NeuroImage: Clinical 18 (2018) 888–896

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Automated versus manual segmentation of brain region volumes in former football players

Jeffrey P. Guenette[a,b], Robert A. Stern[c,d], Yorghos Tripodis[c,e], Alicia S. Chua[e], Vivian Schultz[a,f], Valerie J. Sydnor[a], Nathaniel Somes[a], Sarina Karmacharya[a], Christian Lepage[a], Pawel Wrobel[a,f], Michael L. Alosco[c], Brett M. Martin[g], Christine E. Chaisson[c,g], Michael J. Coleman[a], Alexander P. Lin[a,b,h], Ofer Pasternak[a,b], Nikos Makris[a,b,i], Martha E. Shenton[a,b,j], Inga K. Koerte[a,f,*]

[a] Psychiatry Neuroimaging Laboratory, Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States
[b] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States
[c] BU Alzheimer's Disease and CTE Center, Boston University, Boston, MA, United States
[d] Departments of Neurology, Neurosurgery, and Anatomy & Neurobiology, Boston University School of Medicine, Boston, MA, United States
[e] Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States
[f] Department of Child and Adolescent Psychiatry, Psychosomatic, and Psychotherapy, Ludwig-Maximilian-University, Munich, Germany
[g] Data Coordinating Center, Boston University School of Public Health, Boston, MA, United States
[h] Center for Clinical Spectroscopy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States
[i] Center for Neural Systems Investigations, Massachusetts General Hospital, Boston, MA, United States
[j] VA Boston Healthcare System, Brockton Division, Brockton, MA, United States

ARTICLE INFO

Keywords:
Magnetic resonance imaging
Neuroimaging
Head trauma
Traumatic brain injury
Chronic traumatic encephalopathy

ABSTRACT

Objectives: To determine whether or not automated FreeSurfer segmentation of brain regions considered important in repetitive head trauma can be analyzed accurately without manual correction.

Materials and methods: 3 T MR neuroimaging was performed with automated FreeSurfer segmentation and manual correction of 11 brain regions in former National Football League (NFL) players with neurobehavioral symptoms and in control subjects. Automated segmentation and manually-corrected volumes were compared using an intraclass correlation coefficient (ICC). Linear mixed effects regression models were also used to estimate between-group mean volume comparisons and to correlate former NFL player brain volumes with neurobehavioral factors.

Results: Eighty-six former NFL players (55.2 ± 8.0 years) and 22 control subjects (57.0 ± 6.6 years) were evaluated. ICC was highly correlated between automated and manually-corrected corpus callosum volumes (0.911), lateral ventricular volumes (right 0.980, left 0.967), and amygdala-hippocampal complex volumes (right 0.713, left 0.731), but less correlated when amygdalae (right −0.170, left −0.090) and hippocampi (right 0.539, left 0.637) volumes were separately delineated and also less correlated for cingulate gyri volumes (right 0.639, left 0.351). Statistically significant differences between former NFL player and controls were identified in 8 of 11 regions with manual correction but in only 4 of 11 regions without such correction. Within NFL players, manually corrected brain volumes were significantly associated with 3 neurobehavioral factors, but a different set of 3 brain regions and neurobehavioral factor correlations was observed for brain region volumes segmented without manual correction.

Conclusions: Automated FreeSurfer segmentation of the corpus callosum, lateral ventricles, and amygdala-hippocampus complex may be appropriate for analysis without manual correction. However, FreeSurfer segmentation of the amygdala, hippocampus, and cingulate gyrus need further manual correction prior to performing group comparisons and correlations with neurobehavioral measures.

* Corresponding author at: Psychiatry Neuroimaging Laboratory, 1249 Boylston St., Boston, MA 02215, United States.
E-mail address: ikoerte@bwh.harvard.edu (I.K. Koerte).

https://doi.org/10.1016/j.nicl.2018.03.026
Received 9 November 2017; Received in revised form 2 February 2018; Accepted 21 March 2018
Available online 21 March 2018

1. Introduction

Analysis of brain regional volumes has yielded insight into the pathology and pathophysiology of a variety of neurological and psychiatric diseases including Alzheimer's disease (see reviews by Kantarci and Jack, 2003 and Busatto et al., 2008), schizophrenia (see metaanalysis by Olabi et al., 2011 and reviews by Hulshoff Pol and Kahn, 2008 and Shenton et al., 2010), post-traumatic stress disorder (see reviews by Ahmed-Leitao et al., 2016 and Milani et al., 2017), mild traumatic brain injury (see reviews by Shenton et al., 2012 and Mu et al., 2017) and repetitive head trauma (see reviews by Ng et al., 2014 and Koerte et al., 2015), to name just a few. Accurate and precise volumetric measurements are essential for both reliability and reproducibility. Given the time-consuming nature of manual segmentation, automated segmentation techniques are critical for studies involving large imaging datasets. Moreover, to be useful in the clinical setting, automated segmentation techniques are also critical given that time-consuming manual segmentation by a radiologist for interpretation is not feasible. However, in addition to segmenting the brain in a short period of time, automated segmentation must also provide levels of accuracy and precision that yield results similar to those obtained with manual segmentation, which is currently the gold standard.

