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

approaches for each region, two trained raters manually corrected the image label maps of the amygdala, hippocampus, corpus callosum, and cingulate gyrus. One rater corrected approximately two-thirds of the cases while the second rater corrected the remainder. All cases were then reviewed by a single neuroanatomist for accuracy. A third trained rater manually corrected the image label maps of the lateral ventricles and a radiologist reviewed the corrected lateral ventricle label maps for accuracy. All investigators and raters were blind to group membership at the time of segmentation and review by a trained expert.

To evaluate inter-observer reliability, a fourth trained rater, following the same written directions, corrected the FreeSurfer image label maps of the amygdala, hippocampus, corpus callosum, and cingulate gyrus in 10 randomly chosen subjects, while a radiologist corrected the FreeSurfer image label maps of the lateral ventricles in 10 randomly chosen subjects. A neuroanatomist reviewed the label maps corrected by the trained raters. These individuals were blind to the previously corrected label maps and to group membership at the time of segmentation.

2.3.1. Amygdala and hippocampus

Amygdala and hippocampus volumes were manually corrected based on an approach described by Gurvits et al., 1996. Label maps were corrected on coronal slices, from anterior to posterior, using sagittal slices for verification. Close attention was given to the anterior portion of the amygdala at the height of the frontotemporal junction as these parts were variably included by FreeSurfer. The posterior boundary of the amygdala was defined as the last coronal slice before the appearance of the mammillary bodies. The anterior border of the hippocampus was defined as the image slice where the mammillary bodies first appeared. The posterior boundary of the hippocampus was defined as the coronal slice where the crus of the fornix was last seen. The amygdala and hippocampus were defined and strictly separate volume entities and therefore volume overlap between the structures was not allowed.

Although the hippocampus and amygdala are typically evaluated independently, the posterior aspect of the amygdala partially overlaps anatomically with the hippocampal head (Kiernan, 2012) and the boundary between the structures is generally not visible below the level of the temporal horn (Chera et al., 2009). As such, the amygdala-hippocampal complex was evaluated as an additional variable by adding the volumes of the two regions together for the purpose of analyzing hypothesized FreeSurfer imprecision in discriminating the boundary of these regions.

2.3.2. Cingulate gyrus

The cingulate gyrus was defined as the gyrus superior to the corpus callosum, identified primarily on sagittal images (Wible et al., 1997). The medial aspect of the cingulate gyrus was confirmed on sagittal images by identifying the callosomarginal fissure on a paramidsagittal slice. Then, working from medial to lateral on sagittal images, and using coronal images for verification, voxels extending to the corpus callosum, paracingulate gyrus, or beyond the rostrum of the corpus callosum into Brodmann's Area 25 were excluded.

2.3.3. Corpus callosum

Given high contrast between the high T1 signal of the corpus callosum relative to surrounding tissues, most borders of the corpus callosum were obvious. As drawn by FreeSurfer, the corpus callosum label maps were confined to the 5 most midline sagittal slices.

2.3.4. Lateral ventricles

Given high contrast between the low T1 signal of the ventricles and surrounding brain, most borders of the lateral ventricles were obvious. Working superiorly to inferiorly on the axial images, and using the coronal and sagittal images for confirmation, separate volumes were drawn for the right and left lateral ventricles, respectively. When the septum pellucidum was not clearly visible, a midline strip of 1 voxel in width was not assigned to either ventricle to ensure separation of the volumes. The volumes were terminated at the foramina of Monro and choroid plexus was excluded. Any voxels identified by FreeSurfer as lateral ventricle but extending into the foramina of Monro, third ventricle, or ambient cisterns were removed.

2.4. Neurobehavioral measures

As part of the DETECT protocol, data from the following neurobehavioral tests and self-report measures were converted to age-, gender-, and education-standardized scores and grouped into four factor scores, based on principal component analyses (Alosco et al., 2016). *Mood and Behavior* included the following tests: Apathy Evaluation Scale, Beck Depression Inventory II, Beck Hopelessness Scale, Barratt Impulsivity Scale, Behavior Rating Inventory of Executive Functioning - Adult Version, Center for Epidemiologic Studies - Depression Scale, Hamilton Depression Rating Scale, and the Brown-Goodwin Lifetime History of Aggression. *Attention and Psychomotor Speed* included: Controlled Oral Word Association Test, Delis-Kaplan Executive Function System Color Word Interference Test (inhibition/switching), Trail Making Test, and the Wechsler Adult Intelligence Scale – Revised Digit Symbol test. *Verbal Memory* included: Neuropsychological Assessment Battery (NAB) Story Learning (phrase unit immediate and delayed recall), and NAB List Learning (short and long delayed recall). *Visual Memory* included: Boston Qualitative Scoring System for the Rey-Osterrieth Complex Figure (immediate recall presence and accuracy and delayed recall presence and accuracy).

2.5. Statistical analyses

Statistical Analysis System (SAS) software (SAS version 9.4; SAS Institute Inc., North Carolina, USA) was used for all statistical analyses. Significance was set at a p-value below 0.05.

2.5.1. Part 1: inter-observer reliability of manually-corrected volumes

Intraclass correlation coefficients (ICC) were calculated to compare the two sets of manually-edited segmentation label map volumes for each region. Levels of correlation as classically described by Cicchetti (1994): low with a correlation coefficient below 0.40, fair from 0.40 to 0.59, good from 0.60 to 0.74, and excellent from 0.75 to 1.00.

2.5.2. Part 2: intraclass correlation of automated and manually-corrected segmentation volumes

Intraclass correlations (ICC) were calculated to compare the automated and manually-corrected segmentation label map volumes for each region. Scatter plots revealed a single outlier subject and that subject was excluded from correlation analysis of the lateral ventricles.

2.5.3. Part 3: brain volume differences between former NFL players and controls using automated segmentation with versus without manual correction

Linear mixed effects regression models were used to estimate between-group volume comparisons (Former NFL Players vs Controls) for the amygdalae, hippocampi, corpora callosa, cingulate gyri, and lateral ventricles. Two regression models were used: 1) a regression model with multivariate outcomes being the amygdalae, hippocampi, and cingulate gyri; and 2) a regression model with multivariate outcomes being the corpus callosum, amygdala-hippocampal complex, and lateral ventricles. All models were controlled for age, body-mass index (BMI), and estimated total intracranial volume (eTIV).

2.5.4. Part 4: association of former NFL player brain volumes with neurobehavioral factors using automated segmentation with versus without manual correction

A linear mixed-effects regression model was performed for all brain

Table 1
Intra-class correlation values of inter-observer reliability for manually-corrected volumes.

