PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Can microstructural MRI detect subclinical tissue injury in subjects with asymptomatic cervical spinal cord compression? A prospective cohort study
AUTHORS	Martin, Allan; De Leener, Benjamin; Cohen-Adad, Julien; Cadotte, David; Nouri, Aria; Wilson, Jefferson; Tetreault, Lindsay; Crawley, Adrian; Mikulis, David; Ginsberg, Howard; Fehlings, Michael

VERSION 1 – REVIEW

REVIEWER	Gun Woo Lee Yeungnam University Hospital and Yeungnam University Medical Center, Republic of Korea
REVIEW RETURNED	25-Oct-2017

GENERAL COMMENTS	The subject of the current study is of great interest for spine surgeons that is asymptomatic cervical cord compression and its progression of real degenerative cervical myelopathy. However, I have some of concerned issues in the study. 1. the authors commented that widely held paradigm, mild SC indentation and flattening represent "normal degenerative change". - I think this comment is not perfectual. Many spine physicians said that the change is also related to impending or early-sign of DCM, so they were followed up regularly in outpatient clinic, with taking their sympyoms and checking related signs. Meanwhile, I also agree your comment in part, but that is not improper, I think. So, I suggest the sentence should be revised properly. 2. In real clinic, ASCC is regarded as a important pre-clinical state to progree DCM, so ASCC should be diagnosed as one of disease criteria. In this point, the method for critical ASCC state presented in the current study is paramount for surgeons. So, I recommend that the authors should highlight the importance of ASCC and its possibility aggravating to real pathology, DCM, requiring surgical treatment. 3. radiographic evaluation and parameters should be associated with rater's experience and skill. Particularly in DTI and MRI images, the reliability and agreement is always of great issue. However, in the study, the authors didnot perform agreement and reliability test for the outcomes. So, I recommend performing the reliability and agreement test among inter- and intra-observers for the data outcome. Thanks.

REVIEWER	Jörg Krebs Swiss Paraplegic Centre Clinical Trial Unit Guido A. Zäch Strasse 1 CH-6207 Nottwil Switzerland
REVIEW RETURNED	26-Oct-2017

GENERAL COMMENTS	The authors assessed the presence of asymptomatic cervical spinal cord compression based on the consensus rating of anatomical MRI images in a convenience sample of 40 neurologically intact individuals. The diagnosis of compression based on consensus rating was compared with the diagnosis of compression based on an automated analysis of MRI images. Thirdly, the presence of spinal cord tissue damage based on novel MRI parameters and the association between tissue damage and asymptomatic cord compression were investigated. Finally, the occurrence of cervical myelopathy in the investigated individuals was assessed after 1-2 years. General Remarks There is still much debate and uncertainty regarding the early diagnosis of clinically relevant cervical spinal cord compression. Novel MRI techniques seem to be promising to provide reliable data for diagnosing relevant cord compression. Based on the investigated sample and the available/provided data, the authors should be very cautious with their statements regarding prevalence of asymptomatic spinal cord compression, occurrence of cord damage and association of damage with compression and the occurrence of myelopathy. The authors have investigated a convenience sample of 40 neurologically intact individuals. The reported data regarding cord tissue damage is rather sparse and does not allow to evaluate their validity. The follow-up time is very short and only individuals reporting neurological symptoms during a telephone interview were re-assessed clinically (information bias). The study mainly provides evidence for the validity of the presented automated analysis of MRI images for the diagnosis of cord compression. I commend the authors for making their data available after publication. Title - needs to be revised to reflect the major findings of the study (see general remarks above) Abstract Interventions MRI assessments are not considered interventions. Provide information under Outcome Measures. There is no information regarding consensus diagnosis of compression and tissue damage measurements. Results - diagnostic accuracy > 97%: AUC ≠ accuracy

