



Published in final edited form as:

*Epilepsia*. 2014 September ; 55(9): 1374–1379. doi:10.1111/epi.12694.

## Automated Versus Manual Hippocampal Segmentation in Pre- and Postoperative Epilepsy Patients

Sierra C. Germeyan, BA<sup>1</sup>, David Kalikhman, BS<sup>1</sup>, Lucy Jones, MBChB<sup>1</sup>, and William H. Theodore, MD<sup>1</sup>

<sup>1</sup>Clinical Epilepsy Section, NINDS, NIH, Bethesda, MD

### Abstract

**Objective**—To compare manual and automated pre- and postoperative hippocampal volume measurements in patients with intractable epilepsy.

**Methods**—We studied 34 patients referred to the Clinical Epilepsy Section, NINDS, NIH for evaluation of intractable epilepsy and 21 normal volunteers who received 1.5 or 3 T GE Signa MRI scans. Hippocampal volumes were manually traced on each slice and assembled into three-dimensional volumes by investigators blinded to other data. Automated volumetric measurements were obtained using FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>). Statistical analysis was performed with GraphPad Prism.

**Results**—Automated hippocampal volumes were larger than manual volumes in both patients and normal volunteers,  $p < .05$ . Right to left hemisphere hippocampal ratio and percent of hippocampus resected did not significantly differ by segmentation method. It was not possible to obtain accurate total resection volumes with the automated method.

**Significance**—Values such as side-to-side ratio and percent resected may be more directly translatable between manual and automated methods than absolute measures of volume. Accurate determination of resection volumes is important for studies of the effects of surgery on both seizure control and postoperative neuropsychological deficits. Our preliminary data suggest that FreeSurfer may provide an accurate and simple method for quantitating hippocampal resections. However, it may be less valuable for large or extratemporal resections, or when distortions of normal anatomy are present.

### Keywords

hippocampus; epilepsy; magnetic resonance imaging; segmentation

---

William H. Theodore MD Chief, Clinical Epilepsy Section NINDS NIH Building 10 Room 7D-43 Bethesda MD 20892 Phone: 301-496-1505 Fax: 301-402-2871 theodorw@ninds.nih.gov.

#### Disclosures

None of the authors has any conflict of interest to disclose.

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

## Introduction

FreeSurfer is a well-validated alternative to manual segmentation for control brains as well as brains with several types of pathology.<sup>1</sup> Previous studies have shown that FreeSurfer provides volumetric measurements that correlate with manual segmentation in patients with epilepsy, as well as accurately reveals hippocampal asymmetry in patients as compared to controls.<sup>2,3</sup> However, the accuracy of automated segmentation in temporal lobe epilepsy (TLE) patients is reduced when developmental anomalies or severe atrophy are present.<sup>4</sup> No studies have yet examined the utility of FreeSurfer in temporal lobe epilepsy patients following surgical resection; this study aims to evaluate manual versus automated methods of quantifying postoperative volumes. Accurate determination of resection volumes is important for studies of the effects of surgery on both seizure control and postoperative neuropsychological deficits.<sup>5,6</sup>

## Methods

### Patients

We studied a total of 34 patients referred to the Clinical Epilepsy Section, National Institute of Neurological Disorders and Stroke, NIH for evaluation of intractable mesial temporal lobe epilepsy (Supporting Table 1). Twenty three patients were used in analyses of preoperative volumes. Nine of those 23 patients plus one additional patient were used in analyses of the postoperative contralateral hippocampus. The 10 patients used for contralateral analyses plus another 10 patients were used for analysis of postoperative ipsilateral hippocampi.

All patients were evaluated with neurological examination, ictal video-EEG monitoring, and neuropsychological testing. Positron emission tomography with 18F-fluoro-2-deoxy-d-glucose and subdural electrode placement were performed as clinically indicated. The study was approved by the NIH Combined Neurosciences Institutional Review Board. Informed consent was obtained from all patients.

