Author's response to reviews

Title: Decision-making under explicit risk is impaired in multiple sclerosis: impact of ventricular width and disease disability

Authors:

Ashley D Radomski (adr2@ualberta.ca)
Christopher Power (chris.power@ualberta.ca)
Scot E Purdon (scot.purdon@albertahealthservices.ca)
Derek J Emery (demery@ualberta.ca)
Gregg Blevins (gregg.blevins@ualberta.ca)
Kenneth G Warren (kenneth.warren@ualberta.ca)
Esther Fujiwara (efujiwara@ualberta.ca)

Version: 4 Date: 4 March 2015

Author's response to reviews: see over
26/02/2015

Dear Dr. Geurts,

Please see enclosed our response letter to our previous submission BioMed Central MS: 4965631331494339 - Decision-making is impaired in multiple sclerosis: impact of subcortical atrophy and disease disability. We have addressed each comment in point-form and feel that the changes we made significantly improved the quality of the paper. In addition to the requested changes, we have also become aware of two recently published relevant articles: Muhlert et al. (2014). The grey matter correlates of impaired decision-making in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*. Cogo et al. (2014). Cognitive correlates of under-ambiguity and under-risk decision making in high-functioning patients with relapsing remitting multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, (ahead-of-print), 1-10. We integrated these articles' findings in our revised manuscript. All changes to the manuscript are marked as ‘track changes’ or printed in red with yellow underlay in the word document.

Sincerely,

Esther Fujiwara
Editor's Comments

General
Please reconsider carefully all referral to causation (not warranted). Also, please answer carefully to the first referee's point regarding the method of deep GM atrophy measurement.

Reply: Throughout the manuscript we have made wording changes so that we are no longer implying a causal relationships between cognitive and ventricular width variables. We also now provide substantially more details regarding the methods of assessment of the ventricular measures and their possible interpretation. In summary, we provide an additional file with a table listing all MRI scanning details (Additional File 2), we refer to previous literature regarding the correlations between ventricular width measures and actual volume loss (regional or whole-brain; pages 3-4 of the revised manuscript), we compare some of our absolute measures of ventricular width to previously published ventricular width measures (page 18), and we now provide additional ventricular ratio measures (calculated for our sample) correcting for brain width (Table 5).

We feel the revised manuscript has improved substantially as a result of these changes.

Reviewer 1

Major Comment 1: Various ‘simple’ MRI measures are used to measure (subcortical) atrophy. I understand that it is impossible to run more advanced imaging analysis methods on conventional (2D) clinical MRI data, but I would like to ask the authors to mention results (i.e., correlation coefficients) of previous studies that correlated for example ventricular width to whole brain volume (or subcortical volumes). This way, the authors can better justify the use of ventricular width as a measure for atrophy. If previous studies have not shown that ventricular width perfectly correlates with subcortical volume, I would like to ask the authors to change the term ‘subcortical atrophy’ in something more suitable (e.g., ventricular width) in the text and title.

Response to 1.M1: We appreciate this comment, which mirrors Reviewer 3, minor comment 5. We now have included more targeted literature regarding the validity of using 2-D linear measures to estimate volumetric measures (whole-brain, regional volumes), and explicitly state previous figures on the strength of their association (pages 3-4 in the introduction). There is no perfect match between ventricular size and such volume measures, although previously reported correlations were moderately high (~0.4-0.8). Therefore, we agree with the reviewer’s caution, and made subsequent wording changes in the title and throughout the manuscript omitting the reference to “subcortical atrophy” and referring merely to “ventricular width” where indicated. Nevertheless, we would like to note that many of the consulted papers refer to ‘central atrophy’ when using third ventricle width/intercaudate distance as a proxy for atrophy (e.g., Horakova et al. J Neurol Neurosurg Psychiatry 2008, 79:407-414. Sanchez et al. Eur J Neurol 2008, 15:1091-1099; Tekok-Kilic et al. Neuroimage 2007, 36:1294-1300. Benedict et al. Arch Neurol 2004, 61:226-230). In order to retain and convey to the reader the potential meaning of these ventricular measures, we therefore refer to “central atrophy” in the abstract, and on pages 4, 6, 20-21, 24, 26, and 27 without referring to any specific subcortical locations.
We considered that this compromise is not overstepping the boundaries of our actual methods and results, but would be amenable to further change should this appear necessary.

**Major Comment 2.** The aim of this paper is to investigate whether “problems with decision-making under explicit risk in the GDT are linked to increasing disability and loss of brain volume in the course of MS.” Firstly, I would like to ask the authors to define the term ‘disability’. Secondly, in my opinion, increasing disability and loss of brain volume are not two independent measures. That is, atrophy is a relatively good predictor for physical and cognitive problems in MS; patients with severe atrophy will probably also experience (severe) clinical disability. Perhaps the authors could be more specific concerning their research question.

