UNIVERSITY OF COPENHAGEN

Canada-Denmark MRI scoring system of the spine in patients with axial spondyloarthritis

Updated definitions, scoring rules and inter-reader reliability in a multiple reader setting

Krabbe, Simon; Østergaard, Mikkel; Pedersen, Susanne J.; Weber, Ulrich; Kröber, Georg; Makysmowych, Walter; Lambert, Robert G.W.

Published in:
RMD Open

DOI:
10.1136/rmdopen-2019-001057

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC

ORIGINAL ARTICLE

Canada-Denmark MRI scoring system of the spine in patients with axial spondyloarthritis: updated definitions, scoring rules and inter-reader reliability in a multiple reader setting

Simon Krabbe [1,2] Mikkel Østergaard [1,2] Susanne J Pedersen [1,2] Ulrich Weber [3,4] Georg Kröber [3] Walter Makysmowych [5,6] Robert G W Lambert [7,8]

ABSTRACT

Objective To validate the Canada-Denmark (CANDEN) MRI scoring system for the spine in axial spondyloarthritis with updated lesion definitions.

Methods Lesion definitions in the CANDEN system were updated and illustrated by a consensus set of reference images. Sagittal spine MRIs of 40 patients with axial spondyloarthritis obtained at baseline and at week 52 after initiation of treatment with the tumour necrosis factor inhibitor golimumab were evaluated in unknown chronology by seven readers blinded to all other data.

Results CANDEN MRI spine inflammation score had very good reliability for status scores (single-measure intraclass correlation coefficient (ICC) of 21 reader pairs median of 0.91 (IQR 0.88–0.92)) and change scores (ICC 0.88 (0.86–0.92)). CANDEN MRI spine fat score had good to very good reliability for status scores (ICC 0.79 (0.75–0.86)) and moderate to good reliability for detecting change (ICC 0.59 (0.46–0.73)). CANDEN MRI spine bone erosion score and CANDEN MRI spine new bone formation score had slight to moderate reliability for status scores (ICC 0.38 (0.32–0.52) and 0.39 (0.27–0.49), respectively).

Conclusion The CANDEN MRI spine scoring system allows a comprehensive evaluation of inflammation, fat, bone erosion and new bone formation of the spine in patients with axial spondyloarthritis. It demonstrated very good reliability for detecting change in inflammation, moderate to good reliability for detecting change in fat, and slight to moderate reliability for detecting bone erosions and new bone formation. Studies with longer follow-up or patients with more advanced spinal involvement may be needed to reliably detect change in bone erosion and new bone formation scores.

Trial registration number NCT02011386.

Key messages

What is already known about this subject?
► The Canada-Denmark (CANDEN) scoring system is a comprehensive scoring system that allows evaluation of inflammatory and structural lesions of the spine in patients with axial spondyloarthritis.

What does this study add?
► In a multireader setting with experienced as well as newly trained MRI readers, the CANDEN MRI spine scoring system demonstrated very good reliability for detecting change in inflammation, moderate to good reliability for detecting change in fat, and slight to moderate reliability for detecting bone erosions and new bone formation.
► Four inflammation subscores that capture different anatomical parts of the spine were developed.

How might this impact on clinical practice?
► The scoring system can be used to investigate how drugs with different modes of action influence the individual components of spine inflammation and damage, the mutual relationship between lesions, as well as the evolution of inflammation and damage in the entire spine.

INTRODUCTION

In 2009, the Canada-Denmark (CANDEN) MRI working group published definitions of pathologies, an atlas of inflammatory and structural lesions of the spine, and the reliability for scoring various lesions among the founders of the system (RGWL, WM, SJP and MØ).[1–5] Later, preliminary CANDEN MRI spine scores of inflammation, fat lesion, bone erosion and new bone formation were developed based on these lesion definitions, and they were applied in two studies of patients with axial spondyloarthritis (axSpA) and ankylosing spondylitis (AS) treated with tumour necrosis factor (TNF) inhibitor,[6–8] where significant changes in the scores over

Table 1	MRI lesion definitions
Normal bone marrow signal	The bone marrow signal in the centre of the vertebral body, if normal, constitutes the reference for designation of normal signal. If the usual reference point is not normal, then the closest vertebra with normal marrow signal in the centre of the vertebral body is used as reference.
Central vs lateral slices	*Definitions of central and lateral slices:* Central sagittal slices (referred to as central slices) are slices that include the spinal canal. The pedicle may be partially seen but is not continuous between the vertebral body and the posterior elements. Lateral sagittal slices (referred to as lateral slices) are slices that do not include the spinal canal and where the pedicle is continuous between the vertebral body and the posterior elements, or the slice is lateral to the pedicle. Vertebral body inflammatory lesions and fat lesions are scored on central and lateral slices. Vertebral body bone erosion and new bone formation are scored on central slices only. *Reference point for slice position:* The slice position is defined by the discovertebral unit (DVU) and not by the vertebral body. Therefore, the slice position will always be the same for both endplates of a DVU. The reference structure is the pedicle related to the lower endplate of the DVU. For example, the slice position of both endplates at L1/L2 is defined by the presence/absence of the L2 pedicle on that image. *Reference sequence:* Whether T1-weighted (T1W) or short tau inversion recovery (STIR) is used to designate slice position will depend on the lesion domain that is being evaluated, that is, for fat lesions, bone erosion and new bone formation this would be the T1W sequence, and for inflammatory lesions this would be the STIR sequence.
Slices in the cervical spine	In the cervical spine, facet joints are located laterally to the vertebral body and are not seen on the same sagittal image. Therefore, in the cervical spine, all images that include the vertebral body are defined as central slices. For C7/T1, C7 is scored as a cervical vertebra while T1 is scored as a thoracic vertebra.
Inflammatory lesion	*Inflammatory lesion in vertebral bodies:* Increased signal in bone marrow on STIR/T2FS in a vertebral body. *Corner inflammatory lesion (CIL), subdivided into anterior (aCIL) and posterior (pCIL):* Inflammatory lesion at the vertebral corner on at least one central slice. *Type A CIL:* The inflammatory lesion itself reaches the corner. *Type B CIL:* The inflammatory lesion does not extend to the corner but does extend to both the vertebral endplate and the anterior/posterior vertebral cortex (usually because of the presence of fat or erosion in the corner). If a type B CIL is present, a type A CIL cannot be scored. *Non-corner inflammatory lesion (NIL):* Inflammatory lesion adjacent to the endplate on a central slice but not involving the anterior or posterior vertebral corner. *Lateral vertebral body inflammatory lesion (LIL), subdivided into anterolateral (aLIL) and posterolateral (pLIL):* Inflammatory lesion adjacent to the endplate on a lateral slice without distinction between corner and non-corner lesions. Inflammatory lesions involving the posterior elements of the spine, not the vertebral bodies: *Facet joint inflammatory lesion:* Increased signal in bone marrow on STIR/T2FS in at least one facet of a facet joint. *Transverse process inflammatory lesion:* Increased signal in bone marrow on STIR/T2FS in the transverse process. *Rib inflammatory lesion:* Increased signal in bone marrow on STIR/T2FS in the rib. *Spinous process inflammatory lesion:* Increased signal in bone marrow on STIR/T2FS in the spinous process. *Soft tissue inflammatory lesion:* Increased signal in soft tissue at entheseal attachments on STIR/T2FS.

