ELSEVIER

Alzheimer's & Dementia ■ (2013) 1–11

Alzheimer's & Dementia

Operationalizing protocol differences for EADC-ADNI manual hippocampal segmentation

Marina Boccardi[a], Martina Bocchetta[a,b], Rossana Ganzola[a], Nicolas Robitaille[c], Alberto Redolfi[a], Simon Duchesne[c], Clifford R. Jack, Jr.,[d], Giovanni B. Frisoni[a,*]; for the EADC-ADNI Working Group on The Harmonized Protocol for Hippocampal Volumetry and for the Alzheimer's Disease Neuroimaging Initiative

[a]*Laboratory of Epidemiology, Neuroimaging and Telemedicine, IRCCS - S. Giovanni di Dio – Fatebenefratelli, Brescia, Italy*
[b]*Associazione Fatebenefratelli per la Ricerca, Rome, Italy*
[c]*Department of Radiology, Université Laval and Centre de Recherche de l'Institut universitaire de santé mentale de Québec, Quebec City, Canada*
[d]*Department of Diagnostic Radiology, Mayo Clinic and Foundation, Rochester, MN, USA*

Abstract

Background: Hippocampal volumetry on magnetic resonance imaging is recognized as an Alzheimer's disease (AD) biomarker, and manual segmentation is the gold standard for measurement. However, a standard procedure is lacking. We operationalize and quantitate landmark differences to help a Delphi panel converge on a set of landmarks.

Methods: One hundred percent of anatomic landmark variability across 12 different protocols for manual segmentation was reduced into four segmentation units (the minimum hippocampus, the alveus/fimbria, the tail, and the subiculum), which were segmented on magnetic resonance images by expert raters to estimate reliability and AD-related atrophy.

Results: Intra- and interrater reliability were more than 0.96 and 0.92, respectively, except for the alveus/fimbria, which were 0.86 and 0.77, respectively. Of all AD-related atrophy, the minimum hippocampus contributed to 67%; tail, 24%; alveus/fimbria, 4%; and the subiculum, 5%.

Conclusions: Anatomic landmark variability in available protocols can be reduced to four discrete and measurable segmentation units. Their quantitative assessment will help a Delphi panel to define a set of landmarks for a harmonized protocol.

Keywords: Hippocampus; Hippocampal atrophy; Hippocampal volumetry; Manual segmentation protocol; Harmonization; Anatomic landmark; Alzheimer's disease; Manual tracing; Medial temporal lobes; Atrophy; Degeneration; Magnetic resonance; Neuroimaging; Alzheimer's Disease Neuroimaging Initiative; Standard operating procedures

Project collaborators include George Bartzokis (Department of Psychiatry, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA), John G. Csernansky (Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA), Mony J. de Leon (New York University School of Medicine, Center for Brain Health, New York, NY, USA), Leyla deToledo-Morrell (Department of Neurological Sciences, Rush University, Chicago, IL, USA), Ronald J. Killiany (Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, USA), Stéphane Lehéricy (Center for NeuroImaging Research and Department of Neuroradiology, Université Pierre et Marie Curie-Paris 6, Groupe Hospitalier Pitié-Salpêtrière, Paris, France), Nikolai Malykhin (Department of Biomedical Engineering, Centre for Neuroscience, University of Alberta, Edmonton, Alberta, Canada), Johannes Pantel (Institute of General Practice, University of Frankfurt/Main, Germany), Jens C. Pruessner (McGill Centre for Studies in Aging, Department of Psychiatry, McGill University, Montreal, Quebec, Canada), Hilkka Soininen (Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland), and Craig Watson (Wayne State University School of Medicine, University Health Center, St. Antoine, Detroit, MI, USA).

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). 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.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. This manuscript was approved by the ADNI Data and Publication Committee on June 4, 2012.

*Corresponding author. Tel.: +39 030 3501361; Fax: +39 030 3501592.
E-mail address: gfrisoni@fatebenefratelli.it

1552-5260/$ - see front matter
http://dx.doi.org/10.1016/j.jalz.2013.03.001

1. Introduction

Hippocampal volumetry has been proposed as a diagnostic marker for Alzheimer's disease (AD) by the International Working Group for the new research criteria for the diagnosis of AD and National Institute of Aging–Alzheimer's Association revised diagnostic criteria [1–5]. To obtain such measurement, manual segmentation on T1-weighted, high-resolution magnetic resonance (MR) images is regarded as the gold standard. Although many algorithms for automated hippocampal segmentation exist (cf. [6] for a cursory review), with undeniable usefulness (less time-consuming, high test–retest reliability), manual segmentation [7–10] remains the benchmark that should be used for the validation of the algorithms carrying out automated segmentation [6,11–15].