### Preoperative measurements

All patients received a 1.5-T or 3T GE Signa MRI with fluid attenuated inversion recovery (FLAIR), T1- and T2- weighted images, and 3-D SPGR or MPRage for hippocampal tracing. In 23 patients, three independent manual raters blinded to video-EEG monitoring data traced the hippocampus manually on each slice, using the procedure described previously, separating the anterior head of the hippocampus from the amygdala by the lineal alba and inferior limb of the lateral ventricle.<sup>5</sup> The hippocampus was traced posteriorly to include the gyrus fasciola, and the slices assembled into three-dimensional volumes using MEDx version 3.43.

### Postoperative measurements

Manual rater 3 traced postoperative hippocampal volumes on the side of resection ('ipsilateral') of 20 epilepsy patients by following the same procedure as was used for preoperative volume measurements on all slices in which any part of the hippocampus was

visible. The same rater also manually traced the hippocampus contralateral to the side of resection ('contralateral') in the nonsurgical hemisphere in 10 of those 20 patients to assess interscan reliability.

### FreeSurfer Method

FreeSurfer version 5.3.0 (<http://surfer.nmr.mgh.harvard.edu/>) was used to analyze pre- and postoperative scans of epilepsy patients and normal subjects. As with the manual method, a total of 23 preoperative, 10 nonresected ('contralateral') postoperative, and 20 resected ('ipsilateral') postoperative hippocampal volume measurements were calculated by FreeSurfer for the main analyses. Automated processing included transformation into Talairach coordinates, segmentation of subcortical white matter and deep gray matter volumetric structures and removal of non-brain tissue.<sup>7</sup> The longitudinal processing stream was used for volume estimates.<sup>8</sup>

The difference between a patient's post-resection and pre-resection volumes was used to approximate the amount of tissue resected. These values were calculated from FreeSurfer data output for pre- and postoperative scans of 30 epilepsy patients. Whole brain volume measurements, including estimated total intracranial volume (eTIV) and brain segmentation volume without ventricles (BrainSegNotVent) were compared to obtain estimates of total resection volume (<http://freesurfer.net/fswiki/MorphometryStats>).

Default analysis settings for the standard image processing stream as well as the hippocampal subfield segmentation stream were run on 11 MPRage and 10 SPGR scans of normal volunteers.<sup>9</sup>

Statistical analysis, including t-tests and ANOVAs, was performed with Graphpad Prism version 6.0d.

## Results

### Preoperative

The average hippocampal volumes of 23 epilepsy patients as measured by FreeSurfer were significantly larger than volumes measured for the same 23 patients by three manual raters,  $p < .0001$  (Table 1). The volumes of manual rater 1 were significantly smaller than those of raters 2 and 3,  $p < .05$ . Raters 2 and 3 were not significantly different from each other,  $p > .05$ . The ratio of right to left hippocampal volume was not significantly different between FreeSurfer and any of the three manual raters,  $F(3,90)=0.56$ ,  $p=.64$  (Figure 1).

### Postoperative

**Resection Volumes**—Upon visual inspection of FreeSurfer output, data for 27 of the 30 epilepsy patients was observed to be outside a reasonable range of values. Whether estimated resection volumes were reasonable did not seem to be specific to any given scanner or sequence, nor associated with other possible causes such as resection size.

**Hippocampal Volumes**—Since FreeSurfer was not able to accurately measure total resection volume, we next investigated whether the automated method would accurately

segment hippocampi following resection in order to provide a reliable postoperative measure.

Within each method, there was no significant difference between the contralateral hippocampal volumes measured pre- and post-surgery,  $p > .05$  (Figure 2). Consistent with findings from preoperative scans, FreeSurfer's volumes were significantly larger than manual volumes for both the contralateral and ipsilateral hippocampi,  $p < .05$  (Figure 2, Figure 3).

