

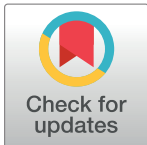
RESEARCH ARTICLE

# Regional analysis of volumes and reproducibilities of automatic and manual hippocampal segmentations

Fabian Bartel<sup>1\*</sup>, Hugo Vrenken<sup>1,2</sup>, Fetsje Bijma<sup>3</sup>, Frederik Barkhof<sup>2,4</sup>, Marcel van Herk<sup>5</sup>, Jan C. de Munck<sup>1</sup>

**1** Department of Physics and Medical Technology, VU University Medical Center, Amsterdam, The Netherlands, **2** Department of Radiology, VU University Medical Center, Amsterdam, The Netherlands, **3** Department of Mathematics, VU University Amsterdam, Amsterdam, The Netherlands, **4** Image Analysis Center, VU University Medical Center, Amsterdam, The Netherlands, **5** Department of Radiotherapy Physics, University of Manchester, Manchester, United Kingdom

\* [f.bartel@vumc.nl](mailto:f.bartel@vumc.nl)



**OPEN ACCESS**

**Citation:** Bartel F, Vrenken H, Bijma F, Barkhof F, van Herk M, de Munck JC (2017) Regional analysis of volumes and reproducibilities of automatic and manual hippocampal segmentations. PLoS ONE 12 (2): e0166785. doi:10.1371/journal.pone.0166785

**Editor:** Vince Grolmusz, Mathematical Institute, HUNGARY

**Received:** June 29, 2016

**Accepted:** November 3, 2016

**Published:** February 9, 2017

**Copyright:** © 2017 Bartel et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The volumes calculated from manual, FreeSurfer and FSL-FIRST segmentations for each hippocampus and for each scan, as well as the whole-hippocampus and regional Jaccard overlap indices for every hippocampus are provided (supporting information files). MRI 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)). Since we do not own the ADNI data used in this study, we do not have permission to redistribute these data ourselves, as is stated in the data use agreement from ADNI (<http://adni.loni>).

## Abstract

### Purpose

Precise and reproducible hippocampus outlining is important to quantify hippocampal atrophy caused by neurodegenerative diseases and to spare the hippocampus in whole brain radiation therapy when performing prophylactic cranial irradiation or treating brain metastases. This study aimed to quantify systematic differences between methods by comparing regional volume and outline reproducibility of manual, FSL-FIRST and FreeSurfer hippocampus segmentations.

### Materials and methods

This study used a dataset from ADNI (Alzheimer's Disease Neuroimaging Initiative), including 20 healthy controls, 40 patients with mild cognitive impairment (MCI), and 20 patients with Alzheimer's disease (AD). For each subject back-to-back (BTB) T1-weighted 3D MPRAGE images were acquired at time-point baseline (BL) and 12 months later (M12). Hippocampi segmentations of all methods were converted into triangulated meshes, regional volumes were extracted and regional Jaccard indices were computed between the hippocampi meshes of paired BTB scans to evaluate reproducibility. Regional volumes and Jaccard indices were modelled as a function of group (G), method (M), hemisphere (H), time-point (T), region (R) and interactions.

### Results

For the volume data the model selection procedure yielded the following significant main effects G, M, H, T and R and interaction effects G-R and M-R. The same model was found for the BTB scans. For all methods volumes reduces with the severity of disease.

Significant fixed effects for the regional Jaccard index data were M, R and the interaction M-R. For all methods the middle region was most reproducible, independent of diagnostic group. FSL-FIRST was most and FreeSurfer least reproducible.

[usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Data\\_Use\\_Agreement.pdf](http://usc.edu/wp-content/uploads/how_to_apply/ADNI_Data_Use_Agreement.pdf)). However, the data can be obtained through procedures and under conditions as described on the ADNI websites (<http://adni.loni.usc.edu/about/committees/> and <http://adni.loni.usc.edu/data-samples/access-data/>). Permission to use manual hippocampus segmentations from this study will not be granted, because these are the property of the Image Analysis Center, VU University Medical Center, and they did not agree to release these data. Permission for these data may be granted by contacting Anne Verhagen ([a.verhagen@vumc.nl](mailto:a.verhagen@vumc.nl)) at the VU University Medical Centrum. All other relevant data are within the paper and its Supporting Information files.

**Funding:** 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.; Cogstate; 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 ([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 Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. 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>.

## Discussion/Conclusion

A novel method to perform detailed analysis of subtle differences in hippocampus segmentation is proposed. The method showed that hippocampal segmentation reproducibility was best for FSL-FIRST and worst for FreeSurfer. We also found systematic regional differences in hippocampal segmentation between different methods reinforcing the need of adopting harmonized protocols.

## Introduction

The hippocampus is an important brain structure that plays a crucial role in episodic memory [1]. For instance, longitudinal decline of hippocampal volume is related to memory impairment and clinical dementia [2,3]. In Alzheimer's disease (AD) and its prodromal phase, mild cognitive impairment (MCI), the hippocampus is affected by amyloid and tau pathology early in the disease course [4,5]. Hippocampal atrophy as measured on T1-weighted volumetric structural magnetic resonance images (MRI) is a sensitive biomarker of AD pathology [6], but can also be a predictive imaging biomarker of MCI [7]. Knowledge of hippocampal shape is also an important aspect in radiotherapy, when prophylactic cranial irradiation (PCI) is used and hippocampal avoidance is executed to limit neurocognitive toxicity [8–12].

Although manual outlining by experts is considered as the gold standard, it requires extensive training and is very labour intensive [13]. Therefore, automatic segmentation tools based on deformable models, single-, multiple- or probabilistic-atlases have been developed over the last decades. V. Dill and colleagues give an excellent overview of semi-automatic and automatic hippocampus segmentation methods [14]. The most commonly used publicly available software tools to the academic community, with active user communities and active support from the developers, are FreeSurfer [Martinos Center for Biomedical Imaging, Harvard-MIT, Boston USA] [15,16] and FSL-FIRST [FMRIB Integrated Registration and Segmentation Tool, University of Oxford, Oxford UK] [17] and therefore we focus on these methods. Previous studies have shown good but not perfect overall agreement for both methods with manual segmentation, given a dice overlap of FreeSurfer and FSL-FIRST segmentation ranging from 74–82% and 79–84% respectively and a good volume correlation of both methods with manual segmentation [16–28]. In a direct comparison, FreeSurfer slightly agreed better with manual segmentation than FSL-FIRST [29–33].

So far, most studies comparing manual and automatic hippocampus segmentations have expressed the performance of hippocampus outline methods in terms of global hippocampal volumes and overlap indices to manual hippocampus segmentation. For instance, Mulder and colleagues compared reproducibility of longitudinal hippocampal volume changes, as determined by manual segmentations, FSL-FIRST and FreeSurfer [33]. However, volumes and volume changes do not contain information about shape and overlap indices only quantify the total amount of agreement of two segmentation methods. It is very likely that some parts of the hippocampal structure are easier to segment than others and therefore to study systematic differences existing global volume and overlap measures need to be extended to regional ones. Following Hackert and colleagues we focus on regional differences along the long axis of hippocampi, computing regional volumes and outline reproducibilities by dividing the hippocampus in three regions, the head, body and tail [34]. Furthermore, different automatic segmentation methods and manual segmentation protocols might be based on different

[edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](https://doi.org/10.1371/journal.pone.0166785.g001). This research was also supported by NIH grants P30 AG010129 and K01 AG030514. HV has received research grants from Novartis, Teva, MerckSerono and Pfizer, and a speaker honorarium from Novartis, but these funders did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. All funds were paid directly to his institution.

**Competing interests:** HV has received research grants from Novartis, Teva, MerckSerono and Pfizer, commercial companies, for this study. There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials.

underlying anatomical definitions. A systematic regional comparison can reveal such differences between methods.

There are a few cross-sectional hippocampus studies using FreeSurfer segmentation which reported that sub-regions undergo differential atrophy in AD [35][36]. These findings further motivate our objective to evaluate regional longitudinal changes in hippocampal volume as determined by different segmentation methods.

To our knowledge, there are no papers reporting reproducibility of hippocampal outlines in a dataset similar to clinical trials. In part the absence of such studies derives from the fact that comparing voxel-wise segmentations obtained from different scans is challenging, because of slightly different positions of the head in the voxel space. Considering these small regional differences between different segmentations, we wish to avoid interpolation errors as much as possible. For that purpose, in this study a surface reconstruction of each hippocampus is derived from the scan to which the labelled segmentation was available in its rawest form. Then, after determining the accurate image registration and applying the corresponding transformation parameters between the reconstructed surfaces overlap measures were computed directly on the surfaces, avoiding interpolation errors as much as possible. Since the limiting factor of these computations is accuracy of the image registration we apply the “full circle method” to test the quality of registration procedures [37].

