Seminars in Arthritis and Rheumatism 79 (2026) 153013

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Structural damage assessment in the spine of patients with axial spondyloarthritis – results from an international OMERACT multi-reader exercise using MRI-based synthetic CT

Simone T. Willesen [a,b,*], Robert G.W. Lambert [c], Walter P. Maksymowych [d], Susanne J. Pedersen [a,b], Jakob M. Møller [e], Kasper K. Gosvig [e], Charles Peterfy [f], Paul Bird [g], Maria S. Stoenoiu [h], Stylianos Georgiadis [a], Mikkel Østergaard [a,b]

[a] Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Rigshospitalet, Glostrup, Denmark
[b] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[c] Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, Alberta, Canada
[d] Department of Medicine, University of Alberta, Edmonton, Canada
[e] Department of Radiology, Copenhagen University Hospital at Herlev Gentofte, Copenhagen, Denmark
[f] Spire Sciences, Inc., Florida, USA
[g] University of New South Wales, Sydney, Australia
[h] Department of Rheumatology, Cliniques Universitaires Saint-Luc, Institut de Recherche Expérimentale et Clinique, Université Catholique de Louvain, Brussels, Belgium

ARTICLE INFO

Keywords:
OMERACT
Axial spondyloarthritis
Outcome measure
Magnetic resonance imaging
Computed tomography
MRI-based synthetic CT
Structural damage

ABSTRACT

Objective: To develop and perform preliminary cross-sectional validation of a magnetic resonance imaging (MRI)-based outcome measure for assessing spinal structural damage in spondyloarthritis clinical trials, with a specific focus on MRI-based synthetic computed tomography (sCT) and related MRI-based techniques.
Methods: Consensus meetings within the Outcome Measures in Rheumatology (OMERACT) MRI in arthritis working group established consensus definitions for spinal new bone formation and scoring rules. Then, low-dose CT and sCT images of twenty patients with axial spondyloarthritis and five healthy controls were independently assessed by seven experienced readers using the agreed-upon definitions for: marginal syndesmophytes, non-marginal syndesmophytes, and osteophytes. sCT images were reconstructed using a deep learning algorithm (BoneMRI v1.8, MRIGuidance B.V., Utrecht, the Netherlands). Sensitivity and specificity of synthetic CT were calculated using low-dose CT as the reference standard, and inter-reader reliability was assessed.
Results: A total mean of 35.5 lesions was scored on both sCT and low-dose CT in all participants. sCT demonstrated an overall good sensitivity (0.77) and excellent specificity ($\geq$ 0.94) with low-dose CT as reference standard. Inter-reader agreement was substantial for both low-dose CT and sCT, with an overall intra-class coefficient of 0.80 (low-dose CT) and 0.86 (sCT), and corresponding kappa values of 0.68 and 0.70, respectively.
Conclusion: This OMERACT international multi-reader exercise established consensus definitions for spinal new bone formation and provided preliminary evidence that sCT meets key OMERACT criteria of domain match and feasibility. This MRI technique for generating CT-like images shows promise as a novel method for assessing spinal structural damage in spondyloarthritis clinical trials.

Introduction

Axial spondyloarthritis (axSpA) is a chronic inflammatory rheumatic disease in which inflammation, pain, and structural spinal damage, particularly in terms of new bone formation (NBF), contribute to long-term and irreversible disability [1]. Preventing structural damage – part of the Outcome Measures in Rheumatology (OMERACT) core domain set for axSpA [2] – and progression therein is a major

* Corresponding author at: Copenhagen Center for Arthritis Research (COPECARE), Center for Rheumatology and Spine Diseases, Copenhagen University Hospital Rigshospitalet, Glostrup, Mailing Address: Center for Rheumatology and Spine Diseases, Copenhagen University Hospital Rigshospitalet – Glostrup, Valdemar Hansens Vej 17, entrance 5, DK-2600 Glostrup, Denmark.
E-mail address: simone.willesen@regionh.dk (S.T. Willesen).

https://doi.org/10.1016/j.semarthrit.2026.153013

Available online 4 June 2026

therapeutic goal; however, the ability to assess structure-modifying effects of therapies remains challenging [3].

In the spine, structural damage (the broad domain [4]) is primarily characterized by pathological new bone formation (the target domain); however, destructive lesions such as erosions can also occur [1]. The main form of spinal new bone formation in axSpA is the development of syndesmophytes (the domain components), which can eventually progress into ankylosis [5]. Radiography using the modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS) [6] is currently the only validated method for assessing spinal structural progression in axSpA [7] and is the preferred instrument for assessment of spinal damage by the Assessment of SpondyloArthritis international Society (ASAS)/OMERACT [8]. However, radiography's low sensitivity to change requires follow-up periods of at least two years [9,10], making it unsuitable for randomized controlled clinical trials.

Magnetic resonance imaging (MRI) is the gold standard for detecting inflammation, but it falls short in reliably detecting smaller bone structures in the spine, such as subtle new bone formation (Fig. 1) [11]. Computed tomography (CT), in contrast, provides excellent detail and is considered the gold standard for bone visualization [12]; however, it is not routinely used in longitudinal studies due to substantial radiation exposure [13].

Adding to these challenges, definitions for different types of spinal new bone formation (NBF) are not consistently applied [14]. Existing studies using radiography to investigate new bone formation in the spine often apply heterogeneous criteria, as reported by Gazel *et al.* [15], making it difficult to compare and distinguish findings across diseases or even within axSpA cohorts.