Brain region	ICC (95% CI)
Corpus callosum	0.956 (0.864–0.989)
Left amygdala	0.715 (0.264–0.913)
Left hippocampus	0.764 (0.345–0.926)
Left amygdala-hippocampal complex	0.863 (0.576–0.959)
Left cingulate cortex[a]	0.574 (0.011–0.858)
Left ventricle	0.990 (0.965–0.997)
Right amygdala	0.887 (0.656–0.968)
Right hippocampus	0.627 (0.104–0.881)
Right amygdala-hippocampal complex	0.919 (0.741–0.977)
Right cingulate cortex	0.856 (0.576–0.959)
Right ventricle	0.989 (0.930–0.994)

[a]A single outlier in the left cingulate cortex skews the results. Excluding the outlier results in an ICC of 0.724 (0.230–0.919), which is comparable to the ICC for the right cingulate gyrus.
CI = confidence interval.
Dark gray background = excellent correlation.
Light gray background = good correlation.
Levels of correlation as classically described by Cicchetti (1994): low with a correlation coefficient below 0.40, fair from 0.40 to 0.59, good from 0.60 to 0.74, and excellent from 0.75 to 1.00.

regions, excluding the lateral ventricles, to compare segmented brain volumes with grouped neurobehavioral factors. A false discovery rate adjustment was then calculated to account for multiple comparisons. A separate linear regression model was performed for the lateral ventricles due to the typical inverse relationship of lateral ventricle size with brain volumes where ventricle size gets larger as brain parenchyma gets smaller with atrophy (Barron et al., 1976). The neurobehavioral factor models were controlled for age, BMI, eTIV, and years of education.

3. Results

3.1. Part 1: inter-observer reliability of manually-corrected volumes

There was excellent inter-observer correlation for the corpus callosum (ICC = 0.956), left hippocampus (ICC = 0.764), left and right amygdala-hippocampal complex (ICC = 0.863, 0.919), left and right lateral ventricle (ICC = 0.990, 0.989), right amygdala (ICC = 0.887), and right cingulate cortex (ICC = 0.856). Good inter-observer correlation was present in all other regions except for the left cingulate cortex (ICC = 0.574) (Table 1), however there was an outlier identified with left cingulate cortex volumes measuring over 2 standard deviations from the mean and if that outlier was removed then the ICC improved to 0.723. Fair or poor inter-observer correlation was not present in any region.

3.2. Part 2: intraclass correlation of automated and manually-corrected segmentation volumes

The correlation coefficients were high between automated and manually-corrected segmentation volumes for corpus callosum (ICC = 0.911) and right and left lateral ventricles (ICC = 0.977, 0.964). Good correlation was demonstrated with the left and right hippocampus (ICC = 0.637, 0.539), left and right amygdala-hippocampal complex (ICC = 0.731, 0.713) and right cingulate cortex (ICC = 0.639). Lower correlation was demonstrated for the left and right amygdala (ICC = −0.090, −0.170) and for the left cingulate cortex (ICC = 0.351) (Table 2). The correlation of automated and manually-corrected volumes was better for the amygdala-hippocampal complex than for either the amygdala or hippocampus alone, indicating that FreeSurfer better estimates the volume of the amygdala-hippocampal complex than either structure independently.

Table 2
Intra-class correlation values of manually-corrected versus automatically-generated volumes.

Brain region	ICC (95% CI)
Corpus callosum	0.911 (0.871–0.937)
Left amygdala	−0.090 (−0.0272–0.098)
Left hippocampus	0.637 (0.515–0.739)
Left amygdala-hippocampal complex	0.731 (0.629–0.807)
Left cingulate cortex	0.351 (0.175–0.504)
Left ventricle	0.967 (0.956–0.979)
Right amygdala	−0.170 (−0.346–0.017)
Right hippocampus	0.539 (0.393–0.660)
Right amygdala-hippocampal complex	0.713 (0.603–0.792)
Right cingulate cortex	0.639 (0.515–0.739)
Right ventricle	0.980 (0.971–0.986)

Dark gray background = excellent correlation ICC full cohort.
Light gray background = good correlation ICC full cohort.
Levels of correlation as classically described by Cicchetti (1994): low with a correlation coefficient below 0.40, fair from 0.40 to 0.59, good from 0.60 to 0.74, and excellent from 0.75 to 1.00.

3.3. Part 3: brain volume differences between former NFL players and controls using automated segmentation with versus without manual correction

There was generally a larger difference between the mean former NFL player and mean control subject brain region volumes when using manually-corrected volumes than when using uncorrected automated segmentation volumes (Table 3), indicating that there is greater

Table 3
Brain volume comparisons of former NFL players and controls using automated segmentation with versus without manual correction.

Region	Mean difference of volume between former NFL players and controls in mL^3 (p)*	
	With manual correction	Without manual correction
Corpus callosum	121.96 (0.265)	129.98 (0.285)
Left amygdala	176.22 (0.005)	159.07 (0.001)
Left hippocampus	158.44 (0.024)	207.12 (0.087)
Left amygdala-hippocampal complex	315.43 (0.031)	366.19 (0.007)
Left cingulate cortex	575.01 (0.036)	286.12 (0.535)
Left ventricle	2300.38 (0.155)	1138.77 (0.474)
Right amygdala	157.43 (0.012)	165.63 (0.018)
Right hippocampus	146.09 (0.032)	86.03 (0.535)
Right amygdala-hippocampal complex	300.88 (0.031)	251.67 (0.045)
Right cingulate cortex	475.28 (0.032)	80.14 (0.827)
Right ventricle	2619.67 (0.118)	1382.15 (0.376)

*p-Value adjusted for multiple comparisons except for corpus callosum and lateral ventricles as those correlations were calculated in separate regression models.
Dark gray background = significant at $p < 0.05$.

variance in the manually-corrected brain volumes, a finding that is addressed in the discussion. There were also a larger number of statistically significant brain region volume differences between Former NFL Players and Controls when using manually-corrected volumes than when using uncorrected automated segmentation volumes (Table 3), indicating that comparisons performed without manual correction of brain volumes would not have identified all significant group differences in this study.

3.4. Part 4: association of former NFL player brain volumes with neurobehavioral factors using automated segmentation with versus without manual correction

No unedited or manually-corrected volumes were associated with the Mood and Behavior factor score. The manually-edited volumes of the cingulate cortex on the left and right were both associated with the Attention and Psychomotor Speed factor score (left: effect size = 31.053, $p = 0.003$; right: effect size = 25.730, $p = 0.003$), but no unedited volumes were associated with this neurobehavioral factor. No unedited or manually-corrected volumes were associated with the Verbal Memory factor score. The manually-corrected volume of the left ventricle was associated with the Visual Memory factor score (effect size = 3.870, $p = 0.047$), however this statistic would not likely hold under multiple comparisons, while the unedited volumes of the left amygdala, left hippocampus, and left amygdala-hippocampal complex were also associated with the Visual Memory score (amygdala: effect size = 0.798, $p = 0.036$; hippocampus: effect size = 2.617, $p = 0.036$; amygdala-hippocampus complex: effect size = 4.147, $p = 0.014$). There was thus no concordance between which manually-corrected volumes and which unedited volumes reached significance when correlating volumes with neurobehavioral factors (Table 4).

