

# Reference standard space hippocampus labels according to the EADC-ADNI Harmonized Protocol: Utility in automated volumetry

Dominik Wolf<sup>a\*</sup>, Martina Bocchetta<sup>b,c</sup>, Gregory M Preboske<sup>d</sup>, Marina Boccardi<sup>e</sup>, Michel J Grothe<sup>f\*</sup>,  
for the Alzheimer's Disease Neuroimaging Initiative<sup>1</sup>

<sup>1</sup> Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](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)

<sup>a</sup> Department of Psychiatry and Psychotherapy, University Medical Center Mainz, Germany

<sup>b</sup> Laboratory of Alzheimer's Neuroimaging and Epidemiology, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

<sup>c</sup> Dementia Research Centre, Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London, UK

<sup>d</sup> Department of Radiology, Mayo Clinic and Foundation, Rochester, MN, USA

<sup>e</sup> LANVIE – Laboratory of Neuroimaging of Aging, University of Geneva, Switzerland

<sup>f</sup> German Center for Neurodegenerative Diseases (DZNE), Rostock, Germany

## \*Corresponding authors

Dr. Dominik Wolf  
Department of Psychiatry and Psychotherapy  
University Medical Center Mainz  
Untere Zahlbacher Str. 8  
55131 Mainz, Germany  
E-Mail: [dominik.wolf@unimedizin-mainz.de](mailto:dominik.wolf@unimedizin-mainz.de)

Dr. Michel Grothe  
German Center for Neurodegenerative Diseases (DZNE)  
Gehlsheimer Str. 20  
18147 Rostock, Germany  
E-Mail: [michel.grothe@dzne.de](mailto:michel.grothe@dzne.de)

## **Abstract**

**Background:** A harmonized protocol (HarP) for manual hippocampal segmentation on MRI has recently been developed by an international EADC-ADNI project. We aimed at providing consensual certified HarP hippocampal labels in Montreal Neurological Institute (MNI) standard space to serve as reference in automated image analyses.

**Methods:** Manual HarP tracings on the high-resolution MNI152 standard space template of four expert certified HarP tracers were combined to obtain consensual bilateral hippocampus labels. Utility of these reference labels is demonstrated in a simple atlas-based morphometry approach for automated calculation of HarP-compliant hippocampal volumes within SPM software.

**Results:** Individual tracings showed very high agreement among the five expert tracers (pair-wise Jaccard indices 0.82-0.87). Automatically calculated hippocampal volumes were highly correlated ( $r_{L/R}=0.89/0.91$ ) with gold standard volumes in the HarP benchmark dataset (N=135 MRIs), with a mean volume difference of 9% (SD 7%).

**Conclusion:** The consensual HarP hippocampus labels in the MNI152 template can serve as a reference standard for automated image analyses involving MNI standard space normalization.

## **Video part 1:**

### **Background of the study**

The hippocampus has been one of the most studied cerebral structures in neuroimaging research during the last decades. Significant volume reductions have been reported in several psychiatric and neurodegenerative disorders, including major depression [1, 2], Parkinson's disease [3], and Alzheimer's disease (AD) [4]. In AD, hippocampal atrophy is one of the core biomarkers in the revised National Institute on Aging-Alzheimer's Association (NIA-AA) diagnostic criteria [4-6]. Moreover, hippocampal atrophy is among the most sensitive markers of disease progression and is regarded as one of the principal biomarkers for the early diagnosis of AD [5, 7]. In addition to volume loss in AD, alterations in metabolism, activity, and microstructural properties within the hippocampus have been reported [8-10].

The gold standard to identify the hippocampus for image analysis is through manual outlining its boundaries on high-resolution structural magnetic resonance imaging (MRI) scans. Historically, widely different measurement protocols have been developed, leading to the application of different anatomical landmarks, and thus different anatomic definitions of the hippocampus across laboratories [11]. As a result, the comparability between studies on diagnostic accuracy and biologic drug efficacy using hippocampal volume is limited. Against that background, a joint European Alzheimer's Disease Consortium (EADC) and Alzheimer's Disease Neuroimaging Initiative (ADNI) effort was carried out to harmonize existing protocols and develop a standard consensual protocol for the manual segmentation of the hippocampus on MRI [12]. The methodological procedure of the development and validation of the Harmonized Protocol for Hippocampal segmentation (HarP) has previously been described in detail. Briefly, the procedure included the following steps: (I) careful

examination, operationalization, and quantification of the differences in hippocampal segmentation among the 12 most commonly used segmentation protocols [13, 14], (II) application of an iterative Delphi procedure polling a group of international hippocampal segmentation experts to define a consensus for a final HarP [15], (III) creation of benchmark hippocampal tracings according to the HarP by five experts in hippocampal segmentation [16], (IV) validation of the HarP in naïve tracers [17], (V) validation of the HarP versus pathological evaluation [18], and VI) production of benchmark labels for the implementation of the HarP into algorithms for automated segmentation [19].