Continued

Table 1   Continued

Fat lesion	Focally increased signal in bone marrow on T1W. *Corner fat lesion (CFAT), subdivided into anterior (aCFAT) and posterior corner fat lesion (pCFAT):* Fat lesion at the vertebral corner on at least one central slice. *Non-corner fat lesion (NFAT):* Fat lesion adjacent to the endplate on any central slice but not involving the anterior or posterior vertebral corner. *Lateral vertebral body fat lesion (LFAT), subdivided into anterolateral (aLFAT) and posterolateral vertebral body fat lesion (pLFAT):* Fat lesion adjacent to the endplate on a lateral slice without distinction between corner and non-corner lesions. *Facet joint fat lesion:* Increased signal in bone marrow on T1W in at least one facet of a facet joint. Fat lesions in the upper cervical spine: To be considered a true fat lesion in the upper part of the cervical spine, a potential lesion has to be very homogeneous and to have a clearly increased signal on T1W that is clearly demarcated from adjacent bone marrow. This applies especially for C2 and C3 lower anterior corners, where a diffuse signal increase on T1W is common in healthy subjects. It may be necessary to use different windowing of the image for optimal visualisation of the upper vs the lower cervical spine.
Bone erosion	Full-thickness loss of dark appearance of cortical bone at its anticipated location and loss of normal bright appearance of adjacent bone marrow on T1W. Vertebral bone erosions are scored on central slices only. *Corner bone erosion, subdivided into anterior and posterior corner erosion:* Bone erosion involving the vertebral corner on a central slice. *Facet joint bone erosion:* Bone erosion involving at least one facet of a facet joint.
New bone formation	Because of the current limitations of MRI, the following definitions are restricted to only those bone spurs within which bone marrow signal is detectable. In the future, MRI may be able to detect mineralisation within the bone spur at an earlier stage. *Bone spur:* Bright signal on T1W extending from the vertebral endplate towards the adjacent vertebra. *Corner bone spur (syndesmophyte) (COS), subdivided into anterior (aCOS) and posterior corner bone spur (pCOS):* Bone spur at the vertebral corner on a central slice. *Non-corner bone spur:* Bone spur involving the endplate, which involves neither the anterior nor the posterior vertebral corner on a central slice. *Ankylosis:* Bright signal on T1W extending from a vertebra and being continuous with the adjacent vertebra. *Corner ankylosis, subdivided into anterior and posterior corner ankylosis:* Ankylosis at the anterior or posterior corner on a central slice. *Non-corner ankylosis:* Ankylosis which involves the endplate, but neither the anterior nor the posterior vertebral corner on a central slice. *Facet joint ankylosis:* Ankylosis at the facet joint.
Assessment of size of inflammatory lesion/fat lesion/bone erosion	Size is only assessed for lesions on central slices in the thoracic and lumbar spine; the slice where the lesion is largest is used. Size of corner lesions is assessed vertically perpendicular to the vertebral endplate and horizontally parallel to the vertebral endplate. Corner lesions are large if they involve 25% or more of the anterior-posterior diameter of the vertebral endplate and/or the height of the vertebral body. Non-corner lesions are large if they involve more than 25% of the vertebral body height. If either new bone formation or bone erosion is distorting the configuration of the corner, size is measured from the point at which the original cortex and endplate would have met. Inflammatory lesions, fat lesions and bone erosions are assessed for size.
Combined corner and non-corner inflammatory lesion/ fat lesion	If a corner inflammatory lesion/fat lesion on any central slice involves more than 50% of the anteroposterior diameter of the vertebra (ie, it extends beyond the midpoint), it is scored as both a corner inflammatory lesion/fat lesion and a non-corner inflammatory lesion/fat lesion. Size of the corner component can only be assessed by vertical extent at the anterior/posterior vertebral cortex (horizontal size cannot be assessed). Size of the non-corner component is assessed by height at the midpoint of the anteroposterior diameter of the vertebra.
Combined central slice and lateral slice inflammatory lesion/ fat lesion	An inflammatory lesion/fat lesion may be observed extending across both central and lateral slices. It is then scored as being both a central (aCIL/pCIL/NIL) and lateral (aLIL/pLIL) inflammatory lesion, or a central (aCFAT/pCFAT/NFAT) and lateral (aLFAT/pLFAT) fat lesion.

T2FS, T2-weigted image with fat saturation.

1–5 years occurred, except for bone erosion. These results suggest that the CANDEN MRI spine scoring system allows a comprehensive, anatomy-based evaluation of both inflammatory and structural lesions, and that it can provide patient-level sum scores for inflammatory lesions, fat lesions, bone erosion and new bone formation that are responsive to detect change.

Several other MRI scoring systems of the spine have been developed. The Ankylosing Spondylitis spine Magnetic Resonance Imaging (ASspiMRI) method provides an activity score of 0–6 and a chronicity score of 0–6 for each discovertebral unit (DVU).[9 10] The Berlin method provides an activity score of 0–3 per DVU and was later amended to allow assessment of inflammation of the posterior segments with a score of 0–2, and erosions, fatty bone marrow deposition and bone proliferation, each with a score of 0–3 per DVU.[11–16] The Spondyloarthritis Research Consortium of Canada (SPARCC) MRI Spine Inflammation Index divides each DVU into four vertebral quadrants, which are assessed for presence/absence of bone marrow oedema with additional scores for intensity and depth. This is done on three consecutive slices, and provides a score of 0–18 for each DVU and has been used for scoring either the 6 worst affected DVUs or all 23 DVUs.[17–21] Assessment of inflammation in pedicles, facet joints, transverse processes, spinous processes and soft tissues in relation to the SPARCC method has been investigated.[22] The Leeds method counts the number of inflammatory lesions in the vertebral bodies, spinous processes, facet joints and soft tissue in the lumbar spine.[23 24] The Aarhus method provides an activity score of 0–3, and an additional score of 1 for costovertebral involvement for each DVU and a chronicity score of 0–9 for each DVU.[25] The ASspiMRI, Berlin and SPARCC scoring methods for inflammatory activity were compared in a multireader study.[26]