It remains unclear to what extent the hippocampal segmentations themselves are reproducible at the most detailed level. Although accuracy of hippocampal segmentations has been investigated by comparing to manual references [17,20–22,27,29], reproducibility of the segmentations has not been investigated on a large population and different groups. Similar to Mulder and colleagues we investigate hippocampus segmentation for different disease groups in different stages and use different segmentation methods [33]. But different to [33], we compare hippocampal volumes and outline reproducibilities in different regions and hemispheres as determined in baseline and follow-up scans. Because of the many factors and possible combination of factors that may influence the response variables, we propose a novel method, based on Akaike Information Criterion (AIC) [38], to select the most suitable statistical model to explain our findings. We test the robustness of this method by performing the same analysis making use of the back-to-back (BTB) scans.

## Materials and methods

### Dataset and MRI acquisition

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](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 magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD).

The dataset used in this study is the same subset of the ADNI dataset that has been used by Mulder and colleagues [33]. MRI data of 80 subjects were selected, of which 20 are control subjects (CTRL), 40 MCI subjects and 20 subjects were diagnosed as AD. MCI subjects were a priori selected based on their cerebrospinal fluid (CSF) profile. For the selection we used the ratio of total tau (t-tau) and Amyloid- $\beta$  1 to 42 peptide ( $A\beta_{1-42}$ ) with an AD-positive cut-off value of  $t\text{-tau}/A\beta_{1-42} \geq 0.39$  determined by Shaw and colleagues [39]. 20 MCI subjects with an AD-positive cut-off value (MCI-P;  $t\text{-tau}/A\beta_{1-42} \geq 0.39$ ) and 20 MCI subjects with an AD-negative

cut-off value ( $MCI-N; t\text{-tau}/A\beta_{1-42} < 0.39$ ) were selected from the database. All healthy controls had a  $t\text{-tau}/A\beta_{1-42} < 0.39$  and all AD's a  $t\text{-tau}/A\beta_{1-42} \geq 0.39$ .

For all subjects four volumetric MRI scans were acquired, two scans at time-point baseline (BL) and two scans one year later, here referred to as M12. Those two MRI BTB scans at each time-point were acquired in a single session, with the acquisition of the second volumetric MRI starting only a few minutes after completing the first acquisition. We refer to these scans as BL-A, BL-B, M12-A, and M12-B. BL scans of all subjects were made between September 2005 and August 2007.

MRI scans were acquired at different locations with 1.5T scanners from various vendors (Philips, Siemens and GE). For every subject the MRI scanner and protocols were the same for each of the four acquisitions. The images were acquired with a 3D T1 weighted magnetization prepared rapid acquisition gradient echo sequence (MPRAGE). All pixels were square and the slice thickness was 1.2mm. The voxel volume ranged from  $1.05\text{mm}^3$  to  $2.03\text{mm}^3$  with a median value of  $1.88\text{mm}^3$ . The MRI scans were visually inspected for their quality and no post-processing other than default scanner corrections were performed. A more detailed description for the MRI acquisition protocol can be found in Jack et al [40].

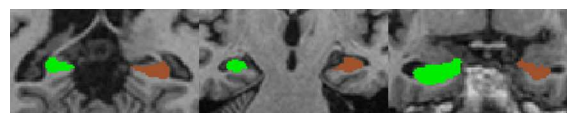
## Hippocampus segmentation

**Manual.** Manual hippocampus segmentations were performed in the Image Analysis Center (IAC, Amsterdam) using their standard operating procedure (SOP) as previously described in [33,41,42]. BL scans were reformatted in a plane perpendicular to the long axis of the left hippocampus, resulting in a pseudo coronal orientation with a slice thickness of 2mm and the original in-plane resolution using sinc interpolation. This procedure was followed independently for all four scans. Rigid body registration was applied to all four (both BL and both M12) reformatted scans to bring them in the same coordinate space for comparison. Three slices of a hippocampus segmentation in pseudo coronal orientation are shown in Fig 1.

Included in the hippocampal formation are the Ammon's horn, dentate gyrus, alveus and fimbria and the subiculum. To summarize hippocampal boundaries, the most posterior slice is chosen such the total length of the crux of the fornix is seen. The medial boundary of the hippocampus is formed by the CSF in the cisterna ambiens and the transverse fissure. The inferior border is formed by the subiculum and the parahippocampal gyrus. The superior border is defined by the CSF of the temporal horn and the alveus. Laterally, the hippocampus is bordered by CSF from the temporal pole of the lateral ventricle. In anterior direction it forms along the amygdala and stops when an additional amount of CSF appears on the medial side of the hippocampus.

One trained expert technician from the IAC segmented the left and right hippocampus of all subjects using a locally developed software package (Show\_Images 3.7.1.0) from the VU University Medical Center (VUmc). The technician was blinded to the diagnosis, but used BL segmentations to segment the follow up M12 scans, as it is part of the workflow of the longitudinal study. However, first and second BTB scans were given in a random order.

**FSL-FIRST.** In [43] and [17] technical details of FSL-FIRST are described. FSL-FIRST is a deformable model based segmentation tool, using shape and appearance models which were constructed from a set of manual segmented subjects provided by the Center for Morphometric



**Fig 1. Hippocampus segmentation in reformatted pseudo coronal orientation.** Brown colour is the left, green the right hippocampus. Left: posterior slice close to the crux of the fornix. Middle: one of the middle slices of the hippocampus. Right: anterior slice with hippocampus next to the amygdala.

doi:10.1371/journal.pone.0166785.g001

Analysis (CMA), Massachusetts General Hospital (MGH), Boston. The manual segmentations were parameterized and described as surface meshes from which a point distribution is modelled. Using observed intensity values from the MR image, FSL-FIRST finds the most probable shape by searching through linear combinations of shape variation modes. FSL-FIRST uses a two-stage affine transformation to a MNI152 standard space of 1mm resolution before performing segmentation. Hippocampus meshes are then converted to labelled voxel region of interests (ROI) after a boundary correction using FAST voxel-wise segmentation software [44]. We used FSL-FIRST v.5.0.4 and the *run\_first\_all* script command, because FSL-FIRST takes adjacent structures into account. The voxelwise labelled hippocampus segmentation produced by FSL-FIRST are in native MRI scan space.

For one subject the FSL-FIRST segmentation failed because of an internal registration problem. To include this subject, we pre-processed it by extracting the subjects brain using BET before running the FSL-FIRST script. The BET extraction corrected the registration problem and enabled us to include this subject.

**FreeSurfer.** In [16] the technical procedure for subcortical segmentation is described in detail. Briefly, FreeSurfer brings the MRI to a conformed 1mm<sup>3</sup> 256<sup>3</sup> space, performs intensity normalization to correct for intensity non-uniformity in the MR image, saves an affine transformation to Talairach space, corrects intensity fluctuations using another normalization and strips the skull leaving only the brain. To apply segmentation labels FreeSurfer transforms the subject's volume to the FreeSurfer atlas and assigns voxels to subcortical structures using prior probabilistic intensity and tissue class information.

We used the FreeSurfer version 5.3 to perform hippocampus segmentations using the longitudinal processing stream. This requires a prior cross-sectional processing of each MRI. FreeSurfer's labelled hippocampus segmentations from the cross-sectional and longitudinal stream were converted back to the native MR image space using the procedure provided by FreeSurfer (*mri\_label2vol*).

**Surface extraction.** All volumetric hippocampi labels from each method were converted to triangulated meshes with the marching cube algorithm to avoid interpolation errors introduced by registrations. Those generated hippocampi meshes were used to compute regional volumes and outline reproducibilities. If the segmentations consisted of multiple connected components the surface reconstruction would also consist of multiple surfaces of which the total volume was taken to correspond to the hippocampus.

## Comparison methods

The marching cubes algorithm applied to the segmented images resulted in closed triangulated surfaces. Regional volumes from surfaces were computed by adopting a fine regular grid enclosing two surfaces *A* and *B*, and by testing for each point whether it was inside either of the surfaces. To speed up these computations, KD trees and some other optimizations were used [45]. The Jaccard index of the surface pair (*A*, *B*), defined as

$$Jacc(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

was approximated as

$$Jacc(A, B) \approx \frac{N(A \cap B)}{N(A \cup B)} \tag{2}$$

where *N(V)* is the number of grid points inside surface *V*. These grid points were derived from a submillimetre mesh that was fine enough to capture all surface details.

