

Training Labels for Hippocampal Segmentation Based on the EADC-ADNI Harmonized Hippocampal Protocol

Marina Boccardi¹, Martina Bocchetta¹, Felix Morency², Masami Nishikawa³, Rossana Ganzola⁴, Michel Grothe⁵, Dominik Wolf⁶, Alberto Redolfi¹, Michela Pievani¹, Andreas Fellgiebel⁶, Hiroshi Matsuda³, Stefan Teipel⁵, Simon Duchesne⁷, Clifford R. Jack Jr.⁸, Giovanni B. Frisoni¹, and Luigi Antelmi¹
¹IRCCS Fatebenefratelli, Brescia, Italy, ²IMEKA, Sherbrooke, Quebec, Canada, ³Kawamura Gakuen Woman's University, Abiko-City, Japan, ⁴Universite Laval, Quebec, Canada, ⁵German Center for Neurodegenerative Diseases, Rostock, Germany, ⁶Johannes Gutenberg-Universität, Mainz, Germany, ⁷Universite Laval, Quebec, Canada, ⁸Mayo Clinic, Rochester, Minnesota, United States

TARGET AUDIENCE - Methodological and clinical researchers alike interested in MRI-based hippocampal volumetry as a neurodegeneration biomarker.

PURPOSE - In the 2008-2013 period, a joint European Alzheimer's Disease Center (EADC) and Alzheimer's Disease Neuroimaging Initiative (ADNI) effort was carried out to provide a consensual, harmonized protocol (HarP) for manual hippocampal segmentation on MRI. The protocol was defined through an evidence based Delphi panel that converged on a consensual definition based on personal experience, data specifically collected for this purpose ¹, and recursive re-evaluation of the choices preferred by the other panelists and reasons thereof ². Although segmentations by different tracers proved to be very reliable ³, manual hippocampal segmentation remains a time consuming task, with added effort required if one is to comply with the HarP, as it is lengthy. The need to train automated algorithms that may segment large datasets based on the HarP is an urgent requirement for clinical trials and a necessary step towards wider clinical acceptance.

Yet, there remains a substantial obstacle to wide uptake of the HarP by automated algorithms. Not unlike new or naïve human raters, most algorithms require a very large sampling of image presentations, physiological variability and pathological effects to learn exemplars and properly generalize the knowledge of hippocampal boundaries to new subjects. However, all of the segmentations generated in the different phases of this project were taken on a very limited number of ADNI subjects ($n=16$ in total).

The purpose of this work was to provide a large sample of benchmark hippocampal segmentations based on the HarP with good balance of typical confounders typically connected to hippocampal volumetry (tracer, subject, side, time point, scanner field strength and manufacturer).

METHODS - 135 ADNI images from different subjects, balanced by main confounders, were segmented by five qualified HarP tracers, with absolute inter-rater intra-class coefficients: left:0.953(95% confidence interval:0.873-0.987), right:0.975(0.920-0.994). Each image was segmented once. Labels were verified and validated as HarP-compliant after centralized quality checks and corrections. Head size was quantified with SIEONAX ⁴. Diagnosis and medial temporal lobe atrophy effects were evaluated with parametric tests and Spearman's rho correlation. Labels were back-transformed in native space and released publicly.

RESULTS - A maximum of two rounds of corrections were required to tracers to achieve full compliance with the HarP for all of the segmented hippocampi. Hippocampal volumes of controls -left: mm³ 3096 (standard deviation:541), right 3997(511)- were over 22% larger than AD patients', and 14% larger than mild cognitive impairment subjects ($p<0.0005$). Volumes significantly correlated with medial temporal lobe atrophy severity at Scheltens' scale ⁵ ($p<0.0005$). Cerebrospinal fluid spaces were larger in controls - left=mm³ 34(41), right=36(40), than in patients – mild cognitive impairment: left=19(13), right=29(20); AD: left=14(18), right=25(34). Five subjects (3.7%) presented with unusual anatomy.

Contours in pseudo-Talairach space were voxelized to the Analyze image format using the ShapeToImage utility found in the LONI ShapeTools library. The resulting Analyze contours were then converted to NIFTI using BrainVisa/Anatomist ⁶. The original image header, including geometry and orientation, was applied to the Nifti contour using FSL ⁷. Finally, a binary hole filling filter was applied on each coronal slices to create the labels. AC-PC pseudo-Talairach voxelized labels were then back-transformed in the native space using the inverse linear transform and trilinear interpolation.

DISCUSSION - UCF labels and Multitracer files, .mnc and .nifti voxelized labels, linear transformations and .mnc reoriented voxelized and interpolated labels in native space are made available at www.hippocampal-protocol.net. List of ADNI subject IDs, image codes and conversion files reporting the orientation function for each MRI for segmentation in AC-PC, as required by the HarP, are also reported.

One of the limitations of this study consists in the lack of longitudinal images of the same subjects; this will not serve algorithms that exploit differences between scans, nor allow for the validation of atrophy rate estimations and other longitudinal behavior (e.g. transitivity, linearity).

A second limitation lies in the segmentation of each hippocampus by a single tracer rather than by more experts as for the previously generated benchmark labels ⁸. It was felt that the very accurate definitions provided by the HarP reduced the range of alternative segmentations that may be considered to be correct for each hippocampus. Nonetheless, some divergence may be considered acceptable due to a certain degree of ambiguity in tissue definition from MRIs, which do not provide perfect visualization of subtle features of brain morphology. Certification criteria that can flexibly account for these ambiguities depending on the different anatomical regions will need to be defined in order to make certification both possible and highly accurate for human tracers and algorithms.

CONCLUSION - With this work, we have provided a relatively large set of benchmark hippocampal segmentations based on the HarP, which covers a wide range of physiological variability. This set is meant to provide the appropriate reference to train human tracers and automated algorithms so that they can generalize the learning and appropriately segment hippocampi of new subjects. This work follows the completion of the HarP project, defining the new standard for the measurement of hippocampal volumetry and its use as a biomarker for Alzheimer's disease.

REFERENCES

- 1.Boccardi M, Bocchetta M, Ganzola R, et al. Operationalizing protocol differences for EADC-ADNI manual hippocampal segmentation. *Alzheimers Dement* 2013.
- 2.Boccardi M, Bocchetta M, Apostolova L, et al. Delphi consensus on landmarks for the manual segmentation of the hippocampus on MRI: preliminary results from the EADC-ADNI Harmonized Protocol Working Group. *Neurology* 2012;78:003.
- 3.Frisoni GB, Jack CR, Bocchetta M, et al. The EADC-ADNI Harmonized Protocol for Hippocampal Segmentation on Magnetic Resonance: Evidence of Validity. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2013.
- 4.Smith SM, Rao A, De Stefano N, et al. Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: cross-validation of BSI, SIENA and SIEONAX. *NeuroImage* 2007;36:1200-1206.
- 5.Scheltens P, Leyls D, Barkhof F, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry* 1992;55:967-972.
- 6.Cointepas Y, Mangin JF, Garner L, Poline JB, Benali H. BrainVISA: Software platform for visualization and analysis of multi-modality brain data. *NeuroImage* 2001;13:S98-S98.
- 7.Smith SM, Zhang Y, Jenkinson M, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage* 2002;17:479-489.
- 8.Bocchetta M, Boccardi M, Ganzola R, et al. Harmonized benchmark labels of the hippocampus on MR: the EADC-ADNI project. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2013.