Reviewer’s report

Title: Disparate voxel based morphometry (VBM) results between SPM and FSL softwares in ALS patients with dementia: Which VBM results to trust?

Version: 1 Date: 25 April 2014

Reviewer: Ali Khan

Reviewer’s report:

The authors present an extension of their previously published voxel-based morphometry comparison (FSL vs SPM) which found disparate results on their dataset (18 ALS patients, 15 controls), this time comparing against Freesurfer, which they use as a surrogate ground truth. The disparity between FSL and SPM is alarming, with SPM revealing 1% atrophic voxels vs 23% for FSL. I have several concerns and questions regarding the methodology, both on how such disparate results were found and how these could be compared against a surface-based cortical thickness method. These would need to be addressed and re-reviewed before I could suggest acceptance.

Major comments:

1. Although it is not the focus of the paper, I must comment on the FSL vs SPM comparison upon which this paper is founded on. In that comparison, the authors tried to use a common set of parameters settings (such as smoothing kernel, subjects used for template generation, non-parametric statistics), and also evaluated how one methods segmentation or registration used in place of the other affected the results. Ultimately, the authors concluded that although the segmentation and registration has an effect on the outcome, it could not account for the huge disparity seen between the two methods (FSL being much more sensitive than SPM). The authors suggested that the threshold-free cluster enhancement (TFCE) used in FSL and not in SPM could be due to the difference, and I am willing to bet that this hypothesis is correct, however it confounds me why the authors did not compare the methods without TFCE. TFCE is a hybrid method that attempts to provide the benefits of cluster-based maps (significance based on the size of suprathresholded clusters), without the need of specifying the setting for the suprathreshold. This would significantly change the number of significant voxels one would observe and puts the FSL and SPM methods inherently on very uneven grounds. The solution for a fair comparison would be to compare the two without TFCE enabled. I admit this should have been addressed in the previous work, but since that has already been published, the authors should at least address this here.

2. My second major concern relates to how the authors were able to compare the surface-based (vertex-based) freesurfer maps with the volumetric (voxel-based) maps from FSL and SPM. This is an important consideration, as there is no one correct way to do this — with the two basic choices to either voxelize the
Freesurfer maps, or map the volume-based maps to the surface. No information at all is provided as to how this was done and how the authors ensured no information was lost in the volume->surface or surface->volume mapping. This needs to be outlined in detail, and I would suggest looking at how the following paper performed the volume/surface comparison to ensure appropriate methodology is used.

NeuroImage. 51: 214-220. PMCID: PMC2862732.

3. Finally, the authors assume Freesurfer to be the gold standard, both for atrophy assessment and cortical thickness assessment. I agree that Freesurfer is the tool of choice for many researchers, however, one caveat with Freesurfer’s accuracy and reproduceability is that the images acquired must be of high enough quality, and the appropriate checks and edits must be made throughout the recon-all workflow to ensure quality control. The authors do not mention whether this was done, or provide any validation whether the cortical surface reconstructions and resulting thickness maps are reliable for their datasets.

4. Furthermore, there are several cortical thickness analysis pipelines openly available, (NITRC is a good resource for locating these), some of which are volume-based instead of surface-based. Given that there exists disparity between VBM approaches, it is reasonable to assume there may be disparity among cortical thickness approaches and thus more than one cortical thickness analysis method should be used.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declare that I have no competing interests.