



Accuracy and bias of automatic hippocampal segmentation in children and adolescents

Annika Herten^{1,2} · Kerstin Konrad^{3,4} · Helga Krinzing³ · Jochen Seitz² · Georg G. von Polier² 

Received: 12 April 2018 / Accepted: 24 November 2018 / Published online: 3 December 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

The hippocampus (Hc) is of great importance in various psychiatric diseases in adults, children and adolescents. Automated Hc segmentation has been widely used in adults, implying sufficient overlap with manual segmentation. However, estimation biases related to the Hc volume have been pointed out. This may particularly apply to children who show age-related Hc volume changes, thus, questioning the accuracy of automated Hc segmentation in this age group. The aim of this study was to compare manual segmentation with automated segmentation using the widely adopted FreeSurfer (FS) and MAGeT-Brain software. In 70 children and adolescents (5–16 years, mean age 10.6 years), T1-weighted images were acquired on one of two identical 3T scanners. Automated segmentation was performed using the FS subcortical segmentation, the FS hippocampal subfields segmentation and the MAGeT-Brain software. In comparison with manual segmentation, volume differences, Dice similarity coefficient (DSC), Bland–Altman plot, intraclass correlation coefficient (ICC) and left–right consistency of automated segmentation were calculated. The average percentage of volume differences (PVD) with manual segmentation was 56.8% for FS standard segmentation, 32.2% for FS subfield segmentation and –15.6% for MAGeT-Brain. The FS Hc subfields segmentation (left/right DSC = 0.86/0.87) and MAGeT-Brain (both hemispheres DSC = 0.91) resulted in a higher volume overlap with manual segmentation compared with the FS subcortical segmentation (DSC = 0.79/0.78). In children aged 5–10.5 years, MAGeT-Brain yielded the highest overlap (DSC = 0.92/0.93). Contrary volume estimation biases were detected in FS and MAGeT-Brain: FS showed larger volume overestimation in smaller Hc volumes, while MAGeT-Brain showed more pronounced volume underestimation in larger Hc volumes. While automated Hc segmentation using FS hippocampal subfields or MAGeT-Brain resulted in adequate volume overlap with manual segmentation, estimation biases compromised the reliability of automated procedures in children and adolescents.

Keywords Pediatric population · FreeSurfer · MAGeT-Brain · Limbic system · Morphometry · Manual segmentation

Introduction

The hippocampus (Hc) plays a central role in memory and learning (Wenger et al. 2014; Squire et al. 2004; Konrad et al. 2009; Moscovitch et al. 2005; Geuze et al. 2005). Differences in hippocampal volume and shape have been detected in a variety of neurological and psychiatric

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00429-018-1802-2>) contains supplementary material, which is available to authorized users.

✉ Georg G. von Polier
gvonpolier@ukaachen.de

¹ Department for Neurosurgery, Universitätsklinikum Essen, Hufelandstraße 55, 45147 Essen, Germany

² Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Clinic Rheinisch-Westfälische Technische Hochschule Aachen, Neuenhofer Weg 21, 52074 Aachen, Germany

³ Child Neuropsychology Section, Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Clinic Rheinisch-Westfälische Technische Hochschule Aachen, Neuenhofer Weg 21, 52074 Aachen, Germany

⁴ JARA-Brain Institute (JBI-II) Molecular Neuroscience and Neuroimaging, Research Centre Jülich, Wilhelm-Johnen-Straße, 52428 Jülich, Germany

disorders in adults, including Alzheimer's disease (Chupin et al. 2009; Nestor et al. 2012; Shen et al. 2010), schizophrenia (Heckers 2001; Csernansky et al. 2002; Okada et al. 2016), temporal lobe epilepsy (Hogan et al. 2006; Akhond-Asl et al. 2011), depression (Ahdidan et al. 2011; Sheline et al. 1996) and obsessive–compulsive disorder (OCD) (Boedhoe et al. 2017). In children and adolescents, the Hc plays a very important role in cognitive development. For example, changes in hippocampal activation during memory-based problem solving have been shown to be associated with cognitive development (Qin et al. 2014). The Hc shows a protracted maturation during childhood and adolescence. It is also influenced by hormonal changes, particularly during adolescence, resulting in a high variability in Hc size, shape and myelination in children and adolescents (Wierenga et al. 2014; Goddings et al. 2014; Giedd et al. 2006). This high variability in Hc size, shape and function during childhood and adolescence may be a reason for the inconsistent findings in many studies examining children and adolescents with psychiatric disorders.

Methodological differences, particularly different brain volume measurement techniques, might also account for these inconsistent findings. To date, manual segmentation has been considered the gold standard for analyzing hippocampal volumes. However, manual segmentation requires a significant amount of time to both master and perform (approximately 30 min per hippocampus) and still shows a degree of variability over time and between raters (Lee et al. 2015; Nugent et al. 2007; Yushkevich et al. 2015). To overcome the limitations of manual segmentation, several semi- and fully automated methods have been developed for hippocampal segmentation (Dill et al. 2015; González-Villà et al. 2016). These methods have been increasingly used in adults (Morey et al. 2009b, 2012; Tae et al. 2008) and to a lesser extent in children and adolescents (Wierenga et al. 2014; Koolschijn et al. 2013; Hoogman et al. 2017), particularly when analyzing large data sets.

Studies comparing automated and manual Hc segmentation exemplify the accuracy of automated measures in adults (Morey et al. 2009a, 2012; Sánchez-Benavides et al. 2010; González-Villà et al. 2016; Tannous et al. 2018). Among the software used to perform automated hippocampus segmentation, FreeSurfer (FS) has been both frequently used and positively evaluated in healthy adults (Morey et al. 2009a; Cherbuin et al. 2009) as well as in different adult clinical populations (Tae et al. 2008; Doring et al. 2011). When compared to other widely used software packages in regards to Hc segmentation accuracy, FS showed favorable results (FS: Hc volume overlap with manual segmentation 82%; ICC = 0.85) compared with FSL (Morey et al. 2009a) and IBASPM (Tae et al. 2008).

The segmentation in automated approaches based on multi-atlas approaches or multiple atlases combined in

a single atlas such as in FS have been considered among the most accurate of methods (Zandifar et al. 2017). The atlases are drawn from young adults, resulting in the best accuracy for this age group. Since the brain reaches approximately 95% of its total size at the age of 6 years (Giedd et al. 1999, 2015), it has been assumed that these segmentation techniques can also be used in children and adolescents. However, the reliability of Hc segmentation in age groups other than middle-aged adults has been recently questioned (Schmidt et al. 2018; Wenger et al. 2014). Both studies reported an age-related bias in Hc volumes segmented with FS (Vers. 5.3/ Version 6.0) in adults aged about 60–70 years that resulted in higher overestimation of Hc volumes in brains with a smaller Hc. The authors discuss a possible impact of a less demarcated Hc in the older age group on the segmentation accuracy of automated pipelines (Wenger et al. 2014). In children, the demarcation of the Hc is more pronounced with increasing age, thus a similar mechanism may compromise the accuracy of automated segmentation in pediatric populations (Schoemaker et al. 2016). Schoemaker et al. compared manual Hc segmentation with automated segmentation using FS (Vers. 4.4) and FSL in children aged 6–11 years. While FS segmentation resulted in more accurate Hc volumes compared to FSL, it also showed a higher overestimation of right Hc volumes in brains with a smaller Hc. The authors questioned the accuracy of automated Hc and amygdala segmentation in children. They also recommended visual inspection and, if necessary, manual correction to ensure validity. However, manual correction requires detailed neuroanatomical skills and can be labor intensive. A second study by (Lyden et al. 2016) in 23 adolescents (15.5–18.7 years) found large volume differences between automated Hc segmentation using NeuroQuant (CorTech Labs, La Jolla, CA), a pipeline that uses FS and age-matched atlases, and manual segmentation (left Hc 82% and right Hc 89%). Additionally, both measures showed relatively low associations using Pearson's correlation ($r = 0.47$, $p < 0.05$ left; $r = 0.53$, $p < 0.01$ right Hc).

Recently, Iglesias et al. presented a novel approach that improves the Bayesian modeling used in FS to augment local image appearance with prior anatomical knowledge (Iglesias et al. 2013). Using Markov chain Monte Carlo techniques to improve Bayesian modeling, this approach yielded better accuracy as well as the possibility to perform Hc subfield segmentation based on a high-resolution atlas (Iglesias et al. 2015). An earlier hippocampus subfields module (Van Leemput et al. 2009) was used in pediatric populations (Teicher et al. 2012), until concerns were published regarding its reliability (Wisse et al. 2014). To date, the accuracy of the Hc subfields approach has not been investigated in children and adolescents.

Multiple automatically generated templates (MAGeT-Brain) is yet another approach used for Hc

subfield segmentation (Pipitone et al. 2014). This multi-atlas approach uses a small atlas library (typically 5 atlases) and improves the segmentation by creating a large template library (typically 21 subjects) built from a subset of the target images, reflecting the neuroanatomy or demographics of the study population. The template library is then used to perform multi-atlas segmentation. By generating a template library from the subjects themselves, the authors report an improved reliability and robustness for various age groups and diseases (Pipitone et al. 2014). The algorithm was positively evaluated in a study investigating Hc segmentation in neonates by creating subject-specific atlases (Guo et al. 2015). MAGeT-Brain has been used in a variety of populations, including children (Decker et al. 2017; Treadway et al. 2015; Sussman et al. 2015). However, the accuracy of this algorithm in Hc segmentation has not yet been evaluated in children and adolescents.

Thus, the aim of the current study was to explore the accuracy of automated segmentation using FS (standard pipeline and hippocampus subfields module) and MAGeT-Brain in children and adolescents aged 5–16 years compared to manual segmentation. Based on previous research, Hc volumes estimated by the FS standard pipeline might show a limited reliability, which might improve with the more elaborate segmentation approaches used in FS subfields and MAGeT-Brain. Given the variability of Hc size and shape throughout development, Hc segmentation accuracy may be associated with age.

