Fully Automated Atlas-Based Hippocampal Volumetry for Detection of Alzheimer's Disease in a Memory Clinic Setting

Per Suppa[a,c], Ulrich Anker[b], Lothar Spies[c], Irene Bopp[b], Brigitte Rüegger-Frey[b], Richard Klaghofer[b], Carola Gocke[d], Harald Hampel[e], Sacha Beck[b,1] and Ralph Buchert[a,1,*]

[a] Department of Nuclear Medicine, Charité, Berlin, Germany
[b] Stadtspital Waid, Zurich, Switzerland
[c] jung diagnostics GmbH, Hamburg, Germany
[d] Medical Prevention Center Hamburg, Hamburg, Germany
[e] Université Pierre et Marie Curie, Paris, France

Handling Associate Editor: Babak Ardekani

Abstract. Hippocampal volume is a promising biomarker to enhance the accuracy of the diagnosis of dementia due to Alzheimer's disease (AD). However, whereas hippocampal volume is well studied in patient samples from clinical trials, its value in clinical routine patient care is still rather unclear. The aim of the present study, therefore, was to evaluate fully automated atlas-based hippocampal volumetry for detection of AD in the setting of a secondary care expert memory clinic for outpatients. One-hundred consecutive patients with memory complaints were clinically evaluated and categorized into three diagnostic groups: AD, intermediate AD, and non-AD. A software tool based on open source software (Statistical Parametric Mapping SPM8) was employed for fully automated tissue segmentation and stereotactical normalization of high-resolution three-dimensional T1-weighted magnetic resonance images. Predefined standard masks were used for computation of grey matter volume of the left and right hippocampus which then was scaled to the patient's total grey matter volume. The right hippocampal volume provided an area under the receiver operating characteristic curve of 84% for detection of AD patients in the whole sample. This indicates that fully automated MR-based hippocampal volumetry fulfills the requirements for a relevant core feasible biomarker for detection of AD in everyday patient care in a secondary care memory clinic for outpatients. The software used in the present study has been made freely available as an SPM8 toolbox. It is robust and fast so that it is easily integrated into routine workflow.

Keywords: Alzheimer's disease, atlas-based segmentation, hippocampal volumetry, magnetic resonance imaging, memory clinic, memory impairment

INTRODUCTION

In the beginning of the 1990s, magnetic resonance (MR)-based hippocampal volumetry was recognized to be useful to support the diagnosis of Alzheimer's disease (AD) [1]. This was triggered by fundamental work of Braak and Braak, who demonstrated that neurofibrillary pathology characteristic for AD usually

[1] Equally contributing as senior authors.

*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.

begins in the medial temporal lobe, particularly in the entorhinal cortex and in the hippocampus [2]. This results in a similar pattern of grey matter (GM) atrophy which can be detected by structural MR imaging. Since then, numerous studies have been published and reviewed [3, 4] to further explore usefulness and implications of MR-based hippocampal volumetry [5–9].

Recently, the interest in this field has been further strengthened by the National Institute on Aging-Alzheimer's Association (NIA-AA) guidelines which suggest hippocampal volumetry as biomarker both in the diagnostic process for early AD [11] and in the evaluation of dementia [12] (see also [10]). To date, assessment of hippocampal atrophy is based on visual inspection and visual rating scales in most institutions [13, 14]. Inter-rater variability, a main limitation of visual rating scales in clinical patient care, might be eliminated by quantitative characterization of hippocampal atrophy. The gold standard for volumetric assessment of the hippocampus is its manual segmentation. However, manual segmentation is also prone to significant variations between centers and operators performing the task [8]. In addition, manual segmentation is very time consuming and, therefore, it is not compatible with the workflow in clinical routine patient care. Fully automated tools eliminate inter-operator variability [15–17]. Hippocampal volumes estimated by fully-automated atlas-based segmentation approaches with rather short computation time are in good quantitative agreement with manual segmentation [18, 19]. Therefore, fully automated methods have the potential to support the diagnosis also in the clinical setting [20].

However, most studies of hippocampal volumetry for the detection of AD have been performed in highly screened populations. For widespread clinical acceptance it is mandatory to validate hippocampal volumetry in populations representing everyday clinical routine in which the presence of other pathologies and comorbidities is the norm rather than the exception.

In light of this, the aim of the present study was to investigate the accuracy of fully automated atlas-based hippocampal volumetry to detect AD in a heterogeneous population of patients of a memory clinic.

PATIENTS AND METHODS

Patients

The memory clinic of the Stadtspital Waid in Zurich is one of the largest specialty clinics in Switzerland with approximately 300 new patients per year

referred from primary care providers for evaluation of subjective or objective memory complaints or other neurocognitive disorders. One-hundred consecutive patients who had presented with memory complaints between 1 January 2010 and 31 December 2010 were included retrospectively in the present study. Patients were included, when (i) a clinical diagnosis was obtained according to the standard diagnostic procedure of the Stadtspital Waid and was clearly stated in the report and (ii) high-resolution MR imaging had been performed. No further selection criteria were applied. In particular, there was no exclusion criterion with respect to the MR image quality. The standard diagnostic procedure of the Stadtspital Waid comprises anamnesis and a caregiver report, clinical examination, routine blood testing, and a battery of standardized and established neuropsychological tests. Diagnoses are made in consensus by an interdisciplinary board using established clinical criteria to identify AD both in subjects with mild cognitive impairment (MCI) [11, 21–24] and in patients with dementia [12].

