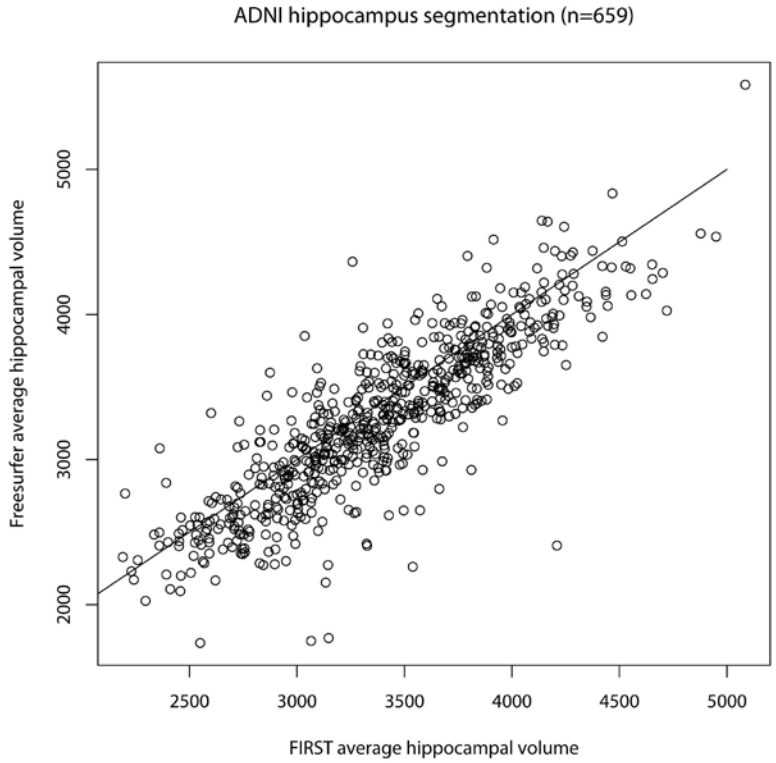
Freesurfer vs. FIRST

ADNI Sample:

We conducted a direct comparison of Freesurfer and FIRST segmentation of the hippocampus in the ADNI sample. Freesurfer volumes were downloaded from the LONI IDA and FIRST segmentation was completed according to the protocol on the ENIGMA website (http://enigma.loni.ucla.edu/protocols/). Upon removing segmentations which contained visually incorrect segmentations, the overlapping number of subjects was n=659 which included healthy elderly subjects, mild cognitively impaired, and Alzheimer's disease patients. The average bilateral hippocampal volume yields Pearson's correlation of 0.86 between segmentation criteria.



Work comparing methods done by others:

In other work, Freesurfer has a higher correlation to manual segmentations than FIRST. The overlap between Freesurfer/FIRST is 0.72 ± 0.037 in controls and 0.67 ± 0.042 in temporal lobe epilepsy patients (Pardoe et al., 2009).

Table 2. Overlap (Dice coefficient) between different hippocampal segmentation methods in controls and left mTLE hippocampi

	Controls (mean ± SD)	mTLE (mean ± SD)
Manual/Freesurfer	0.73 ± 0.028	0.66 ± 0.042
Manual/FSL-FIRST	0.71 ± 0.046	0.62 ± 0.057
Freesurfer/FSL-FIRST	0.72 ± 0.037	0.67 ± 0.042

The overlap is reduced in affected mTLE hippocampi, indicating less agreement between manual and automated methods ($p < 0.05$ for each control/mTLE comparison).

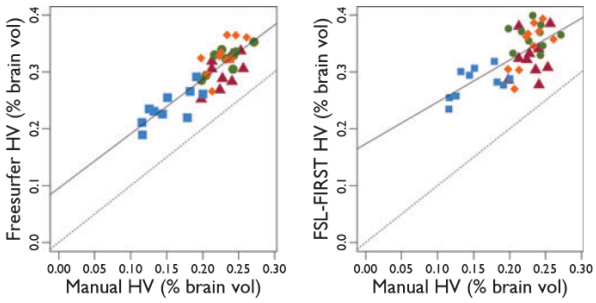


Figure 3.
The relationship between the manual hippocampal volume and the Freesurfer-based (left, r = 0.74, p << 0.01) and FSL-FIRST (right, r = 0.47, p << 0.01) hippocampal volume estimates in left and right hippocampi in controls and mTLE subjects. Blue squares represent ipsilateral mTLE, red triangles represent contralateral mTLE, green circles represent control right, and orange diamonds represent control left hippocampi. The solid line shows a line of best fit for all the hippocampal volumes; the dashed line shows a 1:1 relationship, indicating perfect agreement between the automated and manual techniques.
Epilepsia © ILAE

Freesurfer is more accurate than FIRST according to (Morey et al., 2009a). There was a small controversy about this because the authors did not report enough demographic information (Hasan and Pedraza, 2009), however it has not changed the results (Morey et al., 2009b).

