

# Performance of Hippocampus Volumetry with FSL-FIRST for Prediction of Alzheimer's Disease Dementia in at Risk Subjects with Amnesic Mild Cognitive Impairment

Per Suppa<sup>a,b</sup>, Harald Hampel<sup>c</sup>, Timo Kepp<sup>b</sup>, Catharina Lange<sup>a</sup>, Lothar Spies<sup>b</sup>, Jochen B. Fiebach<sup>d</sup>, Bruno Dubois<sup>c,1</sup> and Ralph Buchert<sup>a,1,\*</sup> for the Alzheimer's Disease Neuroimaging Initiative<sup>2</sup>

<sup>a</sup>*Department of Nuclear Medicine, Charité, Berlin, Germany*

<sup>b</sup>*Jung diagnostics GmbH, Hamburg, Germany*

<sup>c</sup>*Université Pierre et Marie Curie, Institut de la Mémoire et de la Maladie d'Alzheimer & INSERM U1127, Institut du Cerveau et de la Moelle épinière (ICM), Département de Neurologie, Hôpital de la Pitié-Salpêtrière, Paris, France*

<sup>d</sup>*Center for Stroke Research Berlin, Charité, Berlin, Germany*

Accepted 27 December 2015

**Abstract.** MRI-based hippocampus volume, a core feasible biomarker of Alzheimer's disease (AD), is not yet widely used in clinical patient care, partly due to lack of validation of software tools for hippocampal volumetry that are compatible with routine workflow. Here, we evaluate fully-automated and computationally efficient hippocampal volumetry with FSL-FIRST for prediction of AD dementia (ADD) in subjects with amnesic mild cognitive impairment (aMCI) from phase 1 of the Alzheimer's Disease Neuroimaging Initiative. Receiver operating characteristic analysis of FSL-FIRST hippocampal volume (corrected for head size and age) revealed an area under the curve of 0.79, 0.70, and 0.70 for prediction of aMCI-to-ADD conversion within 12, 24, or 36 months, respectively. Thus, FSL-FIRST provides about the same power for prediction of progression to ADD in aMCI as other volumetry methods.

**Keywords:** ADNI, Alzheimer's disease, aMCI-to-Alzheimer disease dementia, amnesic mild cognitive impairment, FSL-FIRST, fully-automated, hippocampal volumetry, magnetic resonance imaging, model-based segmentation, prediction

---

<sup>1</sup>The authors contributed equally to this work 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.

<sup>2</sup>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](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

## INTRODUCTION

Various expert groups including the National Institute on Aging – Alzheimer’s Association (NIA-AA) working group [1], the International Working Group (IWG) [2–4], and the European Federation of the Neurological Societies (EFNS) task force [5] recommend the use of magnetic resonance imaging (MRI)-based hippocampus volume as biomarker for neurodegeneration to complement symptom-based criteria for improved prognostic accuracy in amnesic mild cognitive impairment (aMCI).

Manual segmentation of the hippocampus by an expert is considered the gold standard for hippocampus volumetry. However, manual segmentation is time consuming and therefore hardly compatible with routine workflow in the typical diagnosis setting with high patient throughput. In contrast, fully-automated and computationally efficient software tools do provide the potential to translate hippocampus volumetry into clinical routine. There are several commercial and non-commercial software tools for hippocampus volumetry available, some of which have been evaluated by the European Medicines Agency (EMA) in a process of qualification of hippocampus volume as an imaging biomarker for enrichment of clinical trials in pre-dementia stages of Alzheimer’s disease (AD) [6]. Despite considerable methodological differences, all of the tested tools provided about the same power for prediction of conversion to AD dementia (ADD) in MCI subjects from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) [6]. Our group [7, 8] recently proposed a processing pipeline for fully-automated hippocampus volumetry based on the Statistical Parametric Mapping (SPM) software package (version 8, Wellcome Trust Centre for Neuroimaging, London, UK) [9]. The SPM software is freely available and well documented open source. The prognostic accuracy of the SPM8 processing pipeline in ADNI aMCI subjects was in very good agreement with the EMA results [8]. However, the SPM8 processing pipeline uses (global) stereotactical normalization into template space and a predefined standard hippocampus atlas mask to delineate the hippocampus in template space. The hippocampus mask selected for this application is rather large, in order to account for residual anatomical variability after stereotactical normalization. As a consequence, hippocampus volume estimated by the SPM8 pipeline is considerably larger than hippocampus volume from manual segmentation according to the harmonized

protocol proposed recently [10]. This is a limitation of the atlas-based SPM8 approach.

