Author’s response to reviews

Title: Spondyloarthritis-related and degenerative MRI changes in the axial skeleton - an inter- and intra-observer agreement study

Authors:

Bodil Arnbak (Bodil.Arnbak@slb.regionsyddanmark.dk)
Tue S Jensen (Tue.Secher.Jensen@rsyd.dk)
Claus Manniche (Claus.Manniche@rsyd.dk)
Anna Zejden (annazejd@rm.dk)
Niels Egund (nielegun@rm.dk)
Anne Grethe Jurik (annejuri@rm.dk)

Version: 3 Date: 17 April 2013

Author’s response to reviews: see over
Dear Editors

Please find enclosed our manuscript, “Spondyloarthritis-related and degenerative MRI changes in the axial skeleton - an inter- and intra-observer agreement study” by Arnbak et al., which we would like to submit for publication as a research article in Musculoskeletal disorders, transferred from Arthritis Research and Therapy. We have revised the manuscript according to the reviewer’s report from Arthritis Research and Therapy, please see comments below.

To our knowledge, this is the first agreement study that evaluates a protocol including both spondyloarthritis related and degenerative MRI findings in the axial skeleton. As the greatest challenge in using MRI in the early diagnosis of spondyloarthritis lies within the differentiation between the more frequent degenerative findings and the rare spondyloarthritis related findings, we believe the results from our study also would appeal to the readers of Musculoskeletal Disorders.

We confirm that this manuscript has not been published elsewhere and is not under consideration by another journal.

All authors have approved the manuscript and agree with its submission to Musculoskeletal Disorders. All authors declare they have no competing interests.

We look forward to hearing from you at your earliest convenience.

Best wishes,

Bodil Arnbak
Comments to **Reviewer’s report 1**

**Title:** Spondyloarthritis-related and degenerative MRI changes in the axial skeleton - an inter- and intra-observer agreement study  
**Version:** 2  
**Reviewer number:** 1  
**Referee’s comments to the author(s)**  
Only this section of the report will be returned to the authors. Do not comment in this section on the interest/importance level of the manuscript, or whether or not the manuscript should be accepted.

This is a retrospective study evaluating the inter- and intra-observer reliability for the evaluation of whole body and sacroiliac joint MRI examinations for the inflammatory and degenerative changes of the spine.  
The study is well designed and well written. I have some minor comments.  
**Abstract-OK**  
**Introduction – OK**  
**Materials and Methods:**  
- It is not completely clear from the introduction the need for evaluation of both degenerative and inflammatory changes simultaneously. A more detailed explanation is needed either here or in the introduction.

We agree that additional clarification is needed regarding the inclusion of degenerative findings in the evaluation protocol and we have added a more detailed explanation to the “Background”, page 4, paragraph 1:

> "However, there are still several uncertainties regarding the utility of MRI in the diagnosis of SpA [3], especially in the early stages when the clinical signs of SpA can be difficult to distinguish from non-specific low back pain (LBP) and the MRI signs of SpA can be difficult to distinguish from the much more common findings of degeneration. Signal changes related to degeneration such as Modic changes are an important pitfall in the assessment of SpA [4] and some studies have shown substantial variation in the extent of MRI lesions in the SIJ previously considered to be specific for SpA [5]. Therefore, studies encompassing patients reflecting the target population and using MRI protocol including both SpA-related findings and degenerative changes are needed to validate the utility of this new imaging modality for the diagnosis of SpA."

**Table 1:**  
- The TI parameter is missing for the STIR and SPIR sequences

Information on TI for STIR and SPIR are now included in the text in the method section “Magnetic resonance imaging technique and evaluation” page 6-7, see also the below item.

- The ")" in the header for "TR (ms)" jumped to the next raw.
To reduce the number of tables, Table 1 has been removed from the manuscript. Information from Table 1 is included in the text in the method section “Magnetic resonance imaging technique and evaluation” page 6-7.