The CANDEN MRI spine scoring system is a comprehensive system which permits a detailed description of the involvement of different spinal structures, various topographic parts of the vertebral bodies, the facet joints, the spinous processes, the transverse processes, the ribs and soft tissue. By covering various inflammatory and structural lesion types, the CANDEN MRI spine scoring system may help identify subgroups of patients and different disease trajectories. It also allows investigation of the relation between different lesion types over time at the lesion level and at the patient level,[27–29] and how this association may vary depending on the mode of action of the applied therapy. In this updated version of the scoring system, we have harmonised the definitions and rules across the various lesion types in order to ease the use of the system. Sum scores for inflammation, fat and erosion differ slightly from the preliminary sum scores that were previously published,[6–8] but differences are minor; the decision to make this update of the system was driven by a desire for internally consistent rules, definitions and accompanying atlas.

The objective of this work was to update the lesion definitions including an atlas with consensus reference images, clarify the CANDEN MRI spine scoring system for inflammation and structural lesions, and assess the reliability of the scoring system in a multireader setting.

METHODS
Patient characteristics
Sagittal MRIs of the spine of 40 patients with axSpA obtained at baseline and after 52 weeks in the MANGO (Novel MRI ANd biomarkers in GOlimumab-treated patients with axial spondyloarthritis) cohort study (ClinicalTrials.gov: NCT02011386) were used. All patients were diagnosed with axSpA and fulfilled the imaging arm of the Assessment of SpondyloArthritis international Society (ASAS) axSpA criteria. TNF inhibitor treatment with golimumab 50 mg once every month was initiated in all patients at baseline. The median age of the 40 patients was 33 years (IQR 28–48, range 22–73), 55% were men, 73% were human leucocyte antigen B27 (HLA-B27)-positive, 24 (60%) fulfilled the radiographic part of the modified New York criteria, while 16 (40%) had non-radiographic axSpA, with a median time since diagnosis of 0.7 years (IQR 0.2–3.4) and median symptom duration of 4.6 years (IQR 2.1–9.8). At baseline, 53% had elevated C-reactive protein (CRP) (defined as >5 mg/L), Ankylosing Spondylitis Disease Activity Score (ASDAS) median of 3.6 (IQR 3.1–4.0) and SPARCC sacroiliac joint (SIJ) inflammation median score of 14 (IQR 8–22). At week 52, 21% had elevated CRP, ASDAS median of 1.4 (IQR 0.8–2.2) and SPARCC SIJ inflammation median score of 2 (IQR 0–4).

MRI methodology
MRI of the spine was performed at baseline and at week 52. Spine images were acquired in two halves (upper and lower spine). MRI was performed on a single high-field Philips Medical Systems Ingenia scanner (3.0 T) by sagittal T1-weighted turbo spin echo: repetition time (TR): 663 ms, echo time (TE): 8 ms, matrix: 444×319, field of view (FOV): 400×180 mm, slice thickness 4 mm, interslice gap 0.4 mm, and acquisition duration for upper/lower spine: 2 min 7 s each; and by short tau inversion recovery: TR: 3476 ms, TE: 70 ms, inversion time: 210 ms, matrix: 400×307/400×307, FOV: 400×180 mm, slice thickness 4 mm, interslice gap 0.4 mm, and acquisition duration for upper/lower spine: 5 min 23 s each.

Two readers screened the images for variation in spinal segmentation using C2 as the reference vertebra. Four patients had partial or complete sacralisation of L5 or lumbarisation of S1. In case of lumbarisation of S1, the DVU between S1 and S2 was not scored. Images were read independently by seven readers, of whom five were experienced readers (MØ, SJP, UW, WM and RGWL) and two were newly trained readers with less experience (SK and GK). This was done without knowledge of clinical,

Table 2 Scoring rules

CANDEN MRI spine inflammation score	The following lesions are scored as 0 (absent) or 1 (present): aCIL, pCIL, NIL, aLIL, pLIL, FIL, TIL, RIL, SPIL and STIL. A score of 1 is added for large aCIL and pCIL. Non-corner lesions (NIL) are scored as 0 (absent) or 2 (present), and a score of 2 is added for large non-corner lesions (NIL). The CANDEN MRI spine inflammation score has a total scoring range of 0–614. The vertebral body subscore has a score range of 0–464 (11 cervical endplates from C2 to C7, each with a maximum score of 4, 35 thoracic/lumbar endplates from T1 to S1 each with a maximum score of 12). The posterior elements subscore has a score range of 0–150 (facet joints at all 23 levels from C2/C3 to L5/S1, transverse process at 17 levels from T1 to L5, rib at 12 levels from T1 to T12, spinous process at 23 levels from C2 to L5, soft tissue inflammation at 23 levels from C2/C3 to L5/S1). The CANDEN MRI spine inflammation score may also be divided into the following four subscores: ▶ A, 'Vertebral body corner inflammation subscore', defined as the sum of the anterior and posterior corner lesions, and anterolateral and posterolateral vertebral body lesions (posterolateral vertebral body lesions only at levels T12/L1 to L5/S1), which may predominantly represent enthesitis related to the anterior and posterior longitudinal ligament and the annulus fibrosus (range 0–254). ▶ B, 'Spondylodiscitis subscore', defined as the sum of non-corner vertebral body lesions, which may predominantly represent inflammation from the disc or endplate itself (range 0–162). ▶ C, 'Facet joint inflammation subscore', defined as the sum of facet joint lesions, representing inflammation of the synovial facet joints (range 0–46). ▶ D, 'Posterolateral elements inflammation subscore', defined as the sum of rib, transverse process, spinous process, soft tissue inflammation and posterolateral vertebral body lesions (posterolateral vertebral body only at levels C7/T1 to T11/T12 because pLIL in the thoracic spine is considered to be related to the costovertebral joint), representing inflammation related to these synovial joints and enthesitis of ligaments of the posterior elements of the spine (range 0–152).
CANDEN MRI spine fat score	The following lesions are scored as 0 (absent) or 1 (present): aCFAT, pCFAT, aLFAT, pLFAT and FFAT. A score of 1 is added for large aCFAT and pCFAT. Non-corner lesions (NFAT) are scored as 0 (absent) or 2 (present), and a score of 2 is added for large non-corner lesions (NFAT). The total scoring range for the CANDEN MRI spine fat score is 0–510. The range for the vertebral body subscore is 0–464 (11 cervical endplates each with a maximum score of 4, 35 thoracic/lumbar endplates each with a maximum score of 12). The range for the fat posterior elements subscore is 0–46 (facet joints at all 23 levels).
CANDEN MRI spine bone erosion score	The following lesions are scored as 0 (absent) or 1 (present): aCOBE, pCOBE and FABE. A score of 1 is added for large aCOBE and pCOBE. The total scoring range for the CANDEN MRI spine bone erosion score is 0–208. The range for the vertebral body subscore is 0–162 (11 cervical endplates each with a maximum score of 2, 35 thoracic/lumbar endplates each with a maximum score of 4). The range for the posterior elements subscore is 0–46 (facet joints at all 23 levels).
CANDEN MRI spine new bone formation score	Anterior corner, posterior corner and non-corner ankylosis are scored as 0 (absent) or 6 (present). Anterior corner, posterior corner and non-corner bone spurs are scored as 0 (absent) or 2 (present). Facet joint ankylosis is scored as 0 (absent) or 1 (present). The total scoring range for the CANDEN MRI spine new bone formation score is 0–460. The range for the vertebral body subscore is 0–414 (23 levels with a maximum score of 18). The range for the posterior elements subscore is 0–46 (facet joints at all 23 levels).