Postoperative volumes within each method were significantly smaller than preoperative volumes,  $F(3,76)=83.07$ ,  $p < .01$  (Supporting Table 2, Figure 3). FreeSurfer hippocampal resection volumes were calculated by subtracting postoperative from preoperative hippocampal volumes. Automated resection volumes were significantly larger than manual resection volumes,  $p < .01$  (Supporting Table 2). Manual and automated post-surgery volumes also differed,  $t(38)=2.11$ ,  $p = .041$ , such that postoperative volumes calculated by FreeSurfer were slightly larger than those calculated based on manual tracing (Figure 3).

Percent of hippocampus resected was calculated by dividing the resected volume (preoperative minus postoperative hippocampal volume for each patient) by the pre-surgery volume and multiplying by 100 (Supporting Table 2). Manual and automated percent resected did not significantly differ,  $p = .94$  (Figure 4).

**MRI Sequence Comparison**—To validate the use of both SPGR and MPRage sequences through these analyses, 11 MPRage and 10 SPGR scans of normal volunteers were run through the normal FreeSurfer processing stream and the hippocampal subfield segmentation stream. Average hippocampus size did not differ based on whether the scan was a MPRage ( $1821 \pm 398.7 \text{ mm}^3$ ) or a SPGR ( $1855 \pm 263.1 \text{ mm}^3$ ),  $p = .76$ . Additionally, none of the hippocampal subfields in one hemisphere significantly differed in volume as compared to the corresponding subfield in the other hemisphere, all  $p > .05$  (Figure 5A, 5B).

## Discussion

Our finding that automated volumes are larger than manually traced volumes (Table 1, Figure 2) is consistent with previous literature comparing automated and manual segmentation methods.<sup>3</sup> It is unclear which method produced more accurate results as it is not possible to perform morphometric analyses on the hippocampi to directly measure their volume. Indeed, few studies have measured hippocampal volume directly rather than indirectly through analysis of MRI scans. In one study of 50 adult male cadaver brains, average volume of the dissected left and right hippocampi were 11.839 and 11.713 cubic centimeters respectively, values which are much larger than those calculated by either method used in this study, or reported by other investigators using MRI.<sup>10</sup>

The disparity between manual and automated volumes may be due to differences inherent in these two methods of segmentation. The manual method uses visual identification of hippocampal margins, while FreeSurfer transforms the images to standard coordinates, then utilizes voxel intensities and probabilities to delineate the hippocampus. A study using a

different method found a good correlation between automated and manual hippocampal volume measurements, with a tendency for the former to be larger.<sup>11</sup>

Surgical resections in our cohort rarely extended beyond the anterior two-thirds of the hippocampus. Since manual and automated hippocampal volumes were significantly different pre-surgery but only marginally different post-surgery (Figure 3), the major source of the difference in pre-operative hippocampal volumes was likely variation in measurement of the anterior portion of the hippocampus.

Right to left hippocampal ratios did not significantly differ between manual and automated methods; therefore this may be a more useful measure, as it translates between methods (Figure 1). As well, ratios avoid the problems associated with variations in brain size, and the need to correct for intracranial volume, potentially adding additional variability.<sup>2,11</sup> The lack of difference between hippocampal volume measurements pre and postoperatively on the contralateral side within raters suggests that both manual and automated methods have inter-measurement reliability (Figure 2).

Quantification of resection volume is important for interpretation of surgical outcomes. Manual tracing has been used previously in a study which determined that larger resection volumes were correlated with better postoperative seizure outcome, but not with postoperative neuropsychological results.<sup>5</sup> FreeSurfer is not a feasible method of obtaining accurate total resection volumes. Upon visual inspection, automated segmentation of both pre- and postoperative brains appears reasonable (Supporting Figure 6). However, many of the whole brain volume differences from pre to post-surgery were either the wrong sign, of unreasonable magnitude, or both. The FreeSurfer software was not designed to analyze extremely abnormal brains, and it is likely that errors are made during the process of aligning postoperative brains to the standard atlas. Automated postoperative hippocampal volumes may be more accurate than automated postoperative total intracranial volumes due to the discreteness of the structure, which allows for a smaller margin of error in measurement.