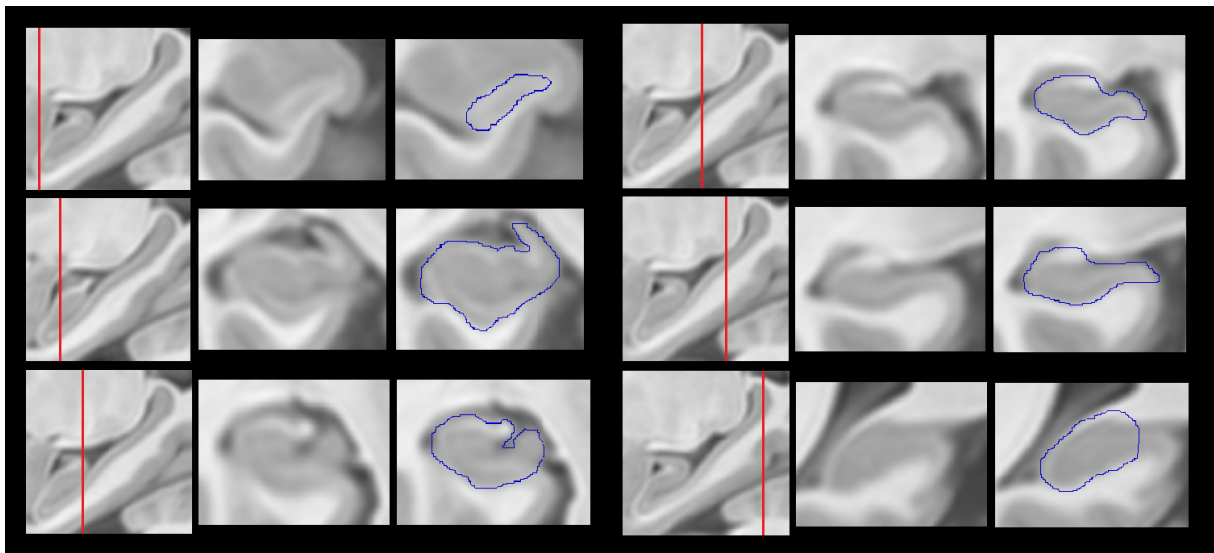
As an extension of the latest milestone of the HarP project, the present work aims at providing high-resolution consensus labels of the HarP in Montreal Neurological Institute (MNI) standard space. These may serve as a graphical 3D reference of the HarP in the standardized human brain, and may be particularly useful for facilitating the use of the HarP definition of hippocampal structure in automated analysis of neuroimaging data, which typically involves MNI standard space normalization in order to make use of the rich anatomic annotations that are available for this image space. Utility of the reference labels is demonstrated in a simple and “easy to use” processing approach for automated calculation of HarP compliant hippocampal volumes (HV) within SPM software, as one of the most widely used software packages for image analysis in neuroimaging research.

## Video part 2

### Methodological procedure

#### 2.1 The development of the standard space HarP reference labels

Four expert tracers who took part in the development of the certified HarP benchmark labels [16, 19] were asked to produce manual HarP tracings on the T1-weighted high-resolution ICBM 2009c Nonlinear Asymmetric MNI152 standard space template (<http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152NLIin2009>). Tracings were performed separately for each hemisphere using the interactive MultiTracer software developed at the Laboratory of Neuroimaging at the University of California (<http://air.bmap.ucla.edu/MultiTracer>). The tracing of one of the four expert tracers is illustrated in Figure 1.



**Figure 1.** HarP conform tracing of the hippocampus on the ICBM 2009c Nonlinear Asymmetric MNI152 standard space template (blue contour) of one of the four expert tracers (selected slices).

In accordance to the development of the certified training labels [19], a two-stage procedure was carried out, consisting of segmentation by the expert tracers and a following quality check by an independent HarP expert who was not involved in the segmentation. In case of possible divergences from the protocol the expert tracers received a written feedback and were allowed to reevaluate their tracings in light of this feedback. Final segmentation contours were converted to 3D volumetric image files in Nifti format as described previously [19], and these were combined to obtain bilateral probabilistic HarP labels encoding the segmentation overlap among the four tracers.

## **2.2 Automated hippocampus volumetry based on the standard space HarP reference labels**

Within the HarP project, a large benchmark set of native-space hippocampal segmentations based on the HarP has been produced, that aims to cover a large range of the anatomical variability in hippocampal structure encountered in aging and AD [19]. This dataset comprises manually segmented MRI data of a total of 135 cognitively normal, mild cognitive impairment (MCI), and AD subjects enrolled in the Alzheimer's Disease Neuroimaging Initiative (ADNI), and has been used in the present study for automated calculation of individual HVs based on the standard space consensual HarP labels. Demographical and clinical data of the subjects are shown in table 1.

**Table 1.** Demographical and clinical data of the diagnostic subgroups

	<b>CN</b>	<b>MCI</b>	<b>AD</b>
N	44	46	45
Age (yr)	76 (7)	75 (8)	75 (8)
Gender (F/M)	22/22	19/27	24/21
Education (yr)	16 (3)	16 (3)	15 (3)
MMSE	29 (1)	26 (3)	21 (2)
Scheltens	1.1 (1.2)	1.8 (1.2)	2.7 (1.2)

Abbreviations: CN = cognitively normal; MCI = mild cognitive impairment; AD = Alzheimer's disease; MMSE = Mini-Mental State Examination; yr = years; F/M = female/male

For the automated calculation of individual HVs, the MRI data were processed using statistical parametric mapping (SPM8, Wellcome Trust Center for Neuroimaging) and the VBM8 toolbox (<http://dbm.neuro.uni-jena.de/vbm>). First, images were segmented into partitions of gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) using the tissue prior free segmentation routine of the VBM8-toolbox. The resulting GM and WM partitions were then high-dimensionally registered to MNI standard space using Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) [20]. Voxel values were modulated to preserve the original amount of GM and WM volume present before normalization. Finally, individual HVs were calculated by summing up the modulated GM and WM voxel values within the consensual standard space HarP labels (probabilistic labels thresholded at 100% overlap). Modulated WM voxel values were included in the HV calculation since the HarP explicitly specifies to include small white matter regions (alveus, fimbria) in hippocampal segmentations.

### **2.3 Correspondence of automatically calculated hippocampal volumes with manually derived gold standard volumes**

Automatically calculated HVs were compared to the volumes of the manually segmented benchmark HarP labels in native space as the gold standard. The latter were calculated by multiplying the number of manually segmented voxels in the volumetric labels with the voxel size of the native space image. Pearson correlation coefficients and mean percentage volume differences between automatically calculated volumes and gold standard volumes were calculated for the total group as well as for diagnostic subgroups separately.