**Response to 1.M2:** We thank the reviewer for his suggestion to add some clarification to this point. We use the term ‘disability’ only as it is reflected by the EDSS score. A note on this definition is now placed into the introduction (page 6 “disability defined as functional impairment according to the EDSS”). We tested both here, whether disability (EDSS) and ventricular enlargement may predict decision making. We further agree with the reviewer that the level of brain atrophy should be related to disability (i.e., EDSS) and may perhaps therefore be considered redundant. However, our results showed: A) ventricular size measures were marginally, but not significantly correlated with the EDSS in our sample (the previous trend correlation was removed, as requested in minor comment 30). B) Psychomotor speed as well as the GDT net-score were both correlated with the ICR/TVR ventricular size measures and with the EDSS. C) Conversely, the Global cognitive function composite score and the GDT shift score were correlated with EDSS but not with ventricular size. This implies that some cognitive functions, i.e., advantageous decision-making, psychomotor speed, but not all cognitive functions (i.e., executive functions, memory functions) were related to the specific ventricular width measures here. Given the coarse nature of the ventricular measures, we do not wish to overstate the importance of these difference in the correlational patterns between EDSS, cognition, and ventricular width. Pages 19/20 nevertheless reiterate the findings and pages 21/22 give a short discussion.

**Major comment 3.** Could the authors please explain why they divided the MS group into subgroups based on EDSS score and MS type, and how they have determined the cut-off values? Please note that correlating EDSS with ventricular width/Game of Dice Task (GDT) in all three subgroups is somewhat circular in terms of reasoning. I would suggest the authors to omit these correlation analyses.

**Reply to 1.M3:** We fully agree with the argument of circularity and have omitted the correlations of EDSS – ventricular size measures within subgroups. MS is not a uniform disorder and subdividing patients based on relapsing or progressive MS subtype reflects major differences in overall neurological status as well as treatment, which could then also differentially affect cognition. However, splitting up the relapsing-remitting subtype further is somewhat arbitrary, as also noted by reviewer 3. Nevertheless, we attempted to do so to illustrate whether there may be a gradient of possible decision-making (and other cognitive) deficits as a function of disability within the RR subtype. Our choice of EDSS cutoff score was motivated such that in the EDSS, a score less than 3 indicates at most, mild disability in 1 functional system or minimal disability in 2 functional systems. Scores 3 and higher include moderate disability levels, which we felt were adequate to distinguish from minimal-mild disability levels. Additionally, our median EDSS score was 3. This is now explicitly spelled out in the paper on pages 7/8. As noted also by reviewer 3, a previous decision-making study by
Kleeberg and others (2004) used an EDSS cut-off of 2.5 instead of 3, i.e., a slightly more restrictive criterion to characterize RR-MS patients with little disability. A single individual in our RR-1 sample would be reclassified as RR-2 using Kleeberg’s cut-off instead of ours. Re-analysing the data with this revised grouping did not affect any of our results using the subgroupings, the new RR-1 group still performed at similarly high levels in the GDT as the healthy controls, and the new RR-2 group still showed reduced GDT performance. For the reasons outlined above, we retained our original sub-grouping.

**Major comment 4.** Imaging data is acquired on different MRI scans. Could the authors please explicitly mention how many different MRI scanners and protocols were used (including the number of patients/HCs per scanner)? Is there a possibility to correct for the type of scanner?

**Reply to 1.M4:** We have now added extensive additional explanation about the scanners, imaging protocols and participants (MS patients and non-MS control patients) at each site. This information is given in detail in two additional files (Additional Files 2 and 4), and referred to in the paper (page 12). With regard to controlling for the scan site/scanner model, we performed and added the following additional analyses to the paper (pages 13/14):

“The use of different imaging facilities, scanner models, and imaging protocols in a retrospective sampling approach, such as the one used here, could conceivably lead to systematic biases. Seven different scan sites were included, with 38 of the 51 scans (24 MS patients, 14 non-MS control patients) from site 1. Sites 2-7 had between 1 and 3 scans per site with a total of 13 scans from sites 2-7 (7 MS patients, 5 non-MS control patients). Furthermore, scans were from six different MR scanners. More than half of the scans (n = 29) were acquired with a Siemens Avanto scanner model (21 MS patients, 8 non-MS control patients), with substantially fewer scans from the other five scanners (details see Additional File 2). The small total number of scans and unequal numbers of scans per site/scanner preclude a direct comparison of the ventricular width measures between each individual site and scanner. In order to allow some comparison and quality control nevertheless, we dichotomised ‘scan site’ and ‘scanner model’, and tested differences in the ventricular width measures between sites (site 1 versus sites 2-7), and between scanners (Siemens Avanto vs. other scanner models). Briefly, none of the comparisons yielded significant differences in any of the six ventricular width measures between scan sites or scanner models (see Additional File 4 for details). Thus, despite the limitations of this retrospective MR sampling method, we did not observe systematic biases in the ventricular measures across scan sites or scanner models.”

Additional File 4 lists the statistical outcome of these comparisons.

We also now include scanner model and scan site as additional covariates in the partial correlations (see page 19/Table 5).