Patients were categorized into three subgroups based on clinical judgment and in line with core clinical criteria: (i) AD, (ii) intermediate AD, and (iii) non-AD. The "AD" subgroup comprised patients with probable AD according to McKhann et al. [12] and patients with MCI consistent with AD according to Albert et al. [11]. The "non-AD" group included patients with subjective cognitive impairment (SCI), patients with MCI inconsistent with AD [11, 23, 24], and patients with dementia due to suspected neurodegenerative disease other than AD, such as frontotemporal dementia, or other neurological or psychiatric disorders, such as Parkinson's disease. All remaining patients were categorized into the "intermediate AD" group. These were patients with dementia or MCI who did not show sufficient clinical evidence of AD (to be categorized into the "AD" group) nor of any other specific disease (to be categorized into the "non-AD" group). The "intermediate AD" category reflects the difficulties of clinical diagnostics in everyday routine in a secondary care expert memory clinic for outpatients. The categorization as "intermediate AD" should not be confused with the diagnosis of possible AD according to [12]; it includes patients with possible AD but also more unclear cases. MR images were not used for the diagnosis except for the exclusion of non-neurodegenerative causes of the cognitive complaints such as stroke or brain tumor. Details of the subgroups are presented in Table 1.

The study protocol fulfilled the requirements of the local ethical committee of Zurich and was approved by the Institutional Review Board (IRB) of the Stadtspital

Table 1
Clinical characterization and results of fully automated MR-based volumetry. Patients with objective memory deficits include patients with either mild cognitive impairment or dementia. Ranges are given in square brackets, standard deviations are given in round brackets. MMSE, Mini-Mental State Examination

	non-AD	Intermediate AD	AD
number of patients	35	21	44
objective memory deficits	22	21	44
subjective cognitive impairment	13	0	0
Age: mean (range) [years]	67 [42–85]	78 [60–92][++++]	79 [64–91][++++]
MMSE: mean (range)	27 [6–30]	25 [16–30]	21 [13–29]
Clock drawing test mean (range)	6 [0–7]	5 [2–7]	4 [0–7]
GMV [ml]	613 (95)	565 (65)[+]	538 (59)[+++]
TIV [ml]	1501 (148)	1493 (166)	1415 (134)[+]
HVL [ml]	3.05 (0.60)	2.57 (0.47)[++]	2.18 (0.42)[+++]
HVR [ml]	3.09 (0.59)	2.55 (0.47)[++]	2.09 (0.41)[+++]
HV [ml]	6.14 (1.17)	5.12 (0.88)[++]	4.27 (0.75)[+++]
HVL_{ad} [ml]	3.05 (0.37)	2.95 (0.44)	2.75 (0.50)[++]
HVR_{ad} [ml]	3.09 (0.37)	2.89 (0.50)	2.62 (0.49)[+++]
HV_{ad} [ml]	6.14 (0.71)	5.84 (0.89)	5.37 (0.93)[+++]
HVL/GMV [per mille]	4.97 (0.54)	4.58 (0.81)	4.08 (0.78)[+++]
HVR/GMV [per mille]	5.02 (0.40)	4.56 (0.87)[+]	3.90 (0.72)[+++]
HV/GMV [per mille]	9.99 (0.86)	9.14 (1.61)[+]	7.98 (1.38)[+++]

[+]$p < 0.05$, [++]$p < 0.005$, [+++]$p < 0.0005$, [++++]$p < 0.00005$ versus non-AD.

Waid. All procedures were done in accord with the Helsinki Declaration of 1975.

Image acquisition

MR imaging had been performed at the Stadtspital Waid with a Siemens Avanto 1.5 T (Siemens Erlangen, Germany) deploying 3D T1-weighted magnetization prepared rapid gradient echo (MPRAGE) with two slightly different acquisition protocols. 89 patients were scanned using protocol A: TR 1900 ms, TE 3.1 ms, TI 1100 ms and a flip angle of 15°. Eleven patients (8 ADs, 2 non-ADs, 1 intermediate AD) were scanned using a second protocol (B): TR 980 ms, TE 2.95 ms, TI 600 ms and a flip angle of 15°. An isotropic voxel grid of 1 mm and 176 sagittal slices were used throughout. All scans were performed without contrast agent. Acquisition time was less than 5 min.

Image segmentation

MR images were segmented and stereotactically normalized to the Montreal Neurological Institute (MNI) space using a combined segmentation and registration approach [25] as implemented in the Statistical Parametric Mapping 8 (SPM8) software package (Wellcome Trust Centre for Neuroimaging, London, UK). Preexisting, freely available prior tissue probability maps for GM, white matter (WM), and cerebrospinal fluid (CSF) were used to assist segmentation and registration [26]. The default setting of the unified segmentation engine was used as described in Arlt et al. [27]. The unified segmentation approach yields stereotactically normalized component images of GM, WM, and CSF with a voxel volume of 1 mm. Modulation was applied to preserve the volumes of the component images after stereotactical normalization.

Computation time for segmentation of a single data set was less than four minutes on an Intel Core 2 Duo CPU with 3.33 GHz and 8 GB RAM.

Volumetry

Hippocampal GM volume (HV) was calculated by multiplying the subject's GM component image with a predefined binary mask from a freely available atlas [28] and then summing over all voxel intensities. Hippocampus masks for the left and the right hemisphere were used separately yielding two sub-volumes for each brain hemisphere, HVL and HVR, respectively. The masks comprise the Cornus ammonis (CA1-CA4, in the following summarized as CA) and Fascia dentata (FD) substructures as defined by Amunts and coworkers [29] and feature an isotropic resolution of 1 mm (Fig. 1). Volumes of the binary hippocampus masks are 6.7 ml and 6.9 ml for the left and right hemisphere, respectively. Total HV was obtained by summing the GM volume within both masks.