To quantify regional specific reproducibility and systematic differences in shape definition, a regional overlap index was computed as follows:

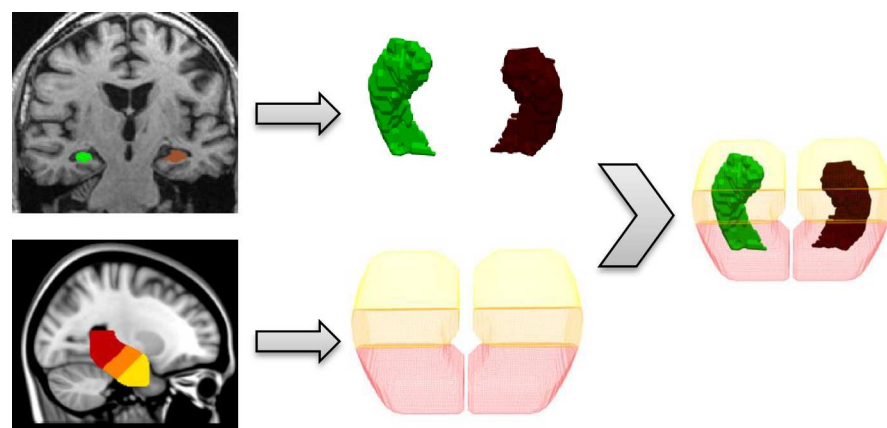
$$Jacc_{ROI} = \frac{|(A \cap B) \cap ROI|}{|(A \cap ROI) \cup (B \cap ROI)|} \quad (3)$$

where ROI represents a region of interest. This equation is an overlap between surfaces  $A$  and  $B$ , both constrained to a third region ROI. To compute regional volumes and Jaccard indices in practice, a hippocampus mask was derived from MNI152 standard-space provided by FSL (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases>, MNI152\_T1\_1mm\_Hipp\_mask\_dil8.nii). This mask was big enough to cover any hippocampus and was split into three parts for each hemisphere along the long hippocampal axis and converted to triangulated meshes resulting into six mesh regions, hereafter named left and right anterior, middle and posterior. The regions have no specific anatomical definition, but they are similar to Hackert and colleagues' regional definition and approximate to an anterior region of 35%, middle region of 45% and posterior region of 20% [34]. To register this six regional hippocampus mask in MNI152 space to each subject image space, we performed a similar procedure as FSL-FIRST, i.e. brain extraction, a two-stage affine registration to MNI152, followed by visual inspection. Fig 2. is a flowchart illustrating the hippocampal mesh conversion and the registration procedure of the six regional mask to the hippocampus mesh. All other triangulated hippocampi segmentation meshes (BL-B, M12-A and M12-B) were rigid body registered to scan BL-A with the registration matrices described in *Registrations and registration quality control*.

### Pre-processing

Before regional volumes and reproducibilities could be computed, MRI scans were mapped to each other so that the segmentations were in the same imaging space. Although BTB scans are very similar to the original, there is still the possibility of subject motion in between the BTB scans, and therefore image registration was also applied between these image pairs.

**Registrations and registration quality control.** Rigid body transformations were used to map BL-B to BL-A, M12-B to M12-A and M12-A scan to BL-A scan. Our registrations were all performed using FSL-FLIRT. To check the quality of these registrations, a consistency test was done on the registration parameters using the full circle method introduced by van Herk and



**Fig 2. Procedure to make a regional analysis.** Top and bottom rows show the conversion from a hippocampus segmentation and the six regional mask to a triangulated mesh respectively. The right part of the figure illustrates the registration procedure to map the six regional hippocampus mask to the left and right hippocampus mesh.

doi:10.1371/journal.pone.0166785.g002

colleagues [37]. By registering images in a cyclic fashion and multiplying all transformation matrices the product should result in the identity matrix, when registration errors were absent. Hence, we computed the “full circle” matrix  $RM = \prod T_{ij}$ , where  $T_{ij}$  is the transformation from image  $i$  to image  $j$ . We analysed four full circles which resulted into residual matrices given by:

$$RM_{Circle1} = T_{BLA-M12B} \times T_{M12B-M12A} \times T_{M12A-BLA} \tag{4}$$

$$RM_{Circle2} = T_{BLA-M12A} \times T_{M12A-BLB} \times T_{BLB-BLA}$$

$$RM_{Circle3} = T_{BLA-M12B} \times T_{M12B-BLB} \times T_{BLB-BLA}$$

$$RM_{Circle4} = T_{M12B-M12A} \times T_{M12A-BLB} \times T_{BLB-M12B}$$

and determined the residual translation and rotation errors as:

$$Translation_{Total} = \|TransVector(RM)\| \tag{5}$$

$$Rotation_{Total} = aCos\left(\frac{Trace(RM) - 1}{2}\right) \tag{6}$$

In addition, the effect of registration errors was directly quantified by computing the Jaccard index between a hippocampal surface and its transformed version obtained by applying RM:

$$Consistency = Jacc(S_{hip}, RM(S_{hip})) \tag{7}$$

The more consistent all registrations, the closer the matrix RM is to the identity, and the higher this Consistency index. Therefore, we use 1- Consistency to quantify the registration error.

**Visual quality control.** Next to the full circle analysis as an additional quality check, we inspected the results of the outline reproducibility analysis and visually reviewed subjects’ registered scan pairs which had low Jaccard indices to be sure that there were no registration errors.

### Statistical analysis

We used linear mixed models for the statistical analysis of the data. The analysis of the regional volume data was performed with the volumes as response variable ( $V$ ). The models consisted of fixed main effects and fixed interaction effects which we selected due to their suspected influence on hippocampal volume and shape. Fixed main effects were segmentation method ( $M$ ) with levels (Manual, FSL-FIRST, FreeSurfer), Group ( $G$ ) with levels (CTRL, MCIN, MCIP, AD), hemisphere ( $H$ ) with levels (Left, Right), region ( $R$ ) with levels (Anterior, Middle, Posterior) and time-point ( $T$ ) with levels (BL, M12). A complete model would include all combinations of pairs, triples, etc. of these effects. To reduce the model complexity we started our search for a physiologically reasonable descriptive model by only considering the following interactions: group-method ( $G-M$ ), group-region ( $G-R$ ), method-region ( $M-R$ ), method-hemisphere ( $M-H$ ), time-point-group ( $T-G$ ), time-point-region ( $T-R$ ), time-point-group-region ( $T-G-R$ ) and group-region-method ( $G-R-M$ ). Individual subject effects ( $S$ ) were modelled as random effects. This yielded the mixed model:

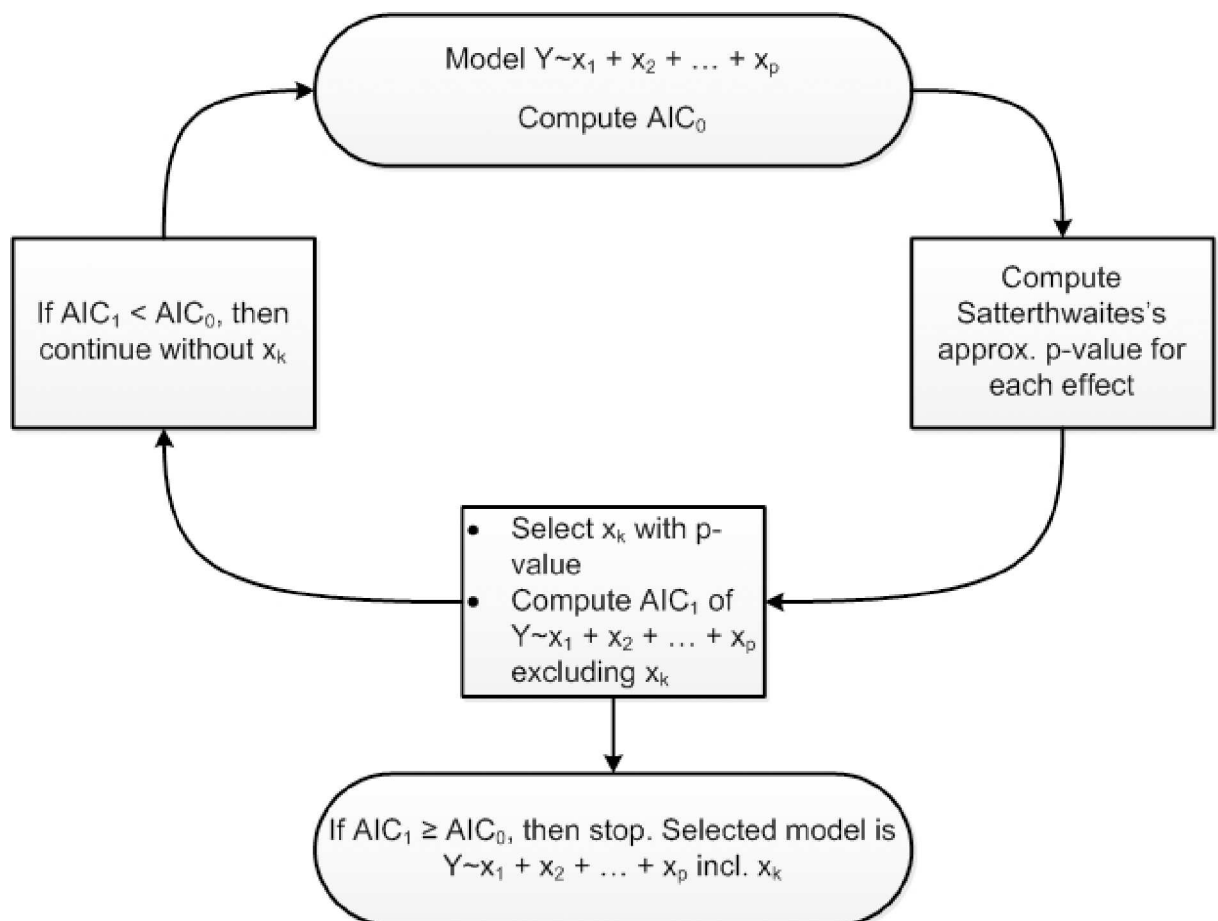
$$V \sim M + G + H + R + T + GM + GR + MR + MH + TG + TR + TGR + GRM + r(S) \tag{8}$$