However, different manual segmentation protocols are available [16,17], leading to volume estimates differing up to 2.5-fold for hippocampal volume in control subjects [16]. Heterogeneity in anatomic definitions and tracing guidelines have hampered comparisons among different studies using hippocampal volumetry for diagnosis or as a surrogate marker for disease progression, and limit its use as a diagnostic marker for clinical diagnosis.

An effort has been undertaken by European Alzheimer's Disease Consortium (EADC) and Alzheimer's Disease Neuroimaging Initiative (ADNI) centers to develop a harmonized protocol for the manual segmentation of the hippocampus on MR images [18–20]). This project aims to achieve a consensus on a gold standard for manual segmentation through a Delphi panel of experts, and then validate it versus local protocols and pathology.

To achieve this goal, the following steps were planned and executed. First, a survey of the most cited protocols in the literature was undertaken to provide the exact range of anatomic landmark variability, after certification of the protocols by authors, who checked and corrected the practical execution of segmentation on scans of a normal subject and of an AD subject. This step was carried out and described in previous work [21,22]. Landmark differences across protocols that contribute significantly to volume estimates concerned the definition of the most posterior slice, the superior border, and the separation of hippocampal tissue from the parahippocampal gyrus at the level of the subiculum, along the hippocampal body [21].

A second step was designed to operationalize the numerous and fuzzy anatomic landmark differences into a limited number of well-defined units, that could undergo quantitative examination. This second step of the project is described specifically as (*i*) the operationalization of protocol differences (i.e., the definition of a limited number of three-dimensional [3D] units able to account for all interprotocol differences in a way that was sufficiently well defined to undergo quantitative investigation) and (*ii*) the estimation, for each unit, of reliability values, contribution to total hippocampal volume, and informative value regarding AD-related atrophy.

During the third step of the project, the information gathered in these previous steps was given to a panel of hippocampus experts to carry out an evidence-based Delphi procedure facilitating a consensual definition [23] of a harmonized protocol. After this definition of a harmonized protocol is achieved, a small group of "master tracers" will segment a set of benchmark images accordingly.

We plan to upload these labels on an interactive web system allowing standard learning, qualification, and periodical certification of the ability of tracers to segment according to the harmonized protocol. Tracers from participating centers, who previously segmented a set of ADNI images based on their local protocols, will then qualify and resegment the same images based on the harmonized protocol, to obtain data for the validation of the harmonized protocol (see the flowchart of the Validation phase [24]). A similar procedure will allow us to validate the protocol versus neuropathological data.

2. Methods

2.1. Operationalization of differences among protocols

This study capitalizes on previous work in which we extracted landmark variability of the 12 most prevalent hippocampal segmentation protocols in the AD literature [21]. We operationalized protocol differences (i.e., we collapsed these differences into a limited number of units, sufficiently well defined to lend themselves to quantitative investigation). In practice, the wide landmark heterogeneity has been reduced [21] and turned into "positive" units consisting of a finite number of elementary blocks named segmentation units (SUs). These SUs actually can be segmented as contiguous labels within the boundaries of the hippocampus on coronal MR images, and with the help of simultaneous 3D visualization, they can thus be measured and tested. SUs have also been modeled as 3D digital objects.

The definition of SU landmarks (see SU Protocol, Supplemental Material) was based on published landmarks definitions, drawn from the previous survey of hippocampal segmentation protocols [21] according to the following criteria: 1) definitions based on internal landmarks were preferred whenever possible, because they are more invariant to image orientation; 2) if different definitions were equally clear, the most frequently used was adopted; otherwise, the clearest definition was chosen.

We segmented SUs (M. Bocchetta and R. G.) as contiguous labels on ADNI MR images to obtain quantitative information about intra- and interrater segmentation reliability, as well as to obtain informative value regarding AD-related atrophy.

2.2. Image selection

For this study, 3D T1-weighted structural MR images at 1.5T were acquired from the ADNI database [25]. The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and