



Providing Choice & Value

Generic CT and MRI Contrast Agents



FRESENIUS
KABI

CONTACT REP

AJNR

**Comparison of Hippocampal Subfield
Segmentation Agreement between 2
Automated Protocols across the Adult Life
Span**

A. Samara, C.A. Raji, Z. Li and T. Hershey

This information is current as
of July 12, 2025.

AJNR Am J Neuroradiol published online 5 August 2021
<http://www.ajnr.org/content/early/2021/08/05/ajnr.A7244>

Comparison of Hippocampal Subfield Segmentation Agreement between 2 Automated Protocols across the Adult Life Span

A. Samara, C.A. Raji, Z. Li, and T. Hershey



ABSTRACT

BACKGROUND AND PURPOSE: The hippocampus is a frequent focus of quantitative neuroimaging research, and structural hippocampal alterations are related to multiple neurocognitive disorders. An increasing number of neuroimaging studies are focusing on hippocampal subfield regional involvement in these disorders using various automated segmentation approaches. Direct comparisons among these approaches are limited. The purpose of this study was to compare the agreement between two automated hippocampal segmentation algorithms in an adult population.

MATERIALS AND METHODS: We compared the results of 2 automated segmentation algorithms for hippocampal subfields (FreeSurfer v6.0 and volBrain) within a single imaging data set from adults ($n = 176$, 89 women) across a wide age range (20–79 years). Brain MR imaging was acquired on a single 3T scanner as part of the IXI Brain Development Dataset and included T1- and T2-weighted MR images. We also examined subfield volumetric differences related to age and sex and the impact of different intracranial volume and total hippocampal volume normalization methods.

RESULTS: Estimated intracranial volume and total hippocampal volume of both protocols were strongly correlated ($r = 0.93$ and 0.9 , respectively; both $P < .001$). Hippocampal subfield volumes were correlated (ranging from $r = 0.42$ for the subiculum to $r = 0.78$ for the cornu ammonis [CA]), all $P < .001$). However, absolute volumes were significantly different between protocols. volBrain produced larger CA1 and CA4-dentate gyrus and smaller CA2-CA3 and subiculum volumes compared with FreeSurfer v6.0. Regional age- and sex-related differences in subfield volumes were qualitatively and quantitatively different depending on segmentation protocol and intracranial volume/total hippocampal volume normalization method.

CONCLUSIONS: The hippocampal subfield volume relationship to demographic factors and disease states should undergo nuanced interpretation, especially when considering different segmentation protocols.

ABBREVIATIONS: CA = cornu ammonis; DG = dentate gyrus; HPSF = hippocampal subfield; ICV = intracranial volume; SR-SL-SM = strata radiatum-lacunosum-moleculare; THV = total hippocampal volume

The hippocampus is a major component of the limbic system, and it is affected in several neurocognitive and neuropsychiatric

disorders from Alzheimer disease to major depressive disorder.^{1,2} Located in the mesial temporal lobes, the hippocampus functions as a computational hub through its extensive afferent and efferent connections with cortical and subcortical structures.³ The hippocampus and hippocampal-related structures sustain a range of cognitive functions in the context of episodic and semantic memory, spatial navigation, planning, and learning.⁴ The hippocampus is additionally divided into distinct cytoarchitectonic regions called subfields, most prominently the dentate gyrus (DG), cornu ammonis (CA) subfields 1–4, and the subiculum.⁵ Distinctive cognitive functions are supported by different subfields,⁶ and subfields are differentially affected in various neuropsychiatric disorders.^{2,7}

An increasing number of in vivo neuroimaging studies have focused on hippocampal subfield (HPSF) involvement in neurologic and psychiatric conditions.^{8,9} The ability to differentiate subfields in vivo provides a unique opportunity to identify

Received February 22, 2021; accepted after revision May 14.

From the Department of Psychiatry (A.S., Z.L., T.H.), Mallinckrodt Institute of Radiology (C.A.R., T.H.), Departments of Neurology (C.A.R., T.H.), and Psychological and Brain Sciences (Z.L.), Washington University School of Medicine, St. Louis, Missouri.

A. Samara and C.A. Raji contributed equally to this work.

A. Samara was supported by National Institute on Drug Abuse (grant No. 5T32DA007261-29). C.A.R. was supported by Washington University in St. Louis, National Institutes of Health KL2 Grant (KL2 TR000450, Institute of Clinical and Translational Sciences Multidisciplinary Clinical Research Career Development Program), and the Radiological Society of North America Research Scholar Grant.

Please address correspondence Cyrus A. Raji, MD, PhD, Mallinckrodt Institute of Radiology, Washington University School of Medicine, 4525 Scott Ave, St. Louis, MO 63110; e-mail: craji@wustl.edu; @cyrusraji

Indicates open access to non-subscribers at www.ajnr.org

Indicates article with online supplemental data.

<http://dx.doi.org/10.3174/ajnr.A7244>

biomarkers for brain diseases like Alzheimer disease.⁹ For example, studies have shown that the HPSFs can be impacted by aging and Alzheimer disease in a regional-specific pattern and can be used as an in vivo biomarker with diagnostic and prognostic significance.¹⁰ Manual segmentation has limited clinical throughput due to the time requirement and the necessity of trained operators. Thus, automated approaches are needed to industrialize the clinical throughput across millions of potential brain MR imaging scans.