Table 2a:
• acronyms should be explained at the bottom of the table

Acronyms are explained at the bottom of Table 1A (Table 2A in the first version).

• It is very helpful to have the appendix file of definitions for the parameters evaluated. However I would shortly add some of the definitions into the Materials and methods part. For example the depth of BMO, intensity of BMO.

The definitions of intensity and depth are added to the explanatory text in the bottom of Tables 1A and B – (Table 2A and 2B in the first version).

• Page 8, line 3: "CI was calculated" should be Confidence interval (CI) was calculated.

Is now corrected.

Results:
• Authors state that only VUs with agreement on the presence of SC were evaluated.

Data regarding for the amount of disagreement regarding these changes is of interest, i.e. how many VUS were excluded from the analysis do to this disagreement.

This information is included in Table 3A, B, C and D, where the number of included observations (endplates, DVU’s or SIJ regions) analysed is listed in the last column. An example for Table 3A: A total of 2208 endplates were evaluated for the presence of signal changes (Type of SC). For “Signal changes in the corner” the agreement analysis between observers A and B was performed on the 97 endplates that both observers had evaluated as having some Type of SC.

• In general there are too many tables, some could be added in an appendix file.

Table 1 and Table 3 are removed and the contents are added to manuscript text. The content from Table 1 is added to “Magnetic resonance imaging technique and evaluation” section, page 6-7 and the content from Table 3 is added to the “Statistical analysis” section, page 9, paragraph 4.

Discussion:
Headers such as "summary of the main results" could be omitted.

The following headers in the discussion are removed:
“Summary of the main results” (page 12)
“Spinal MRI changes” (page 13)
“Changes in the sacroiliac joints” (page 14)
“Global assessment” (page 15).

In general authors compare their results to others but do not give explanations at all to their results. This is missing throughout the discussion. Also the importance of the results and their meaning are not specified. For example, the fact that the general impression of SpA vs. non SpA changes is almost perfect is of critical importance for the clinical and research evaluation of SpA.

We agree that further discussion of the results is necessary and the “Discussion” section is extended with the following paragraphs:

A discussion of the difference between the reliability in the SIJ and spine are added in the “Discussion” section page 13, paragraph 3:

“The tendency of better reliability of the SpA-related findings in the SIJ compared to the spine could be explained by the historically increased focus on SpA-related MRI findings in the SIJ compared to the spine. This is emphasised by the fact that only MRI of the SIJ are currently included in the ASAS criteria for SpA”

Furthermore, a discussion of a recent paper on global assessment is added, page 15, paragraph 2:

“Regarding global assessment, one recent study investigated the inter-observer agreement for global evaluation of MRI of the SIJ in SpA versus non-SpA patients. The kappa value for inter-observer agreement for 5 categories of confidence in the SpA diagnosis were found to be .73(.62-.81) in a cohort of back pain patients referred to a secondary care outpatient clinic in Switzerland due to suspicion of SpA and .74 (.65-.80) in cohorts of back pain patients with anterior uveitis referred to an ophthalmology department in Canada [36]. This is higher than the inter-observer agreement found for global assessment in the current study but with overlapping CI. In general, the spinal MRI findings related to SpA are not as clearly defined as the findings related to the SIJ, which is reflected in the incorporation of only SIJ changes in the ASAS criteria for SpA. Therefore, one reason for the lower agreement in the current study could be that the inclusion of spinal changes in the global assessment increases the uncertainty of the diagnosis.”

The inter-observer agreement of “global assessment of SpA” was found to be 0.61, which is similar to many of the other findings. Therefore, this result is not emphasised.