Methods

Participants

The study included 68 children and adolescents (61 boys and 7 girls; see also Table 1). Ages ranged from 5.0 to 16.4 years (mean age 10.6 years; SD 2.8 years). The data of all subjects were included from two studies, one cross-sectional and one longitudinal study. Subjects included

were either healthy controls ($n = 40$) or had a history of conduct disorder (study 1, $n = 20$) or a history of dyscalculia (study 2, $n = 7$, all girls) thus possibly adding variance in hippocampal volumes and shapes. Three healthy boys were scanned twice, on average after 0.6 years each. One female participant with dyscalculia was excluded due to failure of the segmentation of MAGeT-Brain, resulting in a final sample size of 70 scans. All scans were analyzed as a whole sample and within two subsamples following a median split. The “children” group included 35 scans with participant age from 5.0 to 10.5 years (mean age 8.2 years, SD 1.5), and the “adolescents” group included 35 scans with participant age from 10.6 to 16.4 years (mean age 13.0 years, SD 1.4). Average total intracranial volume (eTIV), as estimated by FS segmentation, was 1,538,880 mm³ (range 1,102,028–1,964,096 mm³; SD 145,420 mm³).

Healthy control subjects did not have a history of neurological or psychiatric conditions and had normal neurodevelopmental and medical history. All subjects were fluent in the German language and had a full scale IQ of 70 or higher, which was assessed using the WISC (Wechsler Intelligence Scale for Children) (Wechsler 1991, 1999). Exclusion criteria included a history of seizure, head injury, current or past substance use and history or current diagnosis of psychosis, trauma, bipolar disorder or pervasive developmental disorder, or the presence of any metal implants (e.g., cochlear implants, vascular surgical clips, cardiac pacemaker or ventricular peritoneal shunts). All participants and their parents or legally appointed guardians gave written informed consent for participation in the MRI study protocols. The studies were approved by the local ethics committee in accordance with the Declaration of Helsinki.

MRI acquisition

All subjects underwent 3T-MRI scanning (Siemens Trio TIM, Erlangen, Germany) at one of two sites (Aachen, 61 participants; Jülich, 10 participants) with identical scanners that were maintained by the same physicist and employed identical sequences. High-resolution T1 weighted images were recorded using a rapid acquisition gradient echo (MP-RAGE) pulse sequence (TE 2.520 ms, TR 1.900 ms, FlipAngle 9°, FOV 256 mm, matrix size = 256 × 256, 176 slices, slice thickness = 1 mm, voxel size = 0.98 × 0.98 × 1 mm³, duration 4:30 min). An initial visual assessment was made by a physician to detect any relevant brain pathology and to ensure the image quality necessary for both manual and automated segmentation. Raw data were converted into Nifti format for manual segmentation and for further automated processing with the FS and MAGeT-Brain software packages.

Table 1 Sample description

	Children	Adolescents
Total scans	35 ^a	35
Age range	5.0–10.5	10.6–16.4
Mean age (SD)	8.2 (1.5)	13.0 (1.4)
Male gender	26	35
Healthy controls	22	19
Conduct disorder	4	16
Dyscalculia ^b	6	0

^aThree boys (healthy controls) were scanned twice after 0.6 years

^bAll females

Manual tracing

Hippocampal volumes were segmented manually using ITKSnap, Version 2.4 (Yushkevich et al. 2006), a freely available software tool that allows simultaneous delineation of borders in all three orthogonal view planes with a 3D reconstruction (Jack et al. 2011; Yushkevich et al. 2009). Segmentation was performed by one trained expert rater, who was blind to the identity, age, sex, and clinical diagnosis of the subjects, as well as to the results of the automated segmentation methods. The coronal view was chosen as the main segmentation plane and was changed to the sagittal or axial plane if necessary (Konrad et al. 2009). All slices were segmented from posterior to anterior and from medial to lateral to avoid a laterality effect (Maltbie et al. 2012). Boundaries of the hippocampus segmentation were chosen based on both the results of a meta-analytic review (Konrad et al. 2009) as well as the intrinsic anatomic properties of the hippocampus (Duvernoy 1988). Full details of our protocol are given in Online Resource 1.

Automated segmentation using FreeSurfer

Automated segmentation of hippocampal volumes was performed on T_1 images using FS, Version 6.0 (<http://surfer.nmr.mgh.harvard.edu/>). HC volumes were estimated using two different pipelines within the FS software: (1) the recon-all pipeline and (2) a hippocampus subfields module. In brief, the recon-all pipeline uses a fully automated Markov random fields approach to identify cortical and subcortical structures based on probabilistic information available from a library of manually segmented images of adult brains (Fischl 2012; Fischl et al. 2002). This procedure yields volumes of all subcortical structures defined in the atlas within a table called aseg.stats. HC volumes derived from this standard procedure are referred to as FS_STD in this study. Second, a segmentation procedure to perform hippocampal subfield segmentation has been developed (Iglesias et al. 2013). The atlases used in this pipeline were built primarily upon ultra-high resolution (~0.1 mm isotropic) ex vivo MRI data. This procedure yields volumes of hippocampal subfields and based on these subfields the volume of the whole hippocampus is calculated and output by FreeSurfer. This segmentation procedure is referred to as FS_SUBF in this study. The volumes of said hippocampal subfields are reported to agree well with those found in histological studies. Full details of the procedure can be found elsewhere (Iglesias et al. 2015). All reconstructed data were visually checked for segmentation accuracy.

Automated segmentation using MAGeT-Brain

Segmentation of the bilateral hippocampus was performed using the publicly available MAGeT-Brain algorithm (Pipitone et al. 2014; Chakravarty et al. 2013; <https://github.com/CobraLab/MAGeT-Brain>). MAGeT-Brain was used with five high-resolution anatomical atlases from adults (Winterburn et al. 2013) that have been validated for use with MAGeT-Brain (Pipitone et al. 2014). A bootstrapping method was used to generate a large set of candidate labels (votes) based on the input atlases for each voxel in the target-segmented image. Then, these candidate labels were fused using a voxel-wise majority vote (Collins and Pruessner 2010). The input atlases were registered to a subset of 21 subjects, an optimal number shown in previous research (Pipitone et al. 2014), to generate a template library. Template library images were chosen based on a representative sampling of subject age, sex and diagnostic group to model neuroanatomical variability within the cohort. All images were converted to the MINC format, and MAGeT-Brain was implemented using tools included in the MINC (<https://bic-mni.github.io/>), particularly a version of ANTs adapted for MINC tools.

All processing was performed on a homogeneous GNU/Linux-based cluster, comprised of 14,400 cores and a total memory size of 75 TB. All reconstructed data were visually checked for segmentation accuracy.

Statistical analysis

All statistical analyses were conducted using SPSS 24 (IBM Corp., Armonk, NY), with an alpha level set to $p < 0.05$ for all analyses.

The percentage volume difference (PVD) was calculated using the following formula: $PVD = [(V_a - V_m)/V_m] \times 100\%$. In the event that the automated method (V_a) and the manual segmentation (V_m) yielded identical volumes, the resulting PVD was zero, indicating equal volumes within the two segmentations. Increasing PVD indicates a greater volume difference between automated and manual segmentation. Positive values indicate an overestimation, and negative values indicate an underestimation of automated segmented volumes relative to manually segmented volumes.

Spatial overlap of segmentations with the manual protocol was assessed using Dice similarity coefficients (DSC; (Dice 1945)). Namely, $DSC = (2 \times |A \cap B|) / (|A| + |B|)$, where A and B are the number of voxels in each individual segmentation and $|A \cap B|$ is the number of voxels included in both segmentations. DSCs range from 0, indicating no agreement (i.e., no overlap), to 1, indicating perfect agreement (i.e., perfect overlap).

To investigate potential differences among age groups, hemisphere and segmentation methods, we conducted two

three-way repeated measures analysis of variance (ANOVA) on (i) PVD and (ii) DSC with segmentation methods (FS_STD, FS_SUBF and MAGeT-Brain) and hemispheres (left, right) as within-subjects factors, and age group (children/adolescents) as between-subjects factor. When the assumption of sphericity was violated, degrees of freedom were corrected using Greenhouse-Geisser procedures. In an attempt to analyze a mediating effect of manually segmented Hc volume on age group effects, we then calculated a repeated measure ANCOVA on DSC with different methods (FS_STD, FS_SUBF and MAGeT-Brain) and hemispheres (left, right) using within-subjects factors, and differences in age and manually segmented hippocampal volume using covariates.

We further investigated the agreement between manual and automated segmentation by means of Bland–Altman plot. This graphical method is used to illustrate differences between two techniques or raters (Shrout and Fleiss 1979). As the manual segmentation is accepted and viewed as the gold standard technique for hippocampal volume estimation, we plotted the difference between automated and manually segmented volumes, as has been suggested previously (Krouwer 2008; Giavarina 2015).

To obtain a measure of consistency between segmentation approaches, we calculated intraclass correlation coefficients (ICCs) using two-way mixed models (Koch 2006). While official guidelines for the interpretation of ICCs are lacking, an ICC of 0.70 is considered the minimum standard for adequate reliability (Terwee et al. 2007). Finally, to assess within-method consistency, Pearson correlations were calculated between volumes of bilateral Hc within the same segmentation technique. Weak left versus right hemisphere correlations would indirectly suggest inconsistency within the assessed segmentation method. Furthermore, if a given automated technique is consistent with manual segmentation, similar correlations between left and right hemisphere volumes are expected. Therefore, the differences in

magnitude between within-method correlations were calculated based on the Fisher r -to- Z transformation (Raghunathan et al. 1996), and the Bonferroni–Holm correction was applied for multiple comparisons.