where  $r()$  indicates a random effect. This model was fitted to the pair of longitudinal A scans and the pair of B scans separately. Then a model selection algorithm was run that selected significant effects amongst fixed effects present in the model. This was done by minimizing the Akaike Information Criterion (AIC) in a backward elimination set up, i.e. least significant terms were dropped from the model until the AIC started to increase. The AIC is a commonly used statistical measure that balances the goodness of fit and model complexity (i.e. number of free parameters). Significance of each term was computed according to an ANOVA analysis with Satterthwaite's approximation for degrees of freedom using R-package lmerTest [46]. The model selection is illustrated with a flowchart in Fig 3.

For the analysis of whole hippocampus outline reproducibilities we transformed the Jaccard index in (2) to  $Jacc^o$  as response variable ( $J$ ) in order to fulfil the assumption of Gaussian errors in the linear model. Fixed main effects were segmentation method ( $M$ ), group ( $G$ ), hemisphere ( $H$ ), time-point ( $T$ ). Fixed interaction effects fitted were group-method ( $G-M$ ) and method-hemisphere ( $M-H$ ). Individual Subject effects ( $S$ ) were modelled as random effects. In all this yielded the mixed model:

$$J \sim G + M + H + T + GM + MH + r(S) \tag{9}$$

The regional hippocampus outline reproducibilities were analysed in a similar way. The Jaccard index in (3) was again transformed to  $Jacc^o$  as response variable in order to meet the



**Fig 3. Flowchart of the model selection procedure.**

doi:10.1371/journal.pone.0166785.g003



Gaussian assumption. Compared to the whole hippocampus analysis we added the fixed main effect Region ( $R$ ) and interaction effects R-M and R-G:

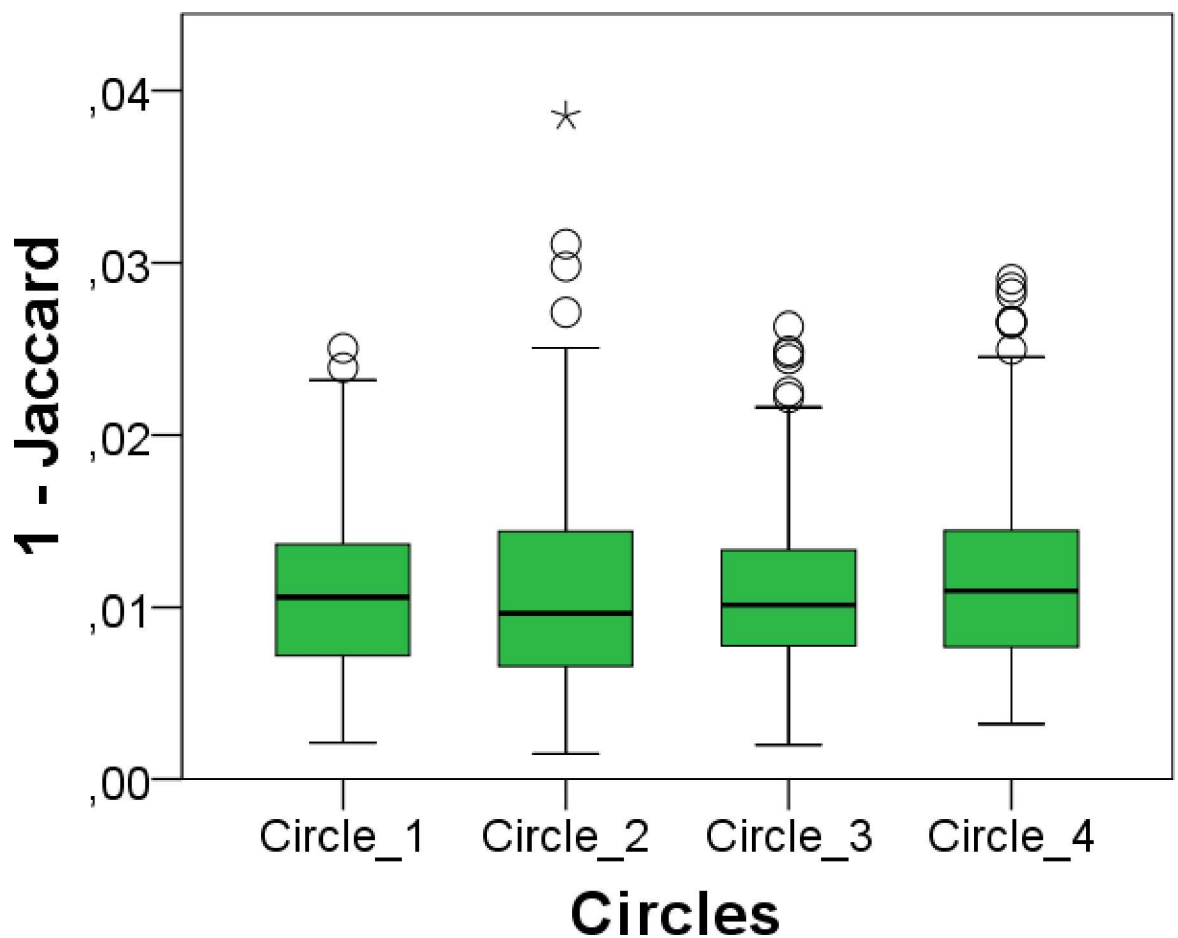
$$J_{ROI} \sim G + M + H + T + R + GM + MH + RM + RG + r(S) \tag{10}$$

The model selection for (9) and (10) was performed using the same algorithm as used for the volume data analysis. For the volume data analysis FreeSurfer's segmentations from the longitudinal stream have been used, but for comparison we also analysed segmentations from the cross-sectional stream. The reproducibility analysis was only performed with FreeSurfer's segmentations from the cross-sectional stream.

## Results

### Registration quality control

Quality of the registrations for all subjects was analysed using the full circle method to evaluate the transitivity error. Taking all subjects into account, for the full circles of the primary analysis, described by equations in (4) the maximum total rotation and translation calculated were 0.12deg and 0.4mm respectively. The mean translation and rotation were 0.01 mm and 0.04 degrees, which is the result of three registration steps, so that each registration will be more



**Fig 4. Quality control for MRI scan registration.** Results obtained from the residual matrix after the full circle approach. Jaccard error computed for circles defined by Eq 7.

doi:10.1371/journal.pone.0166785.g004

accurate than this. In Fig 4, the registration error is plotted in boxplots showing the error for each circle on the basis of Eq (7). In general, all values are quite small, demonstrating the consistency and accuracy of registrations. Additionally, registrations of outliers shown in Fig 4, were reviewed visually and showed no noticeable registration errors, which indicated together with rotation and translation results that all registrations were of good quality.

## Regional hippocampus volume comparison

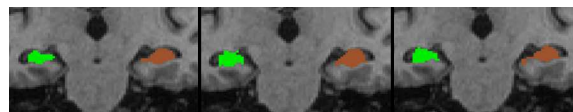
For the regional analysis segmented hippocampi of all segmentation methods, shown in Fig 5, have been processed as described in chapter 2.3. Regional volumes have been extracted and used for our statistical analysis.

