Author's response to reviews

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

Authors:

Venkateswaran Rajagopalan (venkatnrp78@gmail.com)
Erik P Pioro (pioroe@ccf.org)

Version: 2
Date: 16 December 2014

Author's response to reviews: see over
Addressing reviewers comments

Reviewer 1

Comment 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.”

Answer: We thank the reviewer very much for bringing this valuable point to our attention, which we had previously overlooked. When compared to using TFCE, we performed statistical analyses by disabling the TFCE option and, as the reviewer correctly pointed out, found the number voxels with statistical significance to drastically decrease. We addressed the details of the analyses and the results in Introduction page 5 last five lines, methods page 8 last 4 lines, results separate section under the title "Not Using TFCE in FSL " starting page 13, and discussion and conclusion page 15,16 and 17.

Comment 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.”

**Answer:** We thank the reviewer for this suggestion. We did not transform each subject’s cortical thickness and volume surface maps to volume or vice versa. We performed all analyses in respective softwares until the final statistical parametric maps were obtained in the respective softwares. As the reviewer correctly pointed out that there is no one correct method to do this, we voxelized cortical thickness and cortical volume surface statistical parametric maps to MNI space so that Dice similarity index was measured with FSL and SPM statistical parametric maps (which were in MNI space). We believe this approach will result in the least error because converting volume to surface would require transformation of FSL and SPM volume maps and resampling (interpolation) would affect the results. We adopted the approach as suggested by Klein et al to convert surface to volume maps. This information is given in Methods section on page 11 first paragraph.

**Comment 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 reproducibility 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.

**Answer:** Although we do not consider Freesurfer as a gold standard, based on our experience with the software and also from the literature, we have found Freesurfer to be more robust than volume based approaches given on page 16 lines 8-10. We performed quality control checks with both authors evaluating the quality of Freesurfer results (by performing appropriate checks and edits in the entire image processing work flow). As suggested by the reviewer, we have added the above information on page9 last two lines and page 10 first two lines.

**Comment 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.

**Answer:** We thank the reviewer for this suggestion, but to stay within the scope of this paper we could not perform this analysis at the moment. However, on page 16 last 8 lines, we have suggested these types of comparisons for future studies.
Reviewer 2

Comment 1: One concern with the findings is that FSL and SPM applied very different thresholding methods (TFCE vs. voxel-wise), which could explain the results. The authors need to conduct a post-hoc analysis using voxel-wise thresholding for both packages to allow a better comparison.

Answer: We thank the reviewer very much for bringing this valuable point to our attention, which we had previously overlooked. Reviewer 1 also brought up this same point for which we performed statistical analyses and added those results. Please see response to reviewer 1’s comment 1.

Comment 2: It’s not clear how cortical volume was calculated for Freesurfer and cortical thickness was calculated for FSL, SPM. Please clarify this in detail.

Answer: We thank the reviewer for bringing this point to our attention. Reviewer 1 also brought up this same point. Please see response to reviewer 1’s comment 2.

Comment 3: Please explain the DICE similarity index in more detail for readers, as it is a critical factor in the paper. Currently there is only a study cited using this technique, which is not sufficient.

Answer: We thank the reviewer for this suggestion. We have explained the Dice similarity index and cited additional literature on page 10 last paragraph and page 11 first paragraph.

Comment 4: The title should explicitly state the patients have ALS-FTD and not simply ALS with dementia. Please state this more clearly and change the title and abstract accordingly.

Answer: We have made appropriate changes in the title and abstract.

Comment 5: There is virtually no information on the diagnosis of the patients. Were the Strong diagnostic criteria for ALS-FTD applied? Again, this is critical for the characterization of the patients.

Answer: No clinical data had been included because this paper was intended to be more of a technical report extension of a previous paper in which the clinical data had been published. The criteria used to define the FTD are those of Neary, et al – and have now been added to the text and reference list.

Comment 6: Similarly, there are no demographics given for both groups. Please include this information, which will also allow to judge whether any nuisance covariates need to be employed (eg age).

Answer: As indicated in point 5, no clinical data had been included because this paper was intended to be more of a technical report. However, the previous paper in which the clinical data
had been published is now referenced. Age was regressed out in our GLM model, this is mentioned in the methods section.