When brains deviate significantly from normal, FreeSurfer does not necessarily provide accurate results, as its analysis is based on a single template.<sup>11</sup> Even prior to surgery, structures in the temporal lobe of many TLE patients have abnormalities, including atrophy associated with mesial temporal sclerosis. Identification of atrophy contributes to diagnosis and surgical planning. One study found that as compared to manual methods, FreeSurfer was less accurate and overestimated volumes in more atrophic hippocampi.<sup>4</sup> Malrotation also decreased accuracy in automated as compared to manual segmentation.<sup>4</sup>

Recently, an automated segmentation algorithm was developed specifically for adult patients with epilepsy by using a template database comprised of epilepsy patients with a range of pathologies rather than healthy subjects.<sup>11</sup> Following alignment, segmentation is performed using a local ranking strategy.<sup>12</sup> While this method has not yet been tried in epilepsy patients following surgery, it was shown to accurately segment the hippocampus as compared to manual raters, even for hippocampi with substantial atrophy.<sup>12</sup> However, the

segmentations are limited to the entire hippocampus, leaving FreeSurfer as the only currently available automated method of segmenting hippocampal subfields.

The lack of difference between right and left hemisphere hippocampal volumes in normal volunteers indicates that the side-to-side ratios obtained from processing of both SPGRs and MPRAGEs by FreeSurfer are comparable. Therefore, including both scan sequences in analyses does not compromise the validity of results.

These analyses add support to previous findings that automated segmentation of hippocampal volumes in brains of epilepsy patients are comparable to those obtained through manual tracing.<sup>2</sup> While FreeSurfer does not give accurate total resection volumes, it does provide values for the percent of the hippocampus resected that are comparable to those of manual raters.

Automated methods other than FreeSurfer of calculating structural volumes from MRIs are also in use. Analysis of Functional NeuroImages (AFNI) is one software package that is capable of calculating volumes by transforming individual subjects' images to standard atlas spaces where regions of interest are identified based on the stereotaxic coordinate system.<sup>13,14</sup> The validity of AFNI for calculating volumes in postoperative brains has not yet been investigated. Automated measurement of resection volumes in AFNI is not currently possible.

FreeSurfer has many advantages over manual segmentation. It is widely used, observer independent, and substantially decreases the amount of hands-on time necessary to analyze MRIs. Our preliminary data suggest that FreeSurfer may provide an accurate and simple method for quantitating hippocampal resections. However, it may be less valuable for large or extratemporal resections, or when distortions of normal anatomy are present.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

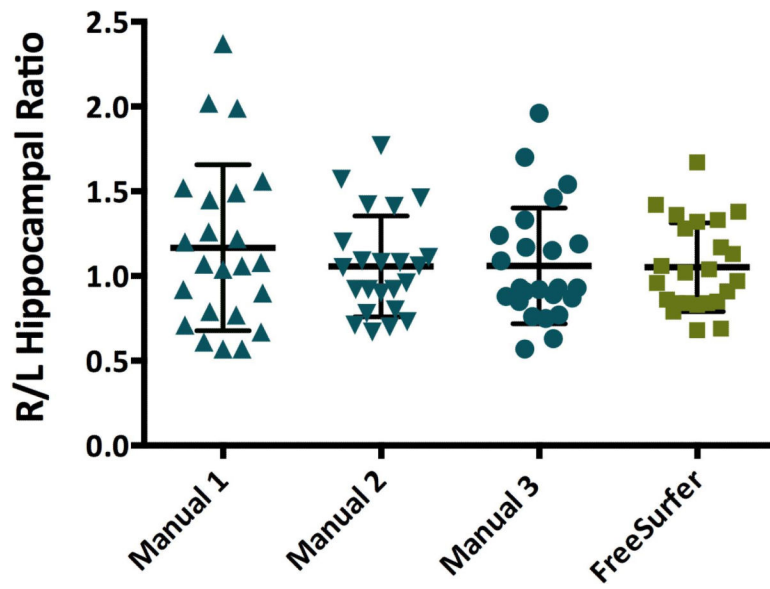
This research was supported by the NINDS Division of Intramural Research.