aCFAT, anterior corner fat lesion; aCIL, anterior corner inflammatory lesion; aCOBE, anterior corner bone erosion; aLFAT, anterolateral vertebral body fat lesion; aLIL, anterolateral vertebral body inflammatory lesion; CANDEN, Canada-Denmark; FABE, facet joint bone erosion; FFAT, facet joint fat lesion; FIL, facet joint inflammatory lesion; NFAT, non-corner fat lesion; NIL, non-corner inflammatory lesion; pCFAT, posterior corner fat lesion; pCIL, posterior corner inflammatory lesion; pCOBE, posterior corner bone erosion; pLFAT, posterolateral vertebral body fat lesion; pLIL, posterolateral vertebral body inflammatory lesion; RIL, rib inflammatory lesion; SPIL, spinous process inflammatory lesion; STIL, soft tissue inflammatory lesion; TIL, transverse process inflammatory lesion.

biochemical and other imaging information and blinded to chronology.

Description of the CANDEN scoring system
Lesion definitions applied in the CANDEN MRI spine scoring system are provided in table 1, and all the rules for scoring lesions and calculating the CANDEN patient-level scores and subscores are provided in table 2. To improve the use of the system, the original set of reference images for all four lesion types (inflammatory lesions, fat lesions,

bone erosion and new bone formation) with definitions of lesions, rules for scoring, and examples of lesions above and below the threshold for scoring were updated and made available at www.carearthritis.com and www.copecare.dk

Statistical analysis
Patient-level sum scores for each lesion type (inflammatory lesions, fat lesions, bone erosion and new bone formation) were calculated based on each reader's

Table 3 Patient-level sum scores and inter-reader reliability of the four lesion types in the CANDEN MRI spine scoring system

Variable	Baseline			Week 52			Change			
	Score	Single-measure ICC	Average-measure ICC	Score	Single-measure ICC	Average-measure ICC	Score	Single-measure ICC	Average-measure ICC	Wilcoxon test, p value
CANDEN MRI spine inflammation score, range 0–614	13.2 (5; 0–132)	0.91 (0.88–0.92)	0.95 (0.94–0.96)	2.4 (1; 0–19)	0.77 (0.67–0.86)	0.87 (0.81–0.92)	−10.8 (−3; −113 to 1)	0.88 (0.86–0.92)	0.94 (0.93–0.96)	<0.001
Vertebral body corner inflammation subscore, range 0–254	8.0 (3; 0–89)	0.93 (0.91–0.95)	0.97 (0.95–0.97)	1.6 (0; 0–14)	0.84 (0.70–0.87)	0.91 (0.82–0.93)	−6.4 (−2; −75 to 1)	0.90 (0.87–0.94)	0.95 (0.93–0.97)	<0.001
Spondylodiscitis subscore, range 0–162	1.8 (0; 0–25)	0.84 (0.74–0.91)	0.91 (0.85–0.95)	0.4 (0; 0–9)	0.66 (0.58–0.69)	0.80 (0.74–0.82)	−1.3 (0; −22 to 0)	0.82 (0.79–0.89)	0.90 (0.88–0.94)	0.003
Facet joint inflammation subscore, range 0–46	1.0 (0; 0–13)	0.78 (0.73–0.80)	0.88 (0.84–0.96)	0.2 (0; 0–2)	0.50 (0.23–0.64)	0.66 (0.37–0.78)	−0.8 (0; −13 to 2)	0.79 (0.75–0.89)	0.88 (0.85–0.94)	0.001
Posterolateral elements inflammation subscore, range 0–152	2.4 (0; 0–19)	0.55 (0.45–0.71)	0.71 (0.62–0.83)	0.2 (0; 0–2)	0.21 (0.16–0.32)	0.34 (0.28–0.48)	−2.2 (0; −18 to 1)	0.55 (0.44–0.70)	0.71 (0.61–0.83)	<0.001
CANDEN MRI spine fat score, range 0–510	6.8 (2; 0–49)	0.79 (0.75–0.86)	0.89 (0.86–0.92)	8.3 (4; 0–57)	0.79 (0.71–0.85)	0.88 (0.83–0.92)	1.5 (0; −4 to 11)	0.59 (0.46–0.73)	0.74 (0.63–0.85)	<0.001
CANDEN MRI spine bone erosion score, range 0–208	0.3 (0; 0–2)	0.38 (0.32–0.52)	0.55 (0.48–0.68)	0.2 (0; 0–2)	0.32 (0.22–0.58)	0.49 (0.36–0.73)	−0.1 (0; −1 to 0)	0.01 (−0.08 to 0.29)	0.01 (−0.17 to 0.45)	0.444
CANDEN MRI spine new bone formation score, range 0–460	3.1 (1; 0–24)	0.39 (0.27–0.49)	0.56 (0.42–0.65)	3.4 (1; 0–27)	0.38 (0.23–0.56)	0.55 (0.38–0.71)	0.2 (0; −1 to 3)	0.00 (0.00–0.00)	0.00 (0.00–0.00)	0.047

Scores are reported as mean (median; range) using the average score of all readers. ICCs are reported as the median (IQR) of 21 pairwise combinations of readers. Wilcoxon test for change between baseline and week 52.
CANDEN, Canada-Denmark; ICC, intraclass correlation coefficient.

scores. The reliability of patient-level sum scores was assessed by pairwise intraclass correlation coefficients (ICC) for status scores and for change scores between all reader pairs, based on two-way single-measure models by absolute agreement and two-way average-measure models by absolute agreement. Similar analyses were also performed separately for the vertebral bodies and posterior elements, as well as for all individual components. Agreement plots in which a reader's scores are plotted against the mean score of all readers allow a graphical assessment of the agreement between multiple readers. The reliability at the lesion level was assessed for status scores and change scores between all reader pairs by pairwise kappa with squared weights. Kappa values and ICC values ≥0.8 were considered to represent very good reliability, ≥0.6 to <0.8 good reliability, ≥0.4 to <0.6 moderate reliability, ≥0.2 to <0.4 slight reliability, and <0.2 poor reliability.

