Reproducibility of brain morphometry results derived at 3T: a multi-center European study comparing the longitudinal and cross-sectional Freesurfer segmentation analyses

Jorge Jovich[1], Moira Marizzoni[2], Roser Sala-Llonch[3], Nuria Bargalló[3], David Bartrés-Faz[3], Jennifer Arnold[4], Jens Benninghoff[4], Jens Wiltfang[4], Luca Roccatagliata[5], Flavio Nobili[5], Christian Zeeh[6], Peter Schonknecht[6], Giada Zoccatelli[7], Franco Alessandrini[7], Alberto Beltramello[7], Hélène GROS-DAGNAC[8], Pierre Payoux[8], Valérie Chanoine[9,10], Jean-Philippe Ranjeva[9,10], Mira Didic[9,10], Melanie Leroy[11], Regis Bordet[11], and Giovanni Frisoni[2]

[1]University of Trento, Center for Mind/Brain Sciences, Mattarello, Trento, Italy, [2]IRCCS San Giovanni di Dio Fatebenefratelli, LENITEM Lab of Epidem., Neuroim. & Telem, Brescia, Lombardia, Italy, [3]Dept. Psychiatry and Clinical Psychobiology, Faculty of Medicine University of Barcelona, Barcelona, Barcelona, Spain, [4]Department of Psychiatry and nuclear medicine, Universitaet Duisburg-Essen, Essen, Essen, Germany, [5]Dept of Neuroscience, Ophthalmology and Genetics, University of Genoa, Genoa, Lombardia, Italy, [6]Dept of Psychiatry and Dept of Neuroradiology, University of Leipzig, Leipzig, Leipzig, Germany, [7]Dept of Neuroradiology, Verona General Hospital, Verona, Verona, Italy, [8]U825 - Plateau Technique IRM, INSERM / Université Paul Sabatier, Toulouse, Toulouse, France, [9]Hôpital La Timone CIC – UPCET, Marseille, Marseille, France, [10]Centre de Resonance Magnetique Biologique et Medicale, Aix Marseille Université, Marseille, Marseille, France, [11]Universitè Lille UL2, Lille, Lille, France

PURPOSE: Pharmacog is an industry-academic European project aimed at identifying reliable biomarkers that are sensitive to disease progression in patients with Mild Cognitive Impairment [1]. Here we present work aimed at implementing standardized procedures to acquire and analyze multi-site structural 3T MRI data for automated brain morphometry. In particular, the goal of this study was to compare the across-session test-retest reproducibility of brain segmentations derived from FreeSurfer using the cross-sectional (CS) and longitudinal (LG) analysis streams [2].

METHODS: Eight 3T MRI sites participated across Italy, Spain, France and Germany. MRI systems include one GE HDxt, two Philips Achieva and five Siemens (two TrioTim, one Verio, one Allegra, and one Skyra) scanners. Most systems used 8 RF channel receive coils, except for the Allegra (birdcage) and Skyra (20 channels) systems. The acquisition protocol (35 min in total) included two structural T1 volumes for brain morphometry per scanning session: 3D MPRAGE, acceleration factor in the range of 1.5 to 2 where possible (GRAPPA, SENSE and ASSET in Siemens, Philips and GE systems, respectively), $1x1x1mm^3$, with TE/TR/TI as recommended by the ADNI project [3]. Each site recruited 5 local healthy volunteers in the age range of the clinical population (55-80 years), who were scanned in two sessions a week apart. Data analysis of each MPRAGE included a visual quality assurance to control for various artifacts followed by full brain automated segmentation with Freesurfer [4]. Both a Bland-Altman analysis and the absolute percent change relative to the means were used to examine the test-retest reliability of volume and cortical thickness estimates from each site. Spatial reproducibility of the segmented volumes across sessions was examined using the Dice coefficients for volume overlap.