Two segmentation protocols that are commonly used are FreeSurfer (<http://surfer.nmr.mgh.harvard.edu>) and volBrain (<https://volbrain.upv.es/index.php>).^{11,12} Between 2013 and 2019, >160 studies applying the FreeSurfer HPSF segmentation protocol in normal development and various neuropsychiatric conditions were published.¹³ Although FreeSurfer is the most widely used software, some concerns about segmentation accuracy in earlier versions of FreeSurfer (v5.1, v5.2, and v5.3) have been previously raised,^{14,15} leading to several improvements in the more recent versions of FreeSurfer using ex vivo and ultra-high-resolution MR imaging.¹¹ An increasing number of studies are using the volBrain protocol as an alternative.^{16,17} The main advantage that volBrain provides over FreeSurfer is its considerably shorter processing time because the segmentation results are produced in approximately 15 minutes compared with several hours for FreeSurfer. The agreement of HPSF volumes from both protocols has never been directly compared in a single study. Such comparison is critical to allow optimal interpretation of results reported by different research groups.

The goal of the current work was to compare the output of the 2 HPSF segmentation protocols, FreeSurfer v6.0 and volBrain, in a large cohort of adults based on T1- and T2-weighted MR imaging. We selected these 2 protocols because FreeSurfer is the most popular software for hippocampal subfield segmentation and volBrain is an increasingly popular alternative due to its considerably shorter processing time. We evaluated the agreement between the 2 protocols in volumetric assessment and investigated the presence of estimation bias in measurements. We also examined qualitative and quantitative subfield differences related to age and sex and the impact of various intracranial volume (ICV) and total hippocampal volume (THV) normalization methods.

MATERIALS AND METHODS

Participants and MR Imaging Acquisition

We used the publicly available IXI Brain Development Dataset (<http://brain-development.org/ixi-dataset/>). This data base includes T1- and T2-weighted brain MR imaging scans of healthy adults with a wide age range. For the current analyses, we included scans that were acquired using the 3T scanner (Philips Healthcare) at Hammersmith Hospital to assess the within-subject agreement of HPSF volumes between protocols. T1-weighted imaging parameters were the following: TR = 9.6 ms, TE = 4.6 ms, number of phase encoding steps = 208, echo-train length = 208, reconstruction diameter = 240.0, acquisition matrix = 208 × 208, flip angle = 8.0°, voxel resolution = 0.9 × 0.9 × 1.2 mm. The T2-weighted parameters were the following: TR = 5725.79 ms, TE = 100.0 ms, number of phase encoding steps = 187, echo-train length = 16, reconstruction diameter = 240.0 mm,

acquisition matrix = 192 × 187, flip angle = 90.0°, voxel resolution = 0.9 × 0.9 × 1.2 mm.

FreeSurfer and volBrain Segmentation

FreeSurfer v6.0 software is one of the most widely used pipelines to obtain HPSF volumes. The FreeSurfer HPSF segmentation module generates a fully automated segmentation based on a probabilistic atlas.¹¹ For each scan, we used the output volume from the standard FreeSurfer processing of T1 MR imaging after correcting for motion, intensity normalization, and skull stripping. The FreeSurfer algorithm detects local variations in MR imaging contrast using a Bayesian inference algorithm and relies on a hippocampus atlas generated by combining manual labels from ex vivo and in vivo whole-brain scans.^{11,18} FreeSurfer uses both T1 and T2 MR imaging to obtain a more reliable segmentation.¹⁹ We used both T1 and T2 MR imaging in the hippocampus subfield segmentation stage to improve tissue contrast and assist in identifying landmarks of the internal hippocampal structure. FreeSurfer generates 12 subfields: parasubiculum, presubiculum, subiculum, CA1, CA3, CA4, granule cell and molecular layer of the dentate gyrus, molecular layer, hippocampus-amygdala transition area, fimbria, hippocampal tail, and hippocampal fissure (definitions of subfield boundaries are detailed in Iglesias et al¹¹). CA2 is always included in the CA3 label, as CA2-CA3. We combined CA4 and the granule cell and molecular layer of the dentate gyri CA4-DG in subsequent analyses.

The volBrain hippocampal subfield segmentation protocol is a new method that consists of a fast multiatlas nonlocal patch-based label fusion.¹² This pipeline is publicly available on a web-based platform.²⁰ volBrain provides the option to use multimodal images (T1 and T2 MR imaging) for improved accuracy of segmentation, which we used in our analysis. The original algorithm is based on the Winterburn atlas,²¹ which produces 5 subfield labels: CA1, CA2/3, CA4/DG, stratum radiatum/stratum lacunosum/stratum moleculare (SR-SL-SM), and subiculum. The processing time is about 15 minutes per scan. An example of both HPSF segmentations is shown in Fig 1. Due to the large number of scans included, we did not review each scan by visual inspection after completion, and we did not perform any manual corrections. However, as a quality control measure, we excluded individuals with >1 subfield volume as an outlying value (outlier defined as > 3 SDs).