Table 4
Association of former NFL Player brain volumes with neurobehavioral factors using automated segmentation with versus without manual correction.

	Effect size with (-value)* determining significance of NFL player brain volume association with neurobehavioral factor score	
	Volumes with manual correction	Volumes without manual correction
Mood and behavior		
Brain region		
Corpus callosum	−0.335 (0.808)	−0.872 (0.674)
Left amygdala	−1.232 (0.270)	−0.629 (0.173)
Left hippocampus	−0.596 (0.591)	−1.024 (0.515)
Left amygdala-hippocampal complex	−2.350 (0.270)	−2.145 (0.206)
Left cingulate cortex	−5.776 (0.591)	2.377 (0.746)
Left ventricle	1.477 (0.418)	1.536 (0.331)
Right amygdala	−1.253 (0.270)	−1.063 (0.173)
Right hippocampus	−0.870 (0.526)	−1.095 (0.543)
Right amygdala-hippocampal complex	−2.632 (0.270)	−2.543 (0.173)
Right cingulate cortex	−4.199 (0.591)	4.288 (0.674)
Right ventricle	1.865 (0.502)	1.988 (0.473)
Attention and psychomotor speed		
Brain region		
Corpus callosum	2.878 (0.182)	2.976 (0.326)
Left amygdala	0.098 (0.993)	0.032 (0.975)
Left hippocampus	1.903 (0.154)	1.389 (0.472)
Left amygdala-hippocampal complex	2.476 (0.265)	1.584 (0.479)
Left cingulate cortex	31.053 (0.003)	22.880 (0.109)
Left ventricle	1.643 (0.563)	1.364 (0.622)
Right amygdala	0.009 (0.993)	1.194 (0.326)
Right hippocampus	2.109 (0.132)	−0.049 (0.975)
Right amygdala-hippocampal complex	2.628 (0.265)	1.543 (0.479)
Right cingulate cortex	25.730 (0.005)	13.305 (0.463)
Right ventricle	1.224 (0.689)	1.123 (0.672)
Verbal memory		
Brain region		
Corpus callosum	−0.515 (0.951)	−0.601 (0.895)
Left amygdala	−0.266 (0.951)	0.143 (0.895)
Left hippocampus	0.090 (0.951)	0.357 (0.895)
Left amygdala-hippocampal complex	−0.239 (0.951)	0.624 (0.895)
Left cingulate cortex	3.929 (0.951)	0.110 (0.988)
Left ventricle	−0.549 (0.653)	−0.394 (0.635)
Right amygdala	−0.413 (0.951)	0.079 (0.988)
Right hippocampus	0.343 (0.951)	1.187 (0.895)
Right amygdala-hippocampal complex	−0.092 (0.951)	1.336 (0.895)
Right cingulate cortex	−0.934 (0.951)	−4.164 (0.895)
Right ventricle	−0.983 (0.792)	−0.923 (0.846)
Visual memory		
Brain region		
Corpus callosum	0.957 (0.785)	1.275 (0.469)
Left amygdala	0.478 (0.785)	0.798 (0.036)
Left hippocampus	1.274 (0.405)	2.617 (0.036)
Left amygdala-hippocampal complex	2.193 (0.405)	4.147 (0.014)
Left cingulate cortex	3.734 (0.817)	2.550 (0.748)
Left ventricle	3.870 (0.047)	3.776 (0.058)
Right amygdala	−0.123 (0.887)	0.810 (0.279)
Right hippocampus	2.236 (0.083)	1.924 (0.241)
Right amygdala-hippocampal complex	2.636 (0.405)	3.056 (0.069)
Right cingulate cortex	−2.457 (0.821)	−9.957 (0.317)
Right ventricle	4.861 (0.098)	4.123 (0.098)

All p-values adjusted for multiple comparisons except for corpus callosum and lateral ventricles as those correlations were calculated in separate regression models.
Dark gray background = significant at $p < 0.05$.
Note: adjustments for multiple comparisons may lead to slightly different p-values in each study depending on the number of comparisons and the method of adjustment.

4. Discussion

The aim of this study was to determine whether or not FreeSurfer automated segmentation can be used accurately and reliably, without the need for manual brain volume editing, in studies of repetitive head impact that examine the cingulate cortex, corpus callosum, amygdala, hippocampus, and/or lateral ventricles.

4.1. Corpus callosum and lateral ventricles

In this study, automated segmentation volumes of the lateral ventricles and corpus callosum demonstrated excellent correlation with manually-corrected volumes. In addition, group comparison statistical inferences (i.e., whether or not volume differences between former NFL players and controls reached statistical significance) were the same whether using automated segmentation or manually-corrected lateral ventricle volumes and corpus callosum volumes. Similarly, statistical inferences when correlating former NFL player volumes with neurobehavioral scores were the same whether using automated segmentation or manually-corrected lateral ventricle volumes and corpus callosum volumes.

The between-group mean volume differences measured with and without manual correction are very similar for the corpus callosum. For the lateral ventricles, there is a slightly larger difference in the mean volumes when measured with and without manual correction. However, the ranges of values are quite large and the difference remains statistically insignificant. Moreover, the magnitude of the mean volume differences between former NFL players and controls is driven in part by the mixed effects model's correlation between the hemispheres and the model's incorporation of and control for the confounders of intracranial volume and age. For example, the absolute raw difference in size of the left lateral ventricle between former NFL players and controls is 1663 mL with manually corrected volumes and 736 mL (0.02%) with automatically segmented volumes, compared to 2300 mL and 1139 mL as calculated with the mixed effects model. The trend toward slightly higher differences measured with manual correction is described below in Section 4.3.

Post-mortem pathology studies have identified corpus callosum abnormalities in multiple diseases with neurodegenerative and psychiatric components including in schizophrenia (Bigelow et al., 1983;

Rosenthal and Bigelow, 1972), multiple sclerosis (Evangelou et al., 2000), traumatic brain injury (Anderson and Bigler, 1994), and chronic traumatic encephalopathy (McKee et al., 2013; Mez et al., 2017). Neuroimaging studies have identified corpus callosum abnormalities in-vivo in both traumatic brain injury (see reviews by Shenton et al., 2012 and Mu et al., 2017) and in repetitive head impacts (see reviews by Ng et al., 2014 and Koerte et al., 2015), both of which are sustained by American football players. Similarly, abnormal enlargement of the lateral ventricles has been described in chronic traumatic encephalopathy (McKee et al., 2013; Mez et al., 2017) and in other diseases that have neurodegenerative features including multiple sclerosis (e.g., Turner et al., 2003), schizophrenia (e.g., Kempton et al., 2010), Alzheimer's disease (e.g., Nestor et al., 2008), and alcoholism (e.g., Fox et al., 1976). Both corpus callosum volume and lateral ventricle volume could therefore potentially serve as in-vivo biomarkers for neurodegenerative processes.