Total grey matter volume (GMV), total white matter volume (WMV), and total cerebrospinal fluid volume (CSFV) were calculated by summing up all voxel
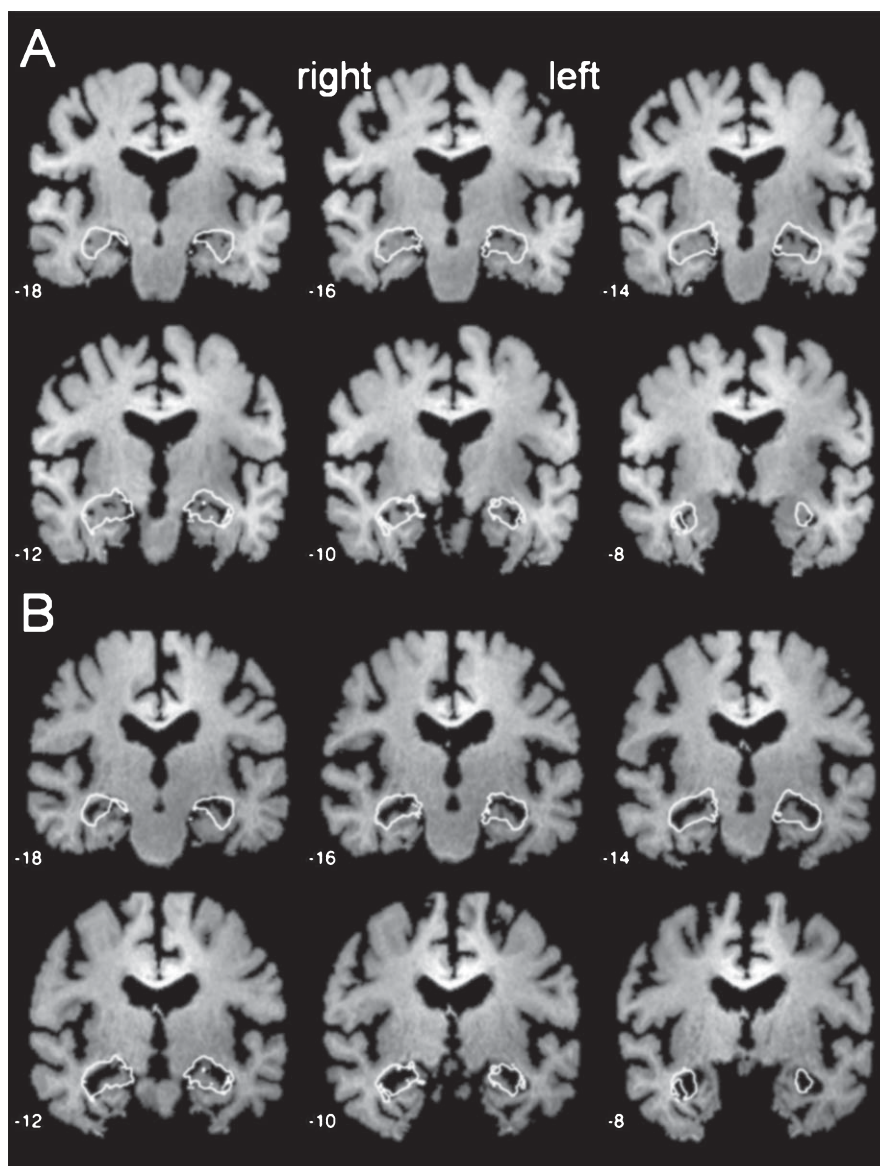
Fig. 1. Coronal views of brain parenchyma normalized to MNI space of two patients in radiological convention (left is right). White contours delineate the region of the hippocampus mask for left and right hemisphere. Numbers are the coordinates of the slices in MNI space. A) 74 years old male non-AD patient with MCI. B) 74 years old male with AD.

intensities of the stereotactically normalized and modulated component images of the corresponding tissue class. Total intracranial volume (TIV) was computed as GMV+WMV+CSFV.

Correction for TIV and age by bilinear fitting

TIV and age are widely used as covariates to minimize additional variance due to inter-subject differences in head-size and GM loss in the hippocampus associated with normal aging [8].

To account for these confounders in the present study, HVL, HVR, and HV were fitted by a bilinear model with age and TIV as independent variables to an independent sample of healthy subjects from a pool of cognitively normal subjects (control group) undergoing whole body MR imaging as part of a check-up program at a medical prevention center in Hamburg, Germany. Subjects were excluded if they had a history of or current neurological or psychiatric disease or if there were abnormal findings in the brain MR image according to visual inspection by an experi-

enced radiologist (C.G.). A total of 218 subjects were included. 3D MPRAGE images had been acquired with a Siemens Avanto 1.5 T using protocol B as specified above. The sample covered a wide range of age from 18 to 85 years (mean age 62 years). HVL, HVR, and HV of an individual patient were then adjusted to the mean age and mean TIV of the control group using the following formulas:

$$HVL_{ad} = HVL + a_{HVL} \cdot (\langle TIV \rangle - TIV)$$
$$+ b_{HVL} \cdot (\langle age \rangle - age)$$
$$HVR_{ad} = HVR + a_{HVR} \cdot (\langle TIV \rangle - TIV)$$
$$+ b_{HVR} \cdot (\langle age \rangle - age)$$
$$HV_{ad} = HV + a_{HV} \cdot (\langle TIV \rangle - TIV)$$
$$+ b_{HV} \cdot (\langle age \rangle - age)$$

TIV and *age* represent TIV and age of the patient. *a*'s and *b*'s are the regression coefficients ($a_{HVL} = 0.0018$ and $b_{HVL} = -0.0107$ ml/year; $a_{HVR} = 0.0017$ and $b_{HVR} = -0.0102$ ml/year; $a_{HV} = 0.0035$ and $b_{HV} = -0.0209$ ml/year) from the bilinear fit to the control group, and ‹age› and ‹TIV› denote mean age and mean TIV in the control group (62 years and 1,464 ml).