The aim of the present study was to evaluate the performance of another freely available software tool, FSL-FIRST, which was developed for anatomical segmentation of subcortical structures including the hippocampus [11]. FSL-FIRST deploys model-based segmentation for accurate anatomical delineation of the hippocampus (detailed description of the method in [11]). Hippocampus volumetry with FSL-FIRST was performed in exactly the same ADNI subjects that had been included in the evaluation of the SPM8 processing pipeline [8]. This allows direct head-to-head comparison of FSL-FIRST and SPM8.

## MATERIAL AND METHODS

### *ADNI subjects*

MRI data used in this study were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). The ADNI was launched in 2003 as a public private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

In brief, 198 aMCIs from phase 1 of the ADNI were included: 95 ADNI aMCIs who had converted to ADD within a period of 36 months (32 after 12 months, 43 between 12 and 24 months, and 20 between 24 and 36 months), and 103 ADNI aMCIs who had remained stable over 36 months. In addition, 137 ADNI normals with 1.5T screening MRI were included as controls. All of them had been documented as normal throughout a period of 36 months after baseline clinical examination. No subjects were excluded based on poor image quality. A more detailed description of eligibility criteria and the characteristics of the included cohorts is given in [8].

Two 3D T1-weighted magnetization prepared rapid gradient echo (3D-MPRAGE) images 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 correction, see <http://adni.loni.usc.edu/methods/mri-analysis/mri-pre-processing/>).

### Hippocampal volumetry

In preparation of the application of FSL-FIRST, all 3D-MPRAGE images were transformed into a common coordinate system. SPM8 rigid-body co-registration to a whole brain template [12] in the Montreal Neurological Institute (MNI) space was used for this preprocessing step. Then, hippocampus segmentation was performed using the FIRST module from the FMRIB's Software Library (FSL; version 5.0; <http://fsl.fmrib.ox.ac.uk/fsl>). The run\_first\_all routine was applied with slight modification to enable a larger search region and a normalized mutual information cost function as described in [13]. The total hippocampus volume (FIRST-HV) was obtained by summing hippocampus volume in left and right hemisphere.

FSL-FIRST uses a hippocampus model comprising shape and intensity information. The model has been derived from 336 T1-weighted MRIs in which the hippocampus was delineated manually. These MRIs were registered globally to MNI space (affine 12 parameter registration). A second, local registration was restricted to subcortical structures. The variation in shape and intensity is modeled by a multivariate Gaussian distribution. A detailed description of the method is given in [11].

### Short-term test-retest stability

Short-term test-retest stability of hippocampus volumetry by FSL-FIRST was assessed by using the repeat MPRAGE scan from the baseline imaging session of the 198 ADNI-aMCI subjects. The difference between the two FIRST-HV estimates was characterized by the signed difference [ml] =  $v_1 - v_2$  and by the relative signed difference [%] =  $200 * (v_1 - v_2) / (v_1 + v_2)$ , where  $v_1$  and  $v_2$  denote FIRST-HV from the first and from the repeat MPRAGE scan within the baseline imaging session. Analysis of variance (ANOVA) was used to test the mean of the signed difference and the mean of the relative signed difference for a group effect (difference between the 4 aMCI subgroups). The Levene test was used to test the variance of the two measures for a group effect.

### Validation against semi-automated segmentation ground truth

FIRST-HV values were correlated with hippocampus volume obtained by a semi-automated method (HV-SNT, Medtronic Surgical Navigation Technolo-

gies, Louisville, CO) which has been shown to provide excellent agreement with manual tracing of the hippocampus [14]. HV-SNT values were available for download from the ADNI homepage for 134 of the 198 ADNI aMCIs ( $n = 68$  stable aMCIs).

### Correction for total intracranial volume and age

For each individual subject, FIRST-HV (from the first baseline MPRAGE scan) was adjusted to mean total intracranial volume (TIV, 1450 ml) and mean age (75.8 years) in the control group based on bilinear regression of FIRST-HV with TIV and age as independent variables in the control group [8]. The adjusted total hippocampus volume is denoted as FIRST-HV<sub>adj</sub>.