ICV Normalization Methods

We examined the effect of different ICV normalization methods on the HPSF volumetric analysis. Total ICV estimation from each protocol was used to correct for the subfields derived by the same protocol. We performed 3 distinct approaches to account for variations in total ICV. These methods were the following: 1) the proportion method (calculated by multiplying each individual subfield-to-ICV ratio with the average ICV for the entire cohort); 2) the residual method (regressing out the effect of ICV before conducting further analysis); and 3) the covariate method (in which we included estimated ICV as a covariate in the regression analyses as described below). In addition, to evaluate regional differences in HPSF, we performed similar correction methods to account for variation in THV.²²

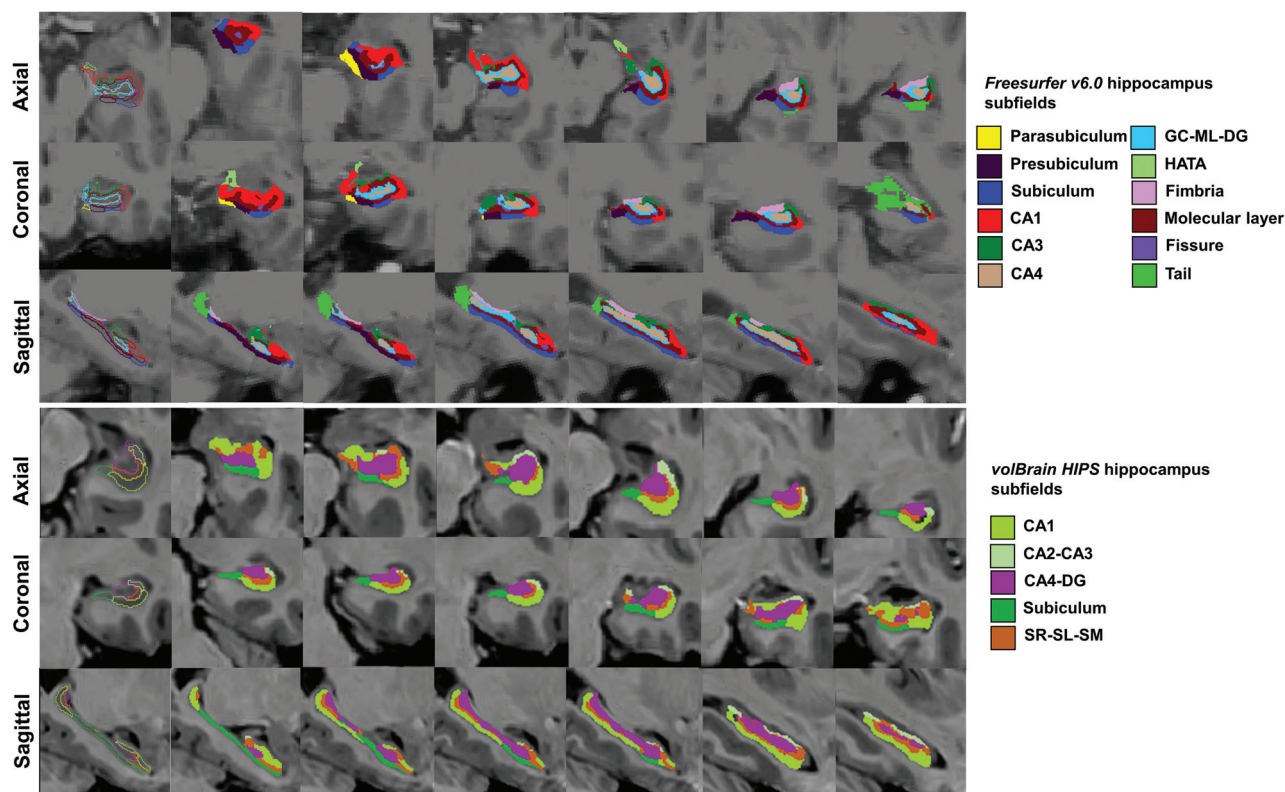


FIG 1. An example of hippocampal subfield segmentation by FreeSurfer (upper row) and volBrain (lower row) shown in axial, coronal, and sagittal sections. GC-ML-DG indicates granule cell and molecular layers of the dentate gyrus; HATA, hippocampus-amygdala transition area.

Statistical Analysis

We conducted all statistical analyses and illustrations using R statistical and computing software (Version 3.6.3; <http://www.r-project.org/>). We combined the right and left hemispheric volumes for each subfield. The THV for each protocol was calculated by combining all subfields (excluding the hippocampal fissure in FreeSurfer segmentation because it represents CSF). We used Pearson r correlations to relate HPSFs between the 2 protocols and paired t -tests to compare the mean differences between the 2 groups. To compare the effects of different ICV normalization methods on the relationship between HPSF and age and sex variables, we conducted multiple linear regression analyses and reported the dependent variable estimates for each ICV/THV normalization method. Additionally, we calculated the effect size of the difference in HPSFs between men and women using the Cohen's D . In regression analyses, the Bonferroni correction for multiple comparisons was used as $P < .0045$ ($P = .05/11$ subfields) for the FreeSurfer analyses, and $P < .01$ ($P = .05/5$ subfields) for the volBrain analyses. Finally, Bland-Altman plots were produced to visualize the disagreement between FreeSurfer and volBrain in terms of absolute, uncorrected hippocampal subfield volumes.

RESULTS

Participants

A total of 176 eligible brain MR imaging scans underwent HPSF segmentation. We subsequently excluded 4 scans because they yielded outlying volumetric values in >1 subfield (2 FreeSurfer

and 2 volBrain). After exclusion, our analyzed sample included 83 men with an age range of 20–79 years (mean = 45 [SD, 16] years) and 89 women with an age range of 21–82 years (mean = 50 [SD, 17] years). A few participants had a single outlier across all subfields (FreeSurfer: CA2-CA3, 1; parasubiculum, 1; hippocampus-amygdala transition area, 1; fimbria, 1; fissure, 2; volBrain: CA1, 1; CA2-CA3, 2; SR-SL-SM, 1; subiculum, 1). We excluded these outliers, but not the entire scans, from subsequent analyses.