Scaling to individual GMV

Total GMV might be used as substitute for both TIV and age as covariate. The rationale for this is that in healthy subjects (i) GMV is expected to scale directly with TIV, and (ii) loss of GM caused by healthy aging can be considered an indirect measure of the brain's age. Correction for GMV was performed by direct scaling, i.e., HVL, HVR, HV of a patient was divided by the patient's GMV yielding the ratios HVL/GMV, HVR/GMV, and HV/GMV. Ratios were specified in per mille. This scaling approach was tested in the control group.

Performance metric and discrimination of AD and non-AD

The different biomarkers were first compared with respect to their potential for the differentiation between AD and non-AD (setup 1). Corresponding receiver operating characteristic (ROC) curves were generated for HVL_{ad}, HVR_{ad}, and HV_{ad} as well as for the ratios HVL/GMV, HVR/GMV, and HV/GMV. The area under the ROC curve (AUC) was calculated as primary performance measure. AUC calculation was performed according to the trapezoidal rule [41]. The 95% confidence intervals for AUCs were determined using the method described by Delong and coworkers [30]. The cut-off point for optimal discrimination is represented by the point on the ROC curve closest to the upper left corner.

Sensitivity, specificity and accuracy are proportions, thus confidence intervals were calculated using standard methods for proportions. The 95% confidence interval was approximated based on the Gaussian law by $p \pm 1.96 \cdot \sqrt{p \cdot \frac{1-p}{N}}$, where p is the value of the proportion and N is the sample size ($N = 79$). Values larger than one were truncated.

Detection of AD or non-AD

All biomarkers were further evaluated for the detection of AD in the whole patient sample (i.e., AD versus intermediate AD and non-AD, setup 2; $N = 100$) and for the detection of non-AD in the whole patient sample (i.e., non-AD versus intermediate and AD, setup 3; $N = 100$).

Statistical analyses

The mean age was compared between the three patient groups using univariate analysis of variance. *Post-hoc* testing was performed by Scheffé's or Tamhane's method depending on the result of Levene's test for equality of variances (which was accepted for p values greater than 0.05). The mean TIV was compared between the three groups using the general linear model with group as fixed factor, gender as random factor, and age as covariate. Further comparisons between two groups were performed using the homoscedastic or heteroscedastic unpaired two-sample *t*-test based on the result of Levene's test (Table 1). For ROC generation and analysis the open source R package pROC was used [31].

RESULTS

There was a highly significant age difference between the groups ($p < 0.0005$). The patients with clinically AD as well as the patients with clinically intermediate AD were significantly older than the non-ADs (Table 1). However, there was a considerable overlap between the three groups with respect to the age range. There was no difference of TIV between the groups after correction for gender and age ($p = 0.149$).

Segmentation of MR images into GM, WM, and CSF worked properly in all cases (i.e., in the patient

population and in the control group) according to visual inspection. The volumetry results for GM are summarized in Table 1. There was a highly significant reduction of total GMV and hippocampal volumes (both with and without correction for head size and/or age) in patients with AD compared to non-ADs. Patients with intermediate AD presented with intermediate GMVs.

Scaling to individual GMV

There were strong positive correlations of hippocampal volumes with TIV in the sample of healthy subjects who had obtained MR imaging as part of a check-up at a prevention center: $r = 0.63$, $r = 0.60$, and $r = 0.64$ for HVL, HVR, and HV, respectively (all $p < 0.000005$). In addition, there were weak but significant negative correlations of hippocampal volumes with age: $r = -0.36$, $r = -0.34$, and $r = -0.36$, for HVL, HVR, and HV, respectively (all $p < 0.00005$).