Results

Analysis of manual segmentation consistency

The ICC for intrarater reliability was 0.93 (Shrout and Fleiss 1979), mean Dice similarity coefficient (DSC, Dice 1945) was 0.977.

Analysis of volume differences among segmentation technique and age group

Percentages of volume differences (PVD) were calculated separately for the left and right hippocampus and are displayed in Table 2. The mean percentage of volume differences between FS-derived volumes (both hemispheres combined) and manual segmentation was 56.8% (SD 21.2) when using the FS_STD routine and 32.2% (SD 17.7) when using the FS_SUBF module. When computing the difference between MAGeT-Brain-derived volumes and manual segmentation, the mean volume difference was -15.2% (SD 11.3). For subsequent analyses, absolute values of volumes differences were used.

The repeated measures ANOVA revealed a significant effect of technique $F(1.09) = 278.2$, $p < 0.001$, a significant interaction of age group \times technique ($F(1.09) = 29.6$, $p < 0.001$) and a main effect of age group ($F(1) = 22.4$, $p = 0.001$). The interaction between technique and hemisphere was not significant ($F(1.48) = 0.5$, $p = 0.50$). Post hoc tests (paired t tests) revealed significant PVD between FS_STD, FS_SUBF and MAGeT-Brain among the children age group (all $p < 0.001$) indicating larger volume differences

Table 2 Comparison of volumes between segmentation methods

	Manual		FreeSurfer (standard)				FreeSurfer (subfields)				MAGeT-Brain			
	Volume	SD	Volume	SD	PVD	SD	Volume	SD	PVD	SD	Volume	SD	PVD	SD
All ($n = 70$)														
L-hippocampus	2773.5	421.2	4233.2	362.0	55.2	20.8	3574.4	324.5	31.0	17.1	2321.1	228.8	-15.0	11.4
R-hippocampus	2795.4	438.3	4350.7	370.5	58.4	21.6	3663.4	327.5	33.4	18.4	2330.0	228.2	-15.3	11.2
Age 5–10.5 ($n = 35$)														
L-hippocampus	2607.8	414.9	4181.7	366.2	63.3	22.6	3514.9	298.8	37.2	18.0	2292.9	233.0	-10.6	12.4
R-hippocampus	2606.7	413.0	4284.6	319.6	67.5	23.6	3613.1	266.7	41.1	18.7	2311.4	220.7	-9.9	11.8
Age 10.6–16 ($n = 35$)														
L-hippocampus	2939.1	362.5	4284.7	355.4	47.1	15.4	3633.9	342.1	24.8	13.8	2349.3	224.3	-19.4	8.3
R-hippocampus	2984.1	382.1	4416.7	409.2	49.3	14.9	3713.7	376.0	25.6	14.5	2348.7	237.3	-20.7	7.5

PVD percentage of volume difference with manual tracing

in FS_STD > FS_SUBF > MAGEt-Brain. Among subjects in the adolescent age group, differences were noted between FS_STD and FS_SUBF ($p < 0.001$) (larger PVD FS_STD > FS_SUBF) and between FS_STD and MAGEt-Brain ($p < 0.001$) (larger PVD FS_SUBF > MAGEt-Brain), but not between FS_SUBF and MAGEt-Brain ($T = 2.0$; $p = 0.05$). This indicates that in both age groups, FS_SUBF

and MAGEt-Brain yielded lower PVD than FS_STD. MAGEt-Brain segmentation resulted in smaller volume differences in the younger age group than FS_SUBF and FS_STD, while no differences occurred in the adolescent age group.

Analysis of overlap (DSC) and associations with age

The dice coefficients of overlap (DSC) of the manual and automated segmentation are shown in Table 3. Average values for FS_STD, FS_SUBF and MAGEt-Brain are 0.785, 0.865 and 0.91. The repeated measures ANOVA revealed a significant effect of technique ($F(1.06) = 162.8$, $p < 0.001$), an interaction between age group and technique ($F(1.05) = 29.7$, $p < 0.001$), as well as a main effect of age group ($F(1) = 16.6$; $p = 0.001$). Post hoc tests (paired t tests) revealed significant differences in DSC between FS_STD, FS_SUBF and MAGEt-Brain among the children age group (all $p < 0.001$). Among subjects in the adolescent age group, differences were noted between FS_STD and FS_SUBF ($p < 0.001$) and between FS_STD and MAGEt-Brain ($p < 0.001$), but not between FS_SUBF and MAGEt-Brain ($T = 0.5$, $p = 0.6$). Figure 1 illustrates the differences in Dice coefficients between FS_SUBF and MAGEt-Brain relative to subject age.

Exploratory analyses with Pearson's correlations of age and manually segmented Hc volumes indicated larger Hc volumes in older subjects ($r = 0.49$; $p < 0.001$). Analyzing the correlations of age and automatically segmented

Table 3 Dice coefficients

	FreeSurfer (STD) Dice (SD)	FreeSurfer (SUBF) Dice (SD)	MAGEt-Brain Dice (SD)
All ($n = 70$)			
L-hippocampus	0.79 (0.062)	0.87 (0.063)	0.91 (0.052)
R-hippocampus	0.78 (0.062)	0.86 (0.065)	0.91 (0.054)
Age 5–10.5 ($n = 35$)			
L-hippocampus	0.76 (0.065)	0.85 (0.064)	0.92 (0.049)
R-hippocampus	0.75 (0.065)	0.83 (0.063)	0.93 (0.052)
Age 10.6–16 ($n = 35$)			
L-hippocampus	0.81 (0.050)	0.89 (0.054)	0.89 (0.051)
R-hippocampus	0.81 (0.046)	0.89 (0.054)	0.88 (0.042)

FreeSurfer (STD) FreeSurfer—Standard segmentation; FreeSurfer (SUBF) FreeSurfer—hippocampus subfields segmentation

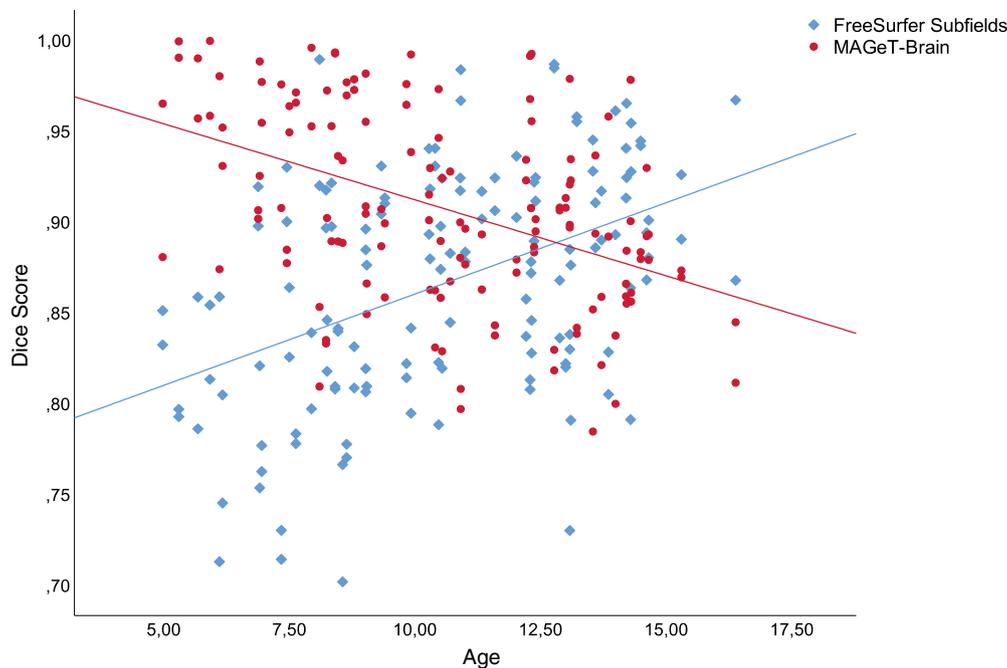


Fig. 1 Dice coefficients of automated segmentation with FreeSurfer subfields and MAGEt-Brain relative to age

Hc volumes (left/right in one model) yields the following numbers: FS_STD: $r=0.189$; $p=0.025$; FS_SUBF $r=0.22$; $p=0.009$; MAGeT-Brain $r=0.172$; $p=0.043$. Thus, while the correlation of hc volume and age holds up, it is significantly smaller than the correlation with manual hc volumes (Fisher r -to- Z transformation: all $p < 0.01$).

To assess the respective influence of age and manually segmented Hc volume on segmentation accuracy as measured with DSC, an ANCOVA with DSC of FS_SUBF (both hemispheres) as dependent variable and age and manually segmented hippocampal volume as covariates was calculated and indicated a main effect of hippocampal volume ($F(1)=158.3$, $p < 0.001$), but not age ($F(1)=0.40$, $p=0.53$). This finding was replicated when using the DSC of FS_STD and MAGeT-Brain, respectively, as dependent variables with similar covariates.

To control for effects of scanner, sex and clinical diagnoses, explorative repeated ANOVAs were conducted with regard to PVD and DSC controlling for age and eTIV. No effects of scanner and diagnoses of conduct disorder were noted (all $p > 0.15$). The six girls with dyscalculia (mean age 8.0 years, SD 0.9 years) showed smaller intracranial volumes (1,309,193 mm³, SD 123,122 mm³) and smaller manual Hc volumes (left Hc 2135 mm³, right Hc 2,164 mm³). Thus, our assumption of added variance with the inclusion of these subjects was confirmed. Excluding (i) the female participants and (ii) the subjects ($n=10$) from site “Jülich” from all analyses, respectively, did not change the results.

Analysis of estimation biases

Bland–Altman graphs, which plotted raw volume differences between manual and automated segmentation, confirmed that FS (both STD and SUBF) yielded larger volumes than manual segmentation, while MAGeT-Brain segmentation resulted in smaller overall Hc volumes (Figs. 2, 3).