Finally, the section “Application of the findings” is also extended, page 14, paragraph 2:

“Earlier publications on the evaluation of SpA-related MRI findings have mainly been focused on grading systems for active and chronic SpA changes as a measurement of disease severity in already diagnosed SpA patients. However, the assessment of each lesion separately creates the potential for additional analysis of the diagnostic and
prognostic value of each individual MRI finding. It also creates the potential for describing the development of the changes in subsequent longitudinal studies and it provides a possibility for analysing location-specific alterations, e.g. to compare MRI changes with pain location. Furthermore, the inclusion of both SpA-related and degenerative changes in the same evaluation protocol facilitate an accessible assessment of MRI findings that could mimic SpA-related findings, assessed under the same standardized evaluation session.”

However, as this study is an agreement study we cannot draw clinical conclusions. The clinical importance will be addressed in future manuscript from the BaPa Cohort.
Comments to **Reviewer’s report 2**

**Title:** Spondyloarthritis-related and degenerative MRI changes in the axial skeleton - an inter- and intra-observer agreement study

**Version:** 2

**Reviewer number:** 2

**Referee’s comments to the author(s)**

Only this section of the report will be returned to the authors. Do not comment in this section on the interest/importance level of the manuscript, or whether or not the manuscript should be accepted.

The authors have conducted a high quality carefully designed analysis of reliability of evaluation of findings on MRI of the spine and sacroiliac joints. The cohort and controls for this study of early disease are appropriate and selection of cases logical. The selection of lesions to be evaluated is reasonably comprehensive. The definition of some of these findings and the grading of them are controversial topics for which universal agreement does not currently exist. However the authors have published extensively in spondyloarthritis and are familiar with the issues.

**Methodology**

The some of the definitions of size require clarification. In the spine, each grade is defined as “... of the subcortical bone area”, or “... of the vertebral plate”. This implies a 2 dimensional definition of anterior-posterior diameter (sagittal) multiplied by right-left diameter (coronal) to create an assessment of area affected. This is possible but would be complex to evaluate on sagittal images only and would result in decreased reader reliability. In fact I wonder if the authors are performing a 1 dimensional assessment of the anterior-posterior diameter only (sagittal diameter) as used by many authors such as Braun et al – which is of course a linear grading not an area grading. If this is the case, then the definitions should be rephrased, for example as:

“Small: <25% of the sagittal diameter of the subcortical bone.”

We agree, this could be clarified and the scoring method has been described in more detail in the method section “Magnetic resonance imaging technique and evaluation”, page 7, paragraph 3:

“An estimate of the total vertebral endplate and subchondral bone marrow areas was based on all sagittal slices creating a “3D like picture” of the changes. The spinal MRI changes assessed are listed in Table 1A. For a detailed definition of the MRI changes assessed, see Additional file 1.”

Similarly in the SI joint, it is not clear how the “subchondral bone area” is divided. The assessment of BMO is performed from a series of transverse images, so is it correct for the reader to assume that the “subchondral bone area” is estimated from the integration of the % extent on individual images multiplied by the % of images involved? The fat infiltration is assessed on coronal images and in order to be able to correlate the two domains, I again assume that this is a true “area” assessment. If this is so, it is very important to be quite clear that size grading is done differently in the SI joint versus the spine. If it is a linear grading of 1
dimension only, then correlation of size between domains may be problematic when the size is assessed in different orientations with different features.

A specification is added regarding the scoring method of the SIJ in the method section “Magnetic resonance imaging technique and evaluation” page 7-8, paragraph 1:

“An estimate of the total cartilaginous and ligamentous joint facets and the adjacent subchondral bone marrow areas was based on all semicoronal and semiaxial slices creating a 3D picture of both joint portions. The MRI changes assessed at the SIJ are listed in Table 1B according to the Danish method described previously [7]. For a detailed definition of the MRI changes assessed, see Additional file 1.”

Subchondral Sclerosis is hard to define on MRI. Sclerosis is only seen indirectly as a loss of signal from other bone marrow tissues. Yes, I agree that “low signal” on all sequences would be required as part of the definition, but compared to what? Some authors would consider muscle or erythropoietic marrow to be “low signal on T1” while others would consider these tissues to be “intermediate signal”. Fibrous tissue can be low signal on all sequences. By not describing any comparator and not providing any illustrations, this definition is too vague for generalized use.