Correlation coefficients were compared between diagnostic subgroups using Steiger's Z-test [21]. Mean percentage differences in HV were compared between subgroups using ANOVAs and Bonferroni post-hoc tests. Moreover, for a graphical illustration of the correspondence between automatically calculated and manually derived HVs, Bland-Altman plots were generated. Normality of data distributions were tested using Kolmogorov-Smirnov tests.

### **2.4 Comparison of diagnostic accuracy of automatically calculated and manually derived hippocampal volumes**

Analysis of variance (ANOVA) and Bonferroni post-hoc tests were applied to compare automatically calculated and manually derived HVs between diagnostic subgroups (CN, MCI, AD). Cohen's d effect sizes were calculated for post-hoc comparisons. Receiver operating characteristic (ROC) analyses and corresponding areas under the curves (AUCs) were determined to assess and compare the discriminatory power of automatically and manually derived HVs using DeLong's test [22]. For these between group analyses, all HVs were normalized by the total intracranial volume (TIV), calculated as the sum of total volumes of the GM, WM, and CSF partitions.

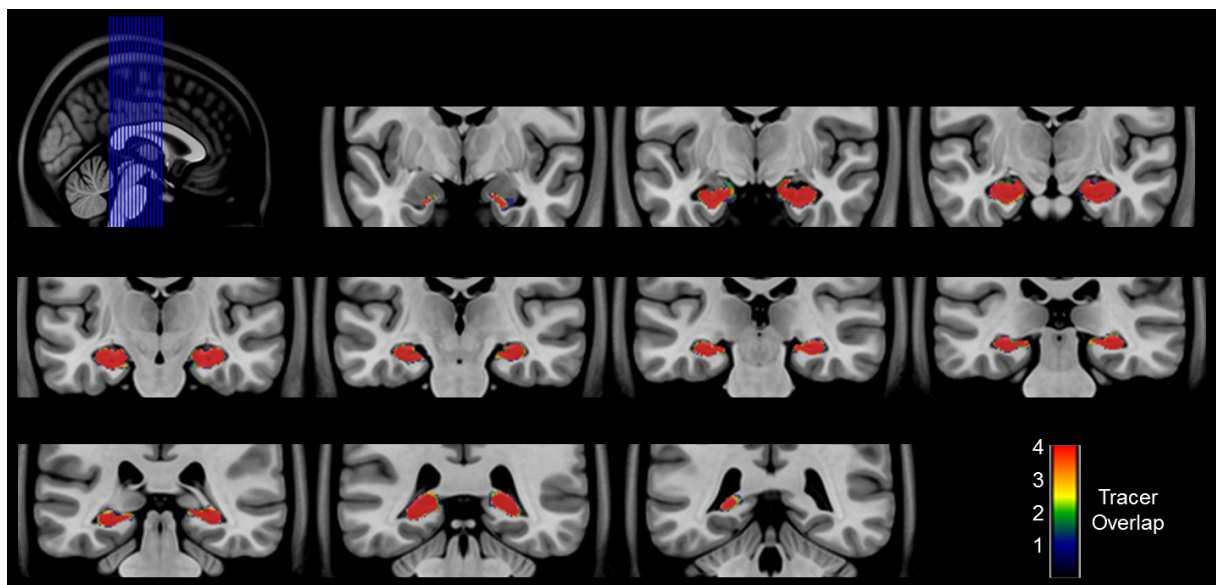


## Video part 3

### Results

#### 3.1 Development of standard space hippocampus reference labels based on HarP criteria

Possible inconsistencies with the HarP were detected in two of the four expert tracings on the high-resolution MNI152 standard space template, both concerning the segmentation of the hippocampal tail on a single posterior section, and the respective tracers decided to modify these segmentations after receiving the written feedback. The final tracings showed very high agreement among tracers (pair-wise Jaccard indices among tracers: 0.82-0.87), and labels were combined into probabilistic HarP labels encoding the segmentation overlap among the four tracers (Figure 2). These labels (one for each hemisphere) are being made publicly available via the official website of the HarP project ([www.hippocampal-protocol.net/SOPs/index.php](http://www.hippocampal-protocol.net/SOPs/index.php)).



**Figure 2.** Probabilistic standard space Harmonized Protocol labels encoding the segmentation overlap among the four expert tracers.

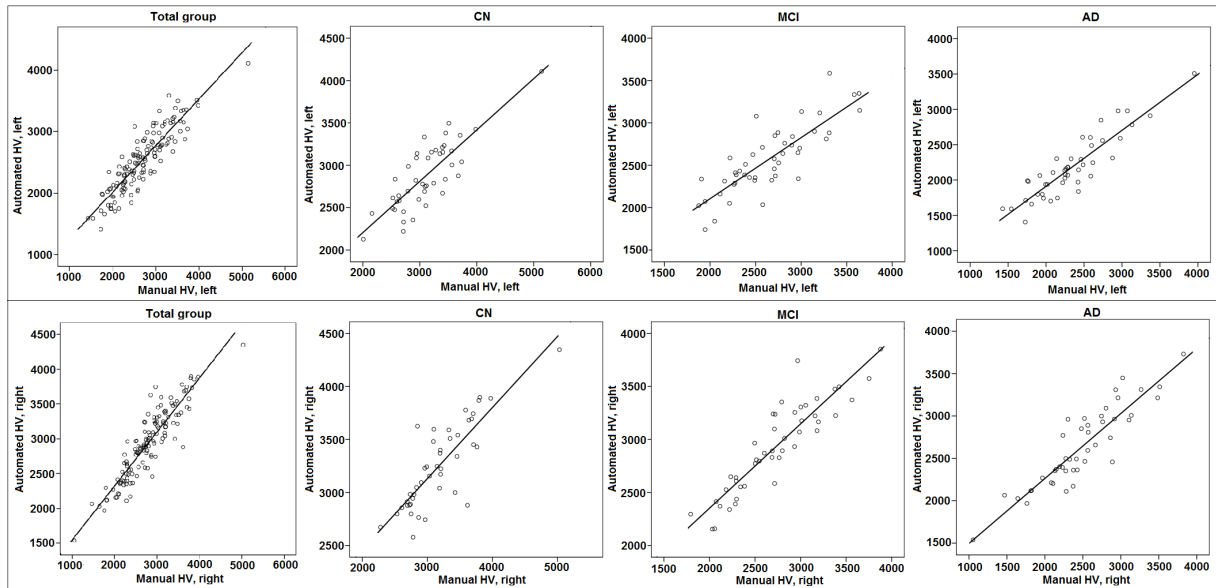
### 3.2 Correspondence between automatically and manually determined HarP hippocampal volumes