Scaling to individual GMV removed both the correlation with TIV ($r = -0.08$, $p = 0.2574$; $r = -0.10$, $p = 0.1235$; and $r = -0.10$, $p = 0.1380$ for HVL/GMV, HVR/GMV, and HV/GMV, respectively) and age ($r = 0.08$, $p = 0.2685$; $r = 0.09$, $p = 0.1792$; and $r = 0.09$, $p = 0.1747$ for HVL/GMV, HVR/GMV, and HV/GMV, respectively).
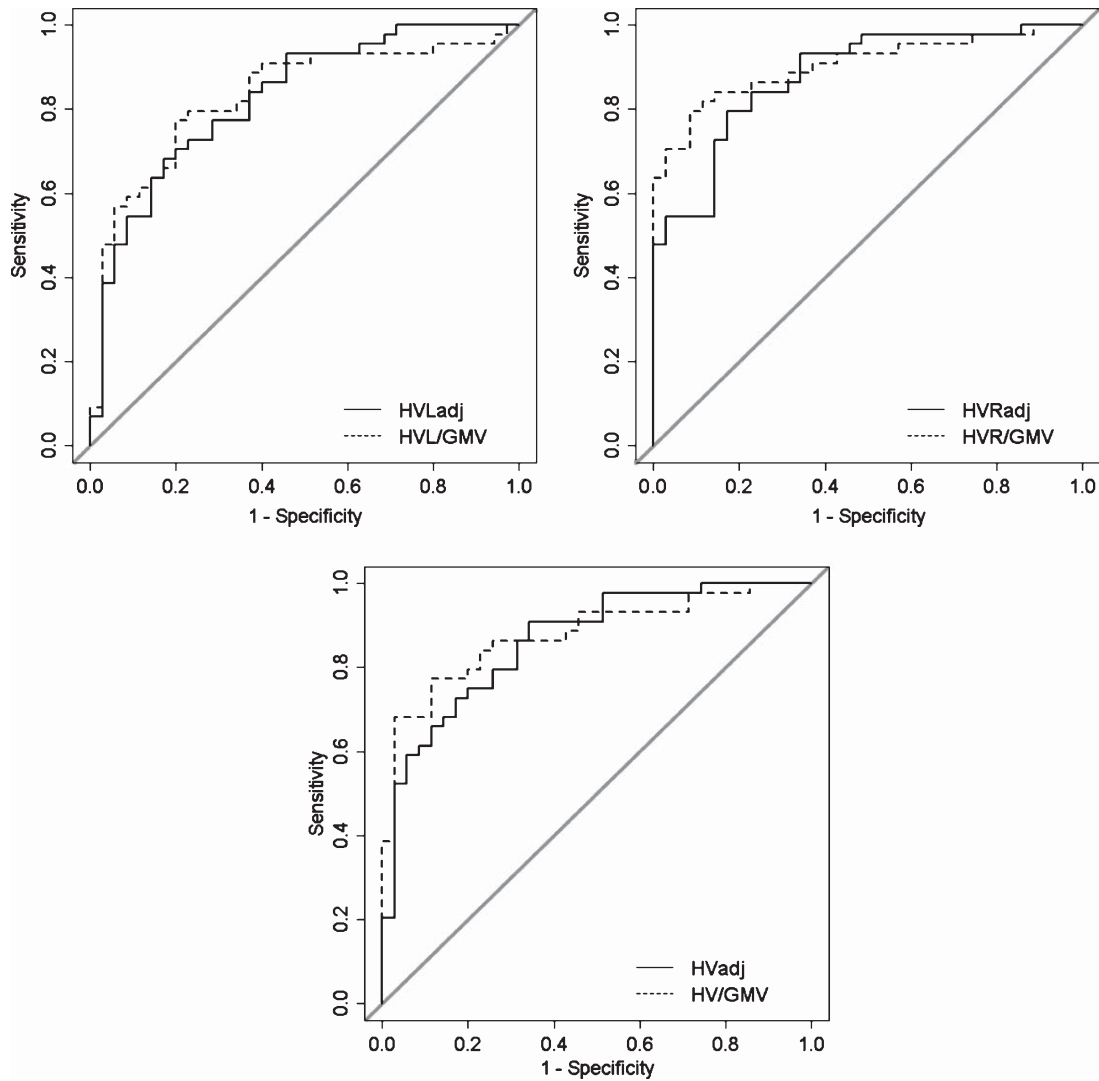

Fig. 2. ROC curves for HVL$_{ad}$ and HVL/GMV (upper left), HVR$_{ad}$ and HVR/GMV (upper right), and HV$_{ad}$ and HV/GMV (bottom) for the discrimination of ADs from non-ADs.

Table 2

Discrimination between AD and non-AD (setup 1). For each biomarker, AUC and the maximum accuracy are given together with sensitivity, specificity and cut-off value at maximum accuracy. 95% confidence intervals are given in brackets. pm is per mille

	AUC	Accuracy	Sensitivity	Specificity	Cut-off
HVL_{ad}	0.82[0.73–0.92]	0.75[0.65–0.85]	0.68[0.58–0.78]	0.83[0.75–0.91]	2.63 ml
HVR_{ad}	0.88[0.80–0.95]	0.81[0.72–0.90]	0.80[0.71–0.89]	0.83[0.75–0.91]	2.77 ml
HV_{ad}	0.86[0.78–0.94]	0.80[0.71–0.89]	0.91[0.85–0.97]	0.66[0.56–0.76]	5.91 ml
HVL/GMV	0.83[0.73–0.92]	0.79[0.70–0.88]	0.77[0.68–0.86]	0.80[0.71–0.89]	4.65 pm
HVR/GMV	0.90[0.84–0.97]	0.85[0.77–0.93]	0.80[0.71–0.89]	0.91[0.85–0.97]	4.54 pm
HV/GMV	0.88[0.80–0.95]	0.83[0.75–0.91]	0.77[0.68–0.86]	0.89[0.82–0.96]	9.15 pm

Table 3

Detection of AD in the whole patient sample (setup 2). pm is per mille

	AUC	Accuracy	Sensitivity	Specificity	Cut-off
HVL_{ad}	0.75[0.66–0.85]	0.71[0.62–0.80]	0.70[0.61–0.79]	0.71[0.62–0.80]	2.69 ml
HVR_{ad}	0.81[0.73–0.90]	0.76[0.68–0.84]	0.75[0.67–0.83]	0.77[0.69–0.85]	2.70 ml
HV_{ad}	0.79[0.70–0.88]	0.75[0.67–0.83]	0.61[0.51–0.71]	0.86[0.79–0.93]	4.95 ml
HVL/GMV	0.77[0.67–0.86]	0.73[0.64–0.82]	0.80[0.72–0.88]	0.66[0.57–0.75]	4.69 pm
HVR/GMV	0.84[0.75–0.92]	0.80[0.72–0.88]	0.80[0.72–0.88]	0.80[0.72–0.88]	4.54 pm
HV/GMV	0.81[0.73–0.90]	0.77[0.69–0.85]	0.66[0.57–0.75]	0.88[0.82–0.94]	8.36 pm