Hippocampal Subfield Volumes

The HPSF volumes produced by both FreeSurfer and volBrain are detailed in the [Table](#). Further correlation analyses were applied only to subfields shared by both FreeSurfer and volBrain segmentations (ie, CA1, CA2-CA3, CA4-DG, and the subiculum). We observed significant correlations between the CA1, CA2-CA3, CA4-DG, and subiculum volumes segmented by both FreeSurfer and volBrain ([Fig 2](#); $P < .001$ for all correlations). Correlation was strongest for CA1 ($r = 0.78$) and weakest for the subiculum volume ($r = 0.42$). However, the 2 protocols produced different average volumes. volBrain yielded larger average CA1 and CA4-DG than FreeSurfer, while FreeSurfer conversely yielded larger CA2-CA3 and subiculum volumes than volBrain ($P < .001$ for all subfields; [Fig 2](#) and [Table](#)). The Bland-Altman plots confirmed that for almost every scan, FreeSurfer generated smaller volumes for CA1 and CA4-DG and larger volumes for CA2-CA3 and the subiculum compared with volBrain ([Fig 3](#)). Furthermore, Bland-Altman plots demonstrated that the size of

Uncorrected hippocampal subfield, total hippocampal, and intracranial volumes measured by FreeSurfer and volBrain protocols^a

	Total (n = 172)		Male (n = 83)		Female (n = 89)		Cohen's D
	Mean	SD	Mean	SD	Mean	SD	
FreeSurfer v6.0							
CA1	1251	193	1321	169	1184	159	0.83
CA2-CA3	472	71	497	51	449	54	0.91
CA4-DG	1113	171	1173	141	1057	135	0.84
Subiculum	807	102	841	101	775	87	0.7
Parasubiculum	121	21	126	16	117	21	0.48
Presubiculum	553	72	581	75	527	59	0.8
Molecular layer	1205	168	1241	136	1172	150	0.48
HATA	136	31	142	19	130	23	0.57
Fimbria	178	67	188	62	168	48	0.36
Tail	999	160	1049	155	952	118	0.7
Fissure ^b	269	71	276	49	262	70	0.23
THV	6839	957	7163	846	6537	758	0.78
eICV	1,563,261	146,983	1,677,235	158,967	1,456,885	231,310	1.11
volBrain, HIPS segmentation protocol							
CA1	1486	286	1574	267	1405	234	0.67
CA2-CA3	264	37	276	43	253	32	0.61
CA4-DG	1420	221	1462	207	1382	260	0.34
SR-SL-SM	938	158	988	173	892	145	0.6
Subiculum	566	87	593	86	541	93	0.58
THV	4689	771	4908	776	4484	665	0.59
eICV	1,374,768	133,467	1,462,729	139,905	1,292,671	126,129	1.28

Note:—HATA indicates hippocampus-amygdala transition area; eICV, estimated intracranial volume; HIPS, hippocampal subfield segmentation protocol.

^a Units are cubic millimeters.

^b Not significantly different. Men had significantly larger values than females in all volumes, except hippocampus fissure ($P < .05$, Student t test). The average values of total and subfield volumes reported here are the sum of right and left hemispheres combined.