	- evidence of tissue injury ($p<0.05$) / stronger difference ($p=0.002$): unclear what p-values refer to - "strong difference": consider rewording Conclusions Be cautious with your statements regarding myelopathy, prevalence of ASCC and the risk of symptomatic myelopathy (see general remarks above). Needs to be revised. Registration - not applicable: cohort studies can and should be registered, too! Strengths and Limitations Strengths - combination of several measures to increase statistical power: this does not increase statistical power directly. Increase in sample size, effect size or decrease in sampling error does. Limitations The authors acknowledge the limitations of the study, however the limitations are not considered in their conclusions. Introduction - re-order the sequence of hypotheses according to abstract - the chosen study design is no suitable to investigate risk factors for the development of myelopathy Methods - follow-up by telephone only and clinical evaluation only when neurological signs were reported: discuss information bias under limitations in the discussion section - what about individuals with metal implants ? - page 5, line 51/51: second ROC of revised diagnostic accuracy is inappropriate: there is an overlap of the criteria used for diagnosis and the same data set was analyzed twice: remove data from manuscript - page 5, line 52/53: Normative values for shape parameters were calculated in uncompressed subjects: provide more details - page 6, assessment of tissue injury: there are several parameters and locations, consensus rating for the location of the MCL and different procedures for normalization which makes it quite challenging for the reader to evaluate the validity: the authors need to elaborate - Statistical Analysis, lines 29-31: A binomial test compared the pattern of differences in ASCC with that in DCM: explain "the pattern of differences"? Results - page7, line 40/41: "appeared" to be invariant: was ANOVA used as described in the Methods section? - figure 2 is misleading! the age of the investigated individuals ranged from 28-79 and the authors report data for the first decade. Remove the figure. - page 9, line 7/8: eight out of ten metrics showed the same direction as previously seen in DCM ($p=0.11$): what does p-value refer to? this is the results and not discussion section, "8/10 metrics": is this relevant? - consider also reporting sensitivity and specificity for the thresholds which may be more meaningful than AUC values for the majority of the readers - table 4: denotes trend ($p<0.1$): significance was set at 0.05: i.e. significance was not reached

	- page 10, line 10/11: does logistic regression yield discrimination? - page 10: results of tissue injury are too sparse; please provide more results - prediction of myelopathy: study design is not appropriate for this analysis, please remove from manuscript Discussion Summary - study design not suitable to make statement concerning prevalence of cord compression - page 11: analysis may suffer from overfitting and must be interpreted with caution: I commend the authors for acknowledging this, but they need to apply this caution in their conclusions Clinical Implications - page 13: revise statements regarding prevalence of ASCC and risk of progression to myelopathy (sample size, short follow-up, selection bias, information bias) - page 13, lines 15-17: do you recommend MRI screening ? Limitations - discuss selection and information bias - authors should consider limitations in their conclusions and the clinical implications section - discuss validity of consensus rating Conclusions - Be cautious with your statements regarding myelopathy, prevalence of ASCC and the risk of symptomatic myelopathy (see general remarks above). Needs to be revised. Miscellaneous - The use of so many abbreviations makes the reading very cumbersome. The number of abbreviations needs to be reduced. - Be aware of colloquialism: e.g. page 6, line 25: “numerical variables used two-tailed Welch’s Test”: variable cannot use a test. - author contributions: all authors claim a contribution to statistical analysis and nine a contribution to the study design. I appreciate the technical complexity of the present study, but contributions to statistical analysis or study design need to be substantial in order to justify authorship according to ICMJE recommendations. Please confirm that all authors have contributed substantially in order to dispel any suspicion regarding honorary authorship.

REVIEWER	Kush Kapur Harvard Medical School USA
REVIEW RETURNED	15-Nov-2017

GENERAL COMMENTS	Comments to the Author: In this study, the authors have demonstrated diagnosis capability of spinal cord shape analysis for detecting asymptomatic spinal cord compression (ASCC). In addition, they have explored detection capability of spinal cord injury similar to the one observed in degenerative cervical myelopathy using multi-parametric quantitative MRI biomarker. The authors have clearly defined their study design and the statistical methods employed by them are mostly adequate.

	However, I believe the authors should address a few "minor" comments in order to improve the overall presentation of their work: a. Page 6: In order to normalize the metrics and to build the logistic regression model for detecting tissue injury, the authors have used backward stepwise selection approach. Did they also explore forward stepwise selection? Were their final models any different? b. The authors haven't performed multiple comparisons in reporting the significance of their findings? They should report adjusted p-values along with the modified text for the significance criterion, i.e. "Significant was set at adjusted p-value<0.05". Table 3 should also include adjusted p-values. It is okay to not adjust for multiple comparisons in case of qMRI measures due to extremely small sample size. But, this should be noted in the discussion and limitation sections c. Plots along with 95% confidence intervals should be included for the reported AUCs on pages 7-8 and Table 2. d. Page 11 discussion "Our results highlight… results standard error of effect estimation of approximately 1/sqrt(n)". This statement should be deleted. The authors have not derived any statistical results to jump to this conclusion and it is not relevant to their current work. It is also unclear why they have used "Post hoc analysis" to describe their secondary findings. "Post hoc analysis" has a very specific meaning in statistical literature. This statement should be modified with proper choice of wording. e. Limitation Section: In case of binary outcome, the sample size of 40 (20 in each group) is not large enough (especially for building logistic regression models). Please modify the starting sentence – "The sample size was large enough….." f. Figure 3 is little unclear – these plots can be replaced with box plots. While they have employed non-parametric procedures to perform the comparisons, they have superimposed the t distributions on the histograms. Is there any specific clinical reason for inclusion of such plots? Similarly, it is not clear why the t9 and t10 scores are included in the qMRI composite score result section (and also mentioned in statistical section).