The agreement for subchondral sclerosis in this study was acceptable (inter-observer agreement, Kappa=0.65…) reflecting that the observers in the study agreed on the definition of “low signal” compared to normal bone marrow. However, we agree that the definition of subchondral sclerosis is challenging and a reference to other tissue types for “low signal” could clarify the definition and perhaps increase agreement. We will keep this in mind for future studies.

The definition of “low-signal” is specified to “low signal compared to normal bone marrow” in the Appendix file, page 4, row 2.

Minor changes I would suggest for the manuscript include:
• Add pixel size to the description of each MRI sequence

Information of matrix and FOV is added to the description of the MRI protocol (page 6)

• SC is not a standard abbreviation and should be replaced throughout with Signal Change

Signal change is now written in full

• VE is a problematic abbreviation – some authors use the VE abbreviation to denote “Vertebral Edge” and I think it would be better written in full as Vertebral Endplate.

Vertebral endplate is now written in full

• Use of “widespread” in the definition of depth – most readers will instinctively
associate “widespread” with the extent of articular surface involvement which here is defined as % area affected. “Depth” and “deep” are clear and unambiguous – i.e. don’t use “widespread”.

We agree that the wording “widespread” can be ambiguous. It has therefore been substituted by “pronounced” in Table 2A and “Pronounced: Oedema extending ≥1cm beneath the joint surface and covering ≥1 cm²” is added to the explanatory text below. In the appendix “widespread” is substituted with “pronounced”, page 4, row 1.
Comments to Reviewer's report 3
Title: Spondyloarthritis-related and degenerative MRI changes in the axial skeleton - an inter- and intra-observer agreement study
Version: 2
Reviewer number: 3
Referee's comments to the author(s)
Only this section of the report will be returned to the authors. Do not comment in this section on the interest/importance level of the manuscript, or whether or not the manuscript should be accepted.

The authors present a manuscript with good ICC for an adapted scoring form for the evaluation of the MRI of spine and SIJ. The additional value of this method is that other findings than SpA parameters are evaluated. A drawback is that it is not clear to the reader how this relates to other (SpA) scoring methods. If this is more efficient or easier to perform/learn it should be stated why.

We agree that this area needs further clarification, and the differences between this method and other evaluations is stressed in the “background” section where the reason for inclusion of degenerative findings is emphasised, page 4, paragraph 1.

“However, there are still several uncertainties regarding the utility of MRI in the diagnosis of SpA [3], especially in the early stages when the clinical signs of SpA can be difficult to distinguish from non-specific low back pain (LBP) and the MRI signs of SpA can be difficult to distinguish from the much more common findings of degeneration. Signal changes related to degeneration such as Modic changes are an important pitfall in the assessment of SpA [4] and some studies have shown substantial variation in the extent of MRI lesions in the SIJ previously considered to be specific for SpA [5]. Therefore, studies encompassing patients reflecting the target population and using a MRI protocol including both SpA-related and degenerative changes are needed to validate the utility of this new imaging modality for the diagnosis of SpA.”

Furthermore, the section “Application of the findings” is extended, page 14, paragraph 2:

“Earlier publications on the evaluation of SpA-related MRI findings have mainly been focused on grading systems for active and chronic SpA changes as a measurement of disease severity in already diagnosed SpA patients. However, the assessment of each lesion separately creates the potential for additional analysis of the diagnostic and prognostic value of each individual MRI finding. It also creates the potential for describing the development of the changes in subsequent longitudinal studies and it provides a possibility for analysing location-specific alterations, e.g. to compare MRI changes with pain location. Furthermore, the inclusion of both SpA-related and degenerative changes in the same evaluation protocol facilitate an accessible assessment of MRI findings that could mimic SpA-related findings, assessed under the same standardized evaluation session.”