Reviewer's report

**Title:** Quantitative and Qualitative Assessment of Structural Magnetic Resonance Imaging Data for Multi-Center Studies

**Version:** 1 **Date:** 8 March 2012

**Reviewer:** Frithjof Kruggel

**Reviewer's report:**

The inclusion of quantitative MRI protocols in multi-center studies has raised the question of the intra- and between-center precision of MRI-based morphological measurements. This submission picks up this question and supposes a protocol to optimize and standardize an anatomical MRI protocol to obtain the highest precision in specific morphometric measurements. Results from a small-scale study in two participating sites are included and discussed.

The topic of this manuscript is scientifically and clinically relevant and of interest to the audience of this journal. The text is (in most parts) clearly written and straightforward to understand (except where noted below). Without a doubt, there is some merit to this study, and questions raised (and partially addressed) are important for future clinical studies including morphometric measures. However, several major and minor issues have to raised:

**Major/general issues:**
1. The use of the term "multi-center" is a bit far-fetched. Only two centers were involved in this study, both with the same scanner hardware, equipment and software. This is most likely the best possible scenario for a "multi-center" study, and hardly realistic for a "real-world" multi-center study. Consider revising your title (and the use of "multi-center" throughout the manuscript) and use a more careful wording about the generalizability of your results.
2. Unless required by this journal, it is suggested to move the "Methods" section before the "Results" section. This is a methodological study, and results cannot be understood without a clear understanding of the methods. As such, a reader has to go forth and back in the
3. A major concern is raised against the conclusions drawn from data included in Table 1 and 2. It is understood that data are based on the examination of a single subject, and just one scan per protocol. Because no information about the precision of the quality measures is included, the evaluation of the results may be arbitrary. The Euler number will depend on (slight) body motion which is assumed to vary between scans. This is even more the case with results in Table 2, where differences between protocol modifications most likely are insignificant. Although the general approach for protocol optimization is viable, data presented here are not convincing in terms of the conclusions drawn. Please, discussion and/or revise.

4. Note that both quality measures (Euler# and CNR) are not independent. This usage of the Euler# is sensitive to local segmentation error at the white/grey matter boundary. Therefore, the measure is dependent on noise and tissue contrast, as measured in CNR. Numbers in Table 2 show that there is a considerable negative correlation between both measures. While tissue contrast and noise are important measures of image quality, segmentation errors as captured by the Euler# are corrected by the algorithms included in FreeSurfer. Consider replacing the Euler# by a more meaningful, independent quality measure.

5. Although the approach taken here has its merits, issues brought up in the introduction remain unanswered: (1) Do longitudinal studies in single subjects yield useful results? Assume a typical (total) brain loss in AD of 0.4%/year - given a retest precision of about 1% as found here, significant results will only be obtained after 3 years. (2) Do multicenter studies yield useful results? The 5% between-center precision may cast a doubt on this. Your thoughts?

Minor/specific issue:
1. Abstract: Consider replacing "software upgrade" by "scanner software upgrade" to avoid possible confusion with an upgrade in morphometry software.
2. p.4, center: "two types of measurements ... quantitative measures, ... and data quality measurements...". Data quality measurements are also quantitative measurements. Consider rephrasing this sentence to avoid
the notion that quality measures result in a "pass/no pass" outcome.
3. p.5, middle: Include a reference to Table 3 here, so that a reader can relate to the imaging protocols.
4. p.5, bottom: Explain your understanding of "artefacts" here. Perhaps, include a reference to Figure 1 here.
5. Table 1 & 2: The rows "mean" and "total CNR" are dependent on the other two rows. Including them in the total score does not make sense.
6. Figure 2: GM segmentation, F2: What happened here? A mean difference of 35% is enormous and should be revised or reported with a comment.
7. Software upgrade & Figure 6: It is unclear how effects of the software upgrade are distinguished from the within-center retest precision. A more throughly statistical analysis is needed here.

**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.