Pearson correlation coefficients and mean percentage differences between automatically calculated HVs and the gold standard values in the total group and in diagnostic subgroups are summarized in Table 2. Scatter plots are shown in Figure 3. Correlation coefficients ranged from  $r = .818$  to  $r = .902$  for the left hippocampus and from  $r = .831$  to  $r = .907$  for the right hippocampus, all being highly significant. Comparisons of correlation coefficients between diagnostic subgroups using Steiger's Z-test showed no significant group differences in the strength of correlation, neither for the left hippocampus nor for the right hippocampus. Mean percentage differences between automatically and manually derived HVs ranged from 8.4% to 9.8% for the left hippocampus and from 7.0% to 9.9% for the right hippocampus. In line with the correlation coefficients, comparisons of percentage volume differences between diagnostic subgroups showed no group differences for the left hippocampus or the right hippocampus.

**Table 2.** Pearson correlation coefficients and mean percentage differences between automatically and manually derived Harmonized Protocol hippocampal volumes

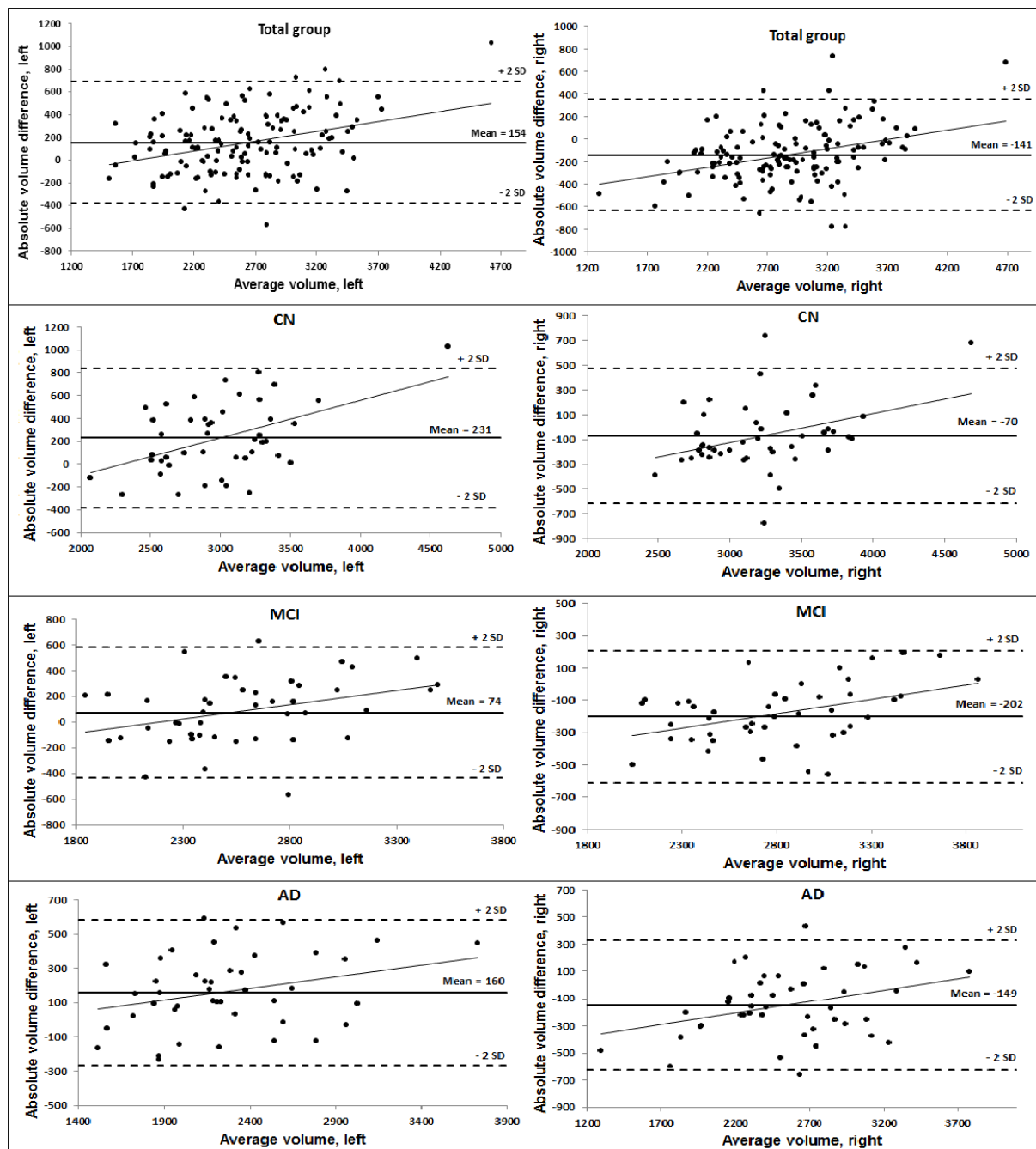
	N		Auto HV (SD)	Manual HV (SD)	r (p)	% diff (SD)
Total	135	left	2540 (497)	2694 (586)	.890 (<.001)	9.3 (6.5)
		right	2939 (497)	2797 (582)	.907 (<.001)	8.5 (6.6)
CN	44	left	2877 (393)	3108 (532)	.818 (<.001)	9.8 (7.1)
		right	3255 (401)	3185 (497)	.831 (<.001)	7.0 (5.3)
MCI	46	left	2570 (401)	2644 (463)	.831 (<.001)	8.4 (5.8)
		right	2933 (423)	2732 (479)	.901 (<.001)	8.7 (6.0)
AD	45	left	2179 (436)	2339 (497)	.902 (<.001)	9.7 (6.7)
		right	2635 (466)	2486 (548)	.900 (<.001)	9.9 (8.0)