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Gun Woo Lee
Institution and Country: Yeungnam University Hospital and Yeungnam University Medical Center, Republic of Korea
Competing Interests: none

The subject of the current study is of great interest for spine surgeons that is asymptomatic cervical cord compression and its progression of real degenerative cervical myelopathy.

However, I have some of concerned issues in the study.

1. the authors commented that widely held paradigm, mild SC indentation and flattening represent "normal degenerative change".
- I think this comment is not perfectual. Many spine physicians said that the change is also related to impending or early-sign of DCM, so they were followed up regularly in outpatient clinic, with taking their sympyoms and checking related signs. Meanwhile, I also agree your comment in part, but that is not improper, I think.

So, I suggest the sentence should be revised properly.

Response: Thank you for your comments, we have revised this sentence.

2. In real clinic, ASCC is regarded as a important pre-clinical state to progree DCM, so ASCC should be diagnosed as one of disease criteria. In this point, the method for critical ASCC state presented in the current study is paramount for surgeons. So, I recommend that the authors should highlight the importance of ASCC and its possibility aggravating to real pathology, DCM, requiring surgical treatment.

Response: We agree with your perspective, but the literature has not characterized the clinical significance of asymptomatic compression, and in our experience we have encountered many clinicians (including radiologists and surgeons) that do not view mild cord compression as a significant entity (for example, our subject that developed myelopathy at follow-up was dismissed by her family physician after an MRI was reported as “normal degenerative changes” in spite of the presence of mild cord compression). The results of this study suggest that ASCC is an important entity and we have arrived at similar conclusions as those that you are stating. However, we cannot make strong assertions given the modest sample size and inherent limitations of the MRI findings (including lack of ground truth data regarding tissue injury).

3. radiographic evaluation and parameters should be associated with rater's experience and skill. Particularly in DTI and MRI images, the reliability and agreement is always of great issue. However, in the study, the authors didnot perform agreement and reliability test for the outcomes. So, I recommend performing the reliability and agreement test among inter- and intra-observers for the data outcome.

Response: We did not report reliability data for 2 reasons: 1) We previously reported reliability data for our methods (Reference 14: Martin et al. AJNR 2017a); and 2) The quantitative data were automatically (objectively) extracted using template-based ROI analysis using the Spinal Cord Toolbox, with the only manual step being a visual confirmation/editing of the segmentation mask. In our opinion, adding reliability data to this manuscript would make it excessively long without adding substantial value (given our previous report).

Thanks.

Reviewer: 2
Reviewer Name: Jörg Krebs
Institution and Country: Swiss Paraplegic Centre Clinical Trial Unit, Switzerland
Competing Interests: None declared

The authors assessed the presence of asymptomatic cervical spinal cord compression based on the consensus rating of anatomical MRI images in a convenience sample of 40 neurologically intact individuals. The diagnosis of compression based on consensus rating was compared with the diagnosis of compression based on an automated analysis of MRI images. Thirdly, the presence of spinal cord tissue damage based on novel MRI parameters and the association between tissue damage and asymptomatic cord compression were investigated. Finally, the occurrence of cervical myelopathy in the investigated individuals was assessed after 1-2 years.

General Remarks
There is still much debate and uncertainty regarding the early diagnosis of clinically relevant cervical spinal cord compression. Novel MRI techniques seem to be promising to provide reliable data for diagnosing relevant cord compression.

Based on the investigated sample and the available/provided data, the authors should be very cautious with their statements regarding prevalence of asymptomatic spinal cord compression, occurrence of cord damage and association of damage with compression and the occurrence of myelopathy. The authors have investigated a convenience sample of 40 neurologically intact individuals. The reported data regarding cord tissue damage is rather sparse and does not allow to evaluate their validity. The follow-up time is very short and only individuals reporting neurological symptoms during a telephone interview were re-assessed clinically (information bias). The study mainly provides evidence for the validity of the presented automated analysis of MRI images for the diagnosis of cord compression.
I commend the authors for making their data available after publication.
Title
- needs to be revised to reflect the major findings of the study (see general remarks above)

Thank you for your comments. We have revised the title.