In all plots, the incorporated regression line highlights a trend between volume difference and manual segmentation volume. Note however, that the volume differences are positive in FreeSurfer and negative in MAGeT-Brain. Therefore, the plots suggest that FS shows a larger overestimation of subjects with smaller Hc volumes. In contrast, MAGeT-Brain shows larger underestimation of subjects with larger Hc volumes.

Exploratory correlation analyses between volume differences and manually segmented Hc volume (see Table 4) indicate negative associations in FS and in MAGeT-Brain; however, the negative values in MAGeT-Brain must be interpreted in the context of overall undersegmentation (i.e., negative volume differences with manual segmentation). Thus, neuroanatomical features may systematically influence automatically segmented Hc volumes in FS and MAGeT-Brain in opposite directions.

Moreover, estimation biases were more pronounced in the children group than in the adolescent group in the FreeSurfer segmentation. In the children group but not in the adolescent group, age was associated with volume differences with medium effect sizes. Calculating a linear regression model to assess the volume differences over age in the whole group revealed decreasing volume differences per year of -49.2 mm³ per year (95% CI: -70 to -28 ; $p < 0.001$) in FS_STD, of -48.4 mm³ per year (95% CI: -68 to -28 ; $p < 0.001$) in FS_SUBF and increasing (more negative) volume differences of -60.1 mm³ per year (95% CI: -78 to -42 ; $p < 0.001$) in MAGeT-Brain.

Intraclass correlation coefficients

Intraclass correlation coefficients (ICC, agreement type) are shown in Table 5. Average values for FS_STD, FS_SUBF and MAGeT-Brain are 0.72, 0.71 and 0.66. ICC values of segmentations performed in FS yielded a slightly larger agreement with the manual comparison than those performed in MAGeT-Brain. However, confidence intervals in all segmentation methods were widely overlapping.

Within-method correlation analysis

Pearson correlations between volumes in the left and right hemispheres obtained from each segmentation technique are summarized in Table 6. Hc volumes within a given segmentation technique were all highly correlated (all $p < 0.001$; $r=0.709$ – 0.874). A significant difference in magnitude between the computed correlations was observed only between within-method correlations of Hc volumes estimated with manual segmentation and the FS_STD procedure (ZPF = 2.6; $p < 0.01$). This result suggests a stronger association between bi-hemispheric volumes when estimated with manual segmentation rather than with FS_STD. No significant differences were found between the strength of within-method correlation of Hc volumes when comparing manual segmentation and FS_SUBF (ZPF = 1.95; $p=0.051$), or manual segmentation and MAGeT-Brain (ZPF = -0.05 ; $p=0.96$).

Discussion

Although previous research in adults has shown sufficient overlap of FS and MAGeT-Brain Hc volume estimates with manual segmentation Hc volumes, the accuracy of these automated segmentation procedures in children and adolescents remains scarcely investigated, particularly given potential differences in brain size and shape over time. Additionally, subcortical structures may develop ahead of cortical structures (Casey et al. 2008). Thus, it is imperative

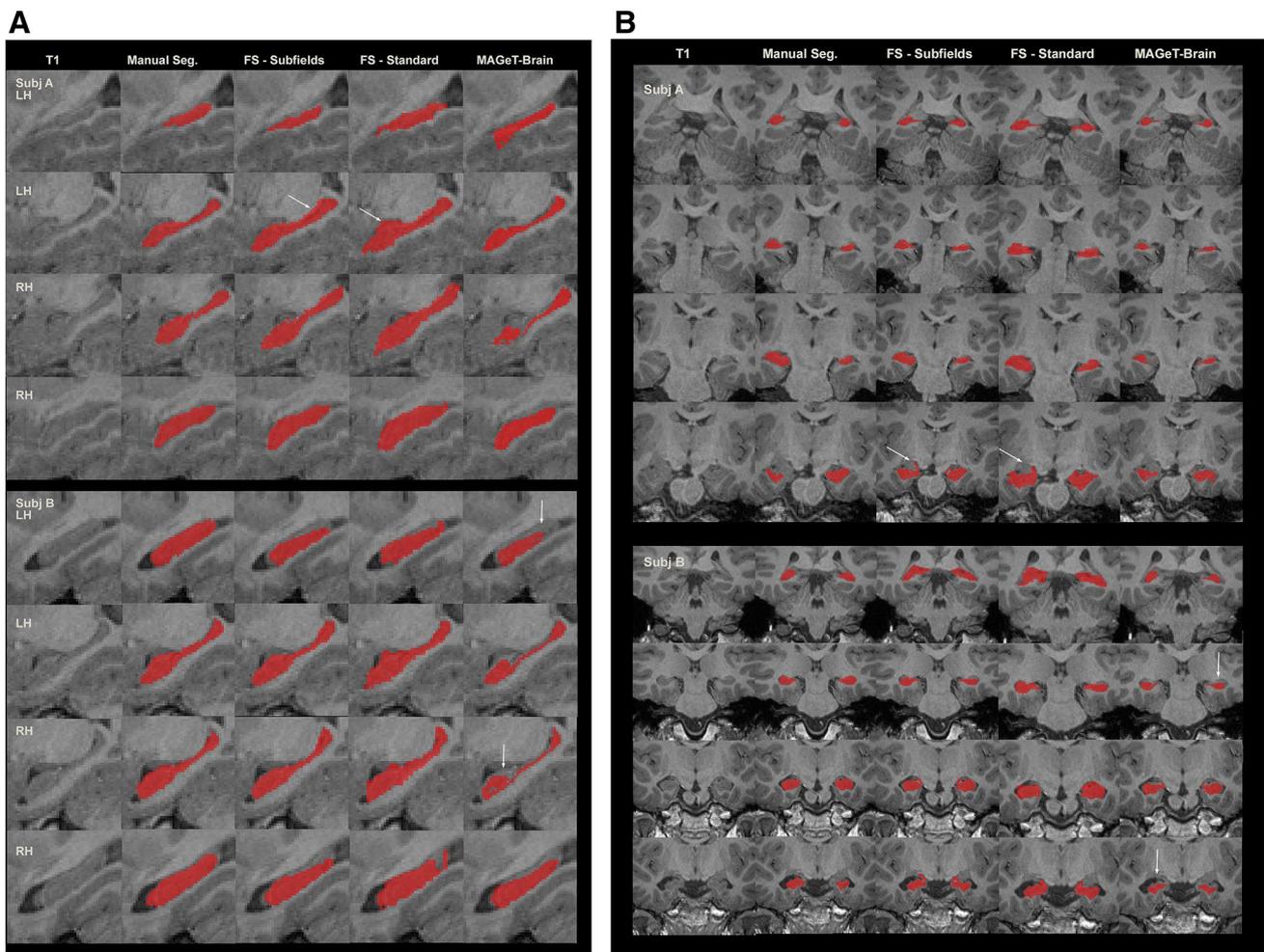


Fig. 3 a Comparison of segmentation inaccuracies in two subjects selected based on low DSC-Scores in FreeSurfer (Subj. A) and MAGeT-Brain (Subj. B). Arrows point to examples of oversegmentation in FreeSurfer (Subj. A) and underestimation in MAGeT-Brain (Subj. B)—sagittal view. **b** Comparison of segmentation inaccura-

cies in two subjects selected based on low DSC-Scores in FreeSurfer (Subj. A) and MAGeT-Brain (Subj. B). Arrows point to examples of oversegmentation in FreeSurfer (Subj. A) and underestimation in MAGeT-Brain (Subj. B)—coronal view

Table 4 Correlation of volume differences with manual volume estimates and age

	FreeSurfer (STD)		FreeSurfer (SUBF)		MAGeT-Brain	
	r_{Vol}	r_{Age}	r_{Vol}	r_{Age}	r_{Vol}	r_{Age}
All ($n = 70$)	-0.58	-0.37	-0.69	-0.38	-0.85	-0.49
Age 5–10.5	-0.62 ^a	-0.29*	-0.73 ^b	-0.28*	-0.84	-0.32*
Age 10.6–16	-0.40 ^a	ns	-0.50 ^b	ns	-0.79	ns

Bilateral Hc included ($n = 140$ hc); all $p < 0.001$ if not indicated otherwise

Note: due to overall negative volume differences with manual segmentation in MAGeT-Brain negative numbers indicate opposite direction compared with FS

* $p < 0.05$, ns not significant, FreeSurfer (STD) FreeSurfer—standard segmentation; FreeSurfer (SUBF) FreeSurfer—hippocampus subfields, Vol manual volume estimate

^{a,b}Comparing r with Fisher r -to- Z : $p < 0.01$

Table 5 Intraclass correlation coefficients

	FreeSurfer (STD) ICC (CI)	FreeSurfer (SUBF) ICC (CI)	MAGeT-Brain ICC (CI)
All ($n=70$)			
L-hippocampus	0.70 (0.52–0.82)	0.71 (0.53–0.82)	0.65 (0.44–0.78)
R-hippocampus	0.74 (0.58–0.84)	0.71 (0.53–0.82)	0.67 (0.47–0.80)
Age 5–10.5 ($n=35$)			
L-hippocampus	0.69 (0.39–0.84)	0.72 (0.45–0.86)	0.65 (0.30–0.82)
R-hippocampus	0.65 (0.30–0.82)	0.70 (0.40–0.85)	0.64 (0.28–0.82)
Age 10.6–16 ($n=35$)			
L-hippocampus	0.71 (0.43–0.85)	0.67 (0.35–0.84)	0.70 (0.40–0.85)
R-hippocampus	0.81 (0.62–0.90)	0.72 (0.45–0.86)	0.79 (0.59–0.89)

ICC intraclass correlation coefficient, agreement type; CI confidence intervals, STD FreeSurfer—Standard segmentation, SUBF FreeSurfer: hippocampus subfields segmentation

to validate these automated methods in children and adolescents. The main finding of this study is an acceptable agreement of both automated and manual segmentation within the FS_SUBF pipeline and the MAGeT-Brain algorithm, as indicated by DSC between 0.86 and 0.91, but less so for the FS_STD pipeline (DSC 0.76–0.81). Notably, Hc segmentation with FS resulted in overestimation of Hc volumes, which was more pronounced in the FS_STD pipeline and in younger subjects, respectively, those with a smaller Hc, as estimated by manual segmentation. Contrary, Hc segmentation within the MAGeT-Brain framework tended to underestimate Hc volumes, particularly in those with larger Hc volumes. The findings on estimation biases thus question the reliability and validity of automated Hc segmentation in children and adolescents.