Abbreviations: Auto HV: automatically determined Harmonized Protocol hippocampal volumes; Manual HV: manually determined Harmonized Protocol hippocampal volumes; CN: cognitively normal; MCI: mild cognitive impairment; AD: Alzheimer's disease; SD: standard deviation; % diff: mean percentage difference. Hippocampal volumes are expressed in mm<sup>3</sup>.



**Figure 3.** Scatterplots between automatically and manually determined Harmonized Protocol hippocampal volumes in the total group and in diagnostic subgroups (i) cognitively normal (CN), (ii) mild cognitive impairment (MCI), and (iii) Alzheimer’s disease (AD), for the left (upper row) and right (lower row) hippocampus. Hippocampal volumes are expressed in  $\text{mm}^3$ . HV: hippocampal volume.

Bland-Altman plots are depicted in Figure 4. The plots showed small deviations of mean difference between automatically and manually determined HarP HVs from zero in the total group as well as in the diagnostic subgroups (above zero for the left hippocampus, below zero for the right hippocampus). Across all groups, mean differences ranged between  $-202$  and  $+231 \text{ mm}^3$ , showing that the differences stay within rather narrow limits. Incorporated regression lines showed small positive trends between the average volume and the volume differences for the total group and for all diagnostics subgroups, indicating a relative overestimation of the automatically determined volumes for small hippocampi and an underestimation for large hippocampi. However, in general this effect was relatively small.



**Figure 4.** Bland-Altman plots. The x-axis represents the average of the automatically and manually determined Harmonized Protocol hippocampal volume measurements. The y-axis represents the difference between the two hippocampal volume measurements (manually determined volumes minus automatically determined volumes). The bolded black line indicates the mean of the differences between the two volume measurements. The dotted lines represent the 95% confidence intervals. The small black lines represent simple regression lines. CN: cognitively normal; MCI: mild cognitive impairment; AD: Alzheimer's disease; SD: standard deviation.

### 3.3 Comparison of diagnostic accuracy of automatically calculated and manually derived hippocampal volumes

ANOVAs demonstrated significant volume differences between the diagnostic subgroups for both TIV-normalized automatically calculated and manually derived HVs (automatically calculated HVs: left:  $F = 32.4$ ,  $p < .001$ , right:  $F = 26.7$ ,  $p < .001$ ; manually derived HVs: left:  $F = 30.6$ ,  $p < .001$ , right:  $F = 24.1$ ,  $p < .001$ ).

Results of Bonferroni post-hoc tests and effect sizes of the group differences (Cohen's  $d$ , AUC) are shown in Table 3. Automatically calculated left and right HVs differed between all diagnostic groups (CN > MCI, CN > AD, MCI > AD), with high effect sizes and diagnostic power (Cohen's  $d$  ranging from 4.5 to 10.9; AUCs ranging from 0.68 to 0.90). Likewise, manually derived left and right HVs differed between all diagnostic groups (CN > MCI, CN > AD, MCI > AD), with Cohen's  $d$  effect sizes ranging from 3.1 to 10.9 and AUCs ranging from 0.64 to 0.89. Differences in the right manually derived HVs between MCI and AD slightly missed statistical significance. However, none of the differences in AUC values between automatically and manually derived HVs were significant.

**Table 3.** Bonferroni post-hoc tests comparing TIV-normalized automatically and manually determined Harmonized Protocol hippocampal volumes between diagnostic subgroups

			Mean HV differences (SD)	Bonferroni Post hoc ( $p$ )	Cohen's $d$	AUC
CN - MCI	norm Auto HV	left	236 (61)	<.001	5.5	0.74
		right	246 (60)	<.001	5.8	0.75
	norm Manual HV	left	349 (69)	<.001	7.1	0.80
		right	339 (72)	<.001	6.6	0.78
CN - AD	norm Auto HV	left	490 (61)	<.001	10.9	0.90
		right	436 (60)	<.001	10.3	0.88
	norm Manual HV	left	539 (70)	<.001	10.9	0.89
		right	496 (73)	<.001	9.4	0.87
MCI - AD	norm Auto HV	left	255 (60)	<.001	5.9	0.75
		right	190 (59)	.005	4.5	0.68
	norm Manual HV	left	190 (69)	.020	3.9	0.69
		right	157 (72)	.095	3.1	0.64

Abbreviations: norm Auto HV: TIV-normalized automatically determined Harmonized Protocol hippocampal volumes; norm Manual HV: TIV-normalized manually determined Harmonized Protocol hippocampal volumes; CN: cognitively normal; MCI: mild cognitive impairment; AD: Alzheimer's disease; SD: standard deviation;

mean HV differences: mean differences of TIV-normalized hippocampal volumes expressed in mm<sup>3</sup>; AUC: Area under receiver operating characteristics curve.