The SPM8-based HV toolbox [7] was used for TIV estimation, although the FSL software provides possibilities to obtain the TIV, for example by using the Brain Extraction Tool (BET) with corrections as described in [15], and the use of different software tools in general means an additional effort for the user. The rationale for using the SPM8-based HV toolbox for TIV estimation was to simplify the comparison with TIV- and age-adjusted hippocampus volume from the SPM8 processing pipeline (avoid additional variability by different TIV estimates) [8]. The HV toolbox is freely available from the SPM website at <http://www.fil.ion.ucl.ac.uk/spm/ext/#HV>.

### ROC analysis

Receiver operating characteristic (ROC) analysis was used to evaluate the power of FIRST-HV<sub>adj</sub> for differentiation between the ADNI aMCI-to-ADD converters and the stable ADNI aMCIs. The area (AUC) under the ROC curve was used as performance measure. The open source R package pROC was deployed for ROC analysis [16].

Cut-off values for estimation of prognostic accuracy were obtained by Youden's method. Accuracy measures were cross-validated to correct for overfitting by using 100 repeats of 20-fold cross-validation [8].

## RESULTS

Delineation of the hippocampi by FSL-FIRST worked properly (according to visual inspection) in all subjects except one normal control. This subject was excluded from the bilinear regression of FIRST-HV in the control group (regression coefficients were

0.0013 ml/ml and  $-0.0561$  ml/year for TIV and age, respectively).

Total processing time including SPM8-based preprocessing (coregistration and TIV estimation) was approximately 15 min per scan on a standard 2.67 GHz CPU with 8 MB cache.

The results on short-term test-retest stability of FIRST-HV are given in Table 1. One stable aMCI subject had to be excluded from this analysis because FSL registration failed for the repeat scan. Test-retest variability did not differ between the aMCI subgroups: neither the mean nor the variance of signed difference or relative signed difference showed a significant group effect (ANOVA/Levene  $p=0.713/0.290$  and  $0.743/0.078$  for signed difference and relative signed difference, respectively).

FIRST-HV showed a strong correlation with the semi-automated HV-SNT method (Pearson's correlation coefficient = 0.86,  $p < 2.2 \times 10^{-16}$ ).

ROC curves for FIRST-HVad are shown in Fig. 1. Maximum AUC of 0.79 was achieved for identification of ADNI aMCIs who converted to ADD within

Table 1

Signed difference and relative signed difference of (uncorrected) FIRST-HV between the first MPRAGE scan and the repeat MPRAGE scan in the same imaging session (mean  $\pm$  standard deviation). The distribution (mean, standard deviation) was not significantly different between the four aMCI subgroups, neither of the signed difference nor of the relative signed difference

	<i>n</i>	Signed difference [ml]	Relative signed difference [%]
stable aMCI subjects	102	$-0.02 \pm 0.18$	$-0.27 \pm 2.65$
aMCI-to-ADD converters within 12 months	32	$0.02 \pm 0.22$	$0.32 \pm 4.11$
aMCI-to-ADD converters between 12 and 24 months	43	$-0.01 \pm 0.17$	$-0.22 \pm 2.83$
aMCI-to-ADD converters between 24 and 36 months	20	$0.01 \pm 0.24$	$0.30 \pm 4.45$
all aMCI subjects	197	$-0.01 \pm 0.19$	$-0.11 \pm 3.16$

12 months. There was a trend to lower AUCs for detection of ADNI aMCIs who converted within 24 (AUC = 0.70) or 36 months (AUC = 0.70). Details are given in Table 2.

## DISCUSSION

Objective of the present study was to qualify the freely available software tool FSL-FIRST for the prediction of aMCI-to-ADD conversion based on hippocampal volume.

For this purpose, we first assessed the short-term test-retest stability of FIRST-HV, the total bilat-

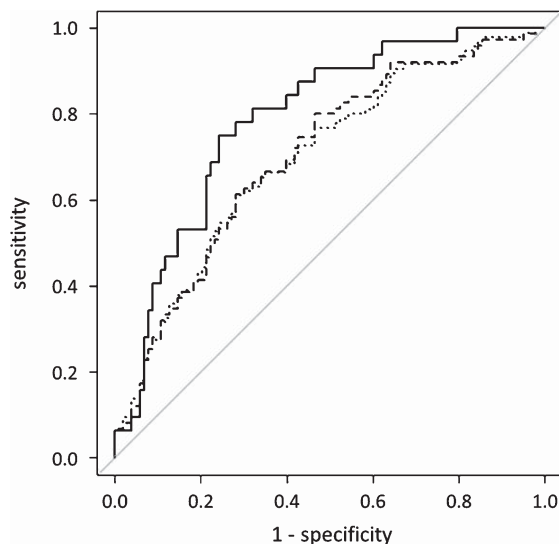


Fig. 1. ROC curve of TIV- and age-corrected total hippocampal volume (FIRST-HVad) for discrimination between ADNI aMCI stable subjects and ADNI aMCI subjects who converted to ADD within 12 (continuous line, AUC = 0.79), 24 (broken line, AUC = 0.70), and 36 months (dotted line, AUC = 0.70).