**Major comment 5.** Frontal Horn Width (FHW), Intercaudate Distance (ICD), and Third Ventricle Width are absolute measures, thereby making it difficult to interpret the differences between groups. Is there a possibility to normalize these values for head size?

**Reply to M1.5:** We agree that these measures could be influenced by overall head size. Therefore, we now provide additional ratio measures for all three absolute ventricular width measures, dividing the raw width of the ventricle by the transverse brain width, as we had done only for the ICD measure before. This is identical to Buetzkueven et al. (2008), who also reported raw width along with ratios for all three measures (FHW, ICD, TVW, i.e., FHR, ICR, TVR). This approach is explained on page 13, outcomes are shown in Tables 4-5. Note that in
order to reduce the total number of correlations (which was also mentioned in minor comment 17), we now only use the ratio measures when correlating with disease-parameters and cognitive functions. Furthermore, the mediation model was rerun using ICR and TVR. Conceptually, the results were identical to those previously reported on the absolute measures.

**Major comment 6.** No data is shown concerning lesion volumes of MS patients. I assume the MRI scans are suitable to measure (T1 and/or T2) lesion volumes. Could the authors please add these measures to the results, as it might influence decision-making and information processing speed?

**Reply to M1.6:** We agree that it would be interesting to inspect the impact of lesion volume on decision-making and that the additional calculation of lesion volume is in principle possible with the type of images we had. However, lesion volume often shows a less pronounced relationship with cognitive changes than (estimates of) atrophy, especially in small samples. Moreover, additional possible measures would be similarly interesting (e.g., brain-parenchymal fraction). Considering the timeline for our revision, we therefore decided to focus on optimizing the present measures and added the following sentences to the limitations section of the discussion (page 26).

“Additional MS-related brain changes (e.g., lesion load) or whole-brain atrophy measures (e.g., brain-parenchymal fraction) would have been of interest to explore in addition to the ventricular width measures. Such measures could address questions surrounding the specificity of different types and regional localization (if any) of MS-related brain pathologies to decision-making in the GDT.”

**Major comment 7.** The authors used published norm scores for all individual neuropsychological tests to calculate z-scores. However, the HCs in this sample also underwent neuropsychological testing, and in my opinion it would be more informative to calculate the z-scores relative to the HCs (instead of using norm scores).

**Reply to M1.7:** We apologize for this misunderstanding as we did in fact generate the composite scores (z-scores) based on study healthy controls’ performance on the neuropsychological tests. The previous table intended to show that our healthy controls performed within the normative ranges of each test. We dropped this table to avoid confusion, and instead we now provide raw means and standard deviations on all individual tests within each neuropsychological composite score (Table 2) in addition to the composite scores, based on study healthy controls’ performance (Table 3 in the main text).

**Minor essential revisions**

**General**

**Minor comment 1.** Please change ‘neurocognitive’ in ‘cognitive’.

**Reply to min1.1:** We have implemented this change and no longer refer to ‘neurocognitive’

**Abstract**

**Minor comment 2.** See the comment above concerning the term ‘subcortical atrophy’
Reply to min1.2: We no longer refer to subcortical atrophy.

Minor comment 3. Please mention which groups were compared

Reply to min1.3: Starting with the abstract, we now state explicitly that there were three groups of participants, healthy controls (neuropsychology only), MS-patients (neuropsychology and MRI), and the non-MS control patients (MRI only). Each table now clearly indicates in the title which groups were compared to each other.

Minor comment 4. Please omit the statement concerning fronto-striatal signalling, as this has not been investigated in the present study.

Reply to min1.4: This section has been removed from the discussion.

Introduction

Minor comment 5. See comment above concerning the correlation between ventricular width and (subcortical) brain volumes

Reply to min1.5: See above reply to M1.1

Minor comment 6. Page 4, line 1-3: could the authors please rephrase this sentence? Additionally, the term ‘neurodegenerative’ already implies ‘brain changes’.

Reply to min1.6: We omitted the word ‘brain’ from the phrase.

Minor comment 7. Page 4, line 4: could the authors please clarify the relevance between the gambling task and ‘blunted physiological reactivity’ in the context of the present study?

Reply to min 1.7: This was a feature of Kleeberg et al’s (2004) study that was non-essential to our discussion. Therefore, it was removed to avoid confusion.

Minor comment 8. Page 4, line 7-9: which tracts were measures using diffusion tensor imaging?

Reply to min 1.8: We have now specified which tracts were measured in this study. Page 5 states: “Roca el al. [23] tested decision-making in the IGT in 12 MS patients in conjunction with diffusion tensor imaging (fractional anisotropy and apparent diffusion coefficient) along frontal lobe white matter bundles (orbito-frontal, fronto-lateral, fronto-medial and gyrus cinguli regions).”

Minor comment 9. Please explicitly mention the research question in the final paragraph before stating the hypothesis.