The authors would like to thank Irene Dustin RN CNP, Daniel Goldenholz MD PhD, John Heiss MD, and John Ostuni PhD.

## References

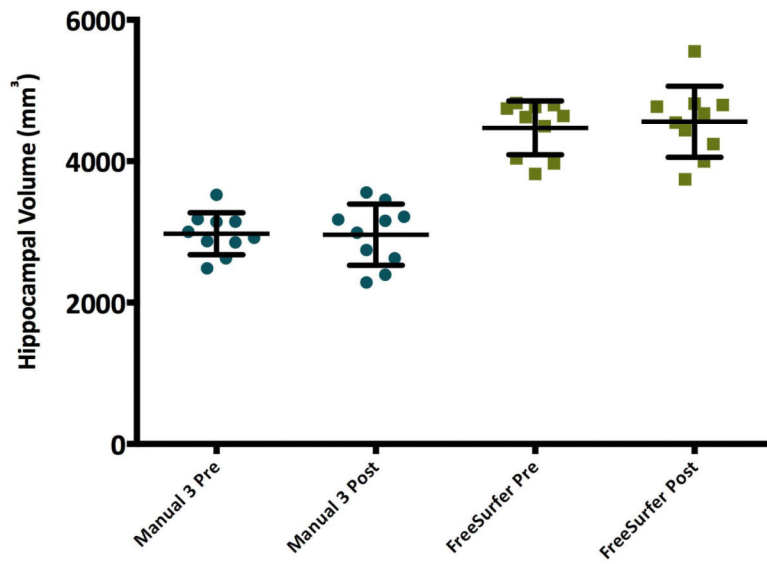
1. Fischl B. FreeSurfer. *NeuroImage*. 2012; 62:774–781. [PubMed: 22248573]
2. McDonald CR, Hagler DJ, Ahmadi ME, et al. Subcortical and cerebellar atrophy in mesial temporal lobe epilepsy revealed by automatic segmentation. *Epilepsy Research*. 2008; 79:130–138. [PubMed: 18359198]
3. Pardoe HR, Pell GS, Abbott DF, et al. Hippocampal volume assessment in temporal lobe epilepsy: How good is automated segmentation? *Epilepsia*. 2009; 50(12):2586–2592. [PubMed: 19682030]
4. Kim H, Chupin M, Colliot O, et al. Automatic hippocampal segmentation in temporal lobe epilepsy: Impact of developmental abnormalities. *NeuroImage*. 2012; 59:3178–3186. [PubMed: 22155377]

5. Shamim S, Wiggs E, Heiss J, et al. Temporal lobectomy: Resection volume, neuropsychological effects, and seizure outcome. *Epilepsy & Behavior*. 2009; 16:311–314. [PubMed: 19703792]
6. Tellez-Zenteno JF, Dhar R, Wiebe S. Long-term seizure outcomes following epilepsy surgery: a systematic review and meta-analysis. *Brain*. 2005; 128:1188–1198. [PubMed: 15758038]
7. Fischl B, Salat D, Busa E, et al. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33(3):341–355. [PubMed: 11832223]
8. Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*. 2012; 61(4):1402–1418. [PubMed: 22430496]
9. Van Leemput K, Bakkour A, Benner T, et al. Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. *Hippocampus*. 2009; 19:549–557. [PubMed: 19405131]
10. Rao BN, Rao KRSP, Rao RR. Morphometric study of hippocampus in adult human brains. *International Journal of Basic and Applied Medical Sciences*. 2012; 2(2):139–143.
11. Winston GP, Cardoso MJ, Williams EJ, et al. Automated hippocampal segmentation in patients with epilepsy: Available free online. *Epilepsia*. 2013; 54(12):2166–2173. [PubMed: 24151901]
12. Cardoso MJ, Leung K, Modat M, et al. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcellation. *Med Image Anal*. 2013; 17:671–684. [PubMed: 23510558]
13. Cox RW. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996; 29:162–173. [PubMed: 8812068]
14. Cox RW. AFNI: What a long strange trip it's been. *NeuroImage*. 2012; 62:743–747. [PubMed: 21889996]