The linear mixed models fitted on the BL-A and M12-A scans on the one hand and those fitted on the BL-B and M12-B scans on the other hand yielded identical selections of fixed effects. That means that in both cases the model selection procedure reduced the model of Eq (8) to the following:

$$V \sim G + M + H + T + R + GR + MR + r(S) \quad (11)$$

We then performed the model selection on all scans BL-A, M12-A, BL-B and M12-B together, and obtained again the same selection of fixed effects. In the sequel, parameter estimates from the combined data set will be mentioned. All fixed main effects and fixed interaction effects in (11) were significant, with the highest p-value (Satterthwaite's approximation) in the selected model of 0.0001082 (main effect Group ( $G$ )), all other p-values were lower. The dropped fixed main effect and fixed interaction effect were insignificant and had a higher Satterthwaite's approximation p-value than 0.05. For the factors hemisphere ( $H$ ) and time-point ( $T$ ) only the main effects are present in the final model and the interaction effects of these dropped. The left hippocampus was on average  $0.0332\text{cm}^3$  smaller than the right hippocampus. Hippocampi from time-point M12 were on average  $0.0326\text{cm}^3$  smaller than from time-point BL. Predictions of the estimated model for the three segmentation methods are shown in Table 1 for the left hemisphere and time-point BL.

Using the average volume difference between left and right ( $0.0332\text{cm}^3$ ) or between BL and M12 ( $0.0326\text{cm}^3$ ) hippocampi, all other predicted volumes can be reconstructed by adding these values to the predicted volumes in Table 1. For example, to obtain the predicted volume from the FSL-FIRST segmentations in the MCIP group of the middle region for the right hippocampus,  $0.0332\text{cm}^3$  need to be added to  $1.207\text{cm}^3$ . Tables for right hippocampus at time-point BL and left and right hippocampus at time-point M12 can be found in the supporting information (Table in S1 Table, Table in S2 Table and Table in S3 Table). The decrease of volume from BL to M12 could be predicted by all methods, but could not be differentiated between different group types. In general, the middle part for both automatic segmentation methods was the largest part, while for manual segmentations the anterior and middle parts seem to be of almost equal size. Moreover, the anterior volume of manual segmentations was systematically bigger than the anterior volume of the automatic segmentations. Also noticeable is that for all three methods the posterior part was predicted to be the smallest part, which is



**Fig 5. Left and right hippocampus segmentations in coronal view.** Left: manual segmentation. Middle: FSL-FIRST segmentation. Right: FreeSurfer segmentation.

doi:10.1371/journal.pone.0166785.g005

**Table 1. Predicted volumes (cm<sup>3</sup>) for the left hippocampus at time-point BL for all segmentation methods.**

Group	CTRL	MCIN	MCIP	AD
<b>Region</b>	<b>Manual Segmentation</b>			
Anterior	1.317	1.171	1.088	1.066
Middle	1.282	1.252	1.120	1.006
Posterior	0.790	0.747	0.731	0.618
	<b>FSL-FIRST Segmentation</b>			
Anterior	1.259	1.113	1.031	1.008
Middle	1.369	1.340	1.207	1.093
Posterior	0.964	0.920	0.905	0.792
	<b>FreeSurfer Segmentation</b>			
Anterior	1.186	1.040	0.957	0.935
Middle	1.324	1.294	1.162	1.048
Posterior	1.000	0.957	0.942	0.828

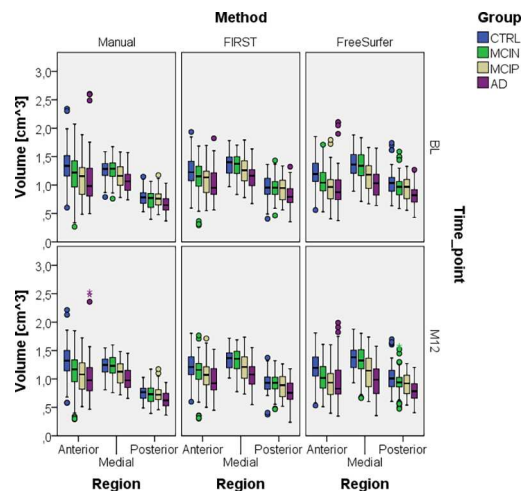
doi:10.1371/journal.pone.0166785.t001

the result of our definition of the ROIs within the mask. Furthermore, the predicted volumes from Table 1 shows that for all methods all three regions showed a decrease in hippocampal volume for increasing severity of disease. Fig 6 illustrates regional hippocampal volume differences for all three methods and regions discriminated in groups and by both time-points, while left and right hippocampi were grouped together.

Following the same procedure, using FreeSurfer’s segmentation from the cross-sectional stream resulted in the same model with very similar predicted volumes, which can be found in the supporting information (Table in S4 Table).

### Whole hippocampus outline reproducibility

The fitted and selected linear mixed model for the hippocampus outline reproducibility only contains the fixed effect method (M), with p-value < 2.2x10<sup>-16</sup>. The predicted Jaccard indices for the three segmentation methods are shown in Table 2. This table shows that FSL-FIRST segmentation is the most and FreeSurfer segmentation the least reproducible.



**Fig 6. Regional volume comparison for all methods and both time-points.** Left and right hippocampus and scan A and B were grouped together.

doi:10.1371/journal.pone.0166785.g006

**Table 2. Predicted Jaccard indices for the whole hippocampus for the different segmentation methods.**

Method	Jaccard
Manual	0.795
FSL-FIRST	0.829
FreeSurfer	0.754

doi:10.1371/journal.pone.0166785.t002

Fig 7 illustrates Jaccard indices of outline reproducibility for all three methods for BL and M12 scan pairs, separated by left and right hippocampus and differentiated into groups. The box-plots show the same tendency as predicted by the mixed model. Even though it was not significant, the boxplots also show a trend that for all methods Jaccard indices decrease with increasing disease severity, and both automatic segmentations show larger variations than manual segmentations. Also, it should be noted that only the automatic segmentations have large outliers.

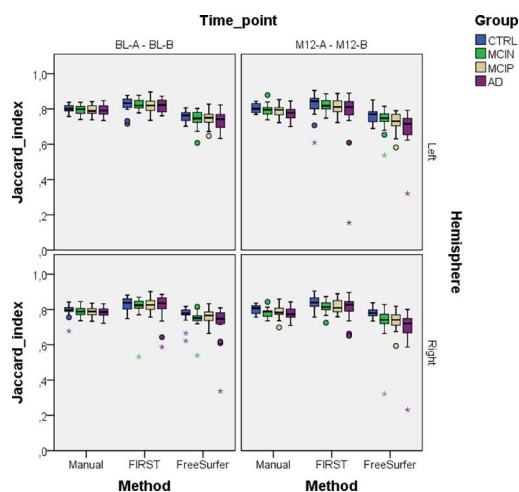
### Regional hippocampus outline reproducibility

Regional hippocampus Jaccard indices have been computed by using Eq (3). The fitted linear mixed models contain as fixed effects the main effects method and region and interaction effect region-method, resulting into the model:

$$J_{ROI} \sim M + R + RM + r(S) \tag{12}$$

The p-values of all three fixed effects in the selected model were similar to the p-values of the analysis of the whole hippocampus outline reproducibility. The predicted Jaccard indices for all method-region combinations are shown in Table 3.

The results related to the segmentation method are similar to that in the whole hippocampus analysis: FSL-FIRST segmentation is most and FreeSurfer segmentation least reproducible. It can also be seen in Fig 8 that for all methods with the severity of the disease in all regions the reproducibility decreased. Additionally, Table 3 shows that the middle region has highest Jaccard indices and the posterior region lowest Jaccard indices.



**Fig 7. Whole hippocampus Jaccard indices for all methods, both time-points and left and right hippocampi to show segmentation reproducibility between BTB scans.**

doi:10.1371/journal.pone.0166785.g007

**Table 3. Predicted Jaccard indices for the regional hippocampus for the different segmentation methods.**

Method	Anterior	Middle	Posterior
Manual	0.794	0.825	0.756
FSL-FIRST	0.829	0.855	0.798
FreeSurfer	0.756	0.784	0.721

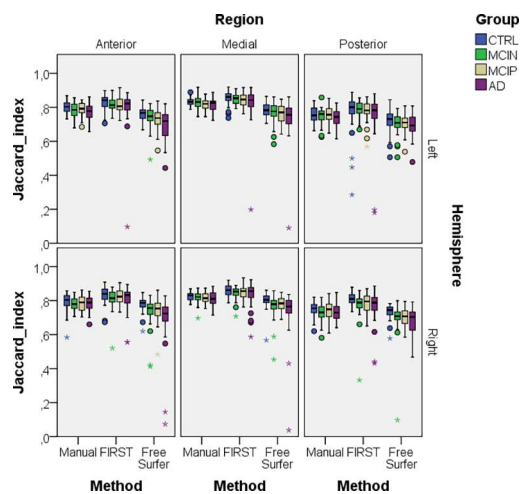
doi:10.1371/journal.pone.0166785.t003

### Discussion and conclusion

With our approach to automatically and precisely extract regional hippocampal volumes and outline reproducibilities from the BTB scans' segmentations we were able to detect systematic differences in volumes among three different segmentation methods and showed that FSL-FIRST was the most reproducible segmentation method.