2
Abstract
Interventions
MRI assessments are not considered interventions. Provide information under Outcome Measures. There is no information regarding consensus diagnosis of compression and tissue damage measurements.

Response: We have revised the abstract to state that there were no interventions and to specify the clinical and MRI assessments as outcome measures. We have also updated the abstract to state that automated diagnosis was compared with consensus diagnosis.

Results
- diagnostic accuracy > 97%: AUC ≠ accuracy

Response: Thank you, we have rephrased the findings.

- evidence of tissue injury ($p<0.05$) / stronger difference ($p=0.002$): unclear what pvalues refer to

Response: These results have been rephrased to explain the comparison.

- “strong difference”: consider rewording

Response: The results have been rephrased to avoid this biased term.

Conclusions
Be cautious with your statements regarding myelopathy, prevalence of ASCC and the risk of symptomatic myelopathy (see general remarks above). Needs to be revised.

Thank you for the comment. We agree and we have revised the conclusions to be more cautious.

Registration
− not applicable: cohort studies can and should be registered, too!

Response: We agree and have revised this statement. We did not seek clinical trial registration for this small pilot study, but will do so for larger confirmatory studies.

Strengths and Limitations

Strengths
- combination of several measures to increase statistical power: this does not increase statistical power directly. Increase in sample size, effect size or decrease in sampling error does.

Response: We have revised the methods and discussion to better explain our approach. Based on the assumption that the 10 MRI metrics are covariant, their combination into a composite score effectively increases the number of measurements and decreases sampling error, thus increasing the statistical power.

Limitations
The authors acknowledge the limitations of the study, however the limitations are not considered in their conclusions.

Response: We have revised the conclusions to be more cautious (as above). Due to space constraints we cannot explicitly repeat the limitations of the study in the conclusions section.

Introduction
- re-order the sequence of hypotheses according to abstract

Response: We agree and have revised the manuscript.

- the chosen study design is no suitable to investigate risk factors for the development of myelopathy

Response: We agree that the study is not powered or designed to detect risk factors and have removed these data from the manuscript.

Methods
- follow-up by telephone only and clinical evaluation only when neurological signs were reported: discuss information bias under limitations in the discussion section

Response: We agree and have added mention of this in the limitations. However, we feel that the diagnosis of clinical myelopathy requires both signs and symptoms, and therefore there is sufficient information to rule it out if the subject denies symptoms over the phone.

- what about individuals with metal implants ?

Response: None of the subjects had previous spinal surgery. Other metallic implants (e.g. dental) were not grounds for exclusion but could bias the results (these effects are not well defined for the techniques employed). We have added mention to the limitations.

3
- page 5, line 51/51: second ROC of revised diagnostic accuracy is inappropriate: there is an overlap of the criteria used for diagnosis and the same data set was analyzed twice: remove data from manuscript

Response: We agree and have removed these data.

- page 5, line 52/53: Normative values for shape parameters were calculated in uncompressed subjects: provide more details

Response: We calculated mean and standard deviation of each shape parameter at each rostro-caudal level to characterize normative values, as a basis for diagnosis of compression. We have elaborated on this in the methods section.

- page 6, assessment of tissue injury: there are several parameters and locations, consensus rating for the location of the MCL and different procedures for normalization which makes it quite challenging for the reader to evaluate the validity: the authors need to elaborate

Response: The 2 raters subjectively assessed each rostrocaudal level for flattening, indentation, and torsion, and also identified the MCL. We have clarified this in the methods.

- Statistical Analysis, lines 29-31: A binomial test compared the pattern of differences in ASCC with that in DCM: explain “the pattern of differences”?

Response: We have clarified this in the methods. We wanted to assess if the 10 MRI metrics deviated in the same direction (increase or decrease) in ASCC compared to uncompressed subjects as previously seen in DCM subjects compared to healthy subjects. The binomial test assessed the null hypothesis (that the pattern of increases/decreases was equivalent to 10 coin tosses).

Results
- page7, line 40/41: “appeared” to be invariant: was ANOVA used as described in the Methods section?

Response: Yes, ANOVA was used and found no differences by rostro-caudal level for solidity and relative rotation. We have rephrased this in the manuscript.

- figure 2 is misleading! the age of the investigated individuals ranged from 28-79 and the authors report data for the first decade. Remove the figure.