Given the excellent correlation of automated corpus callosum and lateral ventricle segmentation volumes generated by FreeSurfer with manually-corrected volumes, and the concordance of results when using either automated or manually-corrected volumes in this study, the corpus callosum and lateral ventricles can probably be reliably segmented by FreeSurfer without the need for manual editing, at least in large-scale studies.

4.2. Amygdala, hippocampus, and cingulate gyrus

In this study, automated segmentation volumes of the amygdala, hippocampus, and cingulate gyrus demonstrated poorer correlation with manually-corrected volumes than was the case for the corpus callosum and the lateral ventricles. Furthermore, statistical inferences differed when comparing volume differences between former NFL players and controls and also when evaluating former NFL player volume correlations with neurobehavioral scores, depending on whether automated segmentation or manually-corrected amygdala, hippocampus, and cingulate gyrus volumes were used. These findings are corroborated by markedly discrepant mean volume differences and estimated neurobehavioral score effect sizes in these brain regions, particularly in the cingulate cortices, when using automated segmentation as opposed to manually-corrected volumes.

The correlations of automated segmentation and manually-corrected volumes of the amygdala and hippocampus were however, somewhat improved when combined into a single amygdala-hippocampal complex volume. Moreover, the automated segmentation and manually-corrected amygdala-hippocampal complex volumes yielded the same statistical study results when comparing the volumes of former NFL players and control subjects and yielded 7 concordant results out of 8 when evaluating the relationships of former NFL player volumes with neurobehavioral scores. Similarly, the discrepancies between the mean volume differences and the estimated neurobehavioral score effect sizes are smaller than those seen when evaluating the amygdala and hippocampus as separate structures.

The amygdala, hippocampus, and cingulate gyrus are of great interest in neuroimaging studies of repetitive head impacts (see reviews by Ng et al., 2014 and Koerte et al., 2015), and mild traumatic brain injury (see reviews by Shenton et al., 2012 and Mu et al., 2017). In chronic traumatic encephalopathy, post-mortem studies have identified deposition of hyperphosphorylated tau in the limbic system, particularly in the amygdala and hippocampus but also to a lesser degree in the cingulate gyrus (McKee et al., 2013; Mez et al., 2017; Omalu et al., 2011). As with the corpus callosum and the lateral ventricle volumes, the volumes of the amygdala, hippocampus, and cingulate gyrus could thus potentially serve as in-vivo biomarkers for diseases with neurodegenerative features.

Mirroring the results of several other studies that demonstrated correlation coefficients often substantially < 0.8, this study demonstrates suboptimal correlation of automated segmentation volumes with manually-corrected volumes of the amygdala and hippocampus (Cherbuin et al., 2009; de Flores et al., 2015; Grimm et al., 2015; Morey et al., 2009; Schoemaker et al., 2016; Wenger et al., 2014). This study also demonstrates that, at least in this cohort of former professional football players with history of exposure to repetitive head impacts, failing to manually edit the amygdala and hippocampus volumes generated by FreeSurfer led to very different brain volume group comparison and neurobehavioral correlation study results. Automated segmentation volumes of the amygdala, hippocampus, and cingulate gyrus are thus not adequate for meaningful study on their own without further manual editing.

Results are substantially improved, however, when the amygdala and hippocampus volumes are combined into an amygdala-hippocampal complex. The combined amygdala-hippocampal complex can thus possibly be segmented by FreeSurfer without the need for manual editing for large-scale studies, provided appropriate quality control is performed.

4.3. Technical considerations

There was greater variance in the sizes of the studied brain structures when the volumes were manually-edited, suggesting that the method of segmentation employed by FreeSurfer, which uses an atlas based approach on a training set to label the likely structural location of each voxel (Fischl et al., 2004), may artificially improve precision. This artificial precision may be related to the inherently normalizing process of stretching a fixed image atlas to brains that differ from the atlas in unique and variable ways that are better captured in the manual correction process. The results of a prior study, which showed that FreeSurfer relatively overestimates the size of smaller hippocampi but not larger hippocampi (Wenger et al., 2014), supports this hypothesis. Although artificial precision could be expected to spuriously increase statistical power, this study demonstrates more statistically significant results with manually-edited volumes, a finding presumably due to better accuracy achieved through the manual editing process because the manual editing process allows for careful evaluation of deep brain structure with low-contrast borders that may not be within the resolving power of FreeSurfer. In addition, as previously described by Wenger et al., 2014, manual segmentation and manual editing typically follow rules that define the often unclear borders between amygdala and the hippocampus and between the tail of the hippocampus and the lateral ventricle whereas FreeSurfer tends to be more inclusive in these areas (Fig. 1). Although Wenger et al. used a set of rules described by Pruessner et al. (2000) and this study used a set of rules described by Gurvits et al. (1996), this study replicates the qualitative results described by Wenger et al.

4.4. Limitations

Just as there are significant differences between the results obtained with FreeSurfer 5 and prior versions of FreeSurfer (Gronenschild et al., 2012), the results of this paper may therefore not apply to other versions of FreeSurfer. Additionally, this study evaluated differences between measured volumes, but did not account for volume overlap or shape and the results may not be translatable to volume overlap or shape (Morey et al., 2009). The DETECT study also included only male subjects and the results of this study may therefore not be generalizable to female subjects. The DETECT study subjects were former professional football players exposed to repetitive head impacts and the conclusions drawn regarding the effect of using automated segmentation versus manually-corrected volumes on group comparison and neurobehavioral study results may not be translatable to studies of subjects with different pathologies. However, the conclusions drawn regarding automated segmentation and manually-corrected volume correlations are likely generalizable. Although the data set for this study appears too small to benefit from the multiple imputation statistical analysis