The manually segmented volumes in this study agree with those reported in previous studies (Schoemaker et al. 2016; Schumann et al. 2004). Schoemaker et al. employed a similar Hc segmentation protocol in children aged 6–11 years, which yielded similar average Hc volumes (left Hc 2746.3 mm³, right Hc 2786.9 mm³) (Schoemaker et al. 2016). Thus, in

both studies, Hc volumes were smaller than those reported in manual Hc segmentation studies of adults, which showed average Hc volumes of approximately 3200–3500 mm³ (left Hc) and 3300–3600 mm³ (right Hc) (Nugent et al. 2013; Wenger et al. 2014; Schmidt et al. 2018).

Hc volumes from the FS_STD pipeline, however, were substantially higher than manual segmented volumes and do not appear to vary considerably among age groups on study: In adults, Hc volumes in the Wenger study were largely similar (left Hc 4275 mm³; right Hc 4244 mm³) to volumes reported in this study as well as those reported in children (left Hc 4378 mm³; right Hc 4195 mm³) (Wenger et al. 2014; Schoemaker et al. 2016). The relative constancy of the Hc volumes using the FS_STD over different age groups with known age-dependent differences in Hc volumes is therefore concerning and may be due to the fact, that for all analyses the same atlas is being used for the FreeSurfer segmentations. In one study in adolescents by (Lyden et al. 2016), age-matched templates instead of the FreeSurfer atlas were used. However, the use of these templates did not seem to largely affect the Hc volumes segmented by FreeSurfer, that were comparable (4546 / 4745 mm³) with those reported in adolescent or adult samples using the same FreeSurfer pipeline but the standard atlas.

Hence, the analysis of volume differences with manual segmentation among the FS_STD pipeline resulted in large volume differences (average 56.8%) in the whole sample. This is in line with results of the Schoemaker et al. study, which reported average volume differences of 56.0% using the same pipeline in an earlier version of FS (Vers. 4.4). Lyden et al. reported even higher volume differences in adolescents (on average 86%) with NeuroQuant, using the same FS pipeline and age-matched templates. However, in this study the manual Hc volumes were smaller (1913/1844 mm³) than reported in this or other studies (Schoemaker et al. 2016). Studying volume differences in young adults, however, has resulted in smaller volume differences between

Table 6 Comparison of inter-hemispheric volume correlation

	Left—right hippocampus
Manual segmentation	0.872
FreeSurfer (standard)	0.709*
FreeSurfer (subfields)	0.761
MAGeT-Brain	0.874

Pearson correlations of left and right hemispheric volumes obtained within a same segmentation method; all $p < 0.001$

* $p < 0.01$; indicates a significant difference in the magnitude of the correlation, as compared with the correlation coefficient obtained with manual segmentation, as defined with the Fisher r -to- Z transform (ZPF) test

FS_STD and manual segmentation of about 25% (Cherbuin et al. 2009) or 22% (Wenger et al. 2014). One possible explanation for the higher volume differences found in children may be that manual segmentation in children resulted in smaller Hc volumes than those found in adults, but Hc volumes using the FS_STD pipeline were similar between children and adults. This discrepancy may partially explain the larger volume differences observed using FS_STD in children and adolescents compared to those reported in adult studies.

Hc volume estimates from FS_SUBF in this study are in close agreement with volumes reported in a group of young and older adults (mean Hc volumes 3363–3565 mm³) (Whelan et al. 2016; Schmidt et al. 2018). Concerning hc volume differences with manual segmentation, we found that the FS_SUBF module resulted in much smaller volume differences than the FS_STD module, particularly in the adolescent group (left Hc 24.7%; right Hc 26.8%). This is in line with findings in adults (mean age 62 years) by (Schmidt et al. 2018) who also indicated much smaller volume differences using the FS_SUBF module over the FS_STD module.

Hc volumes obtained with MAGeT-Brain in other studies are similar to those observed in this study. For example, one study found left Hc volumes of 2418 mm³ in healthy children and adolescents (Sussman et al. 2015). Another study observed volumes of 2476 mm³ (left Hc) and 2535 mm³ (right Hc) in young adults (Patel et al. 2017). Therefore, previous studies of Hc segmentation with MAGeT-Brain resulted in underestimation of Hc volumes (Pipitone et al. 2014) relative to manual segmentation. However, in children and adolescents, this is the first study to report volume differences between both methods.

Moreover, we analyzed the overlap of automated and manually segmented Hc volumes by means of the Dice similarity coefficient (DSC), a widely adopted procedure to compare automated and manual segmentation. Using the FS_STD pipeline, volume overlap in children aged 5–10.5 years in this study was similar to those reported in a study in children aged 2–4 years, with DSC between 0.74 and 0.76 (Lee et al. 2015). Furthermore, we found that volume overlap in adolescents (10.6–16 years) was similar to results reported in adults, such as a DSC of 0.82 in a study by Morey et al. (Morey et al. 2009a). Moreover, we found that applying the FS_SUBF procedure resulted in a significantly improved overlap with manual segmentation, though still relatively large volume differences with manual segmentation have to be considered. An improved overlap with manual segmentation using the FS-SUBF module has recently been reported by (Schmidt et al. 2018) in a large adult sample.

Our results on volume overlap using MAGeT-Brain, particularly in children (DSC of 0.93), were higher than those previously published in adult studies (DSC of 0.82) (Pipitone et al. 2014). This higher volume overlap in our

study compared to the adult study may be due to a smaller discrepancy of manually segmented Hc volumes and Hc volumes segmented with MAGeT-Brain in children and adolescents than in adults: In the study of Pipitone, automated Hc volumes in adults were 2476–2535 mm³ (thus close to those in this study), however manual Hc volumes—though not reported—were likely higher based on other adult literature (Pipitone et al. 2014).

The results of the analyses of volume differences in the Bland–Altman plots confirmed the Hc volume overestimation in FS segmentation opposed to the Hc volume underestimation in MAGeT-Brain segmentation. The negative regression line within the Bland–Altman plots pointed toward an estimation bias in all automated segmentation methods. In FS (both pipelines), a volume overestimation of HC volumes is present in all subjects among both modules. This overestimation is larger in subjects with smaller volumes. This is consistent with previous studies in children (Schoemaker et al. 2016), older adults (60–70 years) (Wenger et al. 2014; Schmidt et al. 2018), and one study including 289 participants aged from 16 to 60 years (Akudjedu et al. 2018) that all report a (negative) bias related to the manually segmented Hc volume. Studies including middle-aged adults, however, have not shown an estimation bias dependent on hippocampal size (Morey et al. 2009a; Wenger et al. 2014). Both FS pipelines are highly dependent on prior information in the adult atlases. Given the higher frequency of smaller Hc size, especially in younger children as shown in our study and previous work, the larger Hc size differences in our sample compared to those in the FS Atlases (Fischl et al. 2002; Iglesias et al. 2015) may partially explain the larger volume differences in subjects with smaller Hc volumes.

An aspect for the larger Hc volumes segmented in FreeSurfer is related to the difference in the segmentation protocols: In our protocol, we included a relatively small section of the subiculum (and pre/ parasubiculum), whereas FreeSurfer includes larger sections of this region in its segmentation protocol. Both segmentation strategies are valid and explain a relatively small, though systematic part of the oversegmentation in FreeSurfer. Contrary to the FS segmentation, the Bland–Altman plot of the MAGeT-Brain segmentation indicated a more pronounced underestimation of Hc volumes in subjects with larger Hc volumes, indicating low volume differences in subjects with small Hc size. Previous research in adults using the same pipeline and atlas did not show an estimation bias (Pipitone et al. 2014).

Performing correlational analysis on volume differences and manually segmented Hc volumes confirmed the negative estimation biases, which are partially explained by age. Wenger et al. found a similar estimation bias range in older adults aged 60–70 years (left Hc $r = -0.55$, $p < 0.001$, right Hc, $r = -0.42$, $p = 0.003$).

Furthermore, our data indicate an increase in Hc volume until adolescence (Wierenga et al. 2014), which explains approximately 24% of the variance in manually segmented Hc volume. While subject age and segmentation accuracy were associated, subsequent analysis shows that this association can be mostly explained (or mediated) by manually segmented Hc size, which is consistent with previous research (Schoemaker et al. 2016).

Analyses of the reliability by means of ICC showed a general low to acceptable reliability. All automated segmentation ICC values were approximately 0.7, which is considered acceptable (Terwee et al. 2007), though this also indicates a limited power to detect relationships of interest. This finding is consistent with the results of Schoemaker et al., who reported ICC values for FS Hc segmentation of 0.68–0.74 in children. In comparison, adult literature indicates much higher reliability, with ICC values of approximately 0.9 (Morey et al. 2009a; Pardoe et al. 2009). While ICC values are penalized by the (partially consistent) volume differences between automated and manual segmentation, the smaller ICC in the children samples compared to adult samples may reflect the larger volume differences (between manual and automated hc volumes) and age/volume-related segmentation biases in the automated segmentation methods.