Table 2

Area (AUC) under the ROC curve, cut-off value determined by the maximum Youden index, and accuracy measures for prediction of aMCI-to-ADD conversion within 12, 24 or 36 months by hippocampal volume corrected for TIV and age (FIRST-HVad). All accuracy measures were cross-validated. Standard deviation is given in round brackets, 95% confidence interval in square brackets

Interval [months]	AUC	Cut-off [ml]	Accuracy	Cross validation			
				Sensitivity	Specificity	PPV	NPV
12	<b>0.79</b> [0.71–0.87]	6.16 (0.01)	0.72 [0.65–0.78]	0.75 [0.68–0.80]	0.71 [0.64–0.77]	0.44 [0.38–0.52]	0.90 [0.85–0.94]
24	<b>0.70</b> [0.63–0.78]	6.58 (0.03)	0.61 [0.55–0.67]	0.64 [0.58–0.70]	0.58 [0.52–0.64]	0.53 [0.47–0.59]	0.69 [0.63–0.74]
36	<b>0.70</b> [0.63–0.77]	6.33 (0.01)	0.65 [0.59–0.70]	0.61 [0.55–0.66]	0.68 [0.62–0.73]	0.64 [0.58–0.69]	0.65 [0.60–0.71]

eral hippocampus volume estimated by FSL-FIRST. Signed difference and relative signed difference of FIRST-HV from the two back-to-back MPRAGE scans of the ADNI baseline imaging session were both very small (Table 1), demonstrating very high test-retest stability of FIRST-HV. The actual estimates of the hippocampus volume (FIRST-HV) were used for the test-retest analysis rather than hippocampus volume corrected for TIV and age (FIRST-HVad) in order to avoid additional variability due to test-retest variability of SPM8-based TIV estimates. FSL-FIRST was applied to exactly the same ADNI aMCI subjects that had previously been included in the evaluation of an SPM8-based processing pipeline [8]. Although both methods are quite distinct methodologically, the predictive accuracy in these subjects was about the same (AUC of FSL-FIRST / SPM8 was 0.79 / 0.78, 0.70 / 0.72, and 0.70 / 0.71 for prediction of aMCI-to-ADD conversion within 12, 24 and 36 months, respectively). The performance of FSL-FIRST is also in good agreement with the performance of other tools evaluated in ADNI MCI subjects. In the EMA study, for example, AUC for prediction of MCI-to-ADD conversion within 24 months ranged from 0.69–0.74 [6]. This suggests that the method used for quantitative estimation of the hippocampus volume has only a small impact on its predictive power. A possible explanation of this might be an intrinsic limitation of the hippocampus volume as prognostic marker in aMCI, that is, an upper threshold for its accuracy considerably below 100%, which also the best volumetry cannot surpass.

FIRST-HV showed high correlation with the semi-automated HV-SNT values which in turn have been shown to provide excellent agreement with manual tracing of the hippocampus [14]. The correlation with HV-SNT was considerably stronger for FSL-FIRST than for the SPM8 processing pipeline (Pearson's correlation coefficient = 0.86 versus 0.72 [8]). This indicates that the model-based FSL-FIRST method indeed achieves more accurate delineation of the individual hippocampus in MPRAGE images than the atlas-based SPM8 approach. However, it should be noted that other methods for fully-automatic hippocampus volumetry might provide even higher agreement with manual delineation. For example, Wolz and co-workers reported very high intraclass correlation between the learning embeddings for atlas propagation (LEAP) method and manual delineation (two-way mixed single measures intraclass correlation coefficient  $ICC(3, 1) = 0.898$ ) [17].

Ahdidan and colleagues, using another multi-atlas method for fully-automated hippocampus volumetry, found very high spatial agreement with manual hippocampal segmentation as measured by the Dice Similarity Coefficient (0.87 on average) [18].