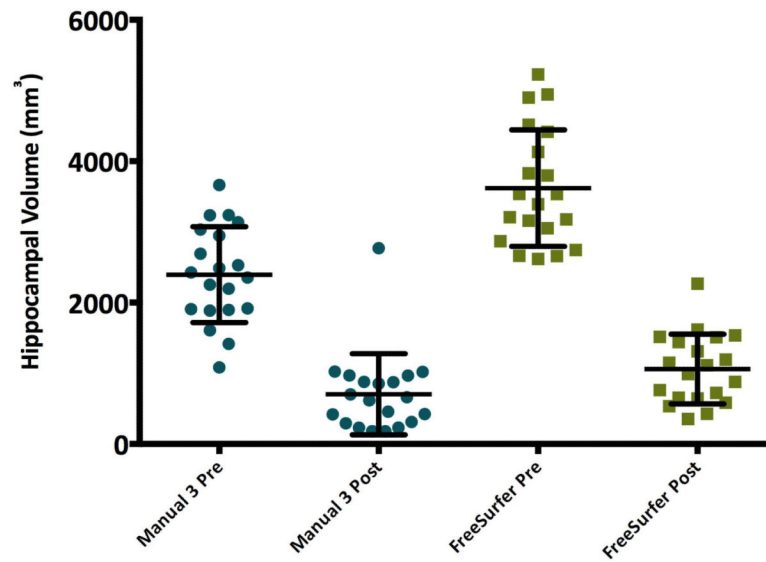


**Figure 1.** Right to left hippocampal volume ratios of 23 patients with epilepsy. No significant difference was found between any of the four independent raters,  $p > .05$ .