The responsiveness for changes in CANDEN MRI spine inflammation score, CANDEN MRI spine fat score, CANDEN MRI spine bone erosion score and CANDEN MRI spine new bone formation score between baseline and week 52 was assessed based on the observed status and change scores. Change at group level was tested for statistical significance using Wilcoxon signed-rank test.
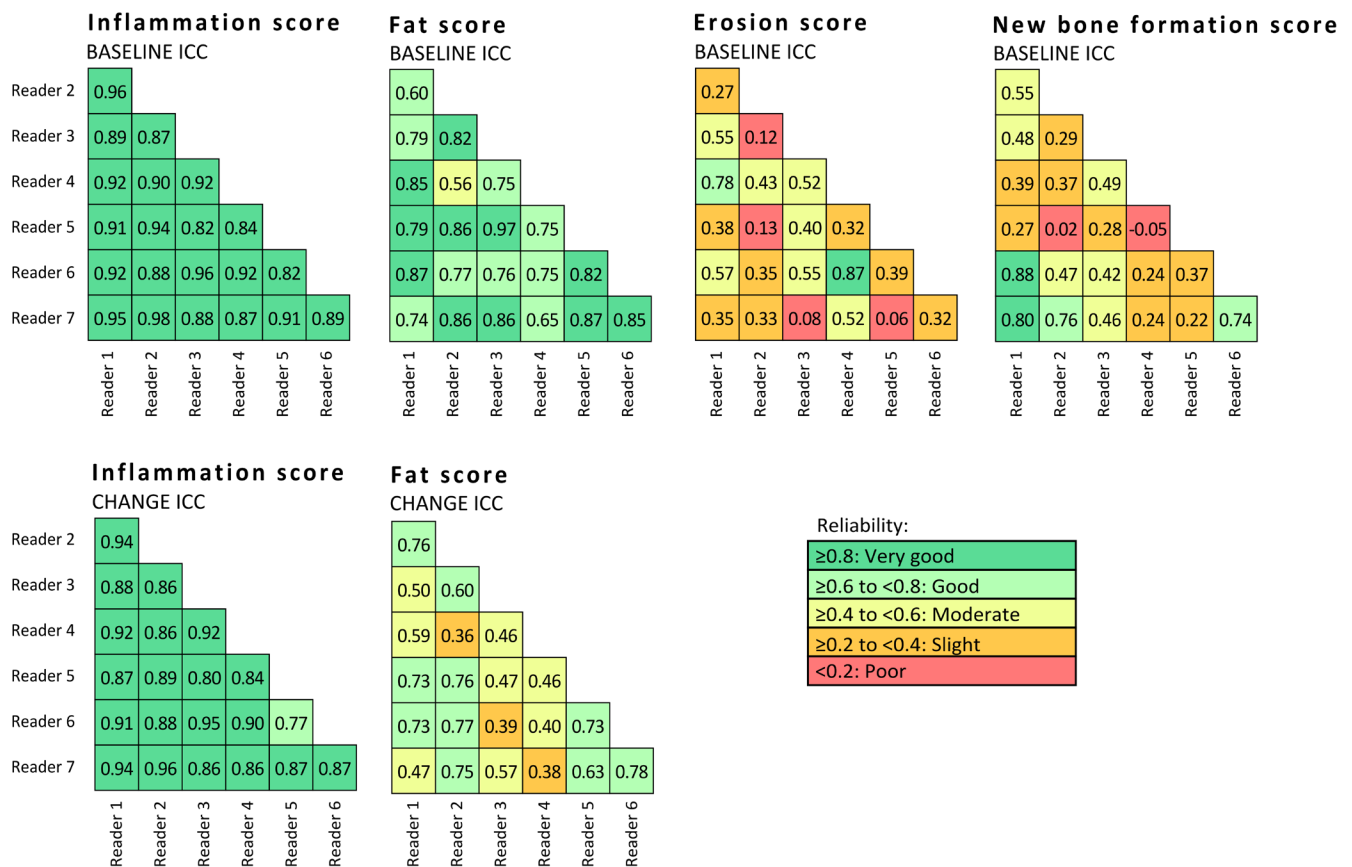
RESULTS

The CANDEN MRI spine inflammation score and all its subscores decreased significantly, the CANDEN MRI spine fat score increased significantly, whereas the CANDEN MRI spine bone erosion score and the CANDEN MRI spine new bone formation score remained largely unchanged (see table 3).

The reliability of the CANDEN MRI spine inflammation score was good to very good, and the reliability of the CANDEN MRI spine fat score was moderate to very good, for status scores and change scores during active treatment after 52 weeks of follow-up (see figure 1). The reliability of the CANDEN MRI spine bone erosion score and the CANDEN MRI spine new bone formation score was slight to moderate for status scores.

All readers had a largely similar mean CANDEN MRI spine inflammation score and CANDEN MRI spine fat score, both regarding baseline score and decrease in score during follow-up. Most readers identified no overall change in CANDEN MRI spine bone erosion score or CANDEN MRI spine new bone formation score during follow-up. Differences between readers' scores tended to be smaller for patients with small mean scores and larger for patients with large mean scores for all four lesion types (see online supplementary figure 1 and online supplementary table 1).

