Reviewer's report

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

Version: 1 Date: 16 January 2012

Reviewer: Frithjof Kruggel

Reviewer's report:

Quantitative biomarkers are needed to rate the progress or remission of disease processes. In longitudinal studies of patients suffering from Multiple Sclerosis (MS), often the total lesion volume is used as such a marker. Manual outlining of the lesion boundaries is still considered as the method delivering the "gold standard" of lesion volume. The are numerous studies that focus on the quality of lesion detection. This submission proposes to replace the often-used similarity index (SI) between two detections of the same lesion with two measures, the detection error (DE, roughly, the disagreement in detecting the same lesion), and the outline error (OE, the disagreement in outlining a region). Properties and performance of both measures are exemplified on a medium sized dataset of 17 scans.

The topic of the manuscript is clinically relevant and within the scope of BMC Medical Imaging. The text is well written and (mostly) straightforward to understand. No major errors were found, except were noted below.

General comments:

The findings of this study are rather obvious. Replacing one useful error measure (SI) by two useful measures (DE, OE) will likely yield a more differentiated assessment of errors. Unfortunately, both measures are not new in terms of the signal detection theory, an area that is well developed during the last 70 years. Thus, on the methodological side, the is little (if any) news here:

1. The detection error (DE) is related to the sensitivity of lesion detection: the smaller a lesion, the more likely it is missed. A correlation with the total lesion load is not expected. Thus, the
discussion in the second paragraph of p.14 is nothing more than a
definition of the detection threshold for a (human) operator.

2. The outline error is (by definition) proportional to the total
length of the lesion boundaries, thus, a correlation with the lesion
volume is expected (as discussed on the top and bottom of p.14).
However, this relationship may only hold for compact lesions
(such as found in MS), but not for more complex-shaped lesions (e.g.,
as found with malignant tumors or strokes). Differences in the shape
characteristics may also explain the observation on p.15, top.

Are these measures clinically useful? This is likely, but perhaps
does not warrant a (full) journal publication.

Specific comments:
1. Abstract: The relation between DOEE, DE and OE is unclear. Does
DOEE correspond to a set of two values per rater?
2. p.9: Was does "semi-automated" contouring tool mean?
3. Fig.4 is redundant w.r.t. the text.
4. Fig.5: Because the minimum number of mismatches is one, extrapolating
the curves beyond the maximum lesion size does not make sense. Consider
revising the curves and adjusting the x range. Else consider expressing
the y axis as a probability (ratio of mismatched vs. total regions at
a specific lesion size).

**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.