## **Video part 4**

### **Discussion**

With the present work, we aimed at providing certified and consensual hippocampal labels in MNI standard space based on the recently published HarP to serve as reference standard for automated image analyses involving MNI standard space normalization. Since these labels represent a detailed graphical representation of the HarP criteria, they may also become a useful reference for training purposes, complementing the written description of the HarP criteria in the published manual.

#### **4.1 Development of consensual HarP labels**

The applied procedure to generate the consensual labels ensured highly accurate HarP-conform hippocampal segmentations, which showed very high agreement among tracers (pair-wise Jaccard indices ranging from 0.82-0.87). Although computed on a much smaller set, this performance was in the upper range of performances provided by the whole sample of certified tracers that took part in the validation of the HarP (Jaccard indices ranging from 0.78-0.85) [23] and that, given the very stable segmentations demonstrated in that validation study [17], can be considered to define the thresholds for certified tracers. Analogously to the labels generated for the web-platform for the training of manual tracers [16], our consensual labels allow a small range of variability admissible for very experienced tracers.

## 4.2 Demonstration of the utility of the consensual HarP labels

Utility of the consensual HarP reference labels has been demonstrated in a simple image processing approach for automated calculation of hippocampal volumes within the SPM software. This automated volumetry approach followed the standard “atlas-based” volumetry approach employed in SPM and several other open-source software packages [24-28], which is easy to use and avoids computationally expensive steps. Given these advantages, this processing approach has been integrated in a freely available SPM toolbox [25, 27]. Within SPM5 a similar approach has been validated as a reliable automated alternative to manual segmentation of the hippocampus, with a mean absolute volume difference compared with manual segmentation of  $11\pm 9\%$  and a correlation coefficient of  $r = 0.83$  [29]. In contrast to more complex multi-atlas approaches [30], this method solely relies on an appropriate representation of the hippocampus in the standard space template used for spatial normalization. Its performance is similar or better compared to values reported for other widely used pipelines for automated determination of hippocampal volume, e.g. Freesurfer ( $r=0.74-0.85$ ) or FSL-FLIRT ( $r=0.47-0.66$ ) [31-34]. The standard atlas-based approach was further optimized in our study by using a fully-deformable high-dimensional non-linear registration algorithm (DARTEL) for spatial normalization, which has previously been shown to increase the performance of atlas-based volumetry approaches [35, 36].

Within the HarP project, a set of reference HarP segmentations has been provided for training and certification of tracers and algorithms. Therefore, 135 structural ADNI scans, balanced by age, medial temporal atrophy, and scanner manufacturer have been chosen [19]. We used the same set of images to compare automatically calculated HVs using the consensual hippocampal labels with manually determined gold standard HVs. In the total group, automatically calculated HVs were highly correlated with manually determined HVs (left:  $r = .89$ ; right:  $r = .91$ ). Moreover, mean percentage volume differences were relatively low (8-



9%) and comparable to previously reported values for SPM-based automated hippocampal volumetry [29, 32]. This result suggests that automated hippocampal volume computation using the consensual HarP labels within a simple atlas-based morphometry approach in SPM produces volumes that are highly comparable to the HarP benchmark volumes. Comparable raw volumes, high correlations, and low absolute mean percentage volume differences between automatically calculated and manually determined HVs have also been found within the diagnostic subgroups CN, MCI, and AD. Bland-Altman plots indicated a certain dependence of the automated method's performance on the size of the hippocampus, i.e. a relative overestimation of the automated volumes for small hippocampi and an underestimation for large hippocampi. This is a known phenomenon in automated volumetry procedures, and even the accuracy of manual delineations is expected to drop in severely atrophied brains, due to substantial changes in the anatomy of the employed landmarks in these subjects. However, overall this effect was relatively small and neither correlation coefficients nor mean volume differences differed significantly between diagnostic groups, indicating a reliable performance even in severely atrophied brains of AD patients.

Analyses of diagnostic accuracy demonstrated a high diagnostic power to discriminate between the subgroups NC, MCI, and AD for both automatically calculated and manually determined HVs. Of note, diagnostic accuracy of automatically calculated HVs was as good as the accuracy of manually determined gold standard HVs, and both were in the range of previously reported values for diagnostic group separation based on automated or manual hippocampal volumetry [25, 37-39]. The distinct advantage of HarP-based volumetry methods is the internationally standardized definition of the hippocampus outlines, which aims at reducing variability of volume estimates across laboratories as a mandatory step for a more widespread use of hippocampal volume as a biomarker in routine clinical settings [5].

The automated quantification of hippocampal volume is one example for the applicability of the consensual HarP labels that is of particular importance to AD biomarker research. However, the consensual HarP labels may also be useful for defining hippocampal regions-of-interest in other automated analysis approaches of structural and functional imaging data, including template based deformation field and shape analysis approaches (REFS) [40], as well as analysis of functional MRI [9], positron-emission tomography (PET) [8], or diffusion-tensor imaging (DTI) data [41]. In general, it is our hope that the provision of the consensual HarP labels in MNI standard space will promote the application of this standardized anatomical definition of the hippocampus in the wider field of neuroimaging research.

### **4.3 Limitations**

The standard “atlas-based” automated volumetry approach used in this study has been validated as a reliable automated alternative to manual segmentation of the hippocampus. However, more complex and more precise volumetry approaches have been developed, most notably advanced techniques based on multi-atlas registration and fusion strategies [30, 42-45]. The standard “atlas-based” automated volumetry approach has been used for the following reasons: first, this SPM-based approach is among the most widely used in the wider field of neuroimaging research and does not require intensive computational work. This is an important precondition to promote a far-reaching usage of the standardized criteria for hippocampal anatomy on structural MRI scans. Second, the HarP criteria can easily be integrated in the applied approach by using a single definition of the HarP hippocampus anatomy in the standard space template used for spatial normalization (the consensual HarP labels). This is not the case for other easily accessible and widely used approaches, such as Freesurfer or FSL-FIRST, which show at best similar performance characteristics.