Almost all components of the CANDEN MRI spine inflammation score, that is, anterior corner, posterior corner, non-corner, anterolateral, posterolateral, transverse process, rib, facet joint and soft tissue inflammatory lesions, decreased significantly during the 52-week

Inflammation score
BASELINE ICC

	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6
Reader 2	0.96					
Reader 3	0.89	0.87				
Reader 4	0.92	0.90	0.92			
Reader 5	0.91	0.94	0.82	0.84		
Reader 6	0.92	0.88	0.96	0.92	0.82	
Reader 7	0.95	0.98	0.88	0.87	0.91	0.89

Fat score
BASELINE ICC

	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6
Reader 2	0.60					
Reader 3	0.79	0.82				
Reader 4	0.85	0.56	0.75			
Reader 5	0.79	0.86	0.97	0.75		
Reader 6	0.87	0.77	0.76	0.75	0.82	
Reader 7	0.74	0.86	0.86	0.65	0.87	0.85

Erosion score
BASELINE ICC

	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6
Reader 2	0.27					
Reader 3	0.55	0.12				
Reader 4	0.78	0.43	0.52			
Reader 5	0.38	0.13	0.40	0.32		
Reader 6	0.57	0.35	0.55	0.87	0.39	
Reader 7	0.35	0.33	0.08	0.52	0.06	0.32

New bone formation score
BASELINE ICC

	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6
Reader 2	0.55					
Reader 3	0.48	0.29				
Reader 4	0.39	0.37	0.49			
Reader 5	0.27	0.02	0.28	-0.05		
Reader 6	0.88	0.47	0.42	0.24	0.37	
Reader 7	0.80	0.76	0.46	0.24	0.22	0.74

Inflammation score
CHANGE ICC

	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6
Reader 2	0.94					
Reader 3	0.88	0.86				
Reader 4	0.92	0.86	0.92			
Reader 5	0.87	0.89	0.80	0.84		
Reader 6	0.91	0.88	0.95	0.90	0.77	
Reader 7	0.94	0.96	0.86	0.86	0.87	0.87

Fat score
CHANGE ICC

	Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	Reader 6
Reader 2	0.76					
Reader 3	0.50	0.60				
Reader 4	0.59	0.36	0.46			
Reader 5	0.73	0.76	0.47	0.46		
Reader 6	0.73	0.77	0.39	0.40	0.73	
Reader 7	0.47	0.75	0.57	0.38	0.63	0.78

Reliability:
≥0.8: Very good	
≥0.6 to <0.8: Good	
≥0.4 to <0.6: Moderate	
≥0.2 to <0.4: Slight	
<0.2: Poor	

Figure 1 Inter-reader reliability, single-measure intraclass correlation coefficient (ICC) values for agreement at the patient level in pairwise combinations of readers.

follow-up. Spinous process inflammation also tended to decrease, although this did not reach statistical significance (see online supplementary table 2).

Similarly, for the CANDEN MRI spine fat score, anterior corner, posterior corner, and anterolateral and posterolateral fat lesions all increased significantly during the 52-week follow-up. Non-corner and facet joint fat lesions tended to increase, although this did not reach statistical significance. None of the components of the CANDEN MRI spine bone erosion score or the CANDEN MRI spine new bone formation score changed significantly during 52-week follow-up, although anterior corner and non-corner bone erosions tended to decrease slightly, and anterior and posterior corner new bone formation tended to increase slightly (see online supplementary table 3).

DISCUSSION

The revised CANDEN MRI spine scoring system demonstrated very good inter-reader agreement for the assessment of inflammatory lesions and moderate to very good inter-reader agreement for the assessment of fat lesions for both status and change scores in this multireader study involving seven readers. The individual components of these scores, whether from the vertebral bodies, that is, anterior corner, posterior corner and non-corner lesions, and anterolateral and posterolateral vertebral body lesions, or from the facet joints (for both inflammation and fat), or from the transverse processes, ribs, spinous processes and soft tissue (for inflammation only), showed responsiveness to change. Inflammatory lesions decreased, while fat lesions increased during treatment, although for some components with a low frequency of lesions, change over time did not reach statistical significance. The proposed four subscores of spinal inflammation, that is, the vertebral body corner inflammation subscore, the spondylodiscitis subscore, the facet joint inflammation subscore and the posterolateral elements inflammation subscore, all decreased significantly during follow-up in this cohort of patients treated with the TNF inhibitor golimumab.

The observed slight to moderate reliability for bone erosion and new bone formation status scores at baseline and their poor reliability for change over time may in part be explained by bone erosion and new bone formation being rather infrequent lesions in this cohort. The cohort consisted of biological treatment-naïve patients with axSpA, where many had only limited or no involvement of the spine, and where little structural change happened during 52 weeks of follow-up. This length of follow-up may be too short a timeframe for detecting structural changes in patients with axSpA, but we were

able to identify a slight progression of new bone formation at the group level. When comparing the reliability for assessing spinal structural lesions across various studies, the proportion of patients who reached different stages in the evolution of disease should be taken into account. Higher reliability may be reached if patients were to be selected based on spinal involvement above a certain threshold at baseline, or in studies with longer follow-up since disease progression is slow. Another contributor to the slight to moderate reliability for bone erosion and new bone formation status scores is that structural lesions are often small and hard to detect on MRI T1-weighted images, where cortical bone provides no MRI signal and contrast and spatial resolution is limited compared with radiography. We have previously shown that the CANDEN MRI spine new bone formation score correlates with clinical examination (Bath Ankylosing Spondylitis Metrology Index) as well as with a radiographic score (modified Stoke Ankylosing Spondylitis Spine Score, mSASSS).[7] Future studies may explore how well structural changes as assessed with MRI correlate with radiography or CT findings.

We have limited data on applying the CANDEN MRI spine scoring system to patients with severe involvement of the spine and with longer follow-up. A small longitudinal study with serial lower spine MRI in patients with AS identified progression of CANDEN MRI spine fat score at 0.4 years of follow-up and CANDEN MRI spine new bone formation score at 0.8 years of follow-up at the group level, with additional progression in CANDEN MRI spine new bone formation also noted at up to 5 years of follow-up, whereas no change in CANDEN MRI spine bone erosion score was observed.[8] This suggests that the CANDEN system can be used for monitoring structural MRI lesions of the spine in patients with axSpA at the group level. Further validation on the ability of the system to capture change in erosion and new bone formation across the entire spectrum of spondyloarthritis is needed.

Fat lesions and bone erosion of the facet joints have been added to the scoring system. The aim is to have a comprehensive scoring system. The image quality of MRI scans obtained today allows such lesions to be visualised, although in this cohort these lesion types were rare. An assessment of fat lesions and bone erosion in the facet joints also allows studies of development of ankylosis of the facet joints over time.

Lateral slices were tentatively assessed for bone erosion, but we decided not to include scoring of bone erosion on lateral slices in the scoring system. Not only in this read-out, but also in the overall experience of the group members, such lesions were considered infrequent, and it is inherently difficult or impossible to assess bone erosion and new bone formation at those locations confidently when only sagittal images are available due to partial volume artefacts. For the same reason, bone spurs and ankylosis are also scored only on central slices, not on lateral slices. It was decided to exclude non-corner erosion from the scoring system, since these lesions were also considered infrequent. Moreover, protrusion of nucleus pulposus through the vertebral endplate (Schmorl's nodes), unrelated to spondyloarthritis, was judged to be a major confounder.