Together, these limitations underline an unmet need for consensus NBF definitions and, most importantly, developing new, alternative methods for sensitive assessment of structural damage in the spine of patients with spondyloarthritis. One promising approach is MRI-based "synthetic CT" (sCT). This technique generates CT-like images without ionizing radiation, enabling high-resolution assessment of bone

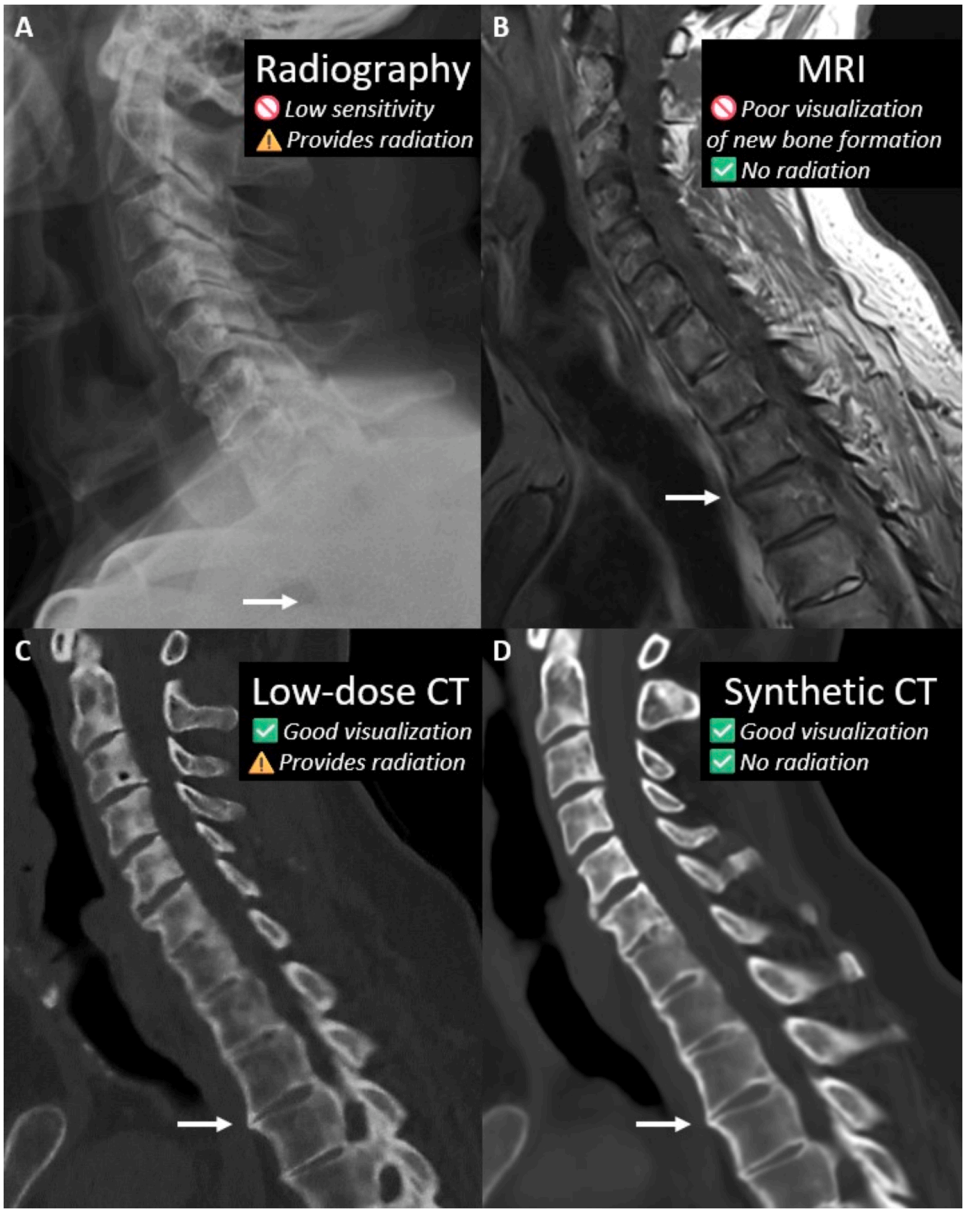

Fig. 1. Radiography (A), magnetic resonance imaging (MRI, B), low-dose computed tomography (CT, C), and MRI-based synthetic CT (D) of the cervical and upper thoracic spine in a patient with axial spondyloarthritis.
Arrows indicate new bone formation at the Th2-Th3 level, clearly visible on both low-dose CT and MRI-based synthetic CT, but not on radiography or MRI.

structure in the spine [16–19] and various joints [16,19–24], and has been shown to be better than radiography in detecting new bone formation in patients with axSpA, with low-dose CT (ldCT) as the reference standard [25].

Against this background, the OMERACT MRI in Arthritis Working Group aims to establish consensus definitions for different types of spinal new bone formation and to develop and validate an MRI-based outcome measure for structural damage in spondyloarthritis clinical trials, with a particular focus on sCT and similar MRI-based techniques [26,27]. Such an instrument is essential for evaluating the structural efficacy of established and emerging therapies in clinical trials. Here we present the results of an international multi-reader exercise aimed at preliminary domain match and feasibility of sCT according to the OMERACT Filter Instrument Selection Algorithm (OFISA) 2.2 framework [4].