In several applications, the quantification of global hippocampal volumes is of limited applicability. For instance, when studying anatomical changes accompanying the development of neurodegenerative diseases or when testing drugs against these diseases, it is well possible that these changes occur in specific regions of the hippocampus and then global measures such as volume would be too coarse to notice them. For clinical applications in radiotherapy where hippocampus avoidance is aimed for, it is insufficient to know that volume of the delineated object is correct, but also accuracy of shape is required. Finally, the need for local shape information is required to determine whether differences in hippocampus segmentation by different methods are caused by hidden systematic differences in the underlying anatomical definitions of the hippocampus.

The present study developed a method to investigate regional effects in shape differences. Confirming with other literature [26,47–51], also our analysis showed a global left and right hippocampus difference. Furthermore, global hippocampal atrophy could be detected, but it could not be distinguished in between groups (G) or regions (R), because the interaction of these with the time-point (T) were not significant. The regional volume analysis showed that both automatic segmentations revealed similar results, while manual segmentations had systematically larger anterior, and smaller middle and posterior volume predictions, which



**Fig 8. Regional hippocampus Jaccard indices for all methods and left and right hippocampi.** We combined both time-points, i.e. BL-A–BL-B and M12-A– M12-B, because both time-points by themselves gave similar results.

doi:10.1371/journal.pone.0166785.g008

indicates that the hippocampus segmentation protocol for manual segmentations is different than the definition of the hippocampus underlying the automatic segmentation methods. Both, FSL-FIRST and FreeSurfer subcortical segmentations are based on manually labelled training data sets following the outline protocol from the Center of Morphometric Analysis (CMA, <http://www.cma.mgh.harvard.edu/>). The intention of both the hippocampal outlining protocol of the CMA and that of Jack and colleagues [41] used in this study for manual segmentation, is to include: dentate gyrus, cornu ammonis, subiculum, fimbria and alveus. Alterations of regional volume distributions among methods shows that with our analysis more subtle differences in segmentation protocols were detectable. Therefore, it would be beneficial to use a standardized protocol like the harmonized protocol for hippocampus volumetry, the outcome of a project to define a standard protocol for hippocampus segmentation [52][53][54].

With our regional volume data we also compared FreeSurfer's results from the cross-sectional and longitudinal stream. For both we obtained the same model with the same selection of fixed effects, only the predicted volumes differed: FreeSurfer's anterior and posterior volume predictions were slightly larger for results from the longitudinal stream. Even though Reuter and colleagues [15] showed an improvement in distinguishing diagnostic groups using the longitudinal stream, with our approach the selected model using either the cross-sectional or longitudinal stream was identical, i.e. neither increased reproducibility nor accelerated decrease of hippocampal volume in AD subjects were found when using the a priori knowledge that scans form a longitudinal series. This might be due to the smaller number of subjects used in this study, as Reuter and colleagues used three times as many non-demented and demented subjects.

At the IAC Amsterdam, technicians undergo yearly reliability trainings with training sets of five cases. In the most recent test sets, the intra-rater variability score of the hippocampal volume—ICC with absolute agreement—was 0.985–0.99 using identical images. Determining the ICC with absolute agreement measure using the BTB dataset of the current study, the technician obtained an ICC of 0.98 and 0.99 for hippocampal volumes of BL-A–BL-B and M12-A–M12-B scans respectively. For FSL-FIRST the ICC was 0.98 and 0.98 and for FreeSurfer it was 0.99 and 0.98 for BL and M12 BTB scans respectively. Even though hippocampal volumes have high correlations, our outline reproducibility analysis showed that comparing volumes alone does not reflect the complete picture of the quality of the outline. We determined outline reproducibilities for the whole hippocampus, but also for anterior, middle and posterior hippocampus sections. For both left and right hippocampus, whole hippocampus and in all three subregions, in all diagnostic groups and at both time points, FSL-FIRST consistently gave significant higher Jaccard indices, followed by manual, followed by FreeSurfer. This confirms the finding of Morey and colleagues, who also found that FSL-FIRST had higher outline reproducibilities than FreeSurfer [30]. However, it should be mentioned that only automatic segmentation methods had large outlier Jaccard indices, as can be seen in Figs 7 and 8. To confirm that these truly resulted from poor segmentations and not by registration errors we visually inspected the MRI scan pairs of these outliers as described in 2.4. No visual noticeable registration errors could be detected, but poor segmentations could be confirmed by inspecting the mesh segmentations of these outliers. The hippocampal volumes of these outliers were also reviewed but they did not show outlier values.

With our regional reproducibility analysis we were also able to determine that for all segmentation methods the middle region had highest Jaccard indices. The middle region shares common borders with the anterior and posterior region, which means the border surface of the middle region to other structures is smaller compared to the anterior and posterior regions. Due to similar grey values the hippocampus is hard to distinguish from adjacent

structures, which means the regions with a larger surface to adjacent structures most probably have a poorer reproducibility, as it can be seen from the anterior and posterior region. It should also be noted that overlap indices in general are sensitive to size differences. The size differences between anterior and posterior parts amounted to between 15 to 20% (Table 1), which could therefore also provide a partial explanation for the observed differences in Jaccard indices.

Given that reproducibility is an important requirement for segmentation methods, FSL-FIRST meets the requirement and exhibits even better results than manual outlining, which is the choice of many clinical trials. Nevertheless, this finding should be treated with care, because outline reproducibility is necessary, but not sufficient to imply that the hippocampus was outlined accurately. In contrast, E. Mulder and colleagues [33] found that FreeSurfer obtains most reproducible volume atrophy measurements compared to manual and FSL-FIRST segmentations. Considering that we found that FreeSurfer has worst outline reproducibilities atrophy measurements FreeSurfers' hippocampus segmentations should be interpreted with care. Furthermore, results show that for all methods and subregions, and for both hemispheres and both time points, AD patients tend to exhibit poorer reproducibilities than healthy controls, while especially FreeSurfers' results have larger decrease in Jaccard indices with disease severity than manual and FIRST segmentations; and only automatic segmentation methods showed extreme Jaccard indices. This finding was not detected as a significant effect by our statistical model because the variation was too large for our sample size. But it is an indication that the training sets of the automatic methods might not be optimized for diseased subjects, which is confirmed by several other studies [20][22][31].

In this study we also proposed a novel method to extract regional Jaccard indices by converting label images to meshes and by using registration parameters on these meshes to map them to a common space. This approach is particularly useful when comparing small structures, because interpolation and registration errors are avoided. The full circle method allowed us to quantitatively estimate registration accuracy by computing rotation and translation components, but we also extended this method to a consistency measure using the Jaccard index. We suggest that this methodology can be a useful tool in other (brain) imaging studies where small structures are compared between scans with different image orientations.

For a better disease understanding and more sophisticated analysis it would be an idea to extend the regional analysis to more specific hippocampal subfields (cornu ammonis fields, dentate gyrus and subiculum). This is an ongoing field of interest and usually high field scanners over 3T with high resolution T2 or proton density sequences are necessary to distinguish boundaries between those regions [55]. We suggest that for the analysis of such datasets the methodology proposed in this study would be particularly suited.

## Supporting information

**S1 Table. Predicted volumes ( $\text{cm}^3$ ) for the right hippocampus at time-point BL for all segmentation methods.**

(DOCX)

**S2 Table. Predicted volumes ( $\text{cm}^3$ ) for the left hippocampus at time-point M12 for all segmentation methods.**

(DOCX)

**S3 Table. Predicted volumes ( $\text{cm}^3$ ) for the right hippocampus at time-point M12 for all segmentation methods.**

(DOCX)

**S4 Table. Volume predictions (cm<sup>3</sup>) for the left hippocampus at time-point BL using FreeSurfer's segmentations for the cross-sectional stream.**  
(DOCX)

## Acknowledgments

The authors thank Felix C. van Dommelen of the Image Analysis Center, VU University Medical Center, Amsterdam, The Netherlands for performing the manual hippocampal volume analyses, and Margo A. Pronk, of the same Image Analysis Center, for assistance in the visual inspection of segmentation outputs. Furthermore, we thank R. Schijndel and R.A. de Jong E. Mulder of the Image Analysis Center, VU University Medical Center, Amsterdam, The Netherlands for their support and previous work. 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.; Cogstate; 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 ([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 Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. This research was also supported by NIH grants P30 AG010129 and K01 AG030514. H. Vrenken has received research grants from Novartis, Teva, MerckSerono and Pfizer, and a speaker honorarium from Novartis. All funds were paid directly to his institution.