Moreover, a within-method correlational analysis of left/right Hc volumes resulted in strong agreement between left and right Hc volumes, but this association was significantly weaker in FS_STD ($r=0.709$) compared to manual segmentation ($r=0.872$). This finding suggests that in children and adolescents, the FS_STD segmentation may be prone to inconsistencies within the same subject. Previous research has been inconsistent in this regard (Schoemaker et al. 2016; Wenger et al. 2014).

Comparing the automated segmentation methods, we found volume overestimation in both FreeSurfer modules and volume underestimation in MAGeT-Brain, thought by a comparatively smaller margin than FreeSurfer. Moreover, higher volume overlap with manual segmentation was noted when using FS_SUBF or MAGeT-Brain, compared to the FS_STD pipeline, in the whole sample as well as both age groups. While the FS_SUBF module and the MAGeT-Brain algorithm yielded comparable overlap in adolescents (10.6–16 years, both DSC 0.89), the Hc segmentation in children was significantly more accurate when using MAGeT-Brain.

Estimation biases affected all automatic segmentation with large effect sizes, most pronounced in MAGeT-Brain. Age-dependent volume differences in FreeSurfer were larger than those published in an adult study of Schmidt et al. (2018) on subjects from 38 to 84 years (FS Vers 6.0.0, hc subfields module: 12 mm³ decrease per year). The difference to the adult study is remarkable and, as discussed earlier, may be due to the more pronounced brain development in

this age period and thus more differences with the (adult) atlases used in FreeSurfer and MAGeT-Brain. Also, a different hc segmentation protocol was used in the study by Schmidt et al.. Moreover, age-dependent estimation biases were detected particularly in the children group and they were more pronounced than in the adolescent group using FreeSurfer. Thus, the data in this study—future replication pending—suggest using MAGeT-Brain in children given the overall higher accuracy with regard to volume difference in this age group also compared to the results of Schoemaker et al. (2016) in this age group (using FreeSurfer Vers 4.4). However, using MAGeT-Brain in a study comparing groups may be limited due to its pronounced estimation bias, also occurring in the children group. In the adolescent age group, the FreeSurfer subfields module may result in better accuracy given the lower volume differences than FS_STD and lower estimation biases than MAGeT-Brain.

Limitations of this study apply to the uneven sex ratio. Female whole brain volumes and subcortical structures including the Hc are smaller than male Hc volumes (Jancke et al. 2015), therefore, the absolute Hc volumes have to be interpreted with reservations. When controlling for intracranial volume, however, the sex differences in Hc volumes disappear (Jancke et al. 2015). Since this study primarily compared volume differences between manual and automated segmentation (within the same subject), these findings should therefore not be compromised by the uneven sex ratio. A limitation of the automated methods is, that the atlases used by FreeSurfer and MAGeT-Brain were based on adult templates. To improve the accuracy of automated segmentation in children and adolescents, the use of age-matched templates may be investigated in future research. Third, acquiring higher resolution T2-weighted MR images helps to improve accuracy of the FS_SUBF and MAGeT-Brain pipelines, particularly concerning Hc subfields segmentation (Iglesias et al. 2015). In this study, we were not able to also assess T2-weighted images, thus reflecting the majority of studies published to date that also used T1-weighted MRI only. Therefore, the results of this study may be better applicable to (previous) studies with similar methodology.

Conclusion

While we demonstrate adequate agreement of automated Hc volume estimates with manual segmentation in children and adolescents using the FS_SUBF module and the MAGeT-Brain algorithm, the reliability of automated segmentation must be questioned in this age group. Both FS and MAGeT-Brain showed a considerable estimation bias, resulting in overestimation of small Hc volumes (FS) and underestimation of large Hc volumes (MAGeT-Brain). This estimation

bias was evident in the whole sample and explained associations among segmentation accuracy and age.

Based on the present findings, using manual segmentation might be preferred over automated segmentation when sample size allows. Exploring large databases, however, may require automated segmentation. In this case, it is recommended to analyze both T1- and T2-weighted images if available. Moreover, our data indicate the use of MAGeT-Brain in children and FreeSurfer subfields module in adolescents over the use of the FS standard pipeline. Given the relatively pronounced estimation bias demonstrated in this study, also related to age, we further suggest investigating samples with narrow age ranges. If our findings are replicated in larger samples, the possibility to statistically correct for age-related bias (maybe even in specific hc subfields) may be a subject of future research to improve the applicability of automated Hc volume estimates among a large age range including children and adolescents.

Funding This work was supported by the Federal Ministry of Education and Research [grant number 01GJ0808]; the “Interdisciplinary center for clinical studies, IZKF RWTH Aachen University” [grant number N4-1]. Automatic segmentations were performed with computing resources granted by RWTH Aachen University under project rwth0151. The authors wish to thank all children and their parents for participating in this study.

Compliance with ethical standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Ahdidan J, Hviid LB, Chakravarty MM, Ravnkilde B, Rosenberg R, Rodell A, Stødkilde-Jørgensen H, Videbech P (2011) Longitudinal MR study of brain structure and hippocampus volume in major depressive disorder. *Acta Psychiatr Scand* 123(3):211–219. <https://doi.org/10.1111/j.1600-0447.2010.01644.x>
- 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–S226. <https://doi.org/10.1016/j.neuroimage.2010.03.066>
- Akudjedu TN, Nabulsi L, Makelyte M, Scanlon C, Hehir S, Casey H, Ambati S, Kenney J, O’Donoghue S, McDermott E, Kilmartin L, Dockery P, McDonald C, Hallahan B, Cannon DM (2018) A comparative study of segmentation techniques for the quantification of brain subcortical volume. *Brain Imaging Behav*. <https://doi.org/10.1007/s11682-018-9835-y>
- Boedhoe PS, Schmaal L, Abe Y, Ameis SH, Arnold PD, Batistuzzo MC, Benedetti F, Beucke JC, Bollettini I, Bose A, Brem S, Calvo A, Cheng Y, Cho KI, Dallspezia S, Denys D, Fitzgerald KD, Fouché JP, Giménez M, Gruner P, Hanna GL, Hibar DP, Hoexter MQ, Hu H, Huyser C, Ikari K, Jahanshad N, Kathmann N, Kaufmann C, Koch K, Kwon JS, Lazaro L, Liu Y, Lochner C, Marsh R, Martínez-Zalacáin I, Mataix-Cols D, Menchón JM, Minuzzi L, Nakamae T, Nakao T, Narayanaswamy JC, Piras F, Pittenger C, Reddy YC, Sato JR, Simpson HB, Soreni N, Soriano-Mas C, Spalletta G, Stevens MC, Szeszko PR, Tolin DF, Venkatasubramanian G, Walitza S, Wang Z, van Wingen GA, Xu J, Xu X, Yun JY, Zhao Q, Thompson PM, Stein DJ, van den Heuvel OA, Group EOW (2017) Distinct subcortical volume alterations in pediatric and adult OCD: a worldwide meta- and mega-analysis. *Am J Psychiatry* 174(1):60–69. <https://doi.org/10.1176/appi.ajp.2016.16020201>
- Casey BJ, Jones RM, Hare TA (2008) The adolescent brain. *Ann N Y Acad Sci* 1124(1):111–126. <https://doi.org/10.1196/annals.1440.010> doi
- Chakravarty MM, Steadman P, van Eede MC, Calcott RD, Gu V, Shaw P, Raznahan A, Collins DL, Lerch JP (2013) Performing label-fusion-based segmentation using multiple automatically generated templates. *Hum Brain Mapp* 34(10):2635–2654. <https://doi.org/10.1002/hbm.22092>
- Cherbuin N, Anstey KJ, Réglade-Meslin C, Sachdev PS (2009) In vivo hippocampal measurement and memory: a comparison of manual tracing and automated segmentation in a large community-based sample. *PLoS One* 4(4):e5265. <https://doi.org/10.1371/journal.pone.0005265>
- Chupin M, Géraud E, Cuingnet R, Boutet C, Lemieux L, Lehericy S, Benali H, Garnero L, Colliot O, Initiative AsDN (2009) Fully automatic hippocampus segmentation and classification in Alzheimer’s disease and mild cognitive impairment applied on data from ADNI. *Hippocampus* 19(6):579–587. <https://doi.org/10.1002/hipo.20626>
- Collins DL, Pruessner JC (2010) Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *Neuroimage* 52(4):1355–1366
- Csernansky JG, Wang L, Jones D, Rastogi-Cruz D, Posener JA, Heydebrand G, Miller JP, Miller MI (2002) Hippocampal deformities in schizophrenia characterized by high dimensional brain mapping. *Am J Psychiatry* 159(12):2000–2006
- Decker AL, Szulc KU, Bouffet E, Laughlin S, Chakravarty MM, Skocic J, de Medeiros CB, Mabbott DJ (2017) Smaller hippocampal subfield volumes predict verbal associative memory in pediatric brain tumor survivors. *Hippocampus*. <https://doi.org/10.1002/hipo.22758>
- Dice LR (1945) Measures of the Amount of Ecologic Association Between Species. *Ecology* 26(3):297–302. <https://doi.org/10.2307/1932409>
- Dill V, Franco AR, Pinho MS (2015) Automated methods for hippocampus segmentation: the evolution and a review of the state of the art. *Neuroinformatics* 13(2):133–150. <https://doi.org/10.1007/s12021-014-9243-4>
- Doring TM, Kubo TT, Cruz LC, Juruena MF, Fainberg J, Domingues RC, Gasparetto EL (2011) Evaluation of hippocampal volume based on MR imaging in patients with bipolar affective disorder applying manual and automatic segmentation techniques. *J Magn Reson Imaging* 33(3):565–572. <https://doi.org/10.1002/jmri.22473>
- Duvernoy HM (1988) Sectional anatomy. In: The human hippocampus. Springer, New York, pp 61–153
- Fischl B (2012) FreeSurfer. *Neuroimage* 62(2):774–781
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van Der Kouwe A, Killiany R, Kennedy D, Klaveness S (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33(3):341–355
- Geuze E, Vermetten E, Bremner JD (2005) MR-based in vivo hippocampal volumetrics: 2. Findings in neuropsychiatric