For the automated calculation of individual HVs, the VBM8 toolbox has been applied. This toolbox spatially normalizes all data to a specific DARTEL-compatible MNI space template, which has been generated by affinely aligning 550 healthy control subjects of the IXI-database ([www.braindevelopment.org](http://www.braindevelopment.org)) to the MNI152 template provided within SPM, followed by high-dimensional inter-subject registration using DARTEL. Thus, although this “IXI550” template generally represents MNI space, it may not perfectly correspond to the official high-resolution ICBM 2009c Nonlinear Asymmetric MNI152 template, which has been used for the development of the consensual standard space HarP labels due to its perfect contrast and anatomical detail. Although the correspondence between automatically and manually determined HarP hippocampal volumes in this study was comparably high, it might be further improved when directly normalizing the images to the ICBM 2009c Nonlinear Asymmetric MNI152 template. However, within standard SPM-based high-dimensional normalization routines this is currently not possible.

#### **4.4 Conclusion**

The present work aimed at providing consensual HarP hippocampal labels in MNI standard space that may serve as a reference standard for automated analyses of neuroimaging data involving MNI standard space normalization. The utility of these labels with respect to the field of AD biomarker research was demonstrated in a simple atlas-based morphometry approach for automated calculation of HarP-compliant hippocampal volumes within widely used SPM software. Public availability of the reference HarP labels in MNI space is expected to foster the use of this internationally standardized definition of hippocampus anatomy within the wider community of neuroimaging researchers in the AD field and beyond.

## **Acknowledgements**

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; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; 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; 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 ([www.fnih.org](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.

## 1. References

- [1] Videbech P, Ravnkilde B. Hippocampal volume and depression: a meta-analysis of MRI studies. *American Journal of Psychiatry*. 2015.
- [2] Sawyer K, Corsentino E, Sachs-Ericsson N, Steffens DC. Depression, hippocampal volume changes, and cognitive decline in a clinical sample of older depressed outpatients and non-depressed controls. *Aging & mental health*. 2012;16:753-62.
- [3] Camicioli R, Moore MM, Kinney A, Corbridge E, Glassberg K, Kaye JA. Parkinson's disease is associated with hippocampal atrophy. *Movement Disorders*. 2003;18:784-90.
- [4] Barnes J, Bartlett JW, van de Pol LA, Loy CT, Scahill RI, Frost C, et al. A meta-analysis of hippocampal atrophy rates in Alzheimer's disease. *Neurobiology of aging*. 2009;30:1711-23.
- [5] Jack CR, Barkhof F, Bernstein MA, Cantillon M, Cole PE, DeCarli C, et al. Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. *Alzheimer's & Dementia*. 2011;7:474-85. e4.
- [6] Frisoni GB, Fox NC, Jack CR, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*. 2010;6:67-77.
- [7] Jack C, Shiung M, Gunter J, O'Brien P, Weigand S, Knopman D, et al. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology*. 2004;62:591-600.
- [8] Mosconi L. Glucose metabolism in normal aging and Alzheimer's disease: methodological and physiological considerations for PET studies. *Clinical and translational imaging*. 2013;1:217-33.
- [9] Huijbers W, Mormino EC, Schultz AP, Wigman S, Ward AM, Larvie M, et al. Amyloid- $\beta$  deposition in mild cognitive impairment is associated with increased hippocampal activity, atrophy and clinical progression. *Brain*. 2015:awv007.
- [10] Pasquini L, Scherr M, Tahmasian M, Meng C, Myers NE, Ortner M, et al. Link between hippocampus' raised local and eased global intrinsic connectivity in AD. *Alzheimer's & Dementia*. 2015;11:475-84.
- [11] Konrad C, Ukas T, Nebel C, Arolt V, Toga AW, Narr K. Defining the human hippocampus in cerebral magnetic resonance images—an overview of current segmentation protocols. *Neuroimage*. 2009;47:1185-95.
- [12] Frisoni GB, Jack CR. HarP: The EADC-ADNI Harmonized Protocol for manual hippocampal segmentation. A standard of reference from a global working group. *Alzheimer's & dementia: the journal of the Alzheimer's Association*. 2015;11:107-10.
- [13] Boccardi M, Ganzola R, Bocchetta M, Pievani M, Redolfi A, Bartzokis G, et al. Survey of protocols for the manual segmentation of the hippocampus: preparatory steps towards a joint EADC-ADNI harmonized protocol. *Journal of Alzheimer's disease: JAD*. 2011;26.
- [14] Boccardi M, Bocchetta M, Ganzola R, Robitaille N, Redolfi A, Duchesne S, et al. Operationalizing protocol differences for EADC-ADNI manual hippocampal segmentation. *Alzheimer's & Dementia*. 2013.
- [15] Boccardi M, Bocchetta M, Apostolova LG, Barnes J, Bartzokis G, Corbetta G, et al. Delphi definition of the EADC-ADNI Harmonized Protocol for hippocampal segmentation on magnetic resonance. *Alzheimer's & Dementia*. 2015;11:126-38.
- [16] Bocchetta M, Boccardi M, Ganzola R, Apostolova LG, Preboske G, Wolf D, et al. Harmonized benchmark labels of the hippocampus on magnetic resonance: The EADC-ADNI project. *Alzheimer's & Dementia*. 2015;11:151-60. e5.
- [17] Frisoni GB, Jack CR, Bocchetta M, Bauer C, Frederiksen KS, Liu Y, et al. The EADC-ADNI Harmonized Protocol for manual hippocampal segmentation on magnetic resonance: Evidence of validity. *Alzheimer's & Dementia*. 2015;11:111-25.
- [18] Apostolova LG, Zarow C, Biado K, Hurtz S, Boccardi M, Somme J, et al. Relationship between hippocampal atrophy and neuropathology markers: A 7T MRI validation study of the EADC-ADNI Harmonized Hippocampal Segmentation Protocol. *Alzheimer's & Dementia*. 2015;11:139-50.
- [19] Boccardi M, Bocchetta M, Morency FC, Collins DL, Nishikawa M, Ganzola R, et al. Training labels for hippocampal segmentation based on the EADC-ADNI harmonized hippocampal protocol. *Alzheimer's & Dementia*. 2015;11:175-83.

