



Comments and Controversies

Rebuttal to Hasan and Pedraza in comments and controversies: “Improving the reliability of manual and automated methods for hippocampal and amygdala volume measurements”

Rajendra A. Morey^{a,b,c,e,*}, Christopher M. Petty^{a,c}, Yuan Xu^{a,f}, Jasmeet Pannu Hayes^{a,b,c},
H. Ryan Wagner II^{b,c}, Darrell V. Lewis^{a,f}, Kevin S. LaBar^{a,b,e}, Martin Styner^{g,h}, Gregory McCarthy^{a,c,d}

^a Duke-UNC Brain Imaging and Analysis Center, Duke University, Durham, NC

^b Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC

^c Mental Illness Research Education and Clinical Center for Post Deployment Mental Health, Durham VA Medical Center, Durham, NC

^d Department of Psychology, Yale University, New Haven, CT

^e Center for Cognitive Neuroscience, Duke University, Durham, NC

^f Department of Pediatrics (Neurology), Duke University, Durham, NC

^g Department of Computer Science, University of North Carolina, Chapel Hill, NC

^h Department of Psychiatry, University of North Carolina, Chapel Hill, NC

ARTICLE INFO

Article history:

Received 15 May 2009

Revised 3 July 2009

Accepted 9 July 2009

Available online 17 July 2009

ABSTRACT

Here we address the critiques offered by Hasan and Pedraza to our recently published manuscript comparing the performance of two automated segmentation programs, FSL/FIRST and FreeSurfer (Morey R, Petty C, Xu Y, Pannu Hayes J, Wagner H, Lewis D, LaBar K, Styner M, McCarthy G. (2009): A comparison of automated segmentation and manual tracing for quantifying of hippocampal and amygdala volumes. *Neuroimage* 45:855–866). We provide an assessment and discussion of their specific critiques. Hasan and Pedraza bring up some important points concerning our omission of sample demographic features and inclusion of left and right hemisphere volumes as independent measures in correlational analyses. We present additional data on demographic attributes of our sample and correlations analyzed separately on left and right hemispheres of the amygdala and hippocampus. While their commentary aids the reader to more critically assess our study, it falls short of substantiating that our omissions ought to lead readers to significantly revise their interpretations. Further research will help to disentangle the advantages and limitations of the various freely-available automated segmentation software packages.

Published by Elsevier Inc.

Rebuttal

Here we address the critiques offered by Hasan and Pedraza to our recently published manuscript comparing the performance of two automated segmentation programs, FSL/FIRST and FreeSurfer (Morey, et al. 2009). An assessment of the specific critiques and discussion follows:

Lack of age and other demographic information

We agree that a clear description of sample characteristics is essential in any scientific report and is an important omission from our original report. Key demographic attributes of our sample of 20 participants are described in Table 1. While we do not believe the particular group of subjects we selected have biased our results, it is

possible one technique may be poorer at segmenting a specific demographic subgroup (e.g. aged) than another owing to the range of subjects who contributed to the atlas. While we agree that basic demographic information of the group is important and may affect absolute volume measures, we did not predict the expert human rater to be tracing the structures differentially with respect to age or other demographic attributes and therefore did not expect them to influence the comparison of methods. We acknowledge the age related changes in volume for certain structures cited by Hasan and Pedraza, however investigation of age related effects would be difficult to assess in the age demographic represented in our sample. This would be an interesting area of further research.

Table 1
Demographic attributes of sample.

Sample size	Mean age [SD]	Male/Female	African American	European American	Native American	Asian American	Other
20	41 [10]	14/6	10	7	1	1	1

* Corresponding author. Duke-UNC Brain Imaging and Analysis Center, Duke University Medical Center, Box 3918, Durham, NC 27710, USA. Fax: +1 919 416 5912.

E-mail address: rajendra.morey@duke.edu (R.A. Morey).

Table 2
Left and right hemisphere correlations.

Measure	Region	FIRST	FreeSurfer	Manual
L–R correlation	Hippocampus	.75	.92	.87
	Amygdala	.69	.65	.63
Correlation with manual tracing	L-hippocampus	.56	.77	1
	R-hippocampus	.76	.87	1
	L, R-hippocampus	.66	.82	1
	L-amygdala	.24	.53	1
	R-amygdala	.23	.65	1
	L, R amygdala	.24	.56	1

Lack of tabulation of individual volumes from manual tracing

It is unclear how tabulating individual participants' volumes would help the interpretation of the results. We are comparing the outcome of two algorithms and therefore the relative differences between methods are much more important than the absolute volumes for individual subjects. Individual volume measures might be useful if a publicly available dataset was used, however this was not the case. Furthermore, manual hippocampal and amygdala segmentations depend on the segmentation protocol and thus it can be misleading to directly compare absolute volumetric measurements.

Introduction of a correlation by not considering left and right structures which ought to be tabulated separately

We agree that because the volumes of left and right structures are generally correlated (Pruessner et al., 2000), this may tend to inflate any correlation that includes left and right as independent measures. We checked the correlation of left and right hemispheres of the hippocampus and amygdala and indeed they are correlated for each of the three methods employed (see Table 2). When left and right hemisphere correlations are assessed separately, the right hemisphere had generally higher correlations with manual tracing than the left hemisphere. The correlations obtained from using left and right hemisphere data as independent measures (as reported in Morey, et al. 2009) led to values intermediate to the separately computed left and right correlations as seen in Table 2. Regarding our power estimates, treating the hemispheres as independent, given the left–right correlations, results in an underestimation of variability.

This leads to overstated inferences (too lenient) and overestimates of power, although it does not change the relative position of the methods. Additional data on laterality is conveyed in the shape analyses presented in Figs. 9 and 10 of the original paper (Morey, et al. 2009). Therefore, in our estimation, this information does not lead to a significant revision of the main conclusions about either of the methods. More conclusive results are certain to be made available from future validation studies.

Evidence of possible sources of difference related to the atlases used in FSL and FreeSurfer with a potentially missed opportunity for further analyses by Morey et al.

Including this point would enhance the completeness of the Discussion section. Atlas differences is one possible source of diverging results between FSL and FreeSurfer. The suggestion for comparing systematic bias of manually delineated volumes to the atlas from FreeSurfer goes beyond the scope of our study as this could be a manuscript unto itself. Indeed, Hasan and Pedraza cite the work of Shattuck, et al. (2008) as a careful study of the influence of atlas selection.

In summary, Hasan and Pedraza bring up some important points concerning our reporting of methods and results. While their commentary aids the reader in more critically assessing our study, it falls short of substantiating that our omissions and methodological shortcomings ought to lead any reader to significantly revise their interpretations. Further research is certain to help disentangle the advantages and limitations of the various freely-available automated segmentation software packages.

References

- Morey, R., Petty, C., Xu, Y., Pannu Hayes, J., Wagner, H., Lewis, D., LaBar, K., Styner, M., McCarthy, G., 2009. A comparison of automated segmentation and manual tracing for quantifying of hippocampal and amygdala volumes. *Neuroimage* 45, 855–866.
- Pruessner, J.C., Li, L.M., Serles, W., Pruessner, M., Collins, D.L., Kabani, N., Lupien, S., Evans, A.C., 2000. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. *Cereb. Cortex* 10 (4), 433–442.
- Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 39 (3), 1064–1080.