Reviewer's report

Title: Decision-making is impaired in multiple sclerosis: impact of subcortical atrophy and disease disability

Version: 2  Date: 24 January 2015

Reviewer: Jesper Hagemeier

Reviewer's report:

This interesting and novel study investigates the previously little explored correlate of (impaired) decision making in MS with an MRI measure. The manuscript provides an interesting point of view and extensive (statistical) analysis, which is novel. Overall, the MS and healthy control results are interesting. However, I have several major concerns, the most important of which pertains to the composition (and size) of the sample (especially RR groups) and whether it is representative of the (MS) population. The unexpected demographic group differences (although non-significant) should be explained, and any strong conclusions should be prefaced with a caution. Also, the MRI protocol should also be elaborated.

Major revisions:

1. Disease duration between the two RR groups is quite different (17 years vs. 11 years), although non-significant, even though group 2 has significantly higher EDSS and MSSS (by definition). This is a somewhat perplexing as one would expect the opposite. The authors also state that "Disease duration and age at onset of MS were similar among patient subgroups". I would argue that disease duration and age are not similar between MS subgroups (but age at onset is similar), and that the reason for non-significance is because of a lack of power to detect difference (n=13, 9 and 10). As it stands, it is hard, if not impossible, to draw any RR-1 vs RR-2 vs SP conclusions, since the group matching is sub-optimal at best and the statistical tests were presumably not corrected for differences. What is the authors' explanation that the more severe EDSS/MSSS group is quite a bit younger and has a much shorter disease duration? Again, one would expect the longer disease duration in RR-2. The authors should address this issue.

2. Also in this context, it is not entirely clear to me whether covariates have been used in the models (I would presume not except for the ventricular width partial correlations, based on the chosen statistical methods as well as the small sample size), the difference at disease duration and age should be addressed and discussed at least as a limitation.

3. The MRI section needs to be greatly elaborated on, or at least a table should be added depicting the different MRI centers/scanner/protocols used. At present, it is not even clear to the reader how many centers were involved in this study "scans were performed in different centers with a variety of protocols". Were
controls all imaged in the same center/scanner? This should also be discussed in the discussion/limitations.

4. The discussion lacks an extensive limitation section, which should be added as there are quite a few limitations. Also, several conclusion are a bit overreaching, such as "both types of neurocognitive deficits were independently explained by the degree of subcortical atrophy involving the thalamus and striatum", even though the actual measurement was width of the ventricles, not actual subcortical atrophy of the thalamus/striatum. Also, the discussion needs more extensive review of MS MRI literature and its relation with for example executive functioning that is of interest to this study. In its current form the discussion is lacking in this regard (neuropsychology literature seems to be well discussed).

Minor comments:

1. Since subcortical atrophy is not explicitly measured, please consider renaming to “ventricular width” or “estimates of subcortical atrophy” or something of the like in the title and throughout.

2. The use of both a healthy control group in some analyses and a non-MS control group in other analyses is somewhat confusing to the reader. It is not completely clear as to why these additional MS vs non-MS ventricular width "Validation" comparisons were performed, they distract from the message of the paper. In addition, several publications have already been cited as to the validity of the measure. Why were these analyses carried out? The results are difficult to interpret since that group is very heterogeneous and therefore warrants further elaboration also in more detail in the discussion.

3. What is the rationale for only using ventricular width as an MRI measure and no other direct volume measurements?

4. Throughout the manuscript mention is made of “decision-making under explicit risk” as an inherent function of the GDT score. In the cited reference (#12), however, no mention is clearly made of “explicit risk”. It is not clear what exactly is meant by “explicit risk”. For the reader who is not familiar with the GDT, this can be confusing. Explicit risk could refer to other matters (such as for example risk of progression of MS), wholly unrelated to GDT.

5. Intra-rater reliability of the width measures that estimate brain atrophy was high (.99), and these width measurements are relatively straightforward and easily done. Are the authors confident that the measures are a reliable estimate of brain/structural volume (as compared to 3D measures)? Especially considering the potential user bias, and inherent limited measurement (length as opposed to an actual volume). An expanded discussion/limitations would be valuable. On P15, line 14 only reference 33 is mentioned, although this reference makes no comparison with other atrophy or volume measures. As such, Martola et al.’s width measures offer good evidence of ventricular widening over the years, but not for the method itself per se. Different reference(s) are needed here to make a claim of validity (e.g. #2).
6. Twelve patients were taking antidepressant and 5 were taking anti-anxiety/sleep medication. In the methods section it is acknowledged this might influence cognition/behavior/decision-making, but this is not elaborated upon in the discussion. Please identify it as a potential confounder and discuss how this could have influenced your conclusions. For example, data medication use is provided in the methods section for RR and SP, but not for RR-1 and RR-2, could this possibly underlay the observed statistical differences?

7. I could not find the a priori rationale for splitting the already small RR group in two based on EDSS <>3. It would seem that splitting may cause more harm than good by reducing the sample sizes and more or less artificially separating the RR group causing unexpected disease duration differences which make interpretation difficult. If taking the lead of Kleeberg et al (page 4 line 4) it seems EDSS 2 was used as a cutoff. Please elaborate.

8. In the moderated regression models (p12, l15) it is mentioned that the effects found in model 2 remain significant in model 3. If I understand this correctly, these are the main effects, as the interaction effects are said to be not significant (line 18). Because of this, the b, SE, t, p-values etc. only seem to confuse the reader as they do not add any relevant information, consider omitting this.

9. Ventricular width measures were correlated with GDT but not with any of the other composite scores or strategy shifts. On p13 l23 it is mentioned that GDT vs. composite scores did not correlate well (as a general remark) and thus additional correlations were carried out. Were additional correlations of individual executive functions scores also carried out vs. the ventricular width measures? The reason for non-significance of ventricular width vs. composite scores may be the exact same (ie that the composite score does not fully embody the sub-scores).

10. Figure 1 and p11. What is the authors explanation/speculation that in the RR-2 group the GDT is so low? Could it be caused by the low number of individuals?

11. Post-hoc tests were corrected for multiple comparisons. Was this also the case for the (likely high number of) correlation analyses?

**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