Methods

Design

Consensus meetings were held within the OMERACT MRI in arthritis working group, during which definitions of NBF pathologies to score were agreed upon, along with reader rules. Published definitions of different types of NBF served as the starting point for the discussions. Group participants were invited to provide comments and suggestions for modifications either orally or in writing, and final decisions were reached by consensus. LdCT and sCT images of twenty patients with axial spondyloarthritis and five healthy controls were independently assessed by seven experienced readers using the consensus NBF definitions for: 1) marginal syndesmophytes, 2) non-marginal syndesmophytes, and 3) osteophytes (see section 2.3, 'consensus definitions'). The lesions to be scored (both bridging and non-bridging) were assessed on sagittal spinal images at each disco-vertebral unit (23 units in total). If a bridging lesion was scored, non-bridging lesions could not be scored, and vice versa. The experienced readers included three musculoskeletal radiologists, three rheumatologists, and one senior scientific radiographer, all with extensive training in reading MR spine images.

Image acquisition

Low-dose CT examination of the spine (cervical, thoracic and lumbar spine) was performed in a photon-counting CT system (Siemens Naeotom Alpha; Siemens Healthineers, Erlangen, Germany). CT examination z-axis coverage was from 1 cm above the dens through the pelvis, with scanning performed in the craniocaudal direction. CT images were reconstructed in the axial plane at 1 mm thickness with 1 mm increments, and in the sagittal and coronal planes at 3 mm slice thickness with 3 mm increments, and subsequently reformatted with 1 mm slice thickness. Automated tube current modulation was used for all examinations, and the kilovoltage peak was set at Sn100 kVp. The scan field of view was 50 cm, and the displayed field of view was set for the patient's body size to include the entire spine. Quantum iteration level 3 was used, the pitch was 0.5, collimation 144 × 0.4, and rotation time 0.5 s. The requested CT-dose index was 1.36 mGy. The mean effective radiation dose was 1.52 mSv (SD 0.27 mSv). Images were reconstructed with a Br64 kernel.

MRI was performed with a 3T system (Vida, Siemens Healthineers, Erlangen, Germany) using a 32-channel spine coil combined with an 18-channel flex covering the pelvis. A 3D dual gradient recalled echo sequence was performed in three sagittal stations covering the spine. Slice thickness was 0.8 mm and in-plane resolution was 0.6 × 0.6 mm. Repetition time, echo times, flip angle, and acquisition time were 7 ms, 2 and 3.5 ms, 10 degrees, and 3:56 min. Synthetic CT was reconstructed using the BoneMRI model v1.8 (MRIGuidance BV, Utrecht, NL). The technique is based on a dual-echo gradient sequence (as described above) and a machine-learning processing pipeline [16].

Consensus definitions

Consensus definitions of three types of pathologies (NBF) to be scored were agreed upon (Figs. 2 and 3):

1. Marginal syndesmophyte: Slender NBF originating from the margin (rim) of the vertebral endplate, with a more vertical orientation at origin.
2. Non-marginal syndesmophyte: NBF that arises away from the margin (rim) of the vertebral endplate, with a more vertical orientation at origin.
3. Osteophyte: NBF originating from the margin (rim) of the vertebral endplate with a more horizontal orientation at origin.

"More vertical" indicates more vertical than 45 degrees compared with the endplate of the vertebra, whereas "more horizontal" indicates more horizontal than 45 degrees compared with the endplate of the vertebra.

Scoring rules

Hierarchy of assessment

When a pathology did not entirely match the predefined categories, the following decision rules were applied in order:

1. Presence or absence of NBF.
   Any NBF should be identified and scored. If a lesion is visible, it must be assigned to one of the definitions above, even if the fit is not perfect.
2. Orientation ("more vertical" vs. "more horizontal")
   When NBF is identified, it should first be classified based on whether its growth pattern appeared predominantly upward/downward (vertical) or sideways (horizontal).

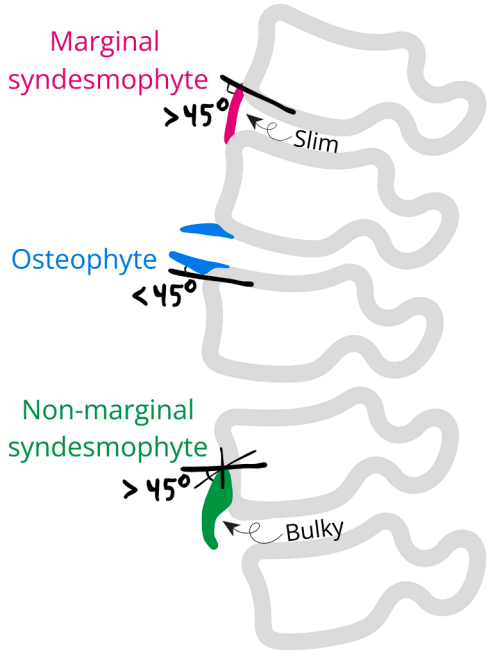

Fig. 2. Schematic drawing of the three defined types of new bone formation (NBF). Pink: Marginal syndesmophyte (bridging) – a slender NBF arising from the margin of the vertebra with a vertical orientation.
Blue: Osteophyte - NBF originating from the vertebral margin with a horizontal direction.
Green: Non-marginal syndesmophyte – a bulky NBF arising away from the margin of the vertebrae with a vertical orientation at origin (> 45 degrees compared with the endplate of the vertebra).

