hippocampal volumetry into routine clinical diagnostic care of subjects presenting with cognitive impairment.

ACKNOWLEDGMENTS

P.S., L.S., J.B.F., and R.B. were supported by the European Regional Development Fund of the European Union (reference 10153407, 10153462, 10153463). P.S. and L.S. are employees of jung diagnostics GmbH.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare;; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/14-2280r1).

REFERENCES

[1] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (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. *Alzheimers Dement* **7**, 270-279.

[2] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, DeKosky ST, Gauthier S, Selkoe D, Bateman R, Cappa S, Crutch S, Engelborghs S, Frisoni GB, Fox NC, Galasko D, Habert MO, Jicha GA, Nordberg A, Pasquier F, Rabinovici G, Robert P, Rowe C, Salloway S, Sarazin M, Epelbaum S, de Souza LC, Vellas B, Visser PJ, Schneider L, Stern Y, Scheltens P, Cummings JL (2014) Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurol* **13**, 614-629.

[3] Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.

[4] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, Delacourte A, Frisoni G, Fox NC,

Galasko D, Gauthier S, Hampel H, Jicha GA, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Sarazin M, de Souza LC, Stern Y, Visser PJ, Scheltens P (2010) Revising the definition of Alzheimer's disease: A new lexicon. *Lancet Neurol* **9**, 1118-1127.

[5] Hort J, O'Brien JT, Gainotti G, Pirttila T, Popescu BO, Rektorova I, Sorbi S, Scheltens P (2010) EFNS guidelines for the diagnosis and management of Alzheimer's disease. *Eur J Neurol* **17**, 1236-1248.

[6] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341-355.

[7] Holland D, Brewer JB, Hagler DJ, Fennema-Notestine C, Dale AM (2009) Subregional neuroanatomical change as a biomarker for Alzheimer's disease. *Proc Natl Acad Sci U S A* **106**, 20954-20959.

[8] Leung KK, Barnes J, Ridgway GR, Bartlett JW, Clarkson MJ, Macdonald K, Schuff N, Fox NC, Ourselin S (2010) Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *Neuroimage* **51**, 1345-1359.

[9] Wolz R, Aljabar P, Hajnal JV, Hammers A, Rueckert D (2010) LEAP: Learning embeddings for atlas propagation. *Neuroimage* **49**, 1316-1325.

[10] Hill DL, Schwarz AJ, Isaac M, Pani L, Vamvakas S, Hemmings R, Carrillo MC, Yu P, Sun J, Beckett L, Boccardi M, Brewer J, Brumfield M, Cantillon M, Cole PE, Fox N, Frisoni GB, Jack C, Kelleher T, Luo F, Novak G, Maguire P, Meibach R, Patterson P, Bain L, Sampaio C, Raunig D, Soares H, Suhy J, Wang H, Wolz R, Stephenson D (2014) Coalition Against Major Diseases/European Medicines Agency biomarker qualification of hippocampal volume for enrichment of clinical trials in predementia stages of Alzheimer's disease. *Alzheimers Dement* **10**, 421-429 e423.

[11] Ashburner J (2012) SPM: A history. *Neuroimage* **62**, 791-800.

[12] Jack CR Jr, Barkhof F, Bernstein MA, Cantillon M, Cole PE, Decarli C, Dubois B, Duchesne S, Fox NC, Frisoni GB, Hampel H, Hill DL, Johnson K, Mangin JF, Scheltens P, Schwarz AJ, Sperling R, Suhy J, Thompson PM, Weiner M, Foster NL (2011) Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. *Alzheimers Dement* **7**, 474-485 e474.

[13] Carmichael OT, Aizenstein HA, Davis SW, Becker JT, Thompson PM, Meltzer CC, Liu Y (2005) Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. *Neuroimage* **27**, 979-990.

[14] Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, McDonald BC (2009) Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Curr Alzheimer Res* **6**, 347-361.

[15] Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, J, Ward LW, Dale C, Felmlee AM, Gunter JP, Hill JL, Killiany DL, Schuff R, Fox-Bosetti N, Lin S, Studholme C, DeCarli C, Krueger CS, Ward G, Metzger HA, Scott GJ, Mallozzi KT, Blezek R, Levy D, Debbins J, Fleisher JP, Albert AS, Green M, Bartzokis R, Glover G, Mugler G, Weiner J, MW (2008) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* **27**, 685-691.

[16] Ashburner J, Friston KJ (2005) Unified segmentation. *Neuroimage* **26**, 839-851.

[17] Lemaitre H, Crivello F, Grassiot B, Alperovitch A, Tzourio C, Mazoyer B (2005) Age- and sex-related effects on the neuroanatomy of healthy elderly. *Neuroimage* **26**, 900-911.

[18] Arlt S, Buchert R, Spies L, Eichenlaub M, Lehmbeck JT, Jahn H (2013) Association between fully automated MRI-based volumetry of different brain regions and neuropsychological test performance in patients with amnestic mild cognitive impairment and Alzheimer's disease. *Eur Arch Psychiatry Clin Neurosci* **263**, 335-344.

[19] Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* **25**, 1325-1335.