Table 1
Comparison of automated measures to manual tracing

Automated measure	Average volume ± SD	% Volume overlap		% Volume difference		Correlation with manual segmentation
		Left	Right	Left	Right	
Hippocampus						
FreeSurfer	4190 ± 526.7	82% ± 1.5	82% ± 2.8	4% ± 2.1	5% ± 1.7	r=0.82, y-intercept=496
FSL-FIRST (threshold = 2)	4193 ± 634.9	79% ± 3.6	80% ± 2.9	4% ± 2.4	4% ± 2.3	r=0.66, y-intercept=1502
FSL-FIRST (threshold = 3)	4843 ± 743.4	78% ± 3.8	79% ± 2.8	6% ± 3.4	7% ± 2.3	r=0.66, y-intercept=1480
FSL-FIRST (version 4.1)	4404 ± 730.1	77% ± 5.9	80% ± 2.4	5% ± 3.2	5% ± 2.8	r=0.66, y-intercept=1480

One study shows that Freesurfer is less accurate than HAMMER (Akhondi-Asl et al., 2010), though it look like one of the segmentations failed completely in several cases in this study and that was not corrected.

Freesurfer is more accurate than Individual Brain Atlases using Statistical Parametric Mapping (IBASPM) (Dewey et al., 2010).

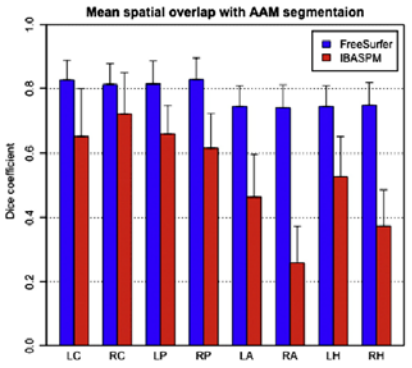


Fig. 1. FreeSurfer achieved significantly higher spatial overlap with AAM segmentation in every structure. Error bars indicate one standard deviation.

Another study shows that Freesurfer is fairly accurate but overestimates the volumes (Sanchez-Benavides et al., 2010).

Table 3
Comparison of manual and automated volumetric measures of hippocampus.

Manual vs. FreeSurfer	Global	Controls	MCI	AD
Left hippocampus				
% Overlap	79% (±7)	81% (±5)	79% (±7)	75% (±9)
% Difference	10% (±7)	9% (±8)	10% (±7)	11% (±12)
Pearson correlation	0.85	0.74	0.82	0.84
Right hippocampus				
% Overlap	77% (±6)	80% (±4)	76% (±6)	74% (±7)
% Difference	11% (±8)	9% (±7)	9% (±8)	13% (±8)
Pearson correlation	0.84	0.74	0.79	0.71

Overlap and difference percentages, and Pearson correlation coefficient.
MCI: Mild Cognitive Impairment; AD: Alzheimer Disease.

References:

Akhondi-Asl, A., Jafari-Khouzani, K., Elisevich, K., Soltanian-Zadeh, H., 2010. Hippocampal volumetry for lateralization of temporal lobe epilepsy: Automated versus manual methods. *Neuroimage*.

Dewey, J., Hana, G., Russell, T., Price, J., McCaffrey, D., Harezlak, J., Sem, E., Anyanwu, J.C., Guttmann, C.R., Navia, B., Cohen, R., Tate, D.F., 2010. Reliability and validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study. *Neuroimage*.

Hasan, K.M., Pedraza, O., 2009. Improving the reliability of manual and automated methods for hippocampal and amygdala volume measurements. *Neuroimage* 48(3), 497-498.

Morey, R.A., Petty, C.M., Xu, Y., Hayes, J.P., Wagner, H.R., 2nd, Lewis, D.V., LaBar, K.S., Styner, M., McCarthy, G., 2009a. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage* 45(3), 855-866.

Morey, R.A., Petty, C.M., Xu, Y., Hayes, J.P., Wagner, H.R., 2nd, Lewis, D.V., Labar, K.S., 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.

Pardoe, H.R., Pell, G.S., Abbott, D.F., Jackson, G.D., 2009. Hippocampal volume assessment in temporal lobe epilepsy: How good is automated segmentation? *Epilepsia* 50(12), 2586-2592.

Sanchez-Benavides, G., Gomez-Anson, B., Sainz, A., Vives, Y., Delfino, M., Pena-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.