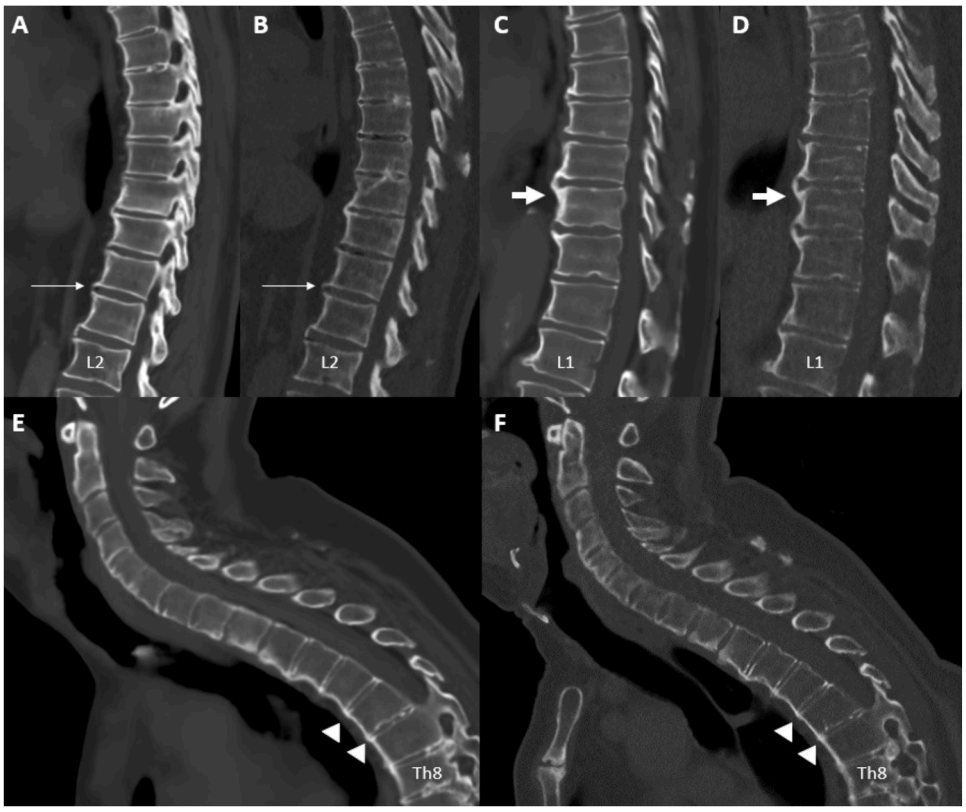

Fig. 3. MRI-based synthetic computed tomography (A, C, E) and low-dose computed tomography (B, D, F) of the thoracolumbar spine (A+B, C+D) and cervical and part of the thoracic spine (E+F) in three patients with axial spondyloarthritis.
Long thin arrows: New bone formation originating from the margin of the vertebra with a horizontal direction at the origin – scored as an osteophyte.
Short arrow: Bulky new bone formation that arises away from the margin (rim) of the vertebral endplate, with a more vertical (> 45 degrees compared with the endplate of the vertebra) orientation at origin – scored as a non-marginal syndesmophyte.
Arrowheads: Slender new bone formation originating from the margin of the vertebral endplate, with a vertical orientation at origin – scored as bridging syndesmophytes.

3. Location (marginal vs. non-marginal)

Lesions arising directly from the vertebral margin should be considered *marginal*, whereas those located away from the margin (on the vertebral surface) should be considered *non-marginal*.

4. Morphology (slender vs. bulky)

Finally, the overall shape of the NBF should be considered (Fig. 3).

Example: a thin, finger-like projection = slender; a thick, broad outgrowth = bulky.

Other rules

5. Two different types of new bone formation, e.g., an osteophyte and a non-marginal syndesmophyte, should not be scored in the same location. If there is both an osteophyte and a syndesmophyte at a location (e.g., in different slices, then the lesion should be scored as a syndesmophyte. If there is both a marginal and a non-marginal syndesmophyte at a location, then it should be scored as a marginal syndesmophyte.
6. To avoid overscoring, a phyte (i.e., a marginal- or non-marginal syndesmophyte or an osteophyte) should only be scored when it is considered definite.
7. Regarding the determination of "more vertical versus more horizontal orientation" in phytes that start away from the margin, then the orientation (</> 45 degrees) at the part that originates away from the margin should be considered, rather than the part that starts at the margin.

Slices/areas to score

The rules below were implemented to avoid the extreme/very lateral slices.

8. All or most of the full anterior-posterior (AP) diameter of the vertebra must be visible (75–100% as a guide, but should not be measured).
9. For the cervical spine: the spinal canal must be visible.
10. For thoracic and lumbar spine: the spinal canal or the medial half of the pedicle must be visible.

Statistical analyses

For each imaging reader, the sum of scored NBF was calculated across spinal segments (cervical, thoracic, and lumbar spinal), and an overall mean (range) per patient was calculated for all participants and the two groups (axSpA and healthy controls) separately. The sensitivity and specificity of sCT were calculated with ldCT as the reference standard. Sensitivity and specificity were calculated for the detection of any type of NBF, as well as separately for marginal syndesmophytes, non-marginal syndesmophytes, and osteophytes.