- disorders. *Mol Psychiatry* 10(2):160–184. <https://doi.org/10.1038/sj.mp.4001579>
- Giavarina D (2015) Understanding Bland Altman analysis. *Biochem Med* 25(2):141–151. <https://doi.org/10.11613/bm.2015.015>
- Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC, Rapoport JL (1999) Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci* 2(10):861–863. <https://doi.org/10.1038/13158>
- Giedd JN, Clasen LS, Lenroot R, Greenstein D, Wallace GL, Ordaz S, Molloy EA, Blumenthal JD, Tossell JW, Stayer C, Samango-Sprouse CA, Shen D, Davatzikos C, Merke D, Chrousos GP (2006) Puberty-related influences on brain development. *Mol Cell Endocrinol* 254(Supplement C):154–162. <https://doi.org/10.1016/j.mce.2006.04.016>
- Giedd JN, Raznahan A, Alexander-Bloch A, Schmitt E, Gogtay N, Rapoport JL (2015) Child psychiatry branch of the national institute of mental health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology* 40(1):43–49. <https://doi.org/10.1038/npp.2014.236>
- Goddings A-L, Mills KL, Clasen LS, Giedd JN, Viner RM, Blakemore S-J (2014) The influence of puberty on subcortical brain development. *NeuroImage* 88(Supplement C):242–251. <https://doi.org/10.1016/j.neuroimage.2013.09.073>
- 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. <https://doi.org/10.1016/j.artmed.2016.09.001>
- Guo T, Winterburn JL, Pipitone J, Duerden EG, Park MT, Chau V, Poskitt KJ, Grunau RE, Synnes A, Miller SP, Mallar Chakravarty M (2015) Automatic segmentation of the hippocampus for pre-term neonates from early-in-life to term-equivalent age. *Neuroimage Clin* 9:176–193. <https://doi.org/10.1016/j.nicl.2015.07.019>
- Heckers S (2001) Neuroimaging studies of the hippocampus in schizophrenia. *Hippocampus* 11(5):520–528. <https://doi.org/10.1002/hipo.1068>
- Hogan RE, Wang L, Bertrand ME, Willmore LJ, Buchholz RD, Nasrif AS, Csernansky JG (2006) Predictive value of hippocampal MR imaging-based high-dimensional mapping in mesial temporal epilepsy: preliminary findings. *AJNR Am J Neuroradiol* 27(10):2149–2154
- Hoogman M, Bralten J, Hibar DP, Mennes M, Zwiers MP, Schieren LSJ, van Hulzen KJE, Medland SE, Shumskaya E, Jahanshad N, Zeeuw P, Szekely E, Sudre G, Wolfers T, Onnink AMH, Dammers JT, Mostert JC, Vives-Gilabert Y, Kohls G, Oberwilling E, Seitz J, Schulte-Ruther M, Ambrosino S, Doyle AE, Hovik MF, Dransdahl M, Tamm L, van Erp TGM, Dale A, Schork A, Conzelmann A, Zierhut K, Baur R, McCarthy H, Yoncheva YN, Cubillo A, Chantiluke K, Mehta MA, Paloyelis Y, Hohmann S, Baumeister S, Bramati I, Mattos P, Tovar-Moll F, Douglas P, Banaschewski T, Brandeis D, Kuntsi J, Asherson P, Rubia K, Kelly C, Martino AD, Milham MP, Castellanos FX, Frodl T, Zentis M, Lesch KP, Reif A, Pauli P, Jernigan TL, Haavik J, Plessen KJ, Lundervold AJ, Hugdahl K, Seidman LJ, Biederman J, Rommelse N, Heslenfeld DJ, Hartman CA, Hoekstra PJ, Oosterlaan J, Polier GV, Konrad K, Vilarroya O, Ramos-Quiroga JA, Soliva JC, Durston S, Buitelaar JK, Faraone SV, Shaw P, Thompson PM, Franke B (2017) Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *Lancet Psychiatry* 4(4):310–319. [https://doi.org/10.1016/S2215-0366\(17\)30049-4](https://doi.org/10.1016/S2215-0366(17)30049-4)
- Iglesias JE, Sabuncu MR, Leemput KV, for the Alzheimer's Disease Neuroimaging I (2013) Improved inference in bayesian segmentation using Monte Carlo sampling: application to hippocampal subfield volumetry. *Medical Image Anal* 17(7):766–778. <https://doi.org/10.1016/j.media.2013.04.005>
- Iglesias JE, Augustinack JC, Nguyen K, Player CM, Player A, Wright M, Roy N, Frosch MP, McKee AC, Wald LL, Fischl B, Van Leemput K, Initiative AsDN (2015) A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. *Neuroimage* 115:117–137. <https://doi.org/10.1016/j.neuroimage.2015.04.042>
- Jack CR, Barkhof F, Bernstein MA, Cantillon M, Cole PE, Decarli C, Dubois B, Duchesne S, Fox NC, Frisoni GB, Hampel H, Hill DL, Johnson K, Mangin JF, Scheltens P, Schwarz AJ, Sperling R, Suhy J, Thompson PM, Weiner M, Foster NL (2011) Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. *Alzheimers Dement* 7(4):474–485.e474. <https://doi.org/10.1016/j.jalz.2011.04.007>
- Jancke L, Merillat S, Liem F, Hanggi J (2015) Brain size, sex, and the aging brain. *Hum Brain Mapp* 36(1):150–169. <https://doi.org/10.1002/hbm.22619>
- Koch GG (2006) Intraclass correlation coefficient. In: Kotz S, Balakrishnan N, Read CB, Vidakovic B (eds) *Encyclopedia of statistical sciences*, 2nd edn, Wiley-Interscience, Hoboken, NJ
- Konrad C, Ukas T, Nebel C, Arolt V, Toga AW, Narr KL (2009) Defining the human hippocampus in cerebral magnetic resonance images—an overview of current segmentation protocols. *Neuroimage* 47(4):1185–1195. <https://doi.org/10.1016/j.neuroimage.2009.05.019>
- Koolschijn PCMP, van Ijzendoorn MH, Bakermans-Kranenburg MJ, Crone EA (2013) Hippocampal volume and internalizing behavior problems in adolescence. *Eur Neuropsychopharmacol* 23(7):622–628. <https://doi.org/10.1016/j.euroneuro.2012.07.001>
- Krouwer JS (2008) Why Bland-Altman plots should use X, not (Y + X)/2 when X is a reference method. *Stat Med* 27(5):778–780. <https://doi.org/10.1002/sim.3086>
- Lee JK, Nordahl CW, Amaral DG, Lee A, Solomon M, Ghetti S (2015) Assessing hippocampal development and language in early childhood: evidence from a new application of the automatic segmentation adapter tool. *Hum Brain Mapp*. <https://doi.org/10.1002/hbm.22931>
- Lyden H, Gimbel SI, Del Piero L, Tsai AB, Sachs ME, Kaplan JT, Margolin G, Saxbe D (2016) Associations between family adversity and brain volume in adolescence: manual vs. automated brain segmentation yields different results. *Front Neurosci* 10:398. <https://doi.org/10.3389/fnins.2016.00398>
- Maltbie E, Bhatt K, Paniagua B, Smith RG, Graves MM, Mosconi MW, Peterson S, White S, Blocher J, El-Sayed M, Hazlett HC, Styner MA (2012) Asymmetric bias in user guided segmentations of brain structures. *Neuroimage* 59(2):1315–1323. <https://doi.org/10.1016/j.neuroimage.2011.08.025>
- Morey RA, Petty CM, Xu Y, Hayes JP, Wagner HR, Lewis DV, LaBar KS, Styner M, McCarthy G (2009a) A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage* 45(3):855–866. <https://doi.org/10.1016/j.neuroimage.2008.12.033>
- Morey RA, Petty CM, Xu Y, Hayes JP, Wagner HR, Lewis DV, LaBar KS, Styner M, McCarthy G (2009b) Rebuttal to Hasan and Pedraza in comments and controversies: “improving the reliability of manual and automated methods for hippocampal and amygdala volume measurements”. *Neuroimage* 48(3):499–500. <https://doi.org/10.1016/j.neuroimage.2009.07.013>
- Morey RA, Gold AL, LaBar KS, Beall SK, Brown VM, Haswell CC, Nasser JD, Wagner HR, McCarthy G, Workgroup M-AM (2012) Amygdala volume changes in posttraumatic stress disorder in a large case-controlled veterans group. *Arch Gen Psychiatry* 69(11):1169–1178. <https://doi.org/10.1001/archgenpsychiatry.2012.50>
- Moscovitch M, Rosenbaum RS, Gilboa A, Addis DR, Westmacott R, Grady C, McAndrews MP, Levine B, Black S, Winocur G,