Reply to min 1.9: We have now added our research aims immediately preceding our hypotheses on page 6 “Our goals with the current study were to test whether the GDT, as a non-speeded measure of decision-making under explicit risk, is impaired in MS, and to evaluate whether GDT performance is linked to increasing disability severity (defined as functional impairment according to the EDSS [27]) and to simple estimates of atrophic brain
changes. Using measures of ventricular width to approximate measures of central brain atrophy, we included MS patients with EDSS scores between 0 and 6.5."

**Methods**

**Minor comment 10.** Were patients relapse and corticosteroid treatment free at least four weeks prior to investigation?

**Reply to min 1.10:** We did not collect this information with respect to timeframe, so we cannot definitely say that this was the case. However, none of the patients was currently experiencing a relapse or had corticosteroid treatment. A cautionary note was added to the limitations section, on page 27 stating:

"Finally, although patients were not currently experiencing relapses and were not treated with corticosteroids at the time of study enrolment, historical information about MS patients’ past relapse rates and preceding treatment with corticosteroids was unfortunately not available at the time of ascertainment of participants. This has to be taken as a caveat of the current study."

**Minor comment 11.** Page 6, line 9: how did the authors actually validate ventricular width measures? Could they please elaborate on this matter?

**Reply to min 1.11:** The validation here referred mainly to the intra- and inter-rater analyses we performed on each of the ventricular measures. These analyses demonstrated that the application of these measures were used in a reliable manner, at least within our study. Since these linear measures have been previously used and validated, we felt that it was appropriate to use them in our sample to investigate if these were related to our task of interest, the GDT, without further validation of the method per se. True validation would require some additional external evidence for atrophy, for example, in form of 3-D high-resolution images, which we did not have available. However, some of the ventricular measures were comparable to published ventricular size measures, as is now stated explicitly on page 18: “Of note, the TVW measured 2.3 mm (median) or 3.3 mm (mean) in our MS group, similar to previously reported mean TVWs in MS cohorts ranging between 3.0 mm [11], 3.12 mm [9], 3.58 (RR-MS), and 5.04 (SP-MS) [2]. Similarly, FHW (32.26 mm) and ICD (12.2 mm) were comparable to previously reported figures (FHW: 33.33 mm, ICD: 12.12 [9]) in MS patients."

**Minor comment 12.** Page 9, line 17-19: please move this paragraph to the beginning of the methods section

**Reply to min1.12:** Page 9, line 17-19 in our previous version of the manuscript was not a distinct paragraph such that we were unsure what exactly needed to be moved to the beginning of the methods section. We would appreciate further clarification as to what type of information should be moved.

**Minor comment 13.** Please omit all analyses in which ventricular width is correlated to EDSS in MS subgroups (i.e., circular reasoning; see Major points for revision).
Reply to min1.13: We have now omitted the correlation analyses between EDSS and MS subgroups.

Minor comment 14. Could the authors please mention how subjects performed the GDT (e.g., on a computer)?

Reply to min1.14: We have now spelled out on page 11 that the GDT is computerized.

Minor comments 15/16. Could the authors please reconsider the cognitive domains and neuropsychological tests that were used to assess these domains? For example, the Digit Span forward is not a test for psychomotor processing speed, but for short-term memory. Additionally, please explicitly mention which tests were used for each composite z-score.

See comment above concerning z-scores of neuropsychological tests.

Reply to min1.15 & 16: The generation of composite scores of neuropsychological performance was driven by theoretical considerations, since the healthy control sample was not sufficiently large to accommodate more sophisticated statistical approaches (such as principal component or factor analyses). Such derivation of composite scores does involve some subjective assignment by the researchers and can be biased. For example, there can be shared functional involvement of tests across composite scores although each test is only counted towards one composite score; this was done to eliminate type 1 errors distribute individual tests into each composite score. Page 10/11 acknowledges this ambiguity: “In order to minimize type-I error, we reduced the number of comparisons with each neuropsychological test by summarizing individual test scores into three composite z-scores: processing speed, memory, and executive functions. These z-scores were formed based on the performance (means and standard deviations) of the healthy controls in our study. Acknowledging that neuropsychological measures, especially executive function tests, are not process-pure and are combinable in multiple ways, we have also included a global cognitive function z-score that was derived by averaging the three composite scores, therefore including all of the individual neuropsychological test scores that constituted each composite score.”

We acknowledge that the forward digit span is also a test of short-term memory, but it is frequently referred to as an attention test as well. The processing speed composite score was not limited to tests of psychomotor speed and we avoid the term psychomotor speed now. Tests subsumed under the processing speed composite broadly reflect motor speed, attention, and cognitive processing speed.

Page 10 now explicitly mentions ‘attention’ tests as part of the processing speed composite.

We now provide an additional table (Table 2) to show exactly which individual tests and test scores were combined into which composite score.

Minor comment 17. Could the authors please indicate whether they have corrected their p-values for multiple comparisons? A large amount of correlations analyses are performed.