The age of the entire cohort ranged from 19 to 79. It seems that you derived the age range from the table of compressed subjects (Supplemental Table 1), but this did not include the uncompressed subjects. We feel that the figure is useful to roughly demonstrate the age relationship of ASCC, although we agree with the comments that this study is not powered to accurately characterize prevalence or the relationship with age. We have modified the text to avoid claims that we have characterized the population prevalence, and we feel that the limitations have now been adequately explained.

- page 9, line 7/8: eight out of ten metrics showed the same direction as previously seen in DCM (p=0.11): what does p-value refer to? this is the results and not discussion section, “8/10 metrics”: is this relevant?

Response: We have modified the methods to better explain this test (please see our response to a related comment below). We devised the binomial test (a priori) as a means of showing subtle changes at the group level. We feel that it is necessary to report the non-significant results of this test, to avoid reporting bias. Furthermore, it was interesting that CSA varied in the opposite direction, which suggests that having a large spinal cord is a predisposing factor for development of ASCC (less space in the canal).

- consider also reporting sensitivity and specificity for the thresholds which may be more meaningful than AUC values for the majority of the readers

Response: We agree and have added sensitivity and specificity for the optimal thresholds.

- table 4: denotes trend ($p<0.1$): significance was set at 0.05: i.e. significance was not reached

Response: The figure legend specifies the differences between trend ($p<0.1$) and significant ($p<0.05$) findings. Since this is an exploratory study, we feel that it may be of interest to the reader to highlight trends, as a point of interest for further investigation.

- page 10, line 10/11: does logistic regression yield discrimination?

Response: Yes, logistic regression yields an ROC curve that can be used to calculate discrimination, as described in the following text:
http://thestatsgeek.com/2014/05/05/area-under-the-roc-curve-assessing-discrimination-in-logistic-regression/

- page 10: results of tissue injury are too sparse; please provide more results

Response: We are not sure what you are specifically referring to by “results of tissue injury are too sparse”– we would be happy to address this if you would kindly clarify your comment.

- prediction of myelopathy: study design is not appropriate for this analysis, please remove from manuscript

Response: We agree that the study is not well powered/designed to detect risk factors and have removed these data from the manuscript.

Discussion
Summary
- study design not suitable to make statement concerning prevalence of cord compression

We agree and have rephrased all statements to avoid the term prevalence and replace it with frequency (in this cohort).

- page 11: analysis may suffer from overfitting and must be interpreted with caution: I commend the authors for acknowledging this, but they need to apply this caution in their conclusions

Response: We agree and have rephrased our conclusions to be appropriately cautious.

Clinical Implications
- page 13: revise statements regarding prevalence of ASCC and risk of progression to myelopathy (sample size, short follow-up, selection bias, information bias)

Response: We agree and have revised the text.

- page 13, lines 15-17: do you recommend MRI screening ?

Response: We have rephrased this statement to suggest that further studies are needed to determine the role of MRI screening.

Limitations
- discuss selection and information bias

Response: We agree and have revised the text.

- authors should consider limitations in their conclusions and the clinical implications section

Response: We agree and have revised the text.

- discuss validity of consensus rating

Response: We agree and have listed this as a limitation.

Conclusions
− Be cautious with your statements regarding myelopathy, prevalence of ASCC and the risk of symptomatic myelopathy (see general remarks above). Needs to be revised.

Response: We agree and have revised the text.

Miscellaneous
− The use of so many abbreviations makes the reading very cumbersome. The number of abbreviations needs to be reduced.

Response: We agree and have attempted to reduce the number of abbreviations, although the nature of the work makes this a challenge.

− Be aware of colloquialism: e.g. page 6, line 25: “numerical variables used twotailed Welch’s Test”: variable cannot use a test.

Response: We agree and have attempted to refine the text.

− author contributions: all authors claim a contribution to statistical analysis and nine a contribution to the study design. I appreciate the technical complexity of the present study, but contributions to statistical analysis or study design need to be substantial in order to justify authorship according to ICMJE recommendations. Please confirm that all authors have contributed substantially in order to dispel any suspicion regarding honorary authorship.

Response: We have reviewed the author contributions and feel that all authors have contributed substantially and met the ICMJE criteria.