- Nadel L (2005) Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *J Anat* 207(1):35–66. <https://doi.org/10.1111/j.1469-7580.2005.00421.x>
- Nestor SM, Gibson E, Gao FQ, Kiss A, Black SE, Initiative ftAsDN (2012) A direct morphometric comparison of five labeling protocols for multi-atlas driven automatic segmentation of the hippocampus in Alzheimer's disease. *Neuroimage* 66C:50–70. <https://doi.org/10.1016/j.neuroimage.2012.10.081>
- Nugent TF, Herman DH, Ordonez A, Greenstein D, Hayashi KM, Lenane M, Clasen L, Jung D, Toga AW, Giedd JN, Rapoport JL, Thompson PM, Gogtay N (2007) Dynamic mapping of hippocampal development in childhood onset schizophrenia. *Schizophr Res* 90(1–3):62–70. <https://doi.org/10.1016/j.schres.2006.10.014>
- Nugent AC, Luckenbaugh DA, Wood SE, Bogers W, Zarate CA, Drevets WC (2013) Automated subcortical segmentation using FIRST: test–retest reliability, interscanner reliability, and comparison to manual segmentation. *Hum Brain Mapp* 34(9):2313–2329. <https://doi.org/10.1002/hbm.22068>
- Okada N, Fukunaga M, Yamashita F, Koshiyama D, Yamamori H, Ohi K, Yasuda Y, Fujimoto M, Watanabe Y, Yahata N, Nemoto K, Hibar DP, van Erp TG, Fujino H, Isobe M, Isomura S, Nat-subori T, Narita H, Hashimoto N, Miyata J, Koike S, Takahashi T, Yamasue H, Matsuo K, Onitsuka T, Iidaka T, Kawasaki Y, Yoshimura R, Suzuki M, Turner JA, Takeda M, Thompson PM, Ozaki N, Kasai K, Hashimoto R (2016) Abnormal asymmetries in subcortical brain volume in schizophrenia. *Mol Psychiatry* 21(10):1460–1466. <https://doi.org/10.1038/mp.2015.209>
- Pardoe HR, Pell GS, Abbott DF, Jackson GD (2009) Hippocampal volume assessment in temporal lobe epilepsy: how good is automated segmentation? *Epilepsia* 50(12):2586–2592. <https://doi.org/10.1111/j.1528-1167.2009.02243.x>
- Patel S, Park MTM, Devenyi GA, Patel R, Masellis M, Knight J, Chakravarty MM (2017) Heritability of hippocampal subfield volumes using a twin and non-twin siblings design. *Hum Brain Mapp* 38:4337–4352
- Pipitone J, Park MT, Winterburn J, Lett TA, Lerch JP, Pruessner JC, Lepage M, Voineskos AN, Chakravarty MM, Initiative AsDN (2014) Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. *Neuroimage* 101:494–512. <https://doi.org/10.1016/j.neuroimage.2014.04.054>
- Qin S, Cho S, Chen T, Rosenberg-Lee M, Geary DC, Menon V (2014) Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nature neuroscience* 17(9):1263–1269. <https://doi.org/10.1038/nn.3788>
- Ragunathan TE, Rosenthal R, Rubin DB (1996) Comparing correlated but nonoverlapping correlations. *Psychol Methods* 1(2):178
- Sánchez-Benavides G, Gómez-Ansón B, Sainz A, Vives Y, Delfino M, Peña-Casanova J (2010) Manual validation of FreeSurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer Disease subjects. *Psychiatry Res* 181(3):219–225. <https://doi.org/10.1016/j.psychres.2009.10.011>
- Schmidt MF, Storrs JM, Freeman KB, Jack CR, Turner ST, Griswold ME, Mosley TH (2018) A comparison of manual tracing and FreeSurfer for estimating hippocampal volume over the adult lifespan. *Hum Brain Mapp* 39(6):2500–2513. <https://doi.org/10.1002/hbm.24017> doi
- Schoemaker D, Buss C, Head K, Sandman CA, Davis EP, Chakravarty MM, Gauthier S, Pruessner JC (2016) Hippocampus and amygdala volumes from magnetic resonance images in children: assessing accuracy of FreeSurfer and FSL against manual segmentation. *Neuroimage* 129:1–14. <https://doi.org/10.1016/j.neuroimage.2016.01.038>
- Schumann CM, Hamstra J, Goodlin-Jones BL, Lotspeich LJ, Kwon H, Buonocore MH, Lammers CR, Reiss AL, Amaral DG (2004) The amygdala is enlarged in children but not adolescents with autism; the hippocampus is enlarged at all ages. *J Neurosci* 24(28):6392–6401. <https://doi.org/10.1523/jneurosci.1297-04.2004>
- Sheline YI, Wang PW, Gado MH, Csernansky JG, Vannier MW (1996) Hippocampal atrophy in recurrent major depression. *Proc Natl Acad Sci USA* 93(9):3908–3913
- Shen L, Saykin AJ, Kim S, Firpi HA, West JD, Risacher SL, McDonald BC, McHugh TL, Wishart HA, Flashman LA (2010) Comparison of manual and automated determination of hippocampal volumes in MCI and early AD. *Brain Imaging Behav* 4(1):86–95. <https://doi.org/10.1007/s11682-010-9088-x>
- Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86(2):420–428
- Squire LR, Stark CE, Clark RE (2004) The medial temporal lobe. *Annu Rev Neurosci* 27:279–306. <https://doi.org/10.1146/annurev.neuro.27.070203.144130>
- Sussman D, Leung RC, Vogan VM, Lee W, Trelle S, Lin S, Cassel DB, Chakravarty MM, Lerch JP, Anagnostou E, Taylor MJ (2015) The autism puzzle: diffuse but not pervasive neuroanatomical abnormalities in children with ASD. *NeuroImage Clin* 8(Supplement C):170–179. <https://doi.org/10.1016/j.nicl.2015.04.008>
- Tae WS, Kim SS, Lee KU, Nam EC, Kim KW (2008) Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology* 50(7):569–581. <https://doi.org/10.1007/s00234-008-0383-9>
- Tannous J, Amaral-Silva H, Cao B, Wu MJ, Zunta-Soares GB, Kazimi I, Zeni C, Mwangi B, Soares JC (2018) Hippocampal subfield volumes in children and adolescents with mood disorders. *J Psychiatr Res* 101:57–62. <https://doi.org/10.1016/j.jpsychires.2018.03.003>
- Teicher MH, Anderson CM, Polcari A (2012) Childhood maltreatment is associated with reduced volume in the hippocampal subfields CA3, dentate gyrus, and subiculum. *Proc Natl Acad Sci USA* 109(9):E563–E572. <https://doi.org/10.1073/pnas.1115396109>
- Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 60(1):34–42
- Treadway MT, Waskom ML, Dillon DG, Holmes AJ, Park MTM, Chakravarty MM, Dutra SJ, Polli FE, Iosifescu DV, Fava M, Gabrieli JDE, Pizzagalli DA (2015) Illness progression, recent stress, and morphometry of hippocampal subfields and medial prefrontal cortex in major depression. *Biol Psychiatry* 77(3):285–294. <https://doi.org/10.1016/j.biopsych.2014.06.018>
- Van Leemput K, Bakkour A, Benner T, Wiggins G, Wald LL, Augustinack J, Dickerson BC, Golland P, Fischl B (2009) Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus* 19(6):549–557. <https://doi.org/10.1002/hipo.20615>
- Wechsler D (1991) Wechsler intelligence scale for children, Third edn. Psychological Corporation, San Antonio
- Wechsler D (1999) Wechsler abbreviated scale of intelligence. Harcourt Assessment, San Antonio
- Wenger E, Märtensson J, Noack H, Bodammer NC, Kühn S, Schaefer S, Heinze HJ, 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(8):4236–4248. <https://doi.org/10.1002/hbm.22473>
- Whelan CD, Hibar DP, van Velzen LS, Zannas AS, Carrillo-Roa T, McMahon K, Prasad G, Kelly S, Faskowitz J, Iglesias JE (2016) Heritability and reliability of automatically segmented human hippocampal formation subregions. *Neuroimage* 128:125–137

- Wierenga L, Langen M, Ambrosino S, van Dijk S, Oranje B, Durston S (2014) Typical development of basal ganglia, hippocampus, amygdala and cerebellum from age 7 to 24. *Neuroimage* 96:67–72. <https://doi.org/10.1016/j.neuroimage.2014.03.072>
- Winterburn JL, Pruessner JC, Chavez S, Schira MM, Lobaugh NJ, Voineskos AN, Chakravarty MM (2013) A novel in vivo atlas of human hippocampal subfields using high-resolution 3 T magnetic resonance imaging. *Neuroimage* 74:254–265. <https://doi.org/10.1016/j.neuroimage.2013.02.003>
- Wisse LE, Biessels GJ, Geerlings MI (2014) A critical appraisal of the hippocampal subfield segmentation package in FreeSurfer. *Front Aging Neurosci* 6:261. <https://doi.org/10.3389/fnagi.2014.00261>
- Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31(3):1116–1128. <https://doi.org/10.1016/j.neuroimage.2006.01.015>
- Yushkevich PA, Avants BB, Pluta J, Das S, Minkoff D, Mechanic-Hamilton D, Glynn S, Pickup S, Liu W, Gee JC, Grossman M, Detre JA (2009) A high-resolution computational atlas of the human hippocampus from postmortem magnetic resonance imaging at 9.4 T. *Neuroimage* 44(2):385–398. <https://doi.org/10.1016/j.neuroimage.2008.08.042>
- Yushkevich PA, Amaral RS, Augustinack JC, Bender AR, Bernstein JD, Boccardi M, Bocchetta M, Burggren AC, Carr VA, Chakravarty MM, Chételat G, Daugherty AM, Davachi L, Ding SL, Ekstrom A, Geerlings MI, Hassan A, Huang Y, Iglesias JE, La Joie R, Kerchner GA, LaRocque KF, Libby LA, Malykhin N, Mueller SG, Olsen RK, Palombo DJ, Parekh MB, Pluta JB, Preston AR, Pruessner JC, Ranganath C, Raz N, Schlichting ML, Schoemaker D, Singh S, Stark CE, Suthana N, Tompary A, Turowski MM, Van Leemput K, Wagner AD, Wang L, Winterburn JL, Wisse LE, Yassa MA, Zeineh MM, (HSG) HSG (2015) Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: towards a harmonized segmentation protocol. *Neuroimage* 111:526–541. <https://doi.org/10.1016/j.neuroimage.2015.01.004>
- Zandifar A, Fonov V, Coupé P, Pruessner J, Collins DL, Initiative AsDN (2017) A comparison of accurate automatic hippocampal segmentation methods. *Neuroimage* 155:383–393. <https://doi.org/10.1016/j.neuroimage.2017.04.018>