Reply min1.17: We corrected for multiple comparisons between subgroups by Dunnett- t-tests against healthy controls, or Bonferroni-correction after U-tests. False-Discovery rate
correction by Benjamini and Hochberg (1995) was now conducted to adjust significance levels of the multiple correlations carried out. This is explained on page 15: “To limit the number of correlations, only ventricular ratio (not absolute ventricular width measures) were correlated with disease parameters and cognitive variables. However, even despite these restrictions and the use of cognitive composite scores instead of individual tests, the total number of correlations carried out here could inflate type-1 error. Thus, we report both, the uncorrected correlation results and those adjusted by false-discovery rate accounting for the number of correlations”.

See also Table 5.

Results

Minor comment 18. Could the authors please be consistent with the methods section in the order in which they present the results?

Reply to min1.18: In the methods section, we have moved the GDT behind the Neuropsychological Test section to reflect the same order in which the results are presented.

Minor comment 19. Please move any interpretation of data to the discussion (e.g., page 11, line 7-8).

Reply to min1.19: We have now removed all interpretation from the results section.

Minor comment 20. Could the authors please provide the actual p-values?

Reply to min1.20: We have now provided actual p-values for all statistics in all tables and in the text. For ease of reading, the only reference to p-value thresholds instead of actual p-values that were retained are those indicating post-hoc test results in more than two groups (see Tables 1-3).

Minor comment 21. Page 12: the authors performed additional analyses to see whether education and premorbid IQ could influence GDT score. Perhaps a part of this paragraph could be moved to the methods section (statistical analysis).

Reply to min1.21: The details of the results of the regression analyses were shortened (see also our reply min3.8, reviewer 3) and we felt that this information is now better readable even within the results section.

Discussion

Minor comment 22. Perhaps it would be convenient if the authors would start their discussion with stating their research aim, and subsequently report their main findings.

Reply to min1.22: We have provided now more detail related to our research aim and main findings in the first paragraph of the Discussion section. Page 21 states:

“The goals of this study were to test whether the GDT, as a non-speeded measure of decision-making under explicit risk, is impaired in MS, and to evaluate whether GDT performance is linked to increasing disability and ventricular width. We further tested whether we could observe correlations between GDT performance, executive functions, and processing speed.
Finally, we assessed whether potential associations between GDT performance and ventricular width were mediated by other cognitive functions.

**Minor comment 23.** On page 17 the authors speculate on the possible changes in brain circuits that might underlie decision-making problems. I think this is very interesting and important, and I would like to challenge the authors to provide additional information concerning this matter. Perhaps the authors can include other neurological diseases that have used more advanced MRI techniques to better understand problems in decision-making.

**Reply to min1.23:** In light of a more advanced MRI study on decision-making in MS (Muhlert et al. 2014) that we became aware of during the revision process, we now provide details on that study which are directly relevant to ours and less speculative than the previous text. See pages 5/6 and 25.

**Minor comment 24.** I would like to ask the authors to mention the strengths and limitations of the present study.

**Reply to min1.24:** We have added a strengths and limitations section on page 25-27.

**Conclusion**

**Minor comment 25.** Page 17, line 21: please weaken the statement concerning “…subcortical atrophy involving the thalamus and striatum”. The measures that were used in this study do not reflect atrophy of the thalamus and striatum (see Major points for revision).

**Reply to min 1.25:** See above **Reply to M1.1**, we have dropped these speculative terms.

**Minor comment 26.** Page 17, line 21: please omit the statement that subcortical atrophy may contribute to disturbances in fronto-striatal signalling related to impaired decision-making, as this cannot be concluded based on the present results.

**Reply to min1.26:** See above **Reply to M1.1**, we have dropped all speculation about the localisation of any of these effects based on the coarse nature of the ventricular width measures.

**Minor comment 27.** Page 18, line 1: based on the present results, I do not believe that decision-making can be a reliable biomarker for atrophy. Especially considering that atrophy can be measured relatively easy and very accurate with MRI.

**Reply to min1.27:** This sentence was dropped.

**Figures**

**Minor comment 28.** Please display in figure 1b the relationship between EDSS score and GDT score for all MS patients (and not for subgroups separately, because of circular reasoning).

**Reply to min1.28:** The figure was adjusted accordingly.
Tables

Minor comment 29. Please provide actual p-values (perhaps instead of test statistics)
Reply to min1.29: see Reply to min1.20

Minor comment 30. Please do not highlight ‘trends’. These results are not statistically significant and cannot be used to draw any conclusions.
Reply to min1.30: We have now removed ‘trend’ results.

Minor comment 31. Table 2: it would be informative if the authors could display and test the raw neuropsychological data.
Reply to min1.31: See new Table 2, and replies to M1.7, min1.15/16.
Reviewer 2

Comment 1
The large number of comorbid neurological diseases and the usage of medication in the patient group should be included as a limitation.

Reply to C2.1
We added the following caveat to the limitations section (page 27):

“Furthermore, it should also be noted that the MS patients were currently taking a number of medications with potential influences on cognitive functions. The sample size precluded any further analyses based on current medication status but in larger cohorts, medication status should be considered as a potentially important covariate. However, we would like to point out that despite the current medications for symptom management and mood, self-report questionnaires with regard to psychological problems did not point to major psychiatric comorbidities in our group as a confounding variable for decision-making performance (see Additional File 1).”