Reviewer: 3
Reviewer Name: Kush Kapur
Institution and Country: Harvard Medical School, USA
Competing Interests: None

In this study, the authors have demonstrated diagnosis capability of spinal cord shape analysis for detecting asymptomatic spinal cord compression (ASCC). In addition, they have explored detection capability of spinal cord injury similar to the one observed in degenerative cervical myelopathy using multi-parametric quantitative MRI biomarker. The authors have clearly defined their study design and the statistical methods employed by them are mostly adequate. However, I believe the authors should address a few "minor" comments in order to improve the overall presentation of their work:

a. Page 6: In order to normalize the metrics and to build the logistic regression model for detecting tissue injury, the authors have used backward stepwise selection approach. Did they also explore forward stepwise selection? Were their final models any different?

Response: Thank you for your comments. We only included the logistic regression modelling to demonstrate a possible approach for future studies or clinical use, since we were aware that the current study is underpowered to develop an accurate prediction model. We selected backward stepwise elimination a priori, but in informal experiments we found similar results (selecting slightly different independent variables with a similar model fit) with forward selection.

b. The authors haven't performed multiple comparisons in reporting the significance of their findings? They should report adjusted p-values along with the modified text for the significance criterion, i.e. "Significant was set at adjusted p-value<0.05". Table 3 should also include adjusted p-values. It is okay to not adjust for multiple comparisons in case of qMRI measures due to extremely small sample size. But, this should be noted in the discussion and limitation sections

Response: We agree and we have addressed this in the methods and limitations including the statement "Statistical correction for multiple comparisons was not performed due to the exploratory nature of this study, but should be incorporated into the design of future confirmatory studies." We did not select an a priori primary outcome measure (since this work was exploratory), but having seen the data it seems that the revised composite score is a reasonably good single measure that could be used for future studies to avoid the need for correction.

c. Plots along with 95% confidence intervals should be included for the reported AUCs on pages 7-8 and Table 2.

Response: We have prepared these plots and included them as an additional figure.

d. Page 11 discussion "Our results highlight… results standard error of effect estimation of approximately 1/sqrt(n)". This statement should be deleted. The authors have not derived any statistical results to jump to this conclusion and it is not relevant to their current work. It is also unclear why they have used "Post hoc analysis" to describe their secondary findings. "Post hoc analysis" has a very specific meaning in statistical literature. This statement should be modified with proper choice of wording.

Response: We agree that this section requires clarification, and we have rephrased both statements, including removing the term "post hoc". However, we think it is important to highlight the utility of composite scores in the context of multiparametric MRI as this is not a common approach in this field. Thus we have revised this section to explain the rationale for calculating a composite score and discuss its use.

e. Limitation Section: In case of binary outcome, the sample size of 40 (20 in each group) is not large enough (especially for building logistic regression models). Please modify the starting sentence – "The sample size was large enough….."

Response: We agree and we have modified the text to state that the small sample size limited our ability to estimate logistic regression model parameters, and larger confirmatory studies are needed. The purpose of performing logistic regression was primarily to identify which MRI measures are the strongest measures, which will be useful for future studies.

f. Figure 3 is little unclear – these plots can be replaced with box plots. While they have employed non-parametric procedures to perform the comparisons, they have superimposed the t distributions on the histograms. Is there any specific clinical reason for inclusion of such plots? Similarly, it is not clear why the t9 and t10 scores are included in the qMRI composite score result section (and also mentioned in statistical section).

Response: We have clarified our approach in the methods and discussion, which we feel is warranted. We assume that the quantitative MRI metrics are normally distributed in uncompressed subjects, but that they are non-normally distributed in subjects with ASCC (based on the hypothesis that subclinical tissue injury occurs to a variable degree between subjects). Thus group comparisons must use non-parametric procedures (Wilcoxon tests) to be valid, but individual ASCC subjects are compared against uncompressed subjects using z or t scores. We felt that displaying the data in histograms was a useful summary of results of individual subjects (which is of critical importance for future clinical use rather than just showing group differences). As for the composite scores, they are an average of 9 or 10 z scores (which are assumed to follow a normal distribution under the null hypothesis, and thus the composite is assumed to follow a t distribution). Of course, this approach is somewhat complex and subject to several assumptions, so we have attempted to explain this appropriately.

VERSION 2 – REVIEW

REVIEWER	Jörg Krebs Clinical Trial Unit Swiss Paraplegic Centre Nottwil Switzerland
REVIEW RETURNED	15-Jan-2018

GENERAL COMMENTS	The authors have addressed my concerns, questions and remarks adequately.

REVIEWER	Kush Kapur Harvard Medical School, USA
REVIEW RETURNED	16-Jan-2018

GENERAL COMMENTS	The authors have addressed all of my statistical concerns.