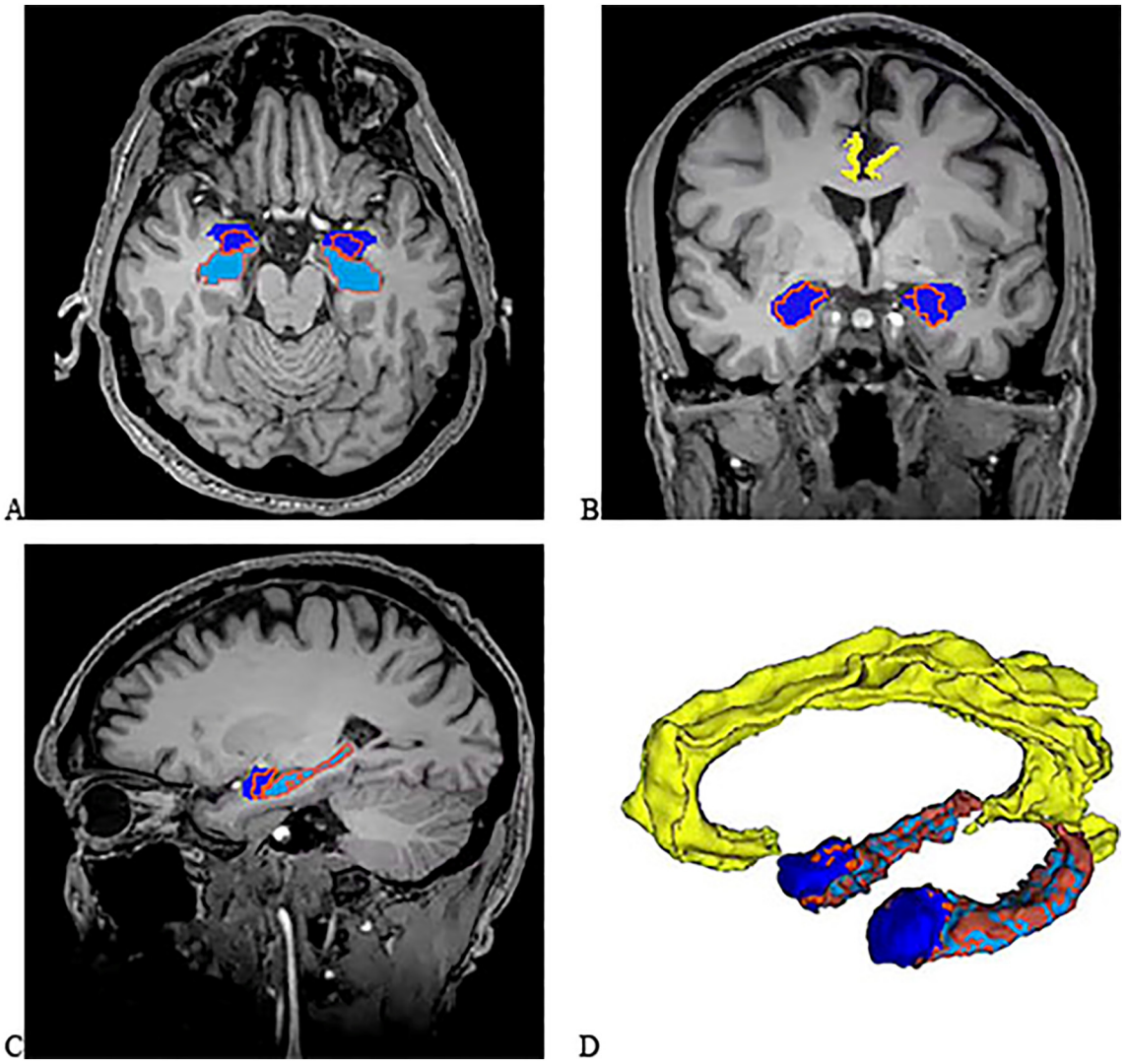

Fig. 1. (A) Axial 2D, (B) Coronal 2D, and (C) Sagittal 2D T1-weighted images of the brain. The manually-edited label map surfaces of the right and left amygdala are dark blue while the manually-edited label map surfaces of the right and left hippocampus are light blue. The outline of the label maps as drawn by FreeSurfer without manual correction are superimposed as orange outlines. The cingulate gyrus label map is presented in yellow. (D) Obliqued 3D reconstruction of the FreeSurfer generated label maps (orange) and manually-edited label maps (blue) created from a randomly selected DETECT study subject using the Model Maker module of Slicer 4.5.0. Note the manually-edited volume of the amygdala is substantially larger than the unedited volume as segmented by FreeSurfer. Also note that the unedited FreeSurfer volume of the hippocampus extends more posterior than the manually-edited volume. The label map was not modified for the 2D image creation but was smoothed for the 3D reconstruction using a sinc filter with a smoothing factor of 10. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

method, which can reduce the number of label maps that require manual correction (Chua et al., 2015), studies with larger data sets may benefit from that approach. The image processing pipeline for this study did not utilize the -T2pial or -FLAIRpial options to optimize pial surface estimations. However, these options are typically useful when there is dura within the brain mask that is not adequately removed with skull stripping. In the utilized pipeline, each brain mask was reviewed and manually edited if necessary, thus likely yielding more precise results than the available automated FreeSurfer options.

5. Conclusions

Automated segmentation using FreeSurfer 5.3 yields excellent correlation with manually-edited volumes of the corpus callosum and lateral ventricles but suboptimal correlation for amygdala, hippocampus, and cingulate gyrus. In addition, using automated segmentation volumes leads to substantially different study results than using manually-corrected volumes when correlating brain volumes with neurobehavioral test scores in this cohort of former professional football players. Study result concordance is improved when the amygdala and hippocampus volumes are combined into an amygdala-hippocampal complex. Automated FreeSurfer-derived segmentation volumes of the corpus callosum and lateral ventricles, and amygdala-hippocampus complex may therefore be suitable for analysis without manual correction, provided appropriate quality control is performed. However, automated FreeSurfer-derived segmentation volumes of the amygdala, hippocampus, and cingulate gyrus should not be utilized for analysis without manual correction until further refinements are made to the FreeSurfer algorithm and appropriately tested.

Author disclosure statement

Dr. Robert Stern is a paid consultant to Avanir Pharmaceuticals, Inc. (Aliso Viejo, CA), Biogen (Cambridge, MA), and Eli Lilly (Indianapolis, IN). He receives royalties for published neuropsychological tests from Psychological Assessment Resources, Inc. (Lutz, FL, USA). He is on the Board of Directors of King-Devick Technologies, Inc. (Chicago, IL), and is a member of the NFL Players Association Mackey-White Health and

Safety Committee. For all other authors, no competing financial interests exist.

Acknowledgments

This study was supported by the NIH (R01 NS 078337 (RAS, YT, CMB, NGF, BMM, CC); R01 NS100952 (IK, APL); F31 NS 081957 (JMS); 1F32NS096803-01 (MLA), P30 AG13846 (RAS, YT, MLA, CMB, BMM, CC); P41 EB015902 (OP); T32GM074905 (IW)), and participant travel was partially funded by JetBlue Airlines, the National Football League, and the NFL Players Association. This study was also partly supported by the Else Kröner-Fresenius Foundation, Germany (IK), by LMU Munich's Institutional Strategy LMUexcellent within the framework of the German Excellence Initiative, and by a VA Merit Award (MES, MC). PW and VS were supported by the German Academic Exchange Service PROMOS award. PW was also supported by an international research scholarship by the medical faculty of the Ludwig-Maximillian University Munich. This work was completed in partial fulfillment of VS's dissertation. CL was supported by the Canadian Institutes of Health Research Frederick Banting and Charles Best Doctoral Award. The sponsors had no role in the design and conduct of the study, the collection, analysis, or the interpretation of the data, nor in the preparation, review, or approval of the manuscript, or the decision of submission for publication.