Comment 2
The authors should specify the consequences of poor decision making in this patient groups’ everyday life. Are MS patients known to make dysfunctional decisions in their everyday life (such as patients with substance abuse or obsessive compulsive disorder)?

Reply to C2.2
The behavioural/cognitive profile of individuals with MS is not generally characterised by impulsive or compulsive choice behaviour as in the psychiatric conditions named by the reviewer. However, decision-making has been studied in a variety of neurological conditions, including with the GDT (Gleichgerrcht et al., 2010), and extending to other MS studies with the GDT (Farez et al., 2014; Cogo et al., 2014). This approach, generally, has aimed at identifying whether cognitive problems (and underlying neuropathologies associated with MS) also affect the quality of decision-making. Individuals with MS are confronted with complex health-related decisions concerning diagnostic and treatment interventions that require proper handling such that a compromised ability to oversee consequences of one’s decisions could have wide-ranging effects. In addition, although not the case in our study, psychiatric comorbidities and fatigue could further influence a person with MS more so than someone without MS to make disadvantageous choices. Along these lines, we added to the conclusions (page 28) the following sentences:

“Individuals with MS are confronted with complex health-related decisions, such as those concerning diagnostic and treatment interventions that require proper handling. A
compromised ability to oversee consequences of one’s decisions could have wide-ranging effects in MS. Our results may imply that providing additional supports in those types of decision situations (e.g., surrogate input, increased time to make a decision), especially for individuals in more advanced stages of the disease, could be beneficial.”

Comment 3
Is IQ usually related to GDT performance as indicated by the usage as a predictor in the analysis?

Reply to C2.3
GDT performance can be moderated by intelligence, at least when assessed with fluid intelligence tests (Brand et al., 2009 J Clin Exp Neuropsychol; Schiebener et al., 2011 J Clin Exp Neuropsychol; Brand and Schiebener, 2013 J Clin Exp Neuropsychol; Donati et al., 2014, J Clin Exp Neuropsychol). To our knowledge, the impact of crystallized intelligence on performance in the GDT, here assessed with the Shipley Vocabulary subtest, has not been tested. However, it is conceivable that certain types of semantic knowledge (e.g., knowledge about probabilities, numerical abilities), which would be reflected in a measure of crystallized intelligence, could also influence performance in the GDT. The main reason we used the premorbid IQ estimate as a potential moderator of the GDT performance was to address the problematic sampling with regard to the (marginal) IQ difference between healthy controls/RR-1 on the one hand, and RR-2/SP on the other hand. Therefore, finding that RR-2 and SP subgroups still had lowered GDT performance regardless of their marginally lower premorbid IQ levels (see the outcome of the moderated regressions, pages 17) lends some credence to the effects we report.
Response to Reviewer 3’s Comments

BioMed Central MS: 4965631331494339 - Decision-making is impaired in multiple sclerosis: impact of subcortical atrophy and disease disability

Major comment 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 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.

Reply to M3.1: We fully agree, it is possible that our sample size precluded findings any statistically significant differences in age and disease duration between subgroups. We have accordingly changed the wording and discuss the possible characteristics and differences between the subgroups.

Pages 8/9: “Disease duration and age at onset of MS were not statistically different among patient subgroups, although it should be noted that the RR-2 group had on average a 6-7 years shorter disease duration than both the RR-1 and the SP subgroup. This disease profile of the RR-1 subgroup, with relatively little functional impairment over long periods of time, may be suggestive of 'benign MS'. The exact criteria and existence of such a subtype is widely debated, and therefore we retain the more neutral RR-1 label here [36-39]. However, support for the suggestion of a more benign disease course in the RR-1 group, comes from the Multiple Sclerosis Severity Score (MSSS; [40]). The MSSS was statistically identical in the RR-2 and SP subgroups but significantly higher in RR-2/SP than in the RR-1 subgroup. That is, the RR-2 and SP subgroups had a more aggressive disease course than the RR-1 subgroup.”

In addition, our main result of a GDT net-score difference between RR-1 versus RR-2/SP subgroups was unaffected by disease duration or age. To address this question, we conducted an ANCOVA on the GDT net-score, with subgroup as a factor and including age and disease duration as covariates. The overall model remained significant (F[4, 27] = 4.98, p = 0.004), with no effects of age (F[4, 27] = 1.49, p = 0.233), or disease duration (F[4, 27] = 1.5, p = 0.232). RR-2 and SP subgroups still significantly underperformed compared to RR-1 after controlling for disease duration and age (age and MS-duration adjusted means in the GDT: RR-1: 14.26, RR-2: -1.79, SP: 4.87, RR-1 vs. RR-2: p < 0.001, RR-1 vs. SP: p = 0.018, RR-2 vs. SP: p = 0.13). In other words, the between-subgroup differences in the GDT became even more pronounced when adjusting for age and disease duration, especially regarding the RR-2 subgroup. Since age and MS-duration were not statistically different between subgroups, we decided not to include this ANCOVA in the main manuscript.