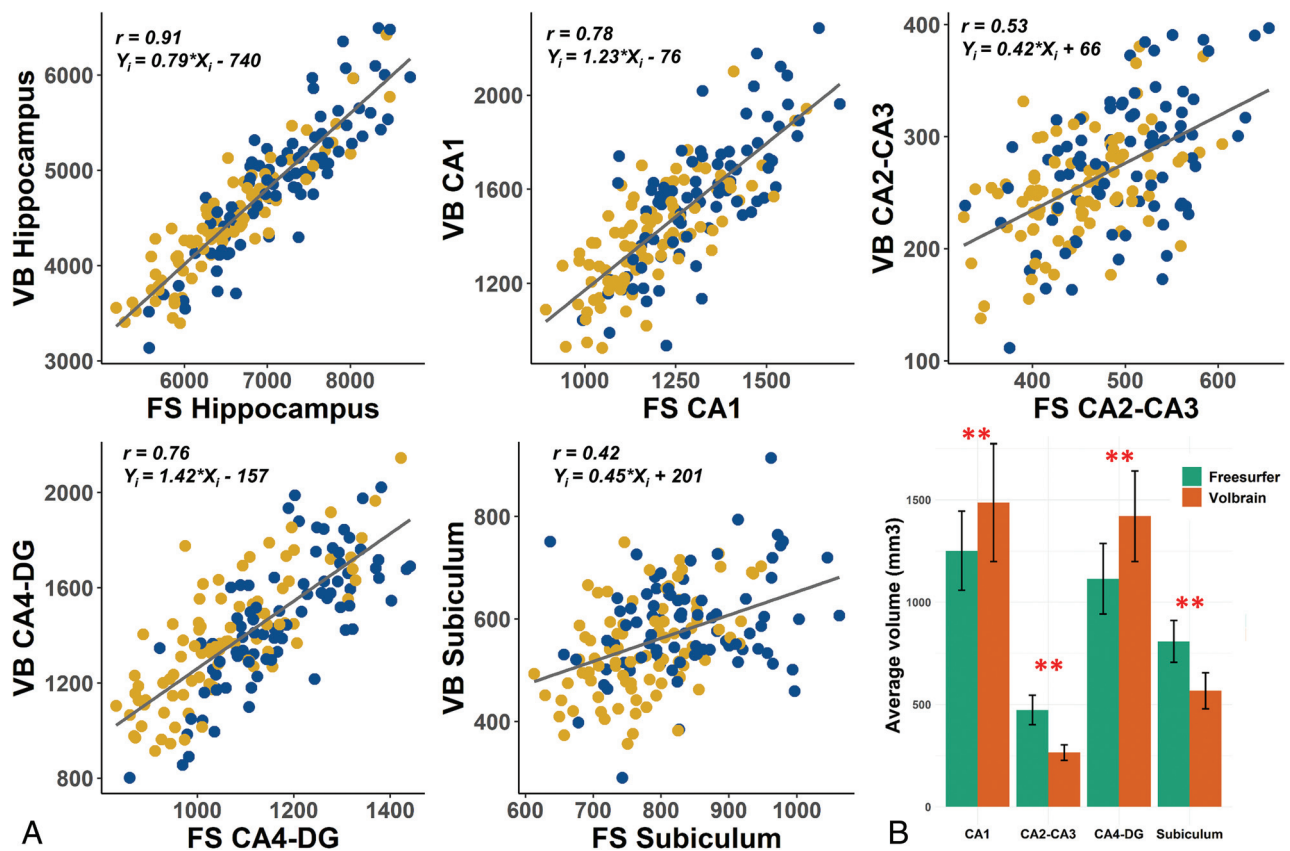


FIG 2. A, Comparison of uncorrected total hippocampus, CA1, CA2-CA3, CA4-DG, and subiculum volumes (cubic millimeters) between FreeSurfer and volBrain (yellow, women; blue, men). Regression lines relating volBrain to FreeSurfer volumes are shown for each subfield. The average values of subfield volumes reported here are the sum of right and left hemisphere volumes combined. All Pearson r correlations are significant ($P < .001$). B, Bar graphs show means (SDs). Double asterisks indicate statistical significance ($P < .001$); FS, FreeSurfer; VB, volBrain.

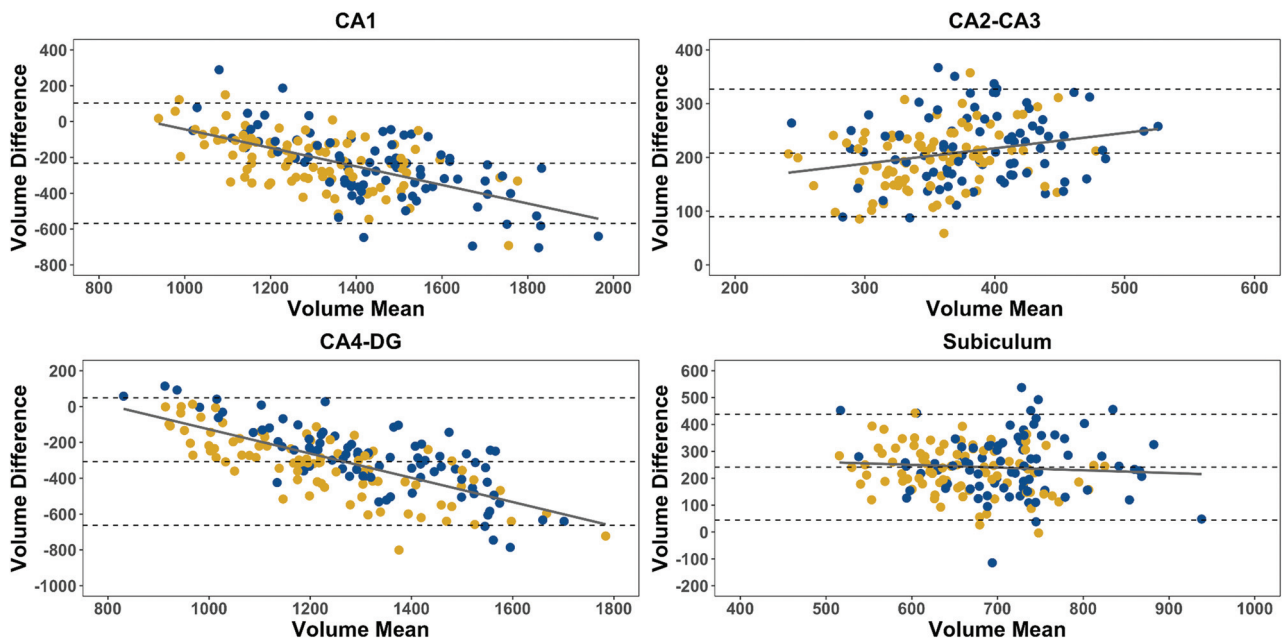


FIG 3. Bland-Altman plots for uncorrected subfield volumes (CA1, CA2-CA3, CA4-DG, and subiculum volumes [cubic millimeters]) generated by FreeSurfer and volBrain (yellow, women; blue, men). Intrасubject volume difference (y-axis) is defined as (FreeSurfer volume) – (volBrain volume). Mean volume is represented on the x-axis. The mean (SD, 1.96) volume difference and 95% confidence intervals are plotted as dashed horizontal lines. Except for the subiculum, all Pearson r correlations are significant ($P < .001$).

the disagreement between the 2 protocols increased for larger volume estimates of CA1 ($r = -0.61$; 95% CI, -0.70 to -0.51), CA2-CA3 ($r = 0.25$; 95% CI, 0.11 – 0.39), and CA4-DG ($r = -0.71$; 95% CI, -0.78 to -0.63). No such relationship was found for the subiculum ($r = -0.08$; 95% CI, -0.23 – 0.07 ; $P = .30$).