References

Ahmed-Leitao, F., Spies, G., van den Heuvel, L., Seedat, S., 2016. Hippocampal and amygdala volumes in adults with posttraumatic stress disorder secondary to childhood abuse or maltreatment: a systematic review. Psychiatry Res. 256, 33–43. http://dx.doi.org/10.1016/j.pscychresns.2016.09.008.

Akhondi-Asl, A., Jafari-Khouzani, K., Elisevich, K., Soltanian-Zadeh, H., 2011. Hippocampal volumetry for lateralization of temporal lobe epilepsy: automated versus manual methods. NeuroImage 54 (Suppl. 1), S218–226. http://dx.doi.org/10.1016/j.neuroimage.2010.03.066.

Alosco, M.L., Jarnagin, J., Tripodis, Y., Platt, M., Martin, B., Chaisson, C.E., Baugh, C.M., Fritts, N.G., Cantu, R.C., Stern, R.A., 2016. Olfactory function and associated clinical correlates in former national football league players. J. Neurotrauma. http://dx.doi.org/10.1089/neu.2016.4536.

Alosco, M.L., Tripodis, Y., Jarnagin, J., Baugh, C.M., Martin, B., Chaisson, C.E., Estochen, N., Song, L., Cantu, R.C., Jeromin, A., Stern, R.A., 2017. Repetitive head impact exposure and later-life plasma total tau in former national football league players. Alzheimers Dement. Amst. Neth. 7, 33–40. http://dx.doi.org/10.1016/j.dadm.2016.11.003.

Anderson, C.V., Bigler, E.D., 1994. The role of caudate nucleus and corpus callosum atrophy in trauma-induced anterior horn dilation. Brain Inj. 8, 565–569. http://dx.doi.org/10.3109/02699059409151008.

Barron, S.A., Jacobs, L., Kinkel, W.R., 1976. Changes in size of normal lateral ventricles during aging determined by computerized tomography. Neurology 26, 1011–1013.

Bernick, C., Banks, S.J., Shin, W., Obuchowski, N., Butler, S., Noback, M., Phillips, M., Lowe, M., Jones, S., Modic, M., 2015. Repeated head trauma is associated with smaller thalamic volumes and slower processing speed: the professional fighters' brain health study. Br. J. Sports Med. 49, 1007–1011. http://dx.doi.org/10.1136/bjsports-2014-093877.

Bigelow, L.B., Nasrallah, H.A., Rauscher, F.P., 1983. Corpus callosum thickness in chronic schizophrenia. Br. J. Psychiatry 142, 284–287. http://dx.doi.org/10.1192/bjp.142.3.284.

Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z., 2004. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. NeuroImage 23, 724–738. http://dx.doi.org/10.1016/j.neuroimage.2004.06.018.

Busatto, G.F., Diniz, B.S., Zanetti, M.V., 2008. Voxel-based morphometry in Alzheimer's disease. Expert. Rev. Neurother. 8, 1691–1702. http://dx.doi.org/10.1586/14737175.8.11.1691.

Chera, B.S., Amdur, R.J., Patel, P., Mendenhall, W.M., 2009. A radiation oncologist's guide to contouring the hippocampus. Am. J. Clin. Oncol. 32, 20–22. http://dx.doi.org/10.1097/COC.0b013e318178e4e8.

Cherbuin, N., Anstey, K.J., Réglade-Meslin, C., Sachdev, P.S., 2009. In vivo hippocampal measurement and memory: a comparison of manual tracing and automated segmentation in a large community-based sample. PLoS One 4, e5265. http://dx.doi.org/10.1371/journal.pone.0005265.

Chua, A.S., Egorova, S., Anderson, M.C., Polgar-Turcsanyi, M., Chitnis, T., Weiner, H.L., Guttmann, C.R.G., Bakshi, R., Healy, B.C., 2015. Using multiple imputation to efficiently correct cerebral MRI whole brain lesion and atrophy data in patients with multiple sclerosis. NeuroImage 119, 81–88. http://dx.doi.org/10.1016/j.neuroimage.2015.06.037.

Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instrument in psychology. Psychol. Assess. 6, 284–290.

Crisco, J.J., Wilcox, B.J., Beckwith, J.G., Chu, J.J., Duhaime, A.-C., Rowson, S., Duma, S.M., Maerlender, A.C., McAllister, T.W., Greenwald, R.M., 2011. Head impact exposure in collegiate football players. J. Biomech. 44, 2673–2678. http://dx.doi.org/10.1016/j.jbiomech.2011.08.003.

Dill, V., Franco, A.R., Pinho, M.S., 2015. Automated methods for hippocampus segmentation: the evolution and a review of the state of the art. Neuroinformatics 13, 133–150. http://dx.doi.org/10.1007/s12021-014-9243-4.

Evangelou, N., Konz, D., Esiri, M.M., Smith, S., Palace, J., Matthews, P.M., 2000. Regional axonal loss in the corpus callosum correlates with cerebral white matter lesion volume and distribution in multiple sclerosis. Brain 123, 1845–1849. http://dx.doi.org/10.1093/brain/123.9.1845.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J.V., Pieper, S., Kikinis, R., 2012. 3D slicer as an image computing platform for the quantitative imaging network. Magn. Reson. Imaging 30, 1323–1341. http://dx.doi.org/10.1016/j.mri.2012.05.001.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex N. Y. N. 1991 (14), 11–22.

de Flores, R., La Joie, R., Landeau, B., Perrotin, A., Mézenge, F., de La Sayette, V., Eustache, F., Desgranges, B., Chételat, G., 2015. Effects of age and Alzheimer's disease on hippocampal subfields: comparison between manual and FreeSurfer volumetry. Hum. Brain Mapp. 36, 463–474. http://dx.doi.org/10.1002/hbm.22640.

Fox, J.H., Ramsey, R.G., Huckman, M.S., Proske, A.E., 1976. Cerebral ventricular enlargement. Chronic alcoholics examined by computerized tomography. JAMA 236, 365–368.

Goddeyne, C., Nichols, J., Wu, C., Anderson, T., 2015. Repetitive mild traumatic brain injury induces ventriculomegaly and cortical thinning in juvenile rats. J. Neurophysiol. 113, 3268–3280. http://dx.doi.org/10.1152/jn.00970.2014.