Major comment 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.

**Reply to M3.2:** We agree the use of covariates should have been explained more explicitly. Partial correlations between ventricular width and EDSS/disease duration were controlled for age, gender, scan site and scanner model. Partial correlations between ventricular width and cognition were controlled for age, gender, delay between scan and test, scan site and scanner model. This is explicitly stated in Table 5.

To clarify, in order to rule out systematic differences in the ventricular measures between scan sites or scanner models, we performed additional analyses (see our reply to **M1.4, reviewer 1**), given in detail in Additional File 4, and referred to in the paper (pages13/14). Even though we found no influence of scan site or scanner model on the ventricular measures in these analyses, both variables were now additionally included as covariates in the partial correlations (Table 5).

Furthermore, to be included into a mediation model, a covariate is required to show bidirectional significant correlations with the predictor (ventricular width measures here) and the outcome (GDT here). Since this was not the case for age, gender, delay between test date and scan date, scan site, or scanner model, we did not include any covariates into the mediation model. This is now spelled out explicitly in the paper, page 20: “Of note, we did not include any additional mediators in this model. Age, gender, scan site, scanner model, or delay between test date and scan date did not show significant bivariate correlations with GDT net-score and with ICR/TVR, and therefore, did not fulfil the necessary criteria for mediation [44]”.

We also added a sentence to the limitations section indicating that a more appropriate sampling would be desirable in order to better match any potential subgroups with regard to demographic and disease parameters (page 26):

“Among the limitations, it should be noted that the results of this study are preliminary based on the small sample size, especially with regard to our analyses in patient subgroups. We intended the division of our RR-MS subgroup to reflect the variability of cognitive and decision-making deficits across the range of disability in MS. Since the correlation between GDT performance and EDSS scores achieves the same outcome, one could alternatively avoid dividing the RR subgroup and treat EDSS (or MSSS) solely parametrically. Again this would be possible only in larger samples. A more targeted inclusion of ‘benign’ MS patients, better-matched in demographic background, would also be helpful to clarify the exact characteristics of MS patients with intact GDT performance.”

**Major comment 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.

**Reply to M3.3:** Kindly refer to our responses to our reply to **1.M4**, to reviewer 1 who expressed similar concerns (see new Additional File 2). The non-MS control patients were not all scanned in the same location, but similar to the MS patients, at different centers. The limitations section now states (page 26): "Evidently, acquiring standardized, high-resolution MR allowing regional and whole-brain volume assessment would be more precise than relying
on retrospective sampling of rather heterogeneous 2-D MRIs. Additional MS-related brain changes (e.g., lesion load) or whole-brain atrophy measures (e.g., brain-parenchymal fraction) would have been of interest to explore in addition to the ventricular width measures. Such measures could address questions surrounding the specificity of different types and regional specificity (if any) of MS-related brain pathologies to decision-making in the GDT.”

**Major comment 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).

**Reply to M3.4:** We fully agree with this comment and added a limitations section to the end of the manuscript (pages 26/27). Also in light of several comments from reviewer 1, we omitted all reference to subcortical atrophy and therefore, concentrate less on the structure-function relationships one should discuss in more detail given better-quality MR data/methods. We would further like to reiterate in this context our reply to min1.23, reviewer 1. In light of a more advanced MRI study on decision-making in MS (Muhlert et al. 2014) that we became aware of during the revision process, we now provide details on that study which are directly relevant to ours and less speculative than the previous text. See pages 5/6 and 25.

**Minor comments:**

**Minor comment 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.

**Reply to min3.1:** kindly refer to our reply to M1.1, reviewer 1, who had similar concerns. No mention of subcortical atrophy is retained in the revised manuscript.

**Minor comment 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.

**Reply to min3.2:** We apologize for the confusion with the included groups. As also reflected in our reply to min1.3, reviewer 1: Starting with the abstract, we now state explicitly that there were three groups of participants, healthy controls (neuropsychology only), MS-patients (neuropsychology and MRI), and the non-MS control patients (MRI only). Each table now clearly indicates in the title which groups were compared to each other. We further added to our explanation for the use of the non-MS neurological control group as follows (page 12: “The main purpose to include ventricular measures from the non-MS control patients was to compare ventricular widths of our MS patient group to another patient population with similar quality retrospectively-sampled clinical MR scans. In past studies, linear ventricular width measures, especially TVW and ICD/ICR have differentiated well between MS patients and
healthy controls [2, 12, 41, 42]. Thus, our approach to include a non-MS patient group as a reference can be considered conservative with regard to ascertaining ventricular widening in the MS group.”

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

Reply to min3.3: The MR scans we had available were 2-D clinical scans from a variety of different sites and scanners, acquired with different scan parameters (see Additional Table 2). Thus, it was not reliably possible to extract any more elaborate measures involving 3-D volumes. We do acknowledge that even with the type of images we had, one could have examined additional parameters, as described also in our reply to M1.6, reviewer 1: Considering the timeline for our revision, we therefore decided to focus on optimizing the present measures and added the following sentences to the limitations section of the discussion (page 26).

