Reviewer's report

Title: Improved Assessment of Multiple Sclerosis Lesion Segmentation Agreement via Detection and Outline Error Estimates

Version: 1 Date: 13 March 2012

Reviewer: Balasrinivasa Sajja

Reviewer's report:

The manuscript titled “Improved Assessment of Multiple Sclerosis Lesion Segmentation Agreement via Detection and Outline Error Estimates” by Wack et al. presented a method for assessing the rater’s ability to detect and outline MS lesions on FLAIR MR images. This is a topic of interest for researchers and clinicians in the field of lesion segmentation and evaluation.

Discretionary Revisions:

Comment: Discussion and Conclusion sections are well written compared to rest of the manuscript.

Minor Essential Revisions:

1. Include a representative FLAIR MR image of a slice that shows MS lesions segmentation performed by the two raters and the differences in their lesion assessment.
2. Page 4, line 9: Udupa et al. (2002, 2006)…. Referencing to articles should be uniform throughout the manuscript.
3. Page 4, line 20: ‘….applicable to ROIs used to measure other hyperintensities,…’. Not necessarily the hyperintensity. It depends on the type of image sequence used. For example, on T1 weighted image, black-holes (matrix destruction) appear as hypo-intense. The presented methodology can be extended for evaluating MS black-hole segmentation too.
4. Page 7, Line 17: ‘ comparing operators’ ROIs (SI=.35-.85).’ What are these numbers?
6. Page 8: in Methods: ‘voxel size was .94x.94x3 mm’. Unit should be mm³.
7. Usually, in the evaluation of MS lesion segmentation methods, SI is analyzed along with other measures false positive (FP) and false negative (FN) lesion classification. Of course, this terminology is meaningful when we have a reference ’ground truth’ for comparison. However, if we consider FP and FN lesion classifications of one rater’s results with respect to the other, it still makes sense. In that case, is OE equal to FP+FN?
8. Page 9: Line 4: ‘Hyperintense lesions were outlined using JIM’s semi-automated contouring tool, except in a few cases which required manual outlining to achieve a proper lesion contouring.’ Since most of the lesions are
outlined using the semi-automated software, is the manual outlining part contributes significantly to disagreement between the raters?

9. This reviewer agrees with the authors that the DE does not depend on the lesion load or even lesion size. However, this reviewer believes that DE would vary with the degree of variation in hyperintensity on FLAIR MR images associated with MS lesions. That is, subtle lesions, irrespective of their sizes, are more likely to be missed by the raters.

10. Page 11, Line 13: “We also observed that rater 1 drew many more mismatched small lesions …. ” Mismatched to what?

11. In Figure 1A: what is the difference between Region 3 and Region 4?

12. All acronyms should be defined only when they are used for the first time. Words like multiple sclerosis (MS), Jaccard index (JI), similarity index (SI) are defined multiple times in the manuscript.

Major Compulsory Revisions:
None.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

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