González-Villà, S., Oliver, A., Valverde, S., Wang, L., Zwiggelaar, R., Lladó, X., 2016. A review on brain structures segmentation in magnetic resonance imaging. Artif. Intell. Med. 73, 45–69. http://dx.doi.org/10.1016/j.artmed.2016.09.001.

Grimm, O., Pohlack, S., Cacciaglia, R., Winkelmann, T., Plichta, M.M., Demirakca, T., Flor, H., 2015. Amygdalar and hippocampal volume: a comparison between manual segmentation, Freesurfer and VBM. J. Neurosci. Methods 253, 254–261. http://dx.doi.org/10.1016/j.jneumeth.2015.05.024.

Gronenschild, E.H.B.M., Habets, P., Jacobs, H.I.L., Mengelers, R., Rozendaal, N., van Os, J., Marcelis, M., 2012. The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. PLoS One 7, e38234. http://dx.doi.org/10.1371/journal.pone.0038234.

Gurvits, T.V., Shenton, M.E., Hokama, H., Ohta, H., Lasko, N.B., Gilbertson, M.W., Orr, S.P., Kikinis, R., Jolesz, F.A., McCarley, R.W., Pitman, R.K., 1996. Magnetic resonance imaging study of hippocampal volume in chronic, combat-related posttraumatic stress disorder. Biol. Psychiatry 40, 1091–1099. http://dx.doi.org/10.1016/S0006-3223(96)00229-6.

Haller, S., Zaharchuk, G., Thomas, D.L., Lovblad, K.-O., Barkhof, F., Golay, X., 2016. Arterial spin labeling perfusion of the brain: emerging clinical applications. Radiology 281, 337–356. http://dx.doi.org/10.1148/radiol.2016150789.

Hulshoff Pol, H.E., Kahn, R.S., 2008. What happens after the first episode? A review of progressive brain changes in chronically ill patients with schizophrenia. Schizophr. Bull. 34, 354–366. http://dx.doi.org/10.1093/schbul/sbm168.

Inglese, P., Amoroso, N., Boccardi, M., Bocchetta, M., Bruno, S., Chincarini, A., Errico, R., Frisoni, G.B., Maglietta, R., Redolfi, A., Sensi, F., Tangaro, S., Tateo, A., Bellotti, R., Alzheimer's Disease Neuroimaging Initiative, 2015. Multiple RF classifier for the hippocampus segmentation: method and validation on EADC-ADNI harmonized hippocampal protocol. Phys. Medica PM Int. J. Devoted Appl. Phys. Med. Biol. Off. J. Ital. Assoc. Biomed. Phys. AIFB 31, 1085–1091. http://dx.doi.org/10.1016/j.ejmp.2015.08.003.

Kantarci, K., Jack, C.R., 2003. Neuroimaging in Alzheimer disease: an evidence-based review. Neuroimaging Clin. N. Am. 13, 197–209.

Kempton, M.J., Stahl, D., Williams, S.C.R., DeLisi, L.E., 2010. Progressive lateral ventricular enlargement in schizophrenia: a meta-analysis of longitudinal MRI studies. Schizophr. Res. 120, 54–62. http://dx.doi.org/10.1016/j.schres.2010.03.036.

Kiernan, J.A., 2012. Anatomy of the temporal lobe. Epilepsy Res. Treat. 2012, 176157. http://dx.doi.org/10.1155/2012/176157.

Koerte, I.K., Lin, A.P., Willems, A., Muehlmann, M., Hufschmidt, J., Coleman, M.J., Green, I., Liao, H., Tate, D.F., Wilde, E.A., Pasternak, O., Bouix, S., Rathi, Y., Bigler, E.D., Stern, R.A., Shenton, M.E., 2015. A review of neuroimaging findings in repetitive brain trauma. Brain Pathol. Zurich Switz. 25, 318–349. http://dx.doi.org/10.1111/bpa.12249.

Laurent, S., Thibaud, J.L., Hordeaux, J., Reyes-Gomez, E., Delisle, F., Blot, S., Colle, M.A., 2010. Chronic traumatic brain injury in a dog. J. Comp. Pathol. 143, 75–80. http://dx.doi.org/10.1016/j.jcpa.2009.12.008.

Makris, N., Gasic, G.P., Kennedy, D.N., Hodge, S.M., Kaiser, J.R., Lee, M.J., Kim, B.W., Blood, A.J., Evins, A.E., Seidman, L.J., Iosifescu, D.V., Lee, S., Baxter, C., Perlis, R.H., Smoller, J.W., Fava, M., Breiter, H.C., 2008. Cortical thickness abnormalities in cocaine addiction—a reflection of both drug use and a pre-existing disposition to drug abuse? Neuron 60, 174–188. http://dx.doi.org/10.1016/j.neuron.2008.08.011.

McKee, A.C., Cantu, R.C., Nowinski, C.J., Hedley-Whyte, E.T., Gavett, B.E., Budson, A.E., Santini, V.E., Lee, H.-S., Kubilus, C.A., Stern, R.A., 2009. Chronic traumatic encephalopathy in athletes: progressive tauopathy after repetitive head injury. J. Neuropathol. Exp. Neurol. 68, 709–735. http://dx.doi.org/10.1097/NEN.