- [20] Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage*. 2007;38:95-113.
- [21] Meng X-L, Rosenthal R, Rubin DB. Comparing correlated correlation coefficients. *Psychological bulletin*. 1992;111:172.
- [22] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988:837-45.
- [23] Duchesne S, Valdivia F, Robitaille N, Mouiha A, Valdivia FA, Bocchetta M, et al. Manual segmentation qualification platform for the EADC-ADNI harmonized protocol for hippocampal segmentation project. *Alzheimer's & Dementia*. 2015;11:161-74.
- [24] Carmichael OT, Aizenstein HA, Davis SW, Becker JT, Thompson PM, Meltzer CC, et al. Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2005;27:979-90.
- [25] Suppa P, Anker U, Spies L, Bopp I, Rügger-Frey B, Klaghofer R, et al. Fully automated atlas-based hippocampal volumetry for detection of Alzheimer's disease in a memory clinic setting. *Journal of Alzheimer's Disease*. 2015;44:183-93.
- [26] Ashburner J. Computational anatomy with the SPM software. *Magn Reson Imaging*. 2009;27:1163-74.
- [27] Suppa P, Hampel H, Spies L, Fiebach JB, Dubois B, Buchert R. Fully Automated Atlas-Based Hippocampus Volumetry for Clinical Routine: Validation in Subjects with Mild Cognitive Impairment from the ADNI Cohort. *Journal of Alzheimer's Disease*. 2015;46:199-209.
- [28] Mak HK, Zhang Z, Yau KK, Zhang L, Chan Q, Chu LW. Efficacy of voxel-based morphometry with DARTEL and standard registration as imaging biomarkers in Alzheimer's disease patients and cognitively normal older adults at 3.0 Tesla MR imaging. *J Alzheimers Dis*. 2011;23:655-64.
- [29] Firbank MJ, Barber R, Burton EJ, O'Brien JT. Validation of a fully automated hippocampal segmentation method on patients with dementia. *Hum Brain Mapp*. 2008;29:1442-9.
- [30] Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: A survey. *Med Image Anal*. 2015;24:205-19.
- [31] Pardoe HR, Pell GS, Abbott DF, Jackson GD. Hippocampal volume assessment in temporal lobe epilepsy: How good is automated segmentation? *Epilepsia*. 2009;50:2586-92.
- [32] Morey RA, Petty CM, Xu Y, Hayes JP, Wagner HR, Lewis DV, et al. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage*. 2009;45:855-66.
- [33] Shen L, Saykin AJ, Kim S, Firpi HA, West JD, Risacher SL, et al. Comparison of manual and automated determination of hippocampal volumes in MCI and early AD. *Brain Imaging Behav*. 2010;4:86-95.
- [34] Sanchez-Benavides G, Gomez-Anson B, Sainz A, Vives Y, Delfino M, Pena-Casanova J. Manual validation of FreeSurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer Disease subjects. *Psychiatry Res*. 2010;181:219-25.
- [35] Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage*. 2009;46:786-802.
- [36] Bergouignan L, Chupin M, Czechowska Y, Kinkingnehun S, Lemogne C, Le Bastard G, et al. Can voxel based morphometry, manual segmentation and automated segmentation equally detect hippocampal volume differences in acute depression? *Neuroimage*. 2009;45:29-37.
- [37] Barnes J, Whitwell JL, Frost C, Josephs KA, Rossor M, Fox NC. Measurements of the amygdala and hippocampus in pathologically confirmed Alzheimer disease and frontotemporal lobar degeneration. *Archives of neurology*. 2006;63:1434-9.
- [38] Du A, Schuff N, Amend D, Laakso M, Hsu Y, Jagust W, et al. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*. 2001;71:441-7.
- [39] Colliot O, Chételat G, Chupin M, Desgranges B, Magnin B, Benali H, et al. Discrimination between Alzheimer disease, Mild Cognitive Impairment, and Normal Aging by Using Automated Segmentation of the Hippocampus 1. *Radiology*. 2008;248:194-201.

- [40] Chételat G, Fouquet M, Kalpouzos G, Denghien I, De La Sayette V, Viader F, et al. Three-dimensional surface mapping of hippocampal atrophy progression from MCI to AD and over normal aging as assessed using voxel-based morphometry. *Neuropsychologia*. 2008;46:1721-31.
- [41] Wolf D, Fischer FU, de Flores R, Chételat G, Fellgiebel A. Differential associations of age with volume and microstructure of hippocampal subfields in healthy older adults. *Human brain mapping*. 2015;36:3819-31.
- [42] Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*. 2009;46:726-38.
- [43] Nestor SM, Gibson E, Gao F-Q, Kiss A, Black SE, Initiative AsDN. A direct morphometric comparison of five labeling protocols for multi-atlas driven automatic segmentation of the hippocampus in Alzheimer's disease. *Neuroimage*. 2013;66:50-70.
- [44] Leung KK, Barnes J, Ridgway GR, Bartlett JW, Clarkson MJ, Macdonald K, et al. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. *Neuroimage*. 2010;51:1345-59.
- [45] Collins DL, Pruessner JC. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *Neuroimage*. 2010;52:1355-66.