RESULTS: Analysis of volume segmentations and their across-session reproducibility were focused on the hippocampus (Hp), amygdale (Amy), caudate (Cau), putamen (Put), pallidum (Pal), thalamus (Thal), and lateral ventricles (Lat). Figure 1 (absolute volume reproducibility error) and Figure 2 (Dice coefficient for spatial overlap) show the across-session test-retest reproducibility analyses for the two FreeSurfer analysis streams (CS and LG), averaged across the eight 3T MRI sites, for each volumetric structures and their average (All). No significant MRI site effects were found for volume reproducibility or Dice coefficient, regardless of analysis (Kruskall-Wallis test, p=0.01). The absolute volume reproducibility and spatial overlap of the LG analysis were significantly better than the CS for all structures, except for the volume reliability in lateral ventricules, which were not significantly different (Wilcoxon rank sum test, p=0.01). The analysis of the cortical thickness analysis was focused on the parahippocampus, fusiform gyrus, superior temporal gyrus, precuneus, superiorparietal gyrus, supramarginal gyrus, lateral occipital gyrus, lingual gyrus and superior
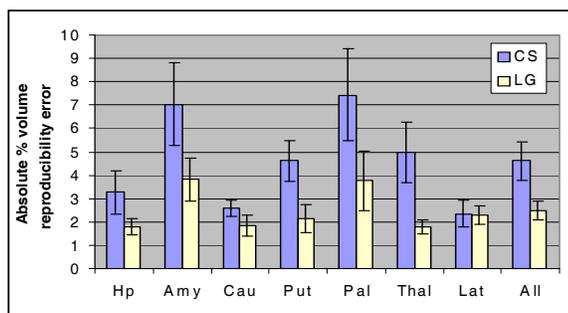

Figure 1: Across-session test-retest percent absolute volume reproducibility errors averaged across eight 3T MRI sites for the cross-sectional (CS) and longitudinal (LG) FreeSurfer segmentation streams.
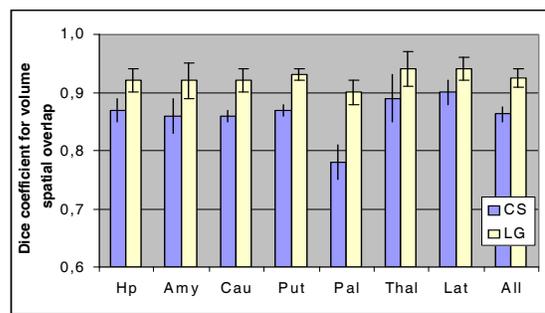

Figure 2: Across-session test-retest Dice coefficients for spatial overlap, averaged across eight 3T MRI sites for the cross-sectional (CS) and longitudinal (LG) FreeSurfer segmentation streams.

frontal gyrus. No MRI site effects were found for the thickness reproducibility, regardless of analysis. No thickness differences were found between the reproducibility errors from two analysis streams, which gave across MRI site mean absolute errors in the range 1.5 – 2.5 %.

DISCUSSION: Within the limitations of the study (40 subjects, 5 per MRI site, 8 MRI sites), we found that the FreeSurfer longitudinal stream gives higher across-session test-retest reproducibility of volumetric segmentations relative to the cross-sectional stream, but no significant improvements for the reproducibility of cortical thickness segmentations. These results extend those of previous studies which evaluated the across-session reproducibility of the method at 1.5T [5] and the within-session reproducibility at 3T with an optimized sequence at a single site [2].

CONCLUSIONS: This is the first study that evaluates on healthy elderly subjects the across-session test-retest reproducibility of brain morphometry segmentations derived from multi-center 3T structural MRI acquisitions using the novel longitudinal FreeSurfer segmentation analysis protocol [2]. The results confirm the advantages of the method giving an overall improvement in the reliability of the measures both within and across MRI sites.

REFERENCES: [1] http://www.alzheimer-europe.org/FR/Research/PharmaCog. [2] Reuter et al., Neuroimage, 2012; 61(4):1402-18. [3] Jack et al., J Magn Reson Imaging. 2008; 27(4):685-91. [4] http://surfer.nmr.mgh.harvard.edu/fswiki. [5] Han et al., Neuroimage 2006; 32(1):180-95.