0b013e3181a9d503.
McKee, A.C., Stern, R.A., Nowinski, C.J., Stein, T.D., Alvarez, V.E., Daneshvar, D.H., Lee, H.-S., Wojtowicz, S.M., Hall, G., Baugh, C.M., Riley, D.O., Kubilus, C.A., Cormier, K.A., Jacobs, M.A., Martin, B.R., Abraham, C.R., Ikezu, T., Reichard, R.R., Wolozin, B.L., Budson, A.E., Goldstein, L.E., Kowall, N.W., Cantu, R.C., 2013. The spectrum of disease in chronic traumatic encephalopathy. Brain J. Neurol. 136, 43–64. http://dx.doi.org/10.1093/brain/aws307.
Mendrik, A.M., Vincken, K.L., Kuijf, H.J., Breeuwer, M., Bouvy, W.H., de Bresser, J., Alansary, A., de Bruijne, M., Carass, A., El-Baz, A., Jog, A., Katyal, R., Khan, A.R., van der Lijn, F., Mahmood, Q., Mukherjee, R., van Opbroek, A., Paneri, S., Pereira, S., Persson, M., Rajchl, M., Sarikaya, D., Smedby, Ö., Silva, C.A., Vrooman, H.A., Vyas, S., Wang, C., Zhao, L., Biessels, G.J., Viergever, M.A., 2015. MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans. Comput. Intell. Neurosci. 2015, 813696. http://dx.doi.org/10.1155/2015/813696.
Mez, J., Daneshvar, D.H., Kiernan, P.T., Abdolmohammadi, B., Alvarez, V.E., Huber, B.R., Alosco, M.L., Solomon, T.M., Nowinski, C.J., McHale, L., Cormier, K.A., Kubilus, C.A., Martin, B.M., Murphy, L., Baugh, C.M., Montenigro, P.H., Chaisson, C.E., Tripodis, Y., Kowall, N.W., Weuve, J., McClean, M.D., Cantu, R.C., Goldstein, L.E., Katz, D.I., Stern, R.A., Stein, T.D., McKee, A.C., 2017. Clinicopathological evaluation of chronic traumatic encephalopathy in players of American football. JAMA 318, 360–370. http://dx.doi.org/10.1001/jama.2017.8334.
Milani, A.C.C., Hoffmann, E.V., Fossaluza, V., Jackowski, A.P., Mello, M.F., 2017. Does pediatric post-traumatic stress disorder alter the brain? Systematic review and meta-analysis of structural and functional magnetic resonance imaging studies. Psychiatry Clin. Neurosci. 71, 154–169. http://dx.doi.org/10.1111/pcn.12473.
Morey, R.A., Petty, C.M., Xu, Y., Hayes, J.P., Wagner, H.R., Lewis, D.V., LaBar, K.S., Styner, M., McCarthy, G., 2009. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. NeuroImage 45, 855–866. http://dx.doi.org/10.1016/j.neuroimage.2008.12.033.
Mu, W., Catenaccio, E., Lipton, M.L., 2017. Neuroimaging in blast-related mild traumatic brain injury. J. Head Trauma Rehabil. 32, 55–69. http://dx.doi.org/10.1097/HTR.0000000000000213.
Næss-Schmidt, E., Tietze, A., Blicher, J.U., Petersen, M., Mikkelsen, I.K., Coupé, P., Manjón, J.V., Eskildsen, S.F., 2016. Automatic thalamus and hippocampus segmentation from MP2RAGE: comparison of publicly available methods and implications for DTI quantification. Int. J. Comput. Assist. Radiol. Surg. http://dx.doi.org/10.1007/s11548-016-1433-0.
Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., Bartha, R., Alzheimer's Disease Neuroimaging Initiative, 2008. Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. Brain J. Neurol. 131, 2443–2454. http://dx.doi.org/10.1093/brain/awn146.
Ng, T.S., Lin, A.P., Koerte, I.K., Pasternak, O., Liao, H., Merugumala, S., Bouix, S., Shenton, M.E., 2014. Neuroimaging in repetitive brain trauma. Alzheimers Res. Ther. 6, 10. http://dx.doi.org/10.1186/alzrt239.
Olabi, B., Ellison-Wright, I., McIntosh, A.M., Wood, S.J., Bullmore, E., Lawrie, S.M., 2011. Are there progressive brain changes in schizophrenia? A meta-analysis of structural magnetic resonance imaging studies. Biol. Psychiatry 70, 88–96. http://dx.doi.org/10.1016/j.biopsych.2011.01.032.
Omalu, B., Bailes, J., Hamilton, R.L., Kamboh, M.I., Hammers, J., Case, M., Fitzsimmons, R., 2011. Emerging histomorphologic phenotypes of chronic traumatic encephalopathy in American athletes. Neurosurgery 69, 173–183. discussion 183. https://doi.org/10.1227/NEU.0b013e318212bc7b.
Pruessner, J.C., Li, L.M., Serles, W., Pruessner, M., Collins, D.L., Kabani, N., Lupien, S., Evans, A.C., 2000. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. Cereb. Cortex N. Y. N 1991 (10), 433–442.
Rosenthal, R., Bigelow, L.B., 1972. Quantitative brain measurements in chronic schizophrenia. Br. J. Psychiatry J. Ment. Sci. 121, 259–264.
Schoemaker, D., Buss, C., Head, K., Sandman, C.A., Davis, E.P., Chakravarty, M.M., Gauthier, S., Pruessner, J.C., 2016. Hippocampus and amygdala volumes from magnetic resonance images in children: assessing accuracy of FreeSurfer and FSL against manual segmentation. NeuroImage 129, 1–14. http://dx.doi.org/10.1016/j.neuroimage.2016.01.038.
Shenton, M.E., Whitford, T.J., Kubicki, M., 2010. Structural neuroimaging in schizophrenia: from methods to insights to treatments. Dialogues Clin. Neurosci. 12, 317–332.
Shenton, M.E., Hamoda, H.M., Schneiderman, J.S., Bouix, S., Pasternak, O., Rathi, Y., Vu, M.-A., Purohit, M.P., Helmer, K., Koerte, I., Lin, A.P., Westin, C.-F., Kikinis, R., Kubicki, M., Stern, R.A., Zafonte, R., 2012. A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. Brain Imaging Behav. 6, 137–192. http://dx.doi.org/10.1007/s11682-012-9156-5.
Stamm, J.M., Koerte, I.K., Muehlmann, M., Pasternak, O., Bourlas, A.P., Baugh, C.M., Giwerc, M.Y., Zhu, A., Coleman, M.J., Bouix, S., Fritts, N.G., Martin, B.M., Chaisson, C., McClean, M.D., Lin, A.P., Cantu, R.C., Tripodis, Y., Stern, R.A., Shenton, M.E., 2015. Age at first exposure to football is associated with altered corpus callosum white matter microstructure in former professional football players. J. Neurotrauma 32, 1768–1776. http://dx.doi.org/10.1089/neu.2014.3822.
Stern, R.A., Tripodis, Y., Baugh, C.M., Fritts, N.G., Martin, B.M., Chaisson, C., Cantu, R.C., Joyce, J.A., Shah, S., Ikezu, T., Zhang, J., Gercel-Taylor, C., Taylor, D.D., 2016. Preliminary study of plasma exosomal tau as a potential biomarker for chronic traumatic encephalopathy. J. Alzheimers Dis. JAD 51, 1099–1109. http://dx.doi.org/10.3233/JAD-151028.
Turner, B., Lin, X., Calmon, G., Roberts, N., Blumhardt, L.D., 2003. Cerebral atrophy and disability in relapsing-remitting and secondary progressive multiple sclerosis over four years. Mult. Scler. Houndmills Basingstoke Engl. 9, 21–27.
Wenger, E., Mårtensson, J., Noack, H., Bodammer, N.C., Kühn, S., Schaefer, S., Heinze, H.-J., Düzel, E., Bäckman, L., Lindenberger, U., Lövdén, M., 2014. Comparing manual and automatic segmentation of hippocampal volumes: reliability and validity issues in younger and older brains. Hum. Brain Mapp. 35, 4236–4248. http://dx.doi.org/10.1002/hbm.22473.
Wible, C.G., Shenton, M.E., Fischer, I.A., Allard, J.E., Kikinis, R., Jolesz, F.A., Iosifescu, D.V., McCarley, R.W., 1997. Parcellation of the human prefrontal cortex using MRI. Psychiatry Res. 76, 29–40.