“Additional MS-related brain changes (e.g., lesion load) or whole-brain atrophy measures (e.g., brain-parenchymal fraction) would have been of interest to explore in addition to the ventricular width measures. Such measures could address questions surrounding the specificity of different types and regional localization (if any) of MS-related brain pathologies to decision-making in the GDT.”

Minor comment 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.

Reply to min3.4: We apologize for this oversight and now state in the abstract:

“Decision making tasks in which the probabilities of an outcome are known, and choice-strategies can be calculated, measure ‘decision-making under explicit risk’.

Early in the paper (page 4), decision making under explicit risk is also defined: “The Game-of-Dice Task (GDT [19]) assesses decision-making under explicit risk and emphasizes the cognitive aspects of decision-making by explicitly laying out winning/ losing probabilities associated with each choice. In contrast to decision-making under explicit risk, in decision-making under ambiguity the odds for each choice option are not explicitly available. Instead, the goal is to implicitly learn choice-outcome contingencies solely by using feedback and through the process of trial-and-error. The most prominent task assessing decision-making under ambiguity is the Iowa Gambling Task (IGT [20]).”

Minor comment 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).

**Reply to min3.5:** We fully agree with this comment, which also matches with reviewer 1, Major comment 1. We kindly refer to our reply M1.1. In brief, we added more targeted literature into the introduction to outline these claims (pages 3-4).

**Minor comment 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?

**Reply to min3.6:** We now added the exact numbers of RR-1, RR-2, and SP patients taking the listed medications (see page 8-9). As the numbers were already small, no additional analyses were carried out to account for potential influences or differences in medications between subgroups. This is noted as a limitation on page 27.

**Minor comment 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.

**Reply to min3.7:** (see also our reply to M1.3, reviewer 1). Pages 7-8 state: “A similar approach was previously used by Kleeberg et al [21]. The rationale for dividing our RR-MS subgroup was to investigate whether there was a gradient of possible decision-making (and other cognitive) deficits as a function of disability within the group of RR-MS patients. Our choice of EDSS cut-off score was motivated such that in the EDSS, a score less than 3 indicates at most, mild disability in one functional system or minimal disability in two functional systems. Scores 3 and higher include moderate disability levels. In addition, the median EDSS in our MS-sample was 3. Therefore, a cut-off of 3 here distinguished RR-MS patients who were experiencing mild-minimal disability from those experiencing moderate or greater disability.”

Kleeberg and others (2004) used an EDSS cut-off of 2.5 instead of 3, i.e., a slightly more restrictive criterion to characterize RR-MS patients with little disability. A single individual in our RR-1 sample would be reclassified as RR-2 using Kleeberg’s cut-off instead of ours. Re-analysing our own data with this revised grouping did not affect any of our results, the new RR-1 group still performed at similarly high levels in the GDT as the healthy controls, and the new RR-2 group still showed reduced GDT performance. For the reasons outlined above, we therefore retained our original sub-grouping.

**Minor comment 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.

**Reply min3.8:** We thank the reviewer for this observation and omitted these details.
Minor comment 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).

Reply min3.9: Some of the ventricular width measures were correlated with the GDT and with the processing speed composite score, but not with the other composite scores or with strategy shifts. Table 5 provides an overview on these findings. We do report on additional correlational analyses for ventricular measures and single tests within the composite scores, as the only reason for us to split the executive functions composite score was to test our a priori assumption of a correlation between the GDT net-score and executive functions. Nevertheless in the revision process we have examined the proposed correlations (i.e., individual executive function tests and ventricular widths). None of these were significant, or trend effects, and are therefore not elaborated on in the manuscript.

Minor comment 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?

Reply: Speculations for the low performance of the RR-2 subgroup has now been expanded on in the discussion, page 22:

“Compared to the RR-1 and the SP patient subgroups, the RR-2 subgroup was most impaired in the GDT and across all of the composite scores. What could be distinguishing characteristics of the RR-2 patients compared to the RR-1 and SP subgroups? As alluded to in the Methods section, some of the RR-1 patients may be considered to have had a ‘benign’ subtype of MS (i.e., long disease duration with relatively low EDSS scores). The conversion from a RR to a SP subtype can be marked by an increase in cognitive impairment [51, 52] accompanied with accelerated degeneration of cerebral grey matter [53]. Assuming that the SP-MS patients were already in a stable state of disease progression (i.e., without experiencing remissions), one could speculate that the apparent increase in cognitive and decision making impairments in the RR-2 group may reflect that these particular patients were currently in a less stable disease state than in both RR-1 and SP patients. However, in light of the very small group size, these possible explanations remain speculative. “

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

Reply min3.11: We would like to kindly refer to our reply to min1.17, reviewer 1. In short, we now adjusted the significance level by false discovery rate (Benjamini and Hchberg, 1995) to account for the large number of correlations.