Table 4

Detection of non-AD in the whole patient sample (setup 3). pm is per mille

	AUC	Accuracy	Sensitivity	Specificity	Cut-off
HVL_{ad}	0.79[0.69–0.88]	0.70[0.61–0.79]	0.65[0.56–0.74]	0.80[0.72–0.88]	2.74 ml
HVR_{ad}	0.83[0.75–0.91]	0.75[0.67–0.83]	0.71[0.62–0.80]	0.83[0.76–0.90]	2.78 ml
HV_{ad}	0.82[0.73–0.90]	0.79[0.71–0.87]	0.86[0.79–0.93]	0.66[0.57–0.75]	5.91 ml
HVL/GMV	0.77[0.67–0.86]	0.75[0.67–0.83]	0.69[0.60–0.78]	0.80[0.72–0.88]	4.65 pm
HVR/GMV	0.82[0.74–0.90]	0.80[0.72–0.88]	0.68[0.59–0.77]	0.91[0.85–0.97]	4.55 pm
HV/GMV	0.80[0.71–0.89]	0.79[0.71–0.87]	0.60[0.50–0.70]	0.97[0.94–1]	8.69 pm

Discrimination of AD and non-AD

ROC curves for HVL_{ad}, HVR_{ad}, and HV_{ad} as well as for HVL/GMV, HVR/GMV, and HV/GMV are shown in Fig. 2. Larger AUCs were obtained for hippocampal volumes scaled to GMV than for the hippocampal volumes adjusted for age and TIV based on the bilinear fit in the control group. The largest AUC was found for HVR/GMV, i.e., the scaled GMV of the right hippocampus (AUC = 0.90). Corresponding accuracy was calculated to be 85% with sensitivity and specificity of 80% and 91%, respectively. Results are summarized in Table 2.

Detection of AD or non-AD

For detection of AD in the whole patient sample (setup 2) HVR/GMV provided an AUC of 0.84. Maximum accuracy and both sensitivity and specificity at maximum accuracy were all 80%. For detection of non-AD (setup 3) HVR/GMV provided an AUC of 0.82 with a maximum accuracy of 80% and corresponding

sensitivity and specificity of 68% and 91%, respectively. Results for setup 2 and setup 3 are summarized in Tables 3 and 4.

DISCUSSION

The present study evaluated freely available software for fully automated atlas-based hippocampal volumetry for the detection of AD in the setting of a secondary care memory clinic for outpatients. The method is based on the combined segmentation and stereotactical normalization approach implemented in the SPM8 software package and predefined masks for left and right hippocampus. The method worked properly in all subjects, i.e., visual inspection of the stereotactically normalized GM component images did not show obvious failures, although no patient was excluded based on technical constraints such as poor MR image quality. This demonstrates the robustness of the method with regard to standard 3D MPRAGE imaging on clinical scanners [42], which is an important prerequisite for use in everyday clinical routine.

On the other hand, it is evident that predefined standard masks for the hippocampus do not allow extraction of hippocampal volumes with the same accuracy as manual segmentation or more sophisticated and hence computationally expensive semi-automatic or automatic methods, such as FreeSurfer [32]. This is due to residual anatomical inter-subject variability after stereotactical normalization, which is more pronounced in case of strongly atrophic brains. However, the primary aim in clinical routine patient care is not most accurate volumetry but robust and readily available parameters that provide good diagnostic accuracy or predictive power. Clerx and co-workers recently have shown that fully automated atlas-based hippocampal volumetry indeed can provide the same power for prediction of AD than manual measurement [33].

Inclusion and exclusion criteria of the present study were relaxed to a minimum to make the patient sample as representative as possible for everyday clinical routine. In this respect this study differs from many previous studies on the use of hippocampal volumetry for the diagnosis of AD investigating rather highly screened patient samples. The present study also differs by the fact that there was no control group of healthy subjects included in the analyses. We do not consider this as a limitation of the present study. It rather reflects clinical routine in the memory clinic setting in which the task is detection of AD among patients with memory complaints and not the differentiation of patients with AD from healthy subjects.

Consensus guidelines require that a diagnostic biomarker to be useful in the clinical setting provides sensitivity for the detection of AD exceeding 80% and specificity for distinguishing AD from other dementias also exceeding 80% [34–36]. We found that GMV of the right hippocampus scaled to the patient's total GMV provided sensitivity and specificity of 80% and 91% for the differentiation of clinically AD from a heterogeneous group of non-AD patients including patients with SCI, non-amnestic MCI, or dementia caused by other neurodegenerative diseases than AD, for example frontotemporal dementia. This demonstrates that the technology is mature and qualifies as a diagnostic marker for use in clinical routine. Sensitivity and specificity for the detection of AD in the whole patient sample were slightly lower, but still reached 80%.

In order to investigate the impact of heterogeneity of the MR acquisition protocol we repeated the analysis using only the subgroup of 89 patients scanned according to protocol (A). As expected, we found a slightly better performance for all biomarkers, e.g., AUC for HVR/GMV of setup 2 improved from 0.84 to 0.85 and AUC for HVR_{ad} from 0.81 to 0.84.