[20] Amunts K, Kedo O, Kindler M, Pieperhoff P, Mohlberg H, Shah NJ, Habel U, Schneider F, Zilles K (2005) Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. *Anat Embryol (Berl)* **210**, 343-352.

[21] Geuze E, Vermetten E, Bremner JD (2005) MR-based *in vivo* hippocampal volumetrics: 1. Review of methodologies currently employed. *Mol Psychiatry* **10**, 147-159.

[22] Hsu YY, Schuff N, Du AT, Mark K, Zhu X, Hardin D, Weiner MW (2002) Comparison of automated and manual MRI volumetry of hippocampus in normal aging and dementia. *J Magn Reson Imaging* **16**, 305-310.

[23] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M (2011) pROC: An open-source package for R and S+to analyze and compare ROC curves. *BMC Bioinformatic* **12**, 77.

[24] Youden WJ (1950) Index for rating diagnostic tests. *Cancer* **3**, 32-35.

[25] Kohavi R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Quebec, Canada, Vol. 2, pp. 1137-1143.

[26] Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29-36.

[27] Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179-194.

[28] Brewer JB (2009) Fully-automated volumetric MRI with normative ranges: Translation to clinical practice. *Behav Neurol* **21**, 21-28.

[29] Brewer JB, Magda S, Airriess C, Smith ME (2009) Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in Alzheimer disease. *AJNR Am J Neuroradiol* **30**, 578-580.

[30] Barnes J, Foster J, Boyes RG, Pepple T, Moore EK, Schott JM, Frost C, Scahill RI, Fox NC (2008) A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *Neuroimage* **40**, 1655-1671.

[31] Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ (1999) Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans Med Imaging* **18**, 712-721.

[32] Warfield SK, Zou KH, Wells WM (2004) Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* **23**, 903-921.

[33] Clerx L, van Rossum IA, Burns L, Knol DL, Scheltens P, Verhey F, Aalten P, Lapuerta P, van de Pol L, van Schijndel R, de Jong R, Barkhof F, Wolz R, Rueckert D, Bocchetta M, Tsolaki M, Nobili F, Wahlund LO, Minthon L, Frolich L, Hampel H, Soininen H, Visser PJ (2013) Measurements of

medial temporal lobe atrophy for prediction of Alzheimer's disease in subjects with mild cognitive impairment. *Neurobiol Aging* **34**, 2003-2013.

[34] Suppa P, Anker U, Spies L, Bopp I, Ruegger-Frey B, Klaghofer R, Gocke C, Hampel H, Beck S, Buchert R (2015) Fully automated atlas-based hippocampal volumetry for detection of Alzheimer's disease in a memory clinic setting. *J Alzheimers Dis* **44**, 183-193.

[35] Huppertz HJ, Kroll-Seger J, Kloppel S, Ganz RE, Kassubek J (2010) Intra- and interscanner variability of automated voxel-based volumetry based on a 3D probabilistic atlas of human cerebral structures. *Neuroimage* **49**, 2216-2224.

[36] Redlich R, Almeida JJ, Grotegerd D, Opel N, Kugel H, Heindel W, Arolt V, Phillips ML, Dannlowski U (2014) Brain morphometric biomarkers distinguishing unipolar and bipolar depression. A voxel-based morphometry-pattern classification approach. *JAMA Psychiatry* **71**, 1222-1230.

[37] Moller C, Dieleman N, van der Flier WM, Versteeg A, Pijnenburg Y, Scheltens P, Barkhof F, Vrenken H (2015) More atrophy of deep gray matter structures in frontotemporal dementia compared to Alzheimer's disease. *J Alzheimers Dis* **44**, 635-647.

[38] Abdulla S, Machts J, Kaufmann J, Patrick K, Kollewe K, Dengler R, Heinze HJ, Petri S, Vielhaber S, Nestor PJ (2014) Hippocampal degeneration in patients with amyotrophic lateral sclerosis. *Neurobiol Aging* **35**, 2639-2645.

[39] Brenowitz WD, Monsell SE, Schmitt FA, Kukull WA, Nelson PT (2014) Hippocampal sclerosis of aging is a key Alzheimer's disease mimic: Clinical-pathologic correlations and comparisons with both alzheimer's disease and non-tauopathic frontotemporal lobar degeneration. *J Alzheimers Dis* **39**, 691-702.

[40] de Flores R, La Joie R, Landeau B, Perrotin A, Mezenge F, de La Sayette V, Eustache F, Desgranges B, Chetelat G (2015) Effects of age and Alzheimer's disease on hippocampal subfields: Comparison between manual and freesurfer volumetry. *Hum Brain Mapp* **36**, 463-474.

[41] Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR Jr, Weiner MW, Thompson PM (2008) Validation of a fully automated 3D hippocampal segmentation method using subjects with Alzheimer's disease mild cognitive impairment, and elderly controls. *Neuroimage* **43**, 59-68.

[42] Apostolova LG, Mosconi L, Thompson PM, Green AE, Hwang KS, Ramirez A, Mistur R, Tsui WH, de Leon MJ (2010) Subregional hippocampal atrophy predicts Alzheimer's dementia in the cognitively normal. *Neurobiol Aging* **31**, 1077-1088.

[43] DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837-845.