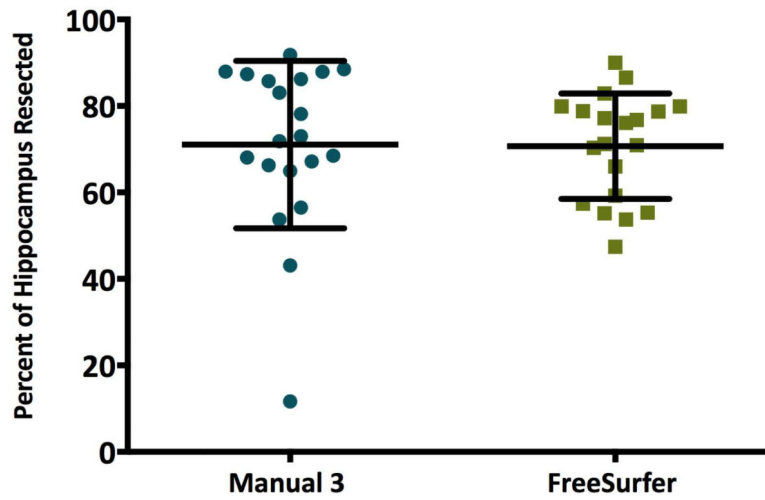




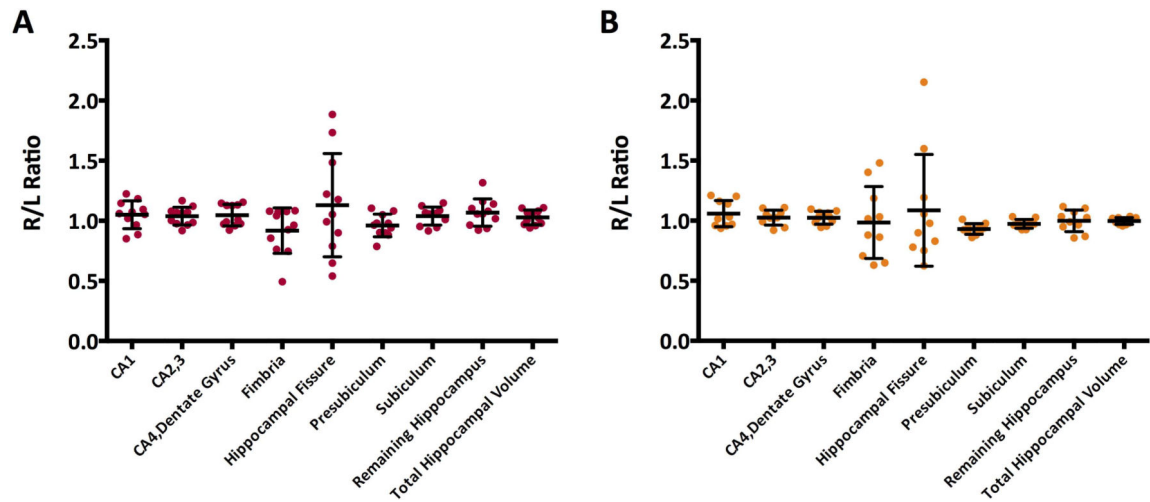
**Figure 2.** Pre- and postoperative volumes for the non-resected (contralateral) hippocampus in 10 patients with temporal lobe epilepsy. No significant pre-post operative difference was found within each rater,  $p > .05$ , while automated volumes were significantly larger than manual volumes both pre- and postoperatively,  $p < .05$ .



**Figure 3.** Manual and automated hippocampal volumes on side of resection (ipsilateral) pre- and postoperatively for 20 temporal lobe epilepsy patients. Within each rater, preoperative volume was significantly larger than postoperative volume,  $p < .01$ . Automated postoperative hippocampal volumes were slightly larger than manual volumes,  $p = .041$ .



**Figure 4.** Percent of hippocampus resected, calculated by subtracting post- from preoperative hippocampal volume, dividing by preoperative volume, and multiplying by 100%. Percent of hippocampus resected did not significantly differ based on segmentation method,  $p = .94$ .



**Figure 5.**

(A) Right to left hippocampal subfield ratios for 11 MPRages. No significant difference is present between the left and right hemispheres for any of the subfields, all  $p$ s  $>.05$ . (B) Right to left hippocampal subfield ratios for 10 SPGRs. No significant difference is present between the left and right hemispheres for any of the subfields, all  $p$ s  $>.05$ .

**Table 1**

Mean and Standard Deviation of Preoperative Hippocampal Volume Measurements for Three Manual Raters and FreeSurfer (n=23).

		<b>Left Hippocampal</b>	<b>Right Hippocampal</b>	<b>Right to Left</b>
		<b>Volume (mm<sup>3</sup>)</b>	<b>Volume (mm<sup>3</sup>)</b>	<b>Hippocampal Ratio</b>
Manual	Rater 1	1790.0 +/- 480.8	1917.0 +/- 429.4	1.167 +/- 0.490
	Rater 2	2440.0 +/- 683.2	2480.0 +/- 629.0	1.057 +/- 0.298
	Rater 3	2386.0 +/- 609.8	2411.0 +/- 549.4	1.065 +/- 0.348
Automated	FreeSurfer	3986.0 +/- 773.3	4078.0 +/- 582.6	1.061 +/- 0.263