THV and total ICV estimations were strongly correlated between FreeSurfer and volBrain ($r = 0.91$ and $r = 0.93$, respectively; $P < .001$). Compared to volBrain, FreeSurfer produced higher THV and ICV volumes ($P < .001$ for both) (Table). Both FreeSurfer and volBrain produced larger uncorrected HPSF volumes in men compared with women, except for the hippocampal fissure (Table).

ICV and THV Normalization Methods

For each HPSF segmented by either FreeSurfer or volBrain, volume values were normalized for both ICV and THV derived by the same protocol using the covariate, proportion, and residual methods. These normalized volumes were then entered, separated by normalization method, into multiple linear regression models with participant age and sex as covariates (Online Supplemental Data).

Both age and sex showed different associations with HPSF volumes depending on the segmentation protocol and normalization method. Marked inconsistency in the statistical significance and magnitude of the regression estimates could be observed in multiple HPSFs. Specifically, in FreeSurfer, CA1, CA2-CA3, CA4-DG, and the presubiculum showed significant negative correlations with age only per ICV covariate and residual methods. However, for the molecular layer and subiculum, the significant positive correlation with age could only be established with THV normalization, but not ICV normalization. Regardless of the ICV/THV normalization method, age positively correlated with

hippocampal fissure and negatively correlated with the fimbria and hippocampal tail. HPSF volumes were consistently higher in women than in men only when the ICV proportion method was used. In volBrain, CA4-DG volume negatively correlated with age across all ICV/THV normalization methods. CA1 and the subiculum negatively correlated with age only when ICV normalization methods were applied. The SR-SL-SM subfield positively correlated with age when THV methods were used. CA4-DG volumes were significantly higher in women using all normalization methods, except for the ICV covariate method.

Moreover, some contradictory findings emerged when comparing the results between the 2 segmentation protocols. Most strikingly, in CA2-CA3, significant regression estimates for age had negative findings in FreeSurfer but positive findings in volBrain. In the subiculum, the estimates were positive in FreeSurfer but negative in volBrain. Additionally, by means of the ICV proportion method, CA2-CA3 and subiculum volumes were significantly larger in women than in men for FreeSurfer but not for volBrain, which showed no significant sex effect.

DISCUSSION

Our study is the first to directly compare the results of 2 commonly used HPSF segmentation protocols, providing new insight to help compare results across different neuroimaging studies. Although the HPSF volumes produced using the 2 protocols were well-correlated, significant differences were observed in absolute volumes. Specifically, volBrain produced larger CA1 and CA4-DG volumes, while FreeSurfer produced larger CA2-CA3 and subiculum volumes. We also observed systematic biases in the HPSF estimations because the absolute volume difference between the 2 protocols increased for larger volume estimates for

CA1, CA2-CA3, and CA4-DG. Finally, we found that different segmentation protocols and ICV/THV normalization methods could lead to inconsistent and sometimes contradictory conclusions regarding the regional effects of age and sex on HPSF volumes.

While absolute volumetric differences exist across the 2 protocols, their results being correlated indicates that they may be interchangeably used for correlational analyses. Some of the inconsistencies between protocols are likely due to differences in the number of HPSF labels (FreeSurfer, $n = 12$; volBrain, $n = 5$) and how the 2 protocols define the HPSF anatomic boundaries. For example, FreeSurfer produces specific labels for the hippocampal tail, fimbria, hippocampus-amygdala transition area, parasubiculum, and presubiculum, while these subfield labels do not exist in volBrain. Yushkevich et al²³ compared the results of 21 HPSF labeling protocols (which did not include volBrain) and concluded that the greatest disagreement was along the CA1/subiculum anatomic boundary and anterior hippocampus. This observation could potentially explain the larger CA1 and smaller subiculum produced by volBrain compared with FreeSurfer.²³ Finally, the correlation between the 2 protocols was more robust for THV than for any HPSF, suggesting greater agreement in the outer hippocampal boundaries than in HPSF labels.

Differences in the age range and acquisition parameters in each algorithm training data set might have also contributed to the observed differences. The generative model for the FreeSurfer protocol was based on 15 ex vivo postmortem brain hemispheres obtained from individuals 60–91 years of age, with some individuals who had mild Alzheimer disease or mild cognitive impairment.¹¹ The brain tissue was scanned using 7T MR imaging at 0.13-mm isotropic resolution on average. On the other hand, the volBrain segmentation protocol relied on the Winterburn atlas data base obtained from 5 healthy individuals 29–57 years of age and acquired at 0.6-mm isotropic resolution.¹² On the basis of the age range differences included in each dataset, it is reasonable to suggest that volBrain might provide more accurate segmentations when applied in younger age groups. In fact, when FreeSurfer segmentation is applied to standard resolution MR imaging (1 mm isotropic), the molecular layer would not be clearly visible and atlas deformation is unlikely to be influenced by this anatomic feature. In this case, fitting of the atlas to internal structure relies on prior encoded information alone.^{11,24}

This issue will introduce bias in a way that underestimates CA1 and CA4-DG volumes in younger individuals because these 2 subfields are susceptible to age-related changes.¹⁵ The Bland-Altman plots support this explanation and show that between-protocol differences in CA1 and CA4-DG volumes increased with higher mean volumes (ie, in younger individuals), while the differences decreased with lower mean volumes (ie, in older individuals). Iglesias et al¹¹ explicitly stated in their original article that the FreeSurfer atlas might include hippocampal atrophy because it was built using delineations in elderly individuals, which could compromise its applicability in younger populations. Nevertheless, we acknowledge the possibility that the differences observed in the Bland-Altman plots could also be attributed to differing segmentation boundaries between the 2 protocols or higher error variances in one segmentation method than in the other.