Inter-reader reliability was assessed using Cohen's Kappa for individual locations (i.e., on "lesion level") and a two-way random effects single-measure model of intraclass correlation coefficient (ICC) based on absolute agreement for sum scores on a patient level across all readers (i.e., on patient level). The following thresholds were used to interpret

Table 1
Participant characteristics.

Characteristics	All participants	Axial spondyloarthritis	Healthy control
n (%)	25 (100%)	20 (80%)	5 (20%)
Sex, female n (%)	9 (36%)	7 (35%)	2 (40%)
Age (years), mean (±SD)	50.0 (14.6)	50.8 (14.6)	46.6 (15.5)
ASAS classification criteria positivity, n (%)	20 (80%)	20 (100%)	0 (0%)
Modified New York Criteria positivity, n (%)	16 (64%)	16 (80%)	0 (0%)
Symptom duration (years), mean (±SD)	23.0 (16.2)	23.0 (16.2)	NA
Time since diagnosis (years), mean (±SD)	9.8 (11.2)	9.8 (11.2)	NA
C-reactive protein (mg/mL), mean (±SD)	8.7 (13.7)	10.1 (15.0)	2.8 (1.7)
HLA-B27 positivity, n (%)	17 (68%)	16 (80%)	1 (20%)
Inflammatory back pain, n (%)	15 (60%)	15 (75%)	0 (0%)
BASDAI (VAS 0–100), mean (±SD)	30.8 (21.2)	37.2 (18.5)	5.1 (6.3)
BASFI, mean (±SD)	25.5 (18.6)	31.5 (15.5)	1.1 (2.3)

NA Not applicable
ASAS Assessment of SpondyloArthritis international Society, *HLA-B27* Human Leukocyte Antigen B27, *BASDAI* Bath Ankylosing Spondylitis Disease Activity Index, *BASFI* Bath Ankylosing Spondylitis Functional Index.

agreement: Kappa values [28]: 0–0.20: poor; 0.21–0.40: fair; 0.41–0.60: moderate; 0.61–0.80: substantial; 0.81–1.00: excellent. ICCs [29]: < 0.40 poor; 0.40–0.59: fair; 0.60–0.74 good; 0.75–1.00 excellent. Statistical analyses were performed using R version 4.4.1 by an experienced statistician.

Table 2
Occurrence (mean per patient), sensitivity, and specificity of marginal syndesmophytes, non-marginal syndesmophytes, and osteophytes, as detected by MRI-based synthetic CT and low-dose CT in the spine in patients with axial spondyloarthritis, and healthy controls.

All readers	All participants, n = 25 **Total** cerv/thor/lumb	Axial spondyloarthritis, n = 20 **Total** cerv/thor/lumb	Healthy control, n = 5 **Total** cerv/thor/lumb
Number of different types of NBF scored on low-dose CT (means per patient)			
Marginal syndesmophytes	**26.6** 5.4/15.4/5.8	**31.9** 6.6/18.5/6.8	**5.2** 0.8/2.9/1.5
Non-marginal syndesmophytes	**1.9** 0.3/1.0/0.6	**2.1** 0.4/1.0/0.7	**1.3** 0.3/0.9/0.1
Osteophytes	**7.0** 1.1/4.6/1.2	**7.1** 1.3/4.6/1.2	**6.3** 0.5/4.7/1.1
Marginal syndesmophyte or non-marginal syndesmophyte	**28.5** 5.8/16.4/6.4	**34** 6.9/19.5/7.6	**6.5** 1.1/3.9/1.6
Any type of new bone formation	**35.5** 6.9/21.0/7.6	**41.1** 8.2/24.1/8.8	**12.9** 1.6/8.6/2.6
Number of different types of NBF scored on synthetic CT (means per patient)			
Marginal syndesmophytes	**28.9** 6.5/16.9/5.5	**35.0** 7.9/20.5/6.6	**4.7** 1.0/2.9/0.8
Non-marginal syndesmophytes	**1.6** 0.3/0.6/0.7	**1.8** 0.3/0.6/0.9	**1.1** 0.1/0.8/0.2
Osteophytes	**4.9** 0.8/2.9/1.2	**5.1** 1.0/2.9/1.2	**4.2** 0.3/2.7/1.2
Marginal syndesmophyte or non-marginal syndesmophyte	**30.6** 6.8/17.6/6.2	**36.7** 8.2/21.0/7.5	**5.8** 1.1/3.7/0.9
Any type of new bone formation	**35.5** 7.6/20.5/7.4	**41.8** 9.2/24.0/8.7	**10.0** 1.4/6.5/2.2
Sensitivity of synthetic CT with low-dose CT as the reference standard			
Marginal syndesmophytes	**0.78** 0.81/0.79/0.75	**0.8** 0.83/0.80/0.77	**0.38** 0.16/0.41/0.51
Non-marginal syndesmophytes	**0.29** 0.20/0.19/0.41	**0.29** 0.22/0.17/0.41	**0.21** 0.12/0.20/0.14
Osteophytes	**0.43** 0.37/0.41/0.57	**0.43** 0.38/0.41/0.56	**0.41** 0.16/0.38/0.57
Marginal syndesmophyte or non-marginal syndesmophyte	**0.79** 0.80/0.78/0.81	**0.81** 0.82/0.80/0.83	**0.45** 0.23/0.51/0.55
Any type of new bone formation	**0.77** 0.76/0.75/0.83	**0.79** 0.79/0.77/0.84	**0.58** 0.28/0.61/0.67
Specificity of synthetic CT with low-dose CT as the reference standard			
Marginal syndesmophytes	**0.95** 0.95/0.94/0.98	**0.94** 0.95/0.93/0.97	**0.98** 0.98/0.98/1.00
Non-marginal syndesmophytes	**0.99** 0.99/0.99/0.98	**0.99** 0.99/0.99/0.98	**0.99** 1.00/0.99/1.00
Osteophytes	**0.98** 0.99/0.98/0.98	**0.98** 0.98/0.98/0.98	**0.98** 1.00/0.98/0.98
Marginal syndesmophyte or non-marginal syndesmophyte	**0.95** 0.95/0.94/0.98	**0.94** 0.95/0.93/0.97	**0.98** 0.98/0.97/0.99
Any type of new bone formation	**0.95** 0.95/0.94/0.97	**0.94** 0.94/0.93/0.97	**0.98** 0.98/0.98/0.99