HV and GMV both were obtained by use of freely available software for fully automated image analysis. Thus, the method could be implemented without cost at any institution. To facilitate clinical use, we bundled the code and composed a SPM8 software toolbox, which is able to compute HVL, HVR, and GMV using the methodology described in this paper. The software (termed "HV") can be freely downloaded (http://www.fil.ion.ucl.ac.uk/spm/ext/#HV) and runs under SPM8's graphical user interface. Total computation time on a PC was less than four minutes. Therefore, the method can be easily integrated in the clinical routine workflow.

The right hippocampus provided better diagnostic accuracy than the left hippocampus. The following two points contributed to this finding (Table 1): (i) in the AD patients (both AD and intermediate AD), the right hippocampus was more strongly affected than the left one, (ii) in the non-AD group, the left hippocampus showed a slightly smaller volume than the right one (although both effects did not reach the level of statistical significance). The first point is in line with a large meta-analysis by Schroeter and co-workers who found statistically significant atrophy in AD patients compared to healthy controls in amygdala, anterior hippocampal formation, uncus, and (trans-)entorhinal area in both hemispheres, whereas the hippocampus (body, tail) showed significant atrophy in the right hemisphere only [37]. The second point most likely is related to the inclusion of patients with frontotemporal dementia in the non-AD group, since frontotemporal dementia causes more pronounced atrophy in most patients in the left hemisphere including the hippocampus [38, 39].

Patients with clinically AD or intermediate AD were significantly older than the non-AD patients (Table 1). This reflects the situation in clinical routine patient care, since the prevalence of AD strongly increases with age compared to most neurodegenerative diseases other than AD, such as frontotemporal dementia, which tend to occur at younger age than AD. However, the significantly different age between the diagnostic groups to be differentiated suggests that age should be taken into account when hippocampal volume is used as a biomarker, since hippocampal volume decreases with healthy aging. Not accounting for age would result in an overestimation of the diagnostic power of hippocampal volumetry. For the patient sample of the present study, uncorrected HVR provided an AUC of

0.92 for discrimination of AD from non-AD patients, i.e., slightly higher than HVR/GMV (AUC = 0.90, Table 2). Therefore, uncorrected hippocampal volume might be used to support the detection of AD in clinical settings in which the patients to be differentiated from AD are known to be younger than the patients with AD.

TIV is another important nuisance variable in hippocampal volumetry, since the hippocampal volume is expected to strongly correlate with TIV. This was confirmed by the present study in a large sample of healthy controls who had obtained MR imaging as part of a check-up program in a medical prevention center.

In order to account for age and TIV, we fitted a bilinear model with age and TIV as independent variables to the hippocampal volumes to an independent control group. The resulting regression functions were used to transform the hippocampal volumes of all patients to the same age and TIV. Alternatively, hippocampal volumes were scaled to individual total GMV. The latter removed the correlation of hippocampal volumes with both age and TIV in the large sample of healthy controls suggesting that GMV can substitute for both age and TIV simultaneously. The simple scaling to GMV resulted in a better diagnostic accuracy than the bilinear fitting to the control group. Scaling to individual GMV does not require a database of healthy controls. This is an important advantage, since it allows widespread use of automated hippocampal volumetry in clinical routine patient care, also in small institutions and practices in which a proper database of healthy controls is not available. Another possible advantage of scaling to individual GMV is that it might reduce the impact of differences in image acquisition which are difficult to avoid in clinical routine (two slightly different acquisition protocols were included in the present study). The rationale is that differences in image acquisition of course do affect tissue segmentation, but similarly in all brain regions so that the effect cancels to some extent when local GMV is scaled to GMV.

Scaling to the GMV has also limitations. In particular, in later stages, when the disease has already spread out and there is substantial loss of GM outside of the hippocampus, scaling to GMV counters the effect of the disease on hippocampal volume. This effect is small in early stages of AD and, therefore, does not limit the use of scaling to GMV for early diagnosis. However, it is expected to result in decreased sensitivity for the detection of moderate-to-severe AD. The impact of GMV scaling should be investigated in further studies, including also patient samples with more advanced disease. Before then, correction for TIV and age based on a bilinear fit in healthy controls might be preferred over GMV scaling in everyday patient care, because it is most likely more robust with respect to atrophy and other pathology outside the hippocampus.

Finally, clinical diagnosis of AD based on core clinical criteria as suggested by current international guidelines [11, 12, 22] was used as gold standard. The clinical diagnosis has sensitivity between 70.9% and 87.3% whereas specificity ranges from 44.3% to 70.8% [40]. The limited accuracy of the clinical diagnosis might have resulted in underestimation of HV diagnostic accuracy.

CONCLUSION

The ratio of right hippocampal to total GMV estimated by freely available software using fully automated atlas-based segmentation fulfills the requirements for a relevant core feasible biomarker for the detection of AD in everyday patient care in a secondary care memory clinic for outpatients. It is easily integrated in the routine workflow.

ACKNOWLEDGMENTS

REFERENCES

[1] Jack CR, Petersen RC, O'Brien PC, Tangalos EG (1992) MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. *Neurology* **42**, 183-188.
[2] Braak H, Braak E (1991) Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* **82**, 239-259.
[3] Ewers M, Sperling RA, Klunk WE, Weiner MW, Hampel H (2011) Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. *Trends Neurosci* **34**, 430-442.
[4] Teipel SJ, Grothe M, Lista S, Toschi N, Garaci FG, Hampel H (2013) Relevance of magnetic resonance imaging for early detection and diagnosis of Alzheimer disease. *Med Clin North Am* **97**, 399-424.
[5] Hasboun D, Chantôme M, Zouaoui A, Sahel M, Deladoeuille M, Sourour N, Duyme M, Baulac M, Marsault C, Dormont D (1996) MR determination of hippocampal volume: Comparison of three methods. *AJNR Am J Neuroradiol* **17**, 1091-1098.
[6] Szentkuti A, Guderian S, Schiltz K, Kaufmann J, Münte TF, Heinze HJ, Düzel E (2004) Quantitative MR analyses of the hippocampus: Unspecific metabolic changes in aging. *J Neurol* **251**, 1345-1353.