When the volBrain protocol was compared with manual segmentation from the Winterburn data base at a standard resolution (0.9 mm isotropic), the average Dice similarity score was 0.66 (ranging between 0.52 for CA2-CA3 and 0.76 for CA4-DG).¹² These findings highlight the inherent limitations of the volBrain protocol. On the other hand, Iglesias et al¹¹ performed a qualitative assessment of the multimodal FreeSurfer segmentation on the Winterburn atlas data base. The authors suggested that direct spatial overlap evaluation (eg, using Dice similarity scores) between the Winterburn manual and FreeSurfer automated segmentations is not possible due to labeling protocol differences. Although the agreement between both segmentations was fair in general, prominent differences were observed in areas poorly supported by the MR imaging contrast (eg, the medial digitation) and regions where the definitions of HPSF boundaries were not similar (eg, the inferior parts). For example, the FreeSurfer subiculum was mostly part of the Winterburn atlas CA1 subfield, while the presubiculum and parasubiculum approximately corresponded to the Winterburn atlas subiculum.¹¹

ICV normalization is a commonly used procedure in neuroimaging research, and it is an important step to account for sex differences and intersubject variations in head size. This step is necessary because relative, rather than absolute, differences in volumes better described the structure-function relationships. Several ICV normalization methods had been described in the literature, including the covariate, proportion, and residual methods. In addition, van Eijk et al²² reported sex-dependent regional differences in HPSF volumes after adjusting for THV. They suggested that the THV normalization could provide additional value over ICV normalization alone.²² When we applied different ICV/THV normalization methods in our study, the most noticeable finding was the marked impact of the choice of normalization method on both the direction and magnitude of estimates of age and sex-related differences. Previous studies have demonstrated marked effects of ICV normalization methods on volumetric assessment of cortical and subcortical structures.^{25,26} For example, different ICV normalization methods produce contradictory results regarding the presence of sex-related volumetric differences.²⁵ Some of these studies have also suggested that the residual method generally provides greater advantages over the proportion and covariate methods,²⁵⁻²⁷ and these recommended residual ICV normalization for volumetric studies of neuroanatomic structures.²⁵ We also noticed a clear trend of larger HPSF volumes in women compared with men when the ICV proportion method was used. This finding is consistent with those in prior studies showing that women have proportionately larger gray matter regions than men,^{26,28} and these differences are potentially due to ICV differences rather than sex.²⁵

The main limitations of this study include a focus on quantitative values for the HPSF volumes without looking at spatial overlap metrics and comparing label segmentations on a voxel-by-voxel basis. Also, we could not directly compare the reliability of HPSF segmentations across different scanners, voxel resolutions, and field strengths (1.5T versus 3T). How this would affect the comparison across protocols remains to be determined. However, prior work had shown that understanding the performance of HPSF

segmentation software at this field strength carries potentially greater clinical significance.²⁹ Additionally, future studies should compare the segmentation results in pathologic conditions like Alzheimer disease. The performance of both protocols could drastically change if applied to MR imaging of patients, when tissue damage could reduce the contrast between the tissues and, consequently, lead to greater variability in segmentation.

CONCLUSIONS

Although automatic segmentation of HPSFs has enabled large-scale in vivo analysis and has increased in popularity in recent years, it is important to interpret the results of these studies with caution. Although volumetric assessment of HPSF derived from FreeSurfer and volBrain are well-correlated, we showed significant differences between the 2 protocols in terms of absolute volumes and estimation bias. These differences could lead to different conclusions about HPSF changes across the adult life span. We also showed that the specific ICV normalization method used could influence the conclusions. Researchers should also be very careful when combining data across different protocols in any meta-analyses. Finally, the findings of our study highlight the need for a standard unified approach for HPSFs in neuroimaging studies.

Disclosures: Amjad Samara—RELATED: Grant: National Institutes on Drug Abuse, Comments: Amjad Samara was supported by National Institute on Drug Abuse (grant No. 5T32DA007261-29). Cyrus A. Raji—UNRELATED: Board Membership: Brainreader ApS; Consultancy: Apollo Health; Expert Testimony: Neurevolution Medical; Grants/Grants Pending: National Institutes of Health KL2, Radiological Society of North American Research & Education Foundation Scholar Grant.* Tamara Hershey—RELATED: Grant: National Institutes of Health*; UNRELATED: Employment: Washington University School of Medicine; Grants/Grants Pending: National Institutes of Health.* *Money paid to the institution.