Although some automated segmentation algorithms have shown potentially promising results (see review by Dill et al., 2015), many often provide suboptimal results (e.g. de Flores et al., 2015; González-Villà et al., 2016; Grimm et al., 2015; Haller et al., 2016; Næss-Schmidt et al., 2016; Schoemaker et al., 2016) and there is thus ongoing research to develop better algorithms (Akhondi-Asl et al., 2011; Inglese et al., 2015; Mendrik et al., 2015).

Neuroimaging volumetry studies routinely utilize freely-available automated segmentation tools such as FreeSurfer (http://surfer.nmr.mgh.harvard.edu; Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA). Some studies of FreeSurfer have shown deficiencies in automated segmentation of the cerebral cortex (Makris et al., 2008), hippocampus (Cherbuin et al., 2009; de Flores et al., 2015; Grimm et al., 2015; Morey et al., 2009; Wenger et al., 2014), and amygdala (Grimm et al., 2015; Morey et al., 2009; Schoemaker et al., 2016), but data regarding the accuracy and precision of FreeSurfer is not readily available for other important and frequently studied regions including the cingulate gyrus, corpus callosum, and lateral ventricles, all areas important in the investigation of repetitive head trauma. Moreover, there are no published data that demonstrate whether study outcome measures are concordant or discordant when using automated segmentation as compared to manual segmentation.

Volumetric analysis of the brain is particularly important in individuals with exposure to repetitive head trauma as there is evidence that repetitive head impacts may result in regional brain atrophy (Bernick et al., 2015; Goddeyne et al., 2015; Laurent et al., 2010; McKee et al., 2009). Players of American football have a particularly high exposure to repetitive head impacts. For example, college American football players sustain a median of 420 head impacts per season and some players sustain over 2400 head impacts per season, as measured by accelerometers (Crisco et al., 2011).

The aim of this study was to determine whether or not FreeSurfer automated segmentation can be used reliably, without the need for manual brain volume editing, in studies of repetitive head impact that investigated the volumes of the cingulate cortex (left and right), corpus callosum, amygdala (left and right), hippocampus (left and right), amygdala-hippocampal complex (left and right), and lateral ventricles (left and right) in retired National Football Players (NFL) and same aged controls without history of contact sports or brain injury.

2. Methods

This study utilized data from the Diagnosing and Evaluating Traumatic Encephalopathy using Clinical Tests (DETECT) study, funded by the National Institutes of Health (NIH). The DETECT study details have been described in prior publications (Alosco et al., 2016, 2017; Stamm et al., 2015; Stern et al., 2016). All study procedures were approved by the Boston University Medical Center Institutional Review Board and all neuroimaging procedures were approved by the Partners Institutional Review Board. All subjects provided written, informed consent.

2.1. Participants and procedure

There were two cohorts in the DETECT study: former NFL players with at least 12 years of organized football experience, at least 2 years of active participation in the NFL, and self-reported declines in cognition, mood, and behavior within 6 months of study commencement; and control subjects with no reported history of participation in organized contact sports or traumatic brain injury. All subjects were male, aged 40 to 69 years, spoke English as their first language, had no contraindication to MR imaging or lumbar puncture, and no history or diagnosis of central nervous system (CNS) disease.

Of the 96 enrolled former NFL player subjects, 10 were excluded due to inadequate or absent neuroimaging data, resulting in a final sample size of 86 former NFL players (age: 55.2 ± 8.0 years). Of these 86 subjects, complete neurobehavioral testing results were available for a total of 76 subjects. Neuroimaging data was available for all 28 control group subjects, 3 of whom were excluded due to image quality and 3 more were excluded due to subsequently identified CNS disease, contact sport participation, or history of mild traumatic brain injury, resulting in a final sample size of 22 control subjects (age: 57.0 ± 6.6 years).

All subjects were evaluated according to the DETECT neurobehavioral and neuroimaging protocol, including neuroimaging, structured psychiatric interview, and neuropsychological testing.

2.2. MRI data acquisition

DETECT neuroimaging was performed at Brigham and Women's Hospital on a 3-Tesla MRI system (Verio, Siemens Healthcare, Erlangen, Germany) with a 32-channel head array and the Syngo MR-B17 software suite. Only the T1-weighted magnetization prepared rapid gradient echo (TR = 1800 ms, TI = 1100 ms, TE = 3.36 ms, voxel size = 1 × 1 × 1 mm, acquisition matrix = 256 × 256, flip angle = 7°) sequence was used for this study.

2.3. Image processing

All T1-weighted images were visually inspected for quality. Brain masks of each subject were generated by FreeSurfer 5.3 (http://surfer.nmr.mgh.harvard.edu; Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA) and corrected manually. Each brain was segmented using T1-weighted images and FreeSurfer 5.3. This process yielded label maps of deep gray matter, white matter, and CSF structures (including the hippocampus, amygdala, corpus callosum, and lateral ventricles). This process also yielded parcellation label maps of the cerebral cortex (including the cingulate gyrus) based on gyral and sulcal structures. The FreeSurfer option for utilizing T2 or FLAIR image contrast to improve pial surface estimations along CSF borders was not used for this study. Estimated total intracranial volumes were also calculated using the automated FreeSurfer method (Buckner et al., 2004).

FreeSurfer segmentation and parcellation maps were then loaded into the Editor module of Slicer 4.5.0 (http://www.slicer.org, Surgical Planning Laboratory, Brigham and Women's Hospital, Boston, Massachusetts, USA) (Fedorov et al., 2012) and overlayed on the aligned T1-weighted images with image interpolation turned off.

Following written directions based on the below-described