[7] Testa C, Laakso MP, Sabattoli F, Rossi R, Beltramello A, Soininen H, Frisoni GB (2004) A comparison between the accuracy of voxel-based morphometry and hippocampal volumetry in Alzheimer's disease. *J Magn Reson Imaging* **19**, 274-282.

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

[9] den Heijer T, van der Lijn F, Koudstaal PJ, Hofman A, van der Lugt A, Krestin GP, Niessen WJ, Breteler MM (2010) A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline. *Brain* **133**, 1163-1172.

[10] Frisoni GB, Hampel H, O'Brien JT, Ritchie K, Winblad B (2011) Revised criteria for Alzheimer's disease: What are the lessons for clinicians? *Lancet Neurol* **10**, 598-601.

[11] 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.

[12] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia 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**, 263-269.

[13] Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein HC, Vermersch P, Kuiper M, Steinling M, Wolters EC, Valk J (1992) Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: Diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry* **55**, 967-972.

[14] Frisoni GB, Fox NC, Jack CR, Scheltens P, Thompson PM (2010) The clinical use of structural MRI in Alzheimer disease. *Nat Rev Neurol* **6**, 67-77.

[15] Morra JH, Tu Z, Apostolova LG, Green AE, Avedissian C, Madsen SK, Parikshak N, Hua X, Toga AW, Jack CR, Weiner MW, Thompson PM, Alzheimer's Disease Neuroimaging, Initiative (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.

[16] Chupin M, Gérardin E, Cuingnet R, Boutet C, Lemieux L, Lehéricy S, Benali H, Garnero L, Colliot O, Alzheimer's Disease Neuroimaging, Initiative (2009) Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* **19**, 579-587.

[17] Kwak K, Yoon U, Lee DK, Kim GH, Seo SW, Na DL, Shim HJ, Lee JM (2013) Fully-automated approach to hippocampus segmentation using a graph-cuts algorithm combined with atlas-based segmentation and morphological opening. *Magn Reson Imaging* **31**, 1190-1196.

[18] Firbank MJ, Barber R, Burton EJ, O'Brien JT (2008) Validation of a fully automated hippocampal segmentation method on patients with dementia. *Hum Brain Mapp* **29**, 1442-1449.

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

[20] 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.

[21] Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, Nordberg A, Bäckman L, Albert M, Almkvist O, Arai H, Basun H, Blennow K, de Leon M, DeCarli C, Erkinjuntti T, Giacobini E, Graff C, Hardy J, Jack C, Jorm A, Ritchie K, van Duijn C, Visser P, Petersen RC (2004) Mild cognitive impairment–beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* **256**, 240-246.

[22] 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.

[23] Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J Intern Med* **256**, 183-194.

[24] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment: Clinical characterization and outcome. *Arch Neurol* **56**, 303-308.

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

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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 Bioinformatics* **12**, 77.

[32] 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.

[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] (1998) Consensus report of the Working Group on: "Molecular and Biochemical Markers of Alzheimer's Disease". The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group. *Neurobiol Aging* **19**, 109-116.

[35] Hampel H, Frank R, Broich K, Teipel SJ, Katz RG, Hardy J, Herholz K, Bokde AL, Jessen F, Hoessler YC, Sanhai WR, Zetterberg H, Woodcock J, Blennow K (2010) Biomarkers for Alzheimer's disease: Academic, industry and regulatory perspectives. *Nat Rev Drug Discov* **9**, 560-574.

[36] Hampel H, Lista S, Khachaturian ZS (2012) Development of biomarkers to chart all Alzheimer's disease stages: The royal road to cutting the therapeutic Gordian Knot. *Alzheimers Dement* **8**, 312-336.

[37] Schroeter ML, Stein T, Maslowski N, Neumann J (2009) Neural correlates of Alzheimer's disease and mild cognitive impairment: A systematic and quantitative meta-analysis involving 1351 patients. *Neuroimage* **47**, 1196-1206.

[38] Muñoz-Ruiz MÁ, Hartikainen P, Koikkalainen J, Wolz R, Julkunen V, Niskanen E, Herukka SK, Kivipelto M, Vanninen R, Rueckert D, Liu Y, Lötjönen J, Soininen H (2012) Structural MRI in frontotemporal dementia: Comparisons between hippocampal volumetry, tensor-based morphometry and voxel-based morphometry. *PLoS One* **7**, e52531.

[39] Boccardi M, Laakso MP, Bresciani L, Galluzzi S, Geroldi C, Beltramello A, Soininen H, Frisoni GB (2003) The MRI pattern of frontal and temporal brain atrophy in fronto-temporal dementia. *Neurobiol Aging* **24**, 95-103.

[40] Beach TG, Monsell SE, Phillips LE, Kukull W (2012) Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. *J Neuropathol Exp Neurol* **71**, 266-273.

[41] Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognit Lett* **27**, 861-874.

[42] Huppertz HJ, Kröll-Seger J, Klöppel 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.