Reviewer's report

Title: Assessment of the impact of the scanner-related factors on brain morphometry analysis with Brainvisa

Version: 1 Date: 16 May 2011

Reviewer: Frithjof Kruggel

Reviewer's report:

Quantitative MRI is increasingly used in Neuroscience, and clinical studies often span multiple centers, in order to recruit sufficient subjects or to increase the statistical power for the sample. However, it was recognized that the reproducability of image-based quantities is limited, and depends on the scanner hardware, the imaging protocol, and the processing software. This submission describes a retrospective analysis of MR imaging data acquired at different sites on the basis of global and local morphometric measures obtained using the BRAINVISA and FSL software tools.

The topic of the manuscript is scientifically relevant, and of interest to the audience of BMC. The text is well written, without major errors (except where noted below), and straightforward to understand for a greater audience in neuroscience. However, several major and minor issues should be dealt with before publication can be considered. Especially, the results section needs considerable revision. At this time, all issues should be considered as compulsory.

Major/general issues:
1. Based on this description it is unclear at which (or whether) spatial normalization was applied in the process. It is assumed that compartment volumes were obtained in individual space, but the probabilistic models most likely require transformation into a common space. Please, revise the manuscript to clarify, and revise your results in terms of whether normalization was applied or not.
2. Brain compartment volumes scale with body height, and males typically have 8% larger volumes than females. In addition, there is age-related loss of brain tissue that amounts to about 5% in 30 years. Rather small
subject groups were studied here. How much is the between-group (1.5T vs 3T) difference explained by sample-related differences?

3. It is unclear whether any confounds were taken care of in the statistical analysis. From the previous point it is strongly suggested to take body height (or weight as a weaker proxy), gender etc into account. Using regression instead of ANOVA may provide more insight because the amount and sign of differences can be reported. Please, discuss and/or revise your analysis.

4. Consider revising the discussion section. Rather than providing a summary of the study results, a deeper insight into the reasons behind those differences would be much acknowledged. Why are results for Center B different? For example, FAST uses local information to classify compartments while the algorithm in BRAINVISA uses histogram information only. Thus, it is expected that FAST has a higher reproducability. A second example: "The researchers need to be aware of the impact of scanner-related factors...and should take such effects into account." Agreed - but how should we take such effects into account? What do these results imply on the calculation for sample sizes in multi-center studies? Are multi-center studies still feasible? In the same vein, it would be useful to relate results and conclusions of this study to several other studies of the within- and between-center differences in quantitative MRI (most of which are not cited here). Please, revise your results and provide more meaningful conclusions.

Minor/specific issues:
1. p.3, bottom: Explain FAST.
2. p.8 and Table 1: Include all information about the imaging protocol (e.g., sequence, coil type).
2. p.8: Where all re-tests performed on the same scanner?
3. Table 3 & 4: Consider transposing thes tables to ease comparison with Table 2.
4. How do you interpret an interaction between centre and visit?
5. Fig. 2: Are volumes given per hemisphere? If yes, please state this clearly. If not, volumes are anatomically unplausible. Why do you split graphs by group (rather than compartment)? A GSI < 1 is anatomically unplausible. Please, check and revise. Using the same GSI range for
both groups would ease the comparison. Did you pool hemispheres here? Is there a side-difference?

7. Fig. 3: What is the rationale of showing confidence intervals here? Are results given for a single (which?) hemisphere? How do you explain the stunning difference of 20% surface difference between groups? Consider using the same scale for all diagrams. If differences are not due to sampling: can we use sulcal surface as a measure in a multi-center study?

8. Fig. 5: Because MRI does not have a definition of an intensity scale, there is little useful information in the bottom figures. Consider finding a better measure to support your observation (e.g., the WM/GM contrast?).

9. Fig. 6: The information displayed here might be more meaningfully reported as a group-related difference obtained from a regression model. Why do you switch the 3T and 1.5T columns here? Again, it is assumed that values correspond to hemispheres. Do you pool results from both hemispheres? Are there significant side differences?

10. Fig. 7 top: What does a single measurement (a dot) mean here? Why are sulcal surfaces reported here much less (on average) than those in Fig. 3? Is a sulcal surface of 10000 mm^2 likely? Please, check and revise.

11. Fig. 7, bottom: What does a single measurement (a dot) mean here? If a dot corresponds to the depth of a specific sulcus, then averaging across sulci is as useful as averaging finger lengths across the five digits. What is the rationale of reporting a median of means? Is a mean geodesic depth up to 60mm in a hemisphere anatomically plausible? In the very end: does this graph reveal any insight?

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.