REFERENCES

- Raji CA, Lopez OL, Kuller LH, et al. Age, Alzheimer disease, and brain structure. *Neurology* 2009;73:1899–905 [CrossRef Medline](#)
- Han KM, Kim A, Kang W, et al. Hippocampal subfield volumes in major depressive disorder and bipolar disorder. *Eur Psychiatry* 2019;57:70–77 [CrossRef Medline](#)
- Battaglia FP, Benchenane K, Sirota A, et al. The hippocampus: hub of brain network communication for memory. *Trends Cogn Sci* 2011;15:310–18 [CrossRef Medline](#)
- Buzsáki G, Moser EI. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat Neurosci* 2013;16:130–38 [CrossRef Medline](#)
- Amaral DG, Witter MP. The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* 1989;31:571–91 [CrossRef Medline](#)
- Dimsdale-Zucker HR, Ritchey M, Ekstrom AD, et al. CA1 and CA3 differentially support spontaneous retrieval of episodic contexts within human hippocampal subfields. *Nat Commun* 2018;9:294 [CrossRef Medline](#)
- Mueller SG, Weiner MW. Selective effect of age, Apo e4, and Alzheimer's disease on hippocampal subfields. *Hippocampus* 2009;19:558–64 [CrossRef Medline](#)
- Apostolova LG, Dutton RA, Dinov ID, et al. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. *Arch Neurol* 2006;63:693–99 [CrossRef Medline](#)
- Hanseeuw BJ, Van Leemput K, Kavec M, et al. Mild cognitive impairment: differential atrophy in the hippocampal subfields. *AJNR Am J Neuroradiol* 2011;32:1658–61 [CrossRef Medline](#)
- Hett K, Ta VT, Catheline G, et al. Alzheimer's Disease Neuroimaging Initiative. Multimodal hippocampal subfield grading for Alzheimer's disease classification. *Sci Rep* 2019;9:13845 [CrossRef Medline](#)

- Iglesias JE, Augustinack JC, Nguyen K, et al. Alzheimer's Disease Neuroimaging Initiative. A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. *Neuroimage* 2015;115:117–37 [CrossRef Medline](#)
- Romero JE, Coupé P, Manjón JV. HIPS: a new hippocampus subfield segmentation method. *Neuroimage* 2017;163:286–95 [CrossRef Medline](#)
- Sāmān PG, Inglesias JE, Gutman B, et al. FreeSurfer-based segmentation of hippocampal subfields: a review of methods and applications, with a novel quality control procedure for ENIGMA studies and other collaborative efforts. *Hum Brain Mapp* 2020 Dec 27. [Epub ahead of print] [CrossRef Medline](#)
- Wisse LE, Biessels GJ, Geerlings MI. A critical appraisal of the hippocampal subfield segmentation package in FreeSurfer. *Front Aging Neurosci* 2014;6:261 [CrossRef Medline](#)
- de Flores R, La Joie R, Landeau B, et al. Effects of age and Alzheimer's disease on hippocampal subfields: comparison between manual and FreeSurfer volumetry. *Hum Brain Mapp* 2015;36: 463–74 [CrossRef Medline](#)
- Moghaddam HS, Aarabi MH, Mehvari-Habibabadi J, et al. Distinct patterns of hippocampal subfield volume loss in left and right mesial temporal lobe epilepsy. *Neurol Sci* 2021;42:1411–21 [CrossRef Medline](#)
- Kim S, Greene DJ, D'Andrea CB, et al. Hippocampal volume in provisional tic disorder predicts tic severity at 12-month follow-up. *J Clin Med* 2020;9:1715 [CrossRef Medline](#)
- Van Leemput K, Bakkour A, Benner T, et al. Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus* 2009;19:549–57 [CrossRef Medline](#)
- Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: challenges, methods, and applications. *Comput Math Methods Med* 2015;2015:450341 [CrossRef Medline](#)
- Manjón JV, Coupé P. volBrain: an online MRI brain volumetry system. *Front Neuroinform* 2016;10:30 [CrossRef Medline](#)
- Winterburn JL, Pruessner JC, Chavez S, et al. A novel in vivo atlas of human hippocampal subfields using high-resolution 3 T magnetic resonance imaging. *Neuroimage* 2013;74:254–65 [CrossRef Medline](#)
- van Eijk L, Hansell NK, Strike LT, et al. Region-specific sex differences in the hippocampus. *Neuroimage* 2020;215:116781 [CrossRef Medline](#)
- Yushkevich PA, Amaral RS, Augustinack JC, et al. Hippocampal Subfields Group (HSG). Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: towards a harmonized segmentation protocol. *Neuroimage* 2015;111:526–41 [CrossRef Medline](#)
- Wisse LE, Chételat G, Daugherty AM, et al. Hippocampal subfield volumetry from structural isotropic 1 mm³ MRI scans: a note of caution. *Hum Brain Mapp* 2021;42:539–50 [CrossRef Medline](#)
- Pintzka CW, Hansen TI, Evensmoen HR, et al. Marked effects of intracranial volume correction methods on sex differences in neuroanatomical structures: a HUNT MRI study. *Front Neurosci* 2015;9:238 [CrossRef Medline](#)
- Voevodskaya O, Simmons A, Nordenskjöld R, et al. Alzheimer's Disease Neuroimaging Initiative. The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Front Aging Neurosci* 2014;6:264 [CrossRef Medline](#)
- Sanfilipo MP, Benedict RH, Zivadinov R, et al. Correction for intracranial volume in analysis of whole brain atrophy in multiple sclerosis: the proportion vs. residual method. *Neuroimage* 2004;22:1732–43 [CrossRef Medline](#)
- Gur RC, Turetsky BI, Matsui M, et al. Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. *J Neurosci* 1999;19: 4065–72 [CrossRef Medline](#)
- Chow N, Hwang KS, Hurtz S, et al. Comparing 3T and 1.5T MRI for mapping hippocampal atrophy in the Alzheimer's Disease Neuroimaging Initiative. *AJNR Am J Neuroradiol* 2015;36:653–60 [CrossRef Medline](#)