Inflammatory and fat lesions have been shown to predict radiographic progression at the same location on conventional spinal radiographs[30 31]; for the CANDEN MRI spine scoring system, this specifically pertains to anterior corner inflammatory lesions and anterior corner fat lesions.[27–29 32] The value of the various MRI spine scores for predicting structural progression, for example, worsening in mSASSS, needs further clarification.

The CANDEN MRI spine scoring system may be used in its entirety, or parts of it may be used selectively, depending on the objectives of the study. In a randomised study with short follow-up time where the aim is to identify a between-group difference for change in spinal inflammation over time, it may be decided to only score inflammatory lesions. To keep scores comparable across studies, the terminology 'CANDEN MRI spine inflammation score', 'CANDEN MRI spine fat score', 'CANDEN MRI spine bone erosion score' and 'CANDEN MRI spine new bone formation score' should be used only when all components are scored and calculated as described in this article.

It may be considered a limitation of the scoring system that it is complex and requires a thorough training of readers as well as experience with musculoskeletal MRI. Not only the rules and definitions of the scoring system must be well understood, but the anatomy of the spine as visualised by MRI as well as the ability to identify the various lesion types seen in patients with spondyloarthritis must also be mastered. The system is not designed for use in daily clinical practice but is intended for research use.

Even though the scoring system is intended to be used in patients diagnosed with spondyloarthritis, such patients may also have concomitant degeneration of intervertebral discs and endplates with related bone marrow changes. We currently propose that in trials and observational cohorts of patients diagnosed with axSpA, all lesions that fulfil the definitions outlined here should be scored. The reader should not make a judgement whether a lesion is likely to be related to spondyloarthritis or not, but should decide whether the lesion fulfils the definitions provided. This gives a higher sensitivity for capturing all spinal lesions that are present, at the cost of a reduced diagnostic specificity from, for example, degeneration of intervertebral discs. Other ways of handling disc degeneration may be considered depending on the research design and the purpose of the study, for example, excluding scores from DVUs that are considered to have loss of at least 50% of normally expected disc height or where disc degeneration is obvious.[33] The issue of degenerative spinal changes needs to be addressed in future studies to find a consistent handling of this major limitation of any MRI spine assessment.

It is possible that some therapeutic interventions may vary in their impact on different components, for

example, entheseal inflammation in the vertebral bodies versus synovial joint inflammation in the facet joints. Using this scoring system, it is possible to calculate four inflammation subscores that may predominantly represent inflammation of different spinal structures, and which may respond differently to various treatments and may have different prognostic significance.

In conclusion, a comprehensive system for the evaluation of inflammation, fat, bone erosion and new bone formation of the spine in patients with axSpA has been updated and validated. The system is designed for assessment of the individual types of MRI lesions and for acquiring total scores for the different types of lesions (inflammation, fat, erosion and new bone formation). Moderate to very good reliability for the assessment of inflammatory lesions and fat lesions for both status scores and change scores and poor to moderate reliability for bone erosion and new bone formation status scores were demonstrated. The scoring system may be used to investigate how drugs with different modes of action influence the individual components of MRI spine inflammation and damage, the mutual relationship between lesions, as well as the evolution of inflammation and damage in the entire spine.

Author affiliations
[1]Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup, Denmark
[2]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[3]Danish Hospital for Rheumatic Diseases, University Hospital of Southern Denmark, Sønderborg, Denmark
[4]Department of Regional Health Research, University Hospital of Southern Denmark, Odense, Denmark
[5]CaRE Arthritis, Edmonton, Alberta, Canada
[6]Department of Medicine, University of Alberta, Edmonton, Alberta, Canada
[7]Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, Alberta, Canada
[8]Medical Imaging Consultants, Edmonton, Alberta, Canada

Patient consent for publication Not required.

Ethics approval The study was approved by the Regional Committee on Health Research Ethics, Capital Region of Denmark, approval number: H1-2013-118. All patients provided written consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Simon Krabbe http://orcid.org/0000-0002-2877-1582
Mikkel Østergaard http://orcid.org/0000-0003-3690-467X
Susanne J Pedersen https://orcid.org/0000-0002-6500-9263
Ulrich Weber http://orcid.org/0000-0001-6701-670X
Georg Kröber https://orcid.org/0000-0001-6265-7081
Walter Makysmowych https://orcid.org/0000-0002-1291-1755
Robert G W Lambert https://orcid.org/0000-0003-0305-3290

REFERENCES
1 Lambert RGW, Pedersen SJ, Maksymowych WP, *et al*. Active Inflammatory Lesions Detected by Magnetic Resonance Imaging in the Spine of Patients with Spondyloarthritis - Definitions, Assessment System, and Reference Image Set. *J Rheumatol Suppl* 2009;84:3–17.
2 Østergaard M, Maksymowych WP, Pedersen SJ, *et al*. Structural Lesions Detected by Magnetic Resonance Imaging in the Spine of Patients with Spondyloarthritis - Definitions, Assessment System, and Reference Image Set. *J Rheumatol Suppl* 2009;84:18–34.
3 Pedersen SJ, Østergaard M, Chiowchanwisawakit P, *et al*. Validation of definitions for active inflammatory lesions detected by magnetic resonance imaging in the spine of patients with spondyloarthritis. *J Rheumatol Suppl* 2009;84:35–8.
4 Chiowchanwisawakit P, Østergaard M, Pedersen SJ, *et al*. Validation of definitions for structural lesions detected by magnetic resonance imaging in the spine of patients with spondyloarthritis. *J Rheumatol Suppl* 2009;84:39–47.
5 Maksymowych WP, Østergaard M, Chiowchanwisawakit P, *et al*. Atlas of Magnetic Resonance Imaging Abnormalities in the Spine in Spondyloarthritis: Definitions, Reliability, Training, and Conceptual Framework. A Report from the Canada (SPARCC) - Denmark International Spondyloarthritis Working Group. *J Rheumatol Suppl* 2009;84:1–2.
6. Pedersen SJ, Zhao Z, Lambert RGW, *et al*. The fat spondyloarthritis spine score (FASSS): development and validation of a new scoring method for the evaluation of fat lesions in the spine of patients with axial spondyloarthritis. *Arthritis Res Ther* 2013;15:R216.
7 Krabbe S, Sørensen IJ, Jensen B, *et al*. Inflammatory and structural changes in vertebral bodies and posterior elements of the spine in axial spondyloarthritis: construct validity, responsiveness and discriminatory ability of the anatomy-based CANDEN scoring system in a randomised placebo-controlled trial. *RMD Open* 2018;4:e000624.
8 Pedersen SJ, Weber U, Said-Nahal R, *et al*. Structural progression rate decreases over time on serial radiography and magnetic resonance imaging of sacroiliac joints and spine in a five-year follow-up study of patients with ankylosing spondylitis treated with tumour necrosis factor inhibitor. *Scand J Rheumatol* 2019;48:185–97.10.1080/03009742.2018.1506822
9 Braun J, van der Heijde D. Imaging and scoring in ankylosing spondylitis. *Best Pract Res Clin Rheumatol* 2002;16:573–604.
10 Braun J, Baraliakos X, Golder W, *et al*. Magnetic resonance imaging examinations of the spine in patients with ankylosing spondylitis, before and after successful therapy with infliximab: evaluation of a new scoring system. *Arthritis & Rheumatism* 2003;48:1126–36.