Hippocampus segmentation by FSL-FIRST clearly failed in only one out of 335 subjects. This demonstrates the robustness of the method, which is an important prerequisite for use in everyday clinical routine. However, it is important to note that this robustness was achieved only after some minor modifications of the original FSL-FIRST pipeline, namely (i) rigid-body transformation of each individual MRI into template space prior to application of FSL-FIRST and (ii) spatial extension of FSL-FIRST's hippocampus search space as has been suggested previously [19].

## CONCLUSION

Hippocampus volumetry with FSL-FIRST provides about the same performance for prediction of aMCI-to-ADD conversion as other tools. This qualifies FSL-FIRST to be added to the list of freely available hippocampus volumetry tools for use in clinical routine. Compared to the atlas-based SPM8 approach, model-based FSL-FIRST provides more accurate anatomical delineation of the hippocampus and, therefore, more accurate estimates.

## ACKNOWLEDGMENTS

The authors 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., T.K., and L.S. are employees of jung diagnostics GmbH.

H.H. is supported by the AXA Research Fund, the Fondation Université Pierre et Marie Curie and the Fondation pour la Recherche sur Alzheimer, Paris, France. The research leading to these results has received funding from the program "Investissements d'avenir" ANR-10-IAIHU-06.

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: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; 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.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. 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 Neuroimaging at the University of Southern California.

Authors' disclosures available online (<http://j-alz.com/manuscript-disclosures/15-0804r1>).

## 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, 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.
- [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, 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.
- [5] Hort J, O'Brien JT, Gainotti G, Pirttila T, Popescu BO, Rektorova I, Sorbi S, Scheltens P, Dementia ESPo (2010) EFNS guidelines for the diagnosis and management of Alzheimer's disease. *Eur J Neurol* **17**, 1236-1248.
- [6] 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 pre-dementia stages of Alzheimer's disease. *Alzheimers Dement* **10**, 421-429 e423.
- [7] 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.
- [8] Suppa P, Hampel H, Spies L, Fiebach JB, Dubois B, Buchert R (2015) Fully-automated atlas-based hippocampus volumetry for clinical routine: Validation in subjects with mild cognitive impairment from the ADNI cohort. *J Alzheimers Dis* **46**, 199-209.
- [9] Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, Ashburner JT, Penny WD (2004) *Human Brain Function*, Academic Press, San Diego.
- [10] Bocchetta M, Boccardi M, Ganzola R, Apostolova LG, Preboske G, Wolf D, Ferrari C, Pasqualetti P, Robitaille N, Duchesne S, Jack CR, Jr., Frisoni GB, EADC-ADNI Working Group on The Harmonized Protocol for Manual Hippocampal Segmentation and for the Alzheimer's Disease Neuroimaging Initiative (2015) Harmonized benchmark labels of the hippocampus on magnetic resonance: The EADC-ADNI project. *Alzheimers Dement* **11**, 151-160 e155.
- [11] Patenaude B, Smith SM, Kennedy DN, Jenkinson M (2011) A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56**, 907-922.
- [12] 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.
- [13] Hibar D, Jahanshad N, Stein J, Winkler A, Toro R, Jenkinson M (2011) Enigma Consortium First Protocol. <http://enigma.ion.ucla.edu/protocols/imaging-protocols/first-protocol>.
- [14] 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.
- [15] Keihaninejad S, Heckemann RA, Fagiolo G, Symms MR, Hajnal JV, Hammers A, Alzheimer's Disease Neuroimaging Initiative (2010) A robust method to estimate the intracranial volume across MRI field strengths (1.5T and 3T). *Neuroimage* **50**, 1427-1437.

- [16] 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.
- [17] Wolz R, Aljabar P, Hajnal JV, Hammers A, Rueckert D, Alzheimer's Disease Neuroimaging Initiative (2010) LEAP: Learning embeddings for atlas propagation. *Neuroimage* **49**, 1316-1325.
- [18] Ahdidan J, Raji CA, DeYoe EA, Mathis J, Noe KO, Rimestad J, Kjeldsen TK, Mosegaard J, Becker JT, Lopez O (2015) Quantitative neuroimaging software for clinical assessment of hippocampal volumes on MR imaging. *J Alzheimers Dis* **49**, 723-732.
- [19] Nugent AC, Luckenbaugh DA, Wood SE, Bogers W, Zarate CA Jr, Drevets WC (2013) Automated subcortical segmentation using FIRST: Test-retest reliability, interscanner reliability, and comparison to manual segmentation. *Hum Brain Mapp* **34**, 2313-2329.