## Author Contributions

**Conceptualization:** FaB JCdeM HV.

**Data curation:** FaB JCdeM HV FrB.

**Formal analysis:** FaB JCdeM HV FeB.

**Funding acquisition:** JCdeM HV FrB.

**Investigation:** FaB JCdeM HV.

**Methodology:** FaB JCdeM HV MvanH FeB.

**Project administration:** JCdeM HV.

**Resources:** FaB JCdeM HV FrB.

**Software:** FaB JCdeM HV.



**Supervision:** JCdeM HV MvanH.

**Validation:** FaB JCdeM HV MvanH FeB FrB.

**Visualization:** FaB.

**Writing – original draft:** FaB.

**Writing – review & editing:** FaB JCdeM HV FeB MvanH FrB.

## References

1. Tulving E, Markowitsch HJ. Episodic and declarative memory: role of the hippocampus. *Hippocampus*. 1998 Jan; 8(3):198–204. doi: [10.1002/\(SICI\)1098-1063\(1998\)8:3<&lt;198::AID-HIPO2&gt;3.0.CO;2-G](https://doi.org/10.1002/(SICI)1098-1063(1998)8:3<&lt;198::AID-HIPO2&gt;3.0.CO;2-G) PMID: [9662134](https://pubmed.ncbi.nlm.nih.gov/9662134/)
2. Mungas D, Harvey D, Reed BR, Jagust WJ, DeCarli C, Beckett L, et al. Longitudinal volumetric MRI change and rate of cognitive decline. *Neurology*. 2005 Aug 22; 65(4):565–71. doi: [10.1212/01.wnl.0000172913.88973.0d](https://doi.org/10.1212/01.wnl.0000172913.88973.0d) PMID: [16116117](https://pubmed.ncbi.nlm.nih.gov/16116117/)
3. den Heijer T, van der Lijn F, Koudstaal PJ, Hofman A, van der Lugt A, Krestin GP, et al. A 10-year follow-up of hippocampal volume on magnetic resonance imaging in early dementia and cognitive decline. *Brain*. 2010 Apr; 133(Pt 4):1163–72. doi: [10.1093/brain/awq048](https://doi.org/10.1093/brain/awq048) PMID: [20375138](https://pubmed.ncbi.nlm.nih.gov/20375138/)
4. Braak H, Braak E, Bohl J. Staging of Alzheimer-related cortical destruction. *Eur Neurol*. 1993 Jan; 33(6):403–8. PMID: [8307060](https://pubmed.ncbi.nlm.nih.gov/8307060/)
5. Mu Y, Gage FH. Adult hippocampal neurogenesis and its role in Alzheimer's disease. *Mol Neurodegener*. 2011 Jan; 6:85. doi: [10.1186/1750-1326-6-85](https://doi.org/10.1186/1750-1326-6-85) PMID: [22192775](https://pubmed.ncbi.nlm.nih.gov/22192775/)
6. Likeman M, Anderson VM, Stevens JM, Waldman AD, Godbolt AK, Frost C, et al. Visual assessment of atrophy on magnetic resonance imaging in the diagnosis of pathologically confirmed young-onset dementias. *Arch Neurol*. 2005; 62(9):1410–5. doi: [10.1001/archneur.62.9.1410](https://doi.org/10.1001/archneur.62.9.1410) PMID: [16157748](https://pubmed.ncbi.nlm.nih.gov/16157748/)
7. Henneman WJP, Sluimer JD, Barnes J, Van Der Flier WM, Sluimer IC, Fox NC, et al. Hippocampal atrophy rates in Alzheimer disease: Added value over whole brain volume measures. *Neurology*. 2009 Mar 17; 72(11):999–1007. doi: [10.1212/01.wnl.0000344568.09360.31](https://doi.org/10.1212/01.wnl.0000344568.09360.31) PMID: [19289740](https://pubmed.ncbi.nlm.nih.gov/19289740/)
8. Aoyama H, Tago M, Kato N, Toyoda T, Kenjyo M, Hirota S, et al. Neurocognitive function of patients with brain metastasis who received either whole brain radiotherapy plus stereotactic radiosurgery or radiosurgery alone. *Int J Radiat Oncol Biol Phys*. 2007 Aug 1; 68(5):1388–95. doi: [10.1016/j.ijrobp.2007.03.048](https://doi.org/10.1016/j.ijrobp.2007.03.048) PMID: [17674975](https://pubmed.ncbi.nlm.nih.gov/17674975/)
9. Chang EL, Wefel JS, Hess KR, Allen PK, Lang FF, Kornguth DG, et al. Neurocognition in patients with brain metastases treated with radiosurgery or radiosurgery plus whole-brain irradiation: a randomised controlled trial. *Lancet Oncol*. 2009 Nov; 10(11):1037–44. doi: [10.1016/S1470-2045\(09\)70263-3](https://doi.org/10.1016/S1470-2045(09)70263-3) PMID: [19801201](https://pubmed.ncbi.nlm.nih.gov/19801201/)
10. Welzel G, Fleckenstein K, Schaefer J, Hermann B, Kraus-Tiefenbacher U, Mai SK, et al. Memory function before and after whole brain radiotherapy in patients with and without brain metastases. *Int J Radiat Oncol Biol Phys*. 2008 Dec 1; 72(5):1311–8. doi: [10.1016/j.ijrobp.2008.03.009](https://doi.org/10.1016/j.ijrobp.2008.03.009) PMID: [18448270](https://pubmed.ncbi.nlm.nih.gov/18448270/)
11. Gondi V, Tolakanahalli R, Mehta MP, Tewatia D, Rowley H, Kuo JS, et al. Hippocampal-sparing whole-brain radiotherapy: a “how-to” technique using helical tomotherapy and linear accelerator-based intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2010 Nov 15; 78(4):1244–52. doi: [10.1016/j.ijrobp.2010.01.039](https://doi.org/10.1016/j.ijrobp.2010.01.039) PMID: [20598457](https://pubmed.ncbi.nlm.nih.gov/20598457/)
12. Oskan F, Ganswindt U, Schwarz SB, Manapov F, Belka C, Niyazi M. Hippocampus sparing in whole-brain radiotherapy. A review. *Strahlentherapie und Onkol Organ der Dtsch Röntgengesellschaft*. [et al]. 2014 Apr; 190(4):337–41.
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. *Adv Alzheimer's Dis*. 2011; 2:111–25.
14. Dill V, Franco AR, Pinho MS. Automated methods for hippocampus segmentation: the evolution and a review of the state of the art. *Neuroinformatics*. 2015 Apr; 13(2):133–50. doi: [10.1007/s12021-014-9243-4](https://doi.org/10.1007/s12021-014-9243-4) PMID: [26022748](https://pubmed.ncbi.nlm.nih.gov/26022748/)
15. Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*. 2012; 61(4):1402–18. doi: [10.1016/j.neuroimage.2012.02.084](https://doi.org/10.1016/j.neuroimage.2012.02.084) PMID: [22430496](https://pubmed.ncbi.nlm.nih.gov/22430496/)

16. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33(3):341–55. PMID: [11832223](#)
17. Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*. 2011; 56(3):907–22. doi: [10.1016/j.neuroimage.2011.02.046](#) PMID: [21352927](#)
18. Tae WS, Kim SS, Lee KU, Nam EC, Kim KW. Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology*. 2008; 50(7):569–81. doi: [10.1007/s00234-008-0383-9](#) PMID: [18414838](#)
19. Cherbuin N, Anstey KJ, Réglade-Meslin C, Sachdev PS. In vivo hippocampal measurement and memory: a comparison of manual tracing and automated segmentation in a large community-based sample. *PLoS One*. 2009; 4(4):e5265. doi: [10.1371/journal.pone.0005265](#) PMID: [19370155](#)
20. Sánchez-Benavides G, Gómez-Ansón B, Sainz A, Vives Y, Delfino M, Peña-Casanova J. Manual validation of FreeSurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer Disease subjects. *Psychiatry Res—Neuroimaging*. 2010; 181(3):219–25.
21. Dewey J, Hana G, Russell T, Price J, McCaffrey D, Harezlak J, et al. Reliability and validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study. *Neuroimage*. 2010 Jul 15; 51(4):1334–44. doi: [10.1016/j.neuroimage.2010.03.033](#) PMID: [20338250](#)
22. Lehmann M, Douiri A, Kim LG, Modat M, Chan D, Ourselin S, et al. Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. *Neuroimage*. 2010; 49(3):2264–74. doi: [10.1016/j.neuroimage.2009.10.056](#) PMID: [19874902](#)
23. 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 Mar; 4(1):86–95. doi: [10.1007/s11682-010-9088-x](#) PMID: [20454594](#)
24. Kim H, Chupin M, Colliot O, Bernhardt BC, Bernasconi N, Bernasconi A. Automatic hippocampal segmentation in temporal lobe epilepsy: Impact of developmental abnormalities. *Neuroimage*. 2012 Feb; 59(4):3178–86. doi: [10.1016/j.neuroimage.2011.11.040](#) PMID: [22155377](#)
25. Germeyan SC, Kalikhman D, Jones L, Theodore WH. Automated versus manual hippocampal segmentation in preoperative and postoperative patients with epilepsy. *Epilepsia*. 2014; 55(9):1–6.
26. Wenger E, Mårtensson J, Noack H, Bodammer NC, Kühn S, Schaefer S, et al. Comparing manual and automatic segmentation of hippocampal volumes: reliability and validity issues in younger and older brains. *Hum Brain Mapp*. 2014 Aug; 35(8):4236–48. doi: [10.1002/hbm.22473](#) PMID: [24532539](#)
27. Grimm O, Pohlack S, Cacciaglia R, Plichta M, Demirakca T, Flor H. Amygdala and hippocampal volume: A comparison between manual segmentation, FreeSurfer and VBM. *J Neurosci Methods*. 2015 Jun; 253:254–61. doi: [10.1016/j.jneumeth.2015.05.024](#) PMID: [26057114](#)
28. Nugent AC, Luckenbaugh DA, Wood SE, Bogers W, Zarate CA, Drevets WC. Automated subcortical segmentation using FIRST: test-retest reliability, interscanner reliability, and comparison to manual segmentation. *Hum Brain Mapp*. 2013 Sep; 34(9):2313–29. doi: [10.1002/hbm.22068](#) PMID: [22815187](#)
29. Morey R a., Petty CM, Xu Y, Pannu Hayes J, Wagner HR, Lewis D V., et al. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage*. 2009; 45(3):855–66. doi: [10.1016/j.neuroimage.2008.12.033](#) PMID: [19162198](#)
30. Morey RA, Selgrade ES, Wagner HR, Huettel SA, Wang L, McCarthy G. Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Hum Brain Mapp*. 2010 Nov; 31(11):1751–62. doi: [10.1002/hbm.20973](#) PMID: [20162602](#)
31. Pardoe HR, Pell GS, Abbott DF, Jackson GD. Hippocampal volume assessment in temporal lobe epilepsy: How good is automated segmentation? *Epilepsia*. 2009 Dec; 50(12):2586–92. doi: [10.1111/j.1528-1167.2009.02243.x](#) PMID: [19682030](#)
32. Doring TM, Kubo TT a, Cruz LCH, Juruena MF, Fainberg J, Domingues RC, et al. Evaluation of hippocampal volume based on MR imaging in patients with bipolar affective disorder applying manual and automatic segmentation techniques. *J Magn Reson Imaging*. 2011; 33(3):565–72. doi: [10.1002/jmri.22473](#) PMID: [21563239](#)
33. Mulder ER, de Jong R a., Knol DL, van Schijndel R a., Cover KS, Visser PJ, et al. Hippocampal volume change measurement: Quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST. *Neuroimage*. 2014; 92:169–81. doi: [10.1016/j.neuroimage.2014.01.058](#) PMID: [24521851](#)
34. Hackert VH, den Heijer T, Oudkerk M, Koudstaal PJ, Hofman A, Breteler MMB. Hippocampal head size associated with verbal memory performance in nondemented elderly. *Neuroimage*. 2002 Nov; 17(3):1365–72. PMID: [12414276](#)

35. Frisoni GB, Ganzola R, Canu E, Rüb U, Pizzini FB, Alessandrini F, et al. Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. *Brain*. 2008; 131(12):3266–76.
36. Gordon BA, Blazey T, Benzinger TLS, Head D. Effects of aging and Alzheimer's disease along the longitudinal axis of the hippocampus. *J Alzheimers Dis*. 2013 Jan; 37(1):41–50. doi: [10.3233/JAD-130011](https://doi.org/10.3233/JAD-130011) PMID: [23780659](https://pubmed.ncbi.nlm.nih.gov/23780659/)
37. van Herk M, de Munck JC, Lebesque J V, Muller S, Rasch C, Touw A. Automatic registration of pelvic computed tomography data and magnetic resonance scans including a full circle method for quantitative accuracy evaluation. *Med Phys*. 1998 Oct; 25(10):2054–67. doi: [10.1118/1.598393](https://doi.org/10.1118/1.598393) PMID: [9800715](https://pubmed.ncbi.nlm.nih.gov/9800715/)
38. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974 Dec; 19(6):716–23.
39. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol*. 2009 Apr; 65(4):403–13. doi: [10.1002/ana.21610](https://doi.org/10.1002/ana.21610) PMID: [19296504](https://pubmed.ncbi.nlm.nih.gov/19296504/)
40. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008 Apr; 27(4):685–91. doi: [10.1002/jmri.21049](https://doi.org/10.1002/jmri.21049) PMID: [18302232](https://pubmed.ncbi.nlm.nih.gov/18302232/)
41. Jack CR. MRI-Based Hippocampal Volume Measurements in Epilepsy. *Epilepsia*. 1994 Dec; 35(s6):S21–9.
42. van de Pol LA, van der Flier WM, Korf ESC, Fox NC, Barkhof F, Scheltens P. Baseline predictors of rates of hippocampal atrophy in mild cognitive impairment. *Neurology*. 2007 Oct 9; 69(15):1491–7. doi: [10.1212/01.wnl.0000277458.26846.96](https://doi.org/10.1212/01.wnl.0000277458.26846.96) PMID: [17923611](https://pubmed.ncbi.nlm.nih.gov/17923611/)
43. Patenaude B. Bayesian Statistical Models of Shape and Appearance for Subcortical Brain Segmentation. *Dep Clin Neurol*. 2007; Doctor of: 247.
44. Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging*. 2001; 20(1):45–57. doi: [10.1109/42.906424](https://doi.org/10.1109/42.906424) PMID: [11293691](https://pubmed.ncbi.nlm.nih.gov/11293691/)
45. Press WH. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. 2007.
46. Satterthwaite FE. An approximate distribution of estimates of variance components. *Biometrics*. 1946 Dec; 2(6):110–4. PMID: [20287815](https://pubmed.ncbi.nlm.nih.gov/20287815/)
47. Barnes J, Foster J, Boyes RG, Pepple T, Moore EK, Schott JM, et al. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *Neuroimage*. 2008; 40(4):1655–71. doi: [10.1016/j.neuroimage.2008.01.012](https://doi.org/10.1016/j.neuroimage.2008.01.012) PMID: [18353687](https://pubmed.ncbi.nlm.nih.gov/18353687/)
48. Basso M, Yang J, Warren L, MacAvoy MG, Varma P, Bronen RA, et al. Volumetry of amygdala and hippocampus and memory performance in Alzheimer's disease. *Psychiatry Res Neuroimaging*. 2006; 146(3):251–61. doi: [10.1016/j.psychres.2006.01.007](https://doi.org/10.1016/j.psychres.2006.01.007) PMID: [16524704](https://pubmed.ncbi.nlm.nih.gov/16524704/)
49. Honeycutt NA, Smith CD. Hippocampal Volume Measurements Using Magnetic Resonance Imaging in Normal Young Adults. *J Neuroimaging*. 1995 Apr; 5(2):95–100. PMID: [7718948](https://pubmed.ncbi.nlm.nih.gov/7718948/)
50. Jack CR Jr., Slomkowski M, Gracon S, Hoover TM, Felmler JP, et al. MRI as a Biomarker of Disease Progression in a Therapeutic Trial of Milameline for AD. *Neurology*. 2003; 60(2):253. PMID: [12552040](https://pubmed.ncbi.nlm.nih.gov/12552040/)
51. Pruessner JC, Li LM, WW S, MM P, Collins DL, NN K, et al. Volumetry of Hippocampus and Amygdala with High-resolution MRI and Three-dimensional Analysis Software: Minimizing the Discrepancies between Laboratories. 2000; 10(4):433–42. PMID: [10769253](https://pubmed.ncbi.nlm.nih.gov/10769253/)
52. 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 Dement*. 2015 Feb; 11(2):126–38.
53. 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. *Alzheimers Dement*. 2015 Feb; 11(2):139–50. doi: [10.1016/j.jalz.2015.01.001](https://doi.org/10.1016/j.jalz.2015.01.001) PMID: [25620800](https://pubmed.ncbi.nlm.nih.gov/25620800/)
54. 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. *Alzheimers Dement*. 2015 Feb; 11(2):111–25. doi: [10.1016/j.jalz.2014.05.1756](https://doi.org/10.1016/j.jalz.2014.05.1756) PMID: [25267715](https://pubmed.ncbi.nlm.nih.gov/25267715/)
55. Flores R de, Joie R La, Chételat G. Structural imaging of hippocampal subfields in healthy aging and Alzheimer's disease. *Neuroscience*. 2015 Aug; 309:29–50. doi: [10.1016/j.neuroscience.2015.08.033](https://doi.org/10.1016/j.neuroscience.2015.08.033) PMID: [26306871](https://pubmed.ncbi.nlm.nih.gov/26306871/)