11 Haibel H, Rudwaleit M, Brandt HC, *et al*. Adalimumab reduces spinal symptoms in active ankylosing spondylitis: clinical and magnetic resonance imaging results of a fifty-two–week open-label trial. *Arthritis Rheum* 2006;54:678–81.

12 Rudwaleit M, Schwarzlose S, Hilgert ES, *et al*. MRI in predicting a major clinical response to anti-tumour necrosis factor treatment in ankylosing spondylitis. *Ann Rheum Dis* 2008;67:1276–81.

13 Song I-H, Hermann K, Haibel H, *et al*. Effects of etanercept versus sulfasalazine in early axial spondyloarthritis on active inflammatory lesions as detected by whole-body MRI (ESTHER): a 48-week randomised controlled trial. *Ann Rheum Dis* 2011;70:590–6.

14 Althoff CE, Sieper J, Song I-H, *et al*. Active inflammation and structural change in early active axial spondyloarthritis as detected by whole-body MRI. *Ann Rheum Dis* 2013;72:967–73.

15 Song I-H, Hermann KG, Haibel H, *et al*. Relationship between active inflammatory lesions in the spine and sacroiliac joints and new development of chronic lesions on whole-body MRI in early axial spondyloarthritis: results of the ESTHER trial at week 48. *Ann Rheum Dis* 2011;70:1257–63.

16 Song I-H, Hermann K-G, Haibel H, *et al*. Inflammatory and fatty lesions in the spine and sacroiliac joints on whole-body MRI in early axial spondyloarthritis—3-Year data of the ESTHER trial. *Semin Arthritis Rheum* 2016;45:404–10.

17 Maksymowych WP, Inman RD, Salonen D, *et al*. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. *Arthritis Rheum* 2005;53:502–9.

18 Maksymowych WP, Dhillon SS, Park R, *et al*. Validation of the spondyloarthritis research Consortium of Canada magnetic resonance imaging spinal inflammation index: is it necessary to score the entire spine? *Arthritis Rheum* 2007;57:501–7.

19 Maksymowych WP, Salonen D, Inman RD, *et al*. Low-Dose infliximab (3 mg/kg) significantly reduces spinal inflammation on magnetic resonance imaging in patients with ankylosing spondylitis: a randomized placebo-controlled study. *J Rheumatol* 2010;37:1728–34.

20 van der Heijde D, Baraliakos X, Gensler LS, *et al*. Efficacy and safety of filgotinib, a selective Janus kinase 1 inhibitor, in patients with active ankylosing spondylitis (TORTUGA): results from a randomised, placebo-controlled, phase 2 trial. *The Lancet* 2018;392:2378–87.

21 van der Heijde D, Cheng-Chung Wei J, Dougados M, *et al*. Ixekizumab, an interleukin-17A antagonist in the treatment of ankylosing spondylitis or radiographic axial spondyloarthritis in patients previously untreated with biological disease-modifying anti-rheumatic drugs (COAST-V): 16 week results of a phase 3 randomised, double-blind, active-controlled and placebo-controlled trial. *The Lancet* 2018;392:2441–51.

22 Maksymowych WP, Crowther SM, Dhillon SS, *et al*. Systematic assessment of inflammation by magnetic resonance imaging in the posterior elements of the spine in ankylosing spondylitis. *Arthritis Care Res* 2010;62:4–10.

23 Marzo-Ortega H, McGonagle D, O'Connor P, *et al*. Efficacy of etanercept in the treatment of the entheseal pathology in resistant spondyloarthropathy. *Arthritis Rheum* 2001;44:2112–7.

24 Marzo-Ortega H, McGonagle D, O'Connor P, *et al*. Baseline and 1-year magnetic resonance imaging of the sacroiliac joint and lumbar spine in very early inflammatory back pain. Relationship between symptoms, HLA-B27 and disease extent and persistence. *Ann Rheum Dis* 2009;68:1721–7.

25 Madsen KB, Jurik AG. MRI grading method for active and chronic spinal changes in spondyloarthritis. *Clin Radiol* 2010;65:6–14.

26 Lukas C, Braun J, van der Heijde D, *et al*. Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: a multireader experiment. *J Rheumatol* 2007;34:862–70.

27 Maksymowych WP, Chiowchanwisawakit P, Clare T, *et al*. Inflammatory lesions of the spine on magnetic resonance imaging predict the development of new syndesmophytes in ankylosing spondylitis: evidence of a relationship between inflammation and new bone formation. *Arthritis Rheum* 2009;60:93–102.

28 Pedersen SJ, Chiowchanwisawakit P, Lambert RGW, *et al*. Resolution of inflammation following treatment of ankylosing spondylitis is associated with new bone formation. *J Rheumatol* 2011;38:1349–54.

29 Chiowchanwisawakit P, Lambert RGW, Conner-Spady B, *et al*. Focal fat lesions at vertebral corners on magnetic resonance imaging predict the development of new syndesmophytes in ankylosing spondylitis. *Arthritis & Rheumatism* 2011;63:2215–25.

30 Baraliakos X, Listing J, Rudwaleit M, *et al*. The relationship between inflammation and new bone formation in patients with ankylosing spondylitis. *Arthritis Res Ther* 2008;10.

31 van der Heijde D, Machado P, Braun J, *et al*. MRI inflammation at the vertebral unit only marginally predicts new syndesmophyte formation: a multilevel analysis in patients with ankylosing spondylitis. *Ann Rheum Dis* 2012;71:369–73.

32 Maksymowych WP, Morency N, Conner-Spady B, *et al*. Suppression of inflammation and effects on new bone formation in ankylosing spondylitis: evidence for a window of opportunity in disease modification. *Ann Rheum Dis* 2013;72:23–8.

33 Pfirrmann CWA, Metzdorf A, Zanetti M, *et al*. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine* 2001;26:1873–8.