CT computed tomography, *synthetic CT* MRI-based synthetic computed tomography, *cerv* cervical spinal segment, *thor* thoracic spinal segment, *lumb* lumbar spinal segment.

Results

Participants

Participant characteristics are summarized in Table 1. Twenty-five participants were included: 20 with axSpA and five healthy controls. In the axSpA group, the mean age was 50.8 ± 14.6 years (standard deviation), 80% fulfilled the modified New York criteria, the mean symptom duration was 23 ± 16.2 years, and the mean time since diagnosis was 9.8 ± 11.2 years. 80% of participants with axSpA and 20% of healthy controls were HLA-B27 positive.

Occurrence of different types of new bone formation

On ldCT, a mean of 26.6 marginal syndesmophytes were scored per participant (range 0–100.6), compared with 28.9 on sCT (range 0–111.0) (Table 2). The mean number of non-marginal syndesmophytes scored was 1.9 on ldCT (range 0–10.1) and 1.6 on sCT (range 0–7.3). For osteophytes, slightly more were scored on ldCT (mean 7.0, range 0–19.7) than on sCT (mean 4.9, range 0–18.4). When considering any type of new bone formation together, the total mean was 35.5 lesions per participant on both ldCT and sCT, with ranges 0.1–103.9 and 0–116.4, respectively.

Sensitivity and specificity

For the detection of any type of NBF in all participants, sCT achieved an overall sensitivity of 0.77 (Table 2). The sensitivity for marginal syndesmophytes was 0.78 across all participants and was higher in the axSpA group compared with controls. The sensitivity of sCT for detecting non-marginal syndesmophytes was 0.29, and 0.43 for detecting osteophytes. Across all lesion types and groups, the specificity of sCT was excellent (≥0.94).

Inter-reader reliability

For any type of NBF in all participants, the ICC for all readers (Table 3) was 0.80 for ldCT and 0.86 for sCT. Agreement was highest for the detection of marginal syndesmophytes (0.80 for ldCT and 0.85 for sCT). ICCs were overall higher in the axSpA group than in the control group.

Cohen's kappa values (Table 3) followed a similar pattern. For any

Table 3
Inter-reader reliability (intraclass correlation coefficient, ICC, and Kappa) for all readers stratified by spinal segment.

All readers ICC	All participants, n = 25 **Total** cerv/thor/lumb	Axial spondyloarthritis, n = 20 **Total** cerv/thor/lumb	Healthy controls, n = 5 **Total** cerv/thor/lumb
Low-dose CT			
Marginal syndesmophytes	**0.80** 0.83/0.75/0.84	**0.78** 0.83/0.73/0.84	**0.48** 0.28/0.45/0.60
Non-marginal syndesmophytes	**0.34** 0.21/0.28/0.45	**0.35** 0.21/0.27/0.44	**0.33** 0.13/0.35/0.13
Osteophytes	**0.67** 0.63/0.67/0.63	**0.69** 0.65/0.69/0.67	**0.58** 0.08/0.56/0.48
Marginal syndesmophytes or non-marginal syndesmophyte	**0.82** 0.84/0.77/0.86	**0.80** 0.84/0.75/0.86	**0.58** 0.34/0.62/0.60
Any type of new bone formation	**0.80** 0.84/0.75/0.85	**0.78** 0.83/0.73/0.85	**0.77** 0.41/0.81/0.79
Synthetic CT			
Marginal syndesmophytes	**0.85** 0.86/0.82/0.89	**0.84** 0.86/0.81/0.89	**0.48** 0.36/0.44/0.40
Non-marginal syndesmophytes	**0.28** 0.22/0.17/0.40	**0.32** 0.22/0.23/0.39	**0.07** 0.10/0.06/0.18
Osteophytes	**0.61** 0.65/0.54/0.63	**0.63** 0.67/0.55/0.65	**0.48** 0.04/0.40/0.59
Marginal syndesmophytes or non-marginal syndesmophyte	**0.86** 0.87/0.82/0.91	**0.85** 0.86/0.80/0.91	**0.52** 0.42/0.49/0.46
Any type of new bone formation	**0.86** 0.87/0.82/0.91	**0.85** 0.86/0.80/0.91	**0.78** 0.56/0.79/0.75
Kappa			
Low-dose CT			
Marginal syndesmophytes	**0.59** 0.62/0.55/0.66	**0.60** 0.63/0.55/0.66	**0.38** 0.20/0.38/0.51
Non-marginal syndesmophytes	**0.23** 0.18/0.18/0.33	**0.24** 0.19/0.18/0.34	**0.16** 0.14/0.16/0.16
Osteophytes	**0.54** 0.38/0.59/0.48	**0.55** 0.40/0.60/0.51	**0.48** 0.12/0.54/0.31
Marginal syndesmophytes or non-marginal syndesmophyte	**0.63** 0.64/0.58/0.72	**0.63** 0.65/0.57/0.73	**0.48** 0.25/0.53/0.52
Any type of new bone formation	**0.68** 0.66/0.65/0.73	**0.67** 0.67/0.64/0.74	**0.66** 0.38/0.72/0.63
Synthetic CT			
Marginal syndesmophytes	**0.66** 0.65/0.64/0.68	**0.66** 0.67/0.64/0.68	**0.33** 0.12/0.40/0.32
Non-marginal syndesmophytes	**0.21** 0.14/0.12/0.31	**0.24** 0.16/0.14/0.32	**0.04** 0.00/0.05/0.05
Osteophytes	**0.45** 0.42/0.45/0.45	**0.45** 0.45/0.45/0.44	**0.43** 0.06/0.44/0.47
Marginal syndesmophytes or non-marginal syndesmophyte	**0.67** 0.67/0.63/0.74	**0.67** 0.68/0.63/0.75	**0.40** 0.21/0.45/0.39
Any type of new bone formation	**0.70** 0.67/0.67/0.77	**0.69** 0.68/0.66/0.78	**0.63** 0.29/0.71/0.60

ICC intraclass correlation coefficient, *CT* computed tomography, *synthetic CT* MRI-based synthetic computed tomography, *cerv* cervical spinal segment, *thor* thoracic spinal segment, *lumb* lumbar spinal segment.

type of NBF in all participants, kappas were 0.68 for ldCT and 0.70 for sCT. As with ICCs, the kappa values were highest for marginal syndesmophytes (0.59 and 0.66, respectively), and slightly superior in the axSpA group compared with controls.

Discussion

In this first international multi-reader exercise, sCT demonstrated good sensitivity and excellent specificity versus ldCT for detecting new bone formation in the spine of patients with axSpA. Its performance in identifying new bone formation was comparable in numbers to that of ldCT, which is considered the gold standard for assessing structural damage. These findings support the large potential of sCT as a radiation-free structural imaging outcome measure in clinical trials. In addition, the MRI in arthritis working group reached agreement on consensus definitions of new bone formation, which contributed to very good inter-reader reliability for both overall new bone formation and marginal syndesmophytes.

When developing a new imaging instrument under the OFISA 2.2 framework, the first two steps require evidence of domain match, or truth, showing that the instrument accurately measures the target domain – in this case, new bone formation [2] – and feasibility, demonstrating that the method is practical to use [4]. To our knowledge, this is the first study to provide preliminary evidence for these two criteria in evaluating sCT as a potential new core imaging outcome measure in axSpA. For domain match, sCT demonstrated good sensitivity and specificity for detecting new bone formation, a key feature of structural damage in axSpA, when using ldCT as the reference standard. The substantial to excellent inter-reader reliability demonstrates that different readers consistently agreed on their assessments, thus indicating good content validity, and underlining the method's reproducibility for use in future clinical trials. Regarding feasibility, sCT is generated from a standard MRI sequence and adds only eight minutes of scan time. When acquired in addition to standard MRI sequences, it also offers a valuable "two-in-one" advantage: visualization of inflammation, and thus disease activity (another core domain in axSpA), together with structural damage assessment, all without radiation exposure. These features support feasibility in terms of time and operational practicality. Furthermore, sCT captures the entire spine, including the thoracic segment, which is poorly visualized by radiography due to superimposition. This represents another advantage of sCT, as the thoracic spine was the spinal segment most frequently affected by all types of NBF.

A major strength of this first sCT reading exercise in the MRI in arthritis working group is its international multi-reader design, which enhances generalizability and reduces the risks of local reading bias. The direct head-to-head comparison with ldCT provides a rigorous benchmark for diagnostic performance. The main limitation is the relatively small sample size and cross-sectional study design. Both the sensitivity of sCT compared with ldCT and the inter-reader reliability for detecting non-marginal syndesmophytes were low. A plausible explanation is that the presence of non-marginal syndesmophytes among the participants was relatively scarce, making the reliability highly susceptible to small discrepancies between modalities (affecting sensitivity) and readers. This effect would likely have been reduced if a larger number of such lesions had been present. In contrast, marginal syndesmophytes represented the most abundant type of new bone formation in the cohort, and both the sensitivity of sCT for detecting these lesions and the inter-reader reliability were good to excellent, further supporting this explanation. 80% of the axSpA population fulfilled the modified New York criteria, increasing the likelihood of detecting spinal NBF. However, the predominance of patients with advanced disease may limit the generalizability of the findings to patients with early disease, who are often the primary target population in interventional trials aiming to prevent or reduce the development of NBF. Additional studies are needed to confirm the construct validity and to evaluate the intrareader reliability and ability of sCT and similar methods to detect structural progression over time (responsiveness) and to differentiate between treatment groups (discrimination) [30]. The next steps may enable this new imaging method and related scoring procedure (in combination "the imaging instrument" [31]) to progress further toward endorsement as an OMERACT MRI outcome measure, following the framework outlined in OFISA 2.2 [4]. In conclusion, this OMERACT MRI in arthritis multi-reader exercise established consensus definitions for spinal new bone formation and demonstrated preliminary evidence that sCT meets key OMERACT criteria of domain match and feasibility. By generating CT-like images from MRI, sCT provides a practical, radiation-free tool for assessing structural damage in axSpA. Confirmation in larger longitudinal studies will be essential to validate and ultimately endorse this new approach as an imaging outcome measure in spondyloarthritis clinical trials.

Funding

This work was supported by the Danish Rheumatism Association (the Salary of STW, grant number R230-A8503).

Data statement

Data are available upon reasonable request.

CRediT authorship contribution statement

Simone T. Willesen: Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Robert G.W. Lambert:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Walter P. Maksymowych:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Susanne J. Pedersen:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Jakob M. Møller:** Writing – review & editing, Visualization, Resources, Methodology, Investigation, Data curation, Conceptualization. **Kasper K. Gosvig:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Charles Peterfy:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Paul Bird:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Maria S. Stoenoiu:** Writing – review & editing, Visualization, Methodology, Investigation, Data curation. **Stylianos Georgiadis:** Writing – review & editing, Visualization, Validation, Software, Methodology, Formal analysis, Data curation. **Mikkel Østergaard:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

Simone Tromborg Willesen reports financial support was provided by The Danish Rheumatism Association. Mikkel Østergaard reports a relationship with AbbVie that includes: consulting or advisory, funding grants, and speaking and lecture fees. Mikkel Østergaard reports a relationship with BMS that includes: consulting or advisory, funding grants, and speaking and lecture fees. Mikkel Østergaard reports a relationship with Merck that includes: consulting or advisory, funding grants, and speaking and lecture fees. Mikkel Østergaard reports a relationship with Novartis that includes: consulting or advisory, funding grants, and speaking and lecture fees. Mikkel Østergaard reports a relationship with Eli Lilly and Company that includes: consulting or advisory, funding grants, and speaking and lecture fees. Mikkel Østergaard reports a relationship with UCB that includes: consulting or advisory, funding grants, and speaking and lecture fees. Mikkel

Østergaard reports a relationship with Celgene that includes: consulting or advisory and speaking and lecture fees. Mikkel Østergaard reports a relationship with Galapagos that includes: consulting or advisory and speaking and lecture fees. Mikkel Østergaard reports a relationship with Gilead that includes: consulting or advisory and speaking and lecture fees. Mikkel Østergaard reports a relationship with Janssen that includes: consulting or advisory and speaking and lecture fees. Mikkel Østergaard reports a relationship with MEDAC that includes: consulting or advisory and speaking and lecture fees. Mikkel Østergaard reports a relationship with Pfizer that includes: consulting or advisory and speaking and lecture fees. Mikkel Østergaard reports a relationship with Sandoz that includes: consulting or advisory and speaking and lecture fees. Susanne Juhl Pedersen reports a relationship with Innovation Fund Denmark that includes: funding grants. Susanne Juhl Pedersen reports a relationship with Nordic Bioscience that includes: funding grants. Susanne Juhl Pedersen reports a relationship with AbbVie that includes: consulting or advisory and speaking and lecture fees. Susanne Juhl Pedersen reports a relationship with Novartis that includes: consulting or advisory and speaking and lecture fees. Susanne Juhl Pedersen reports a relationship with MSD that includes: consulting or advisory and speaking and lecture fees. Susanne Juhl Pedersen reports a relationship with Pfizer that includes: consulting or advisory and speaking and lecture fees. Susanne Juhl Pedersen reports a relationship with UCB that includes: consulting or advisory and speaking and lecture fees. Stylianos Georgiadis reports a relationship with Novartis that includes: funding grants. Stylianos Georgiadis reports a relationship with UCB that includes: funding grants. Robert G. W. Lambert reports a relationship with CARE arthritis that includes: consulting or advisory. Robert G. W. Lambert reports a relationship with Image Analysis Group that includes: consulting or advisory. Robert G. W. Lambert reports a relationship with AbbVie that includes: speaking and lecture fees. Walter Maksymowych reports a relationship with CARE arthritis that includes: consulting or advisory. Paul Bird reports a relationship with Pfizer that includes: consulting or advisory and speaking and lecture fees. Paul Bird reports a relationship with Eli Lilly and Company that includes: consulting or advisory and speaking and lecture fees. Paul Bird reports a relationship with Novartis that includes: consulting or advisory and speaking and lecture fees. Paul Bird reports a relationship with BMS that includes: consulting or advisory and speaking and lecture fees. Paul Bird reports a relationship with Celgene that includes: consulting or advisory and speaking and lecture fees. Charles Peterfy reports a relationship with AbbVie that includes: funding grants. Charles Peterfy reports a relationship with AstraZeneca that includes: funding grants. Charles Peterfy reports a relationship with Eli Lilly and Company that includes: funding grants. Charles Peterfy reports a relationship with SetPoint that includes: funding grants. Charles Peterfy reports a relationship with Sonoma that includes: funding grants. Charles Peterfy reports a relationship with AmMax that includes: funding grants. Charles Peterfy reports a relationship with Abbisko that includes: funding grants. Charles Peterfy reports a relationship with Daiichi Sankyo Inc that includes: funding grants. Charles Peterfy reports a relationship with Deciphera that includes: funding grants. Charles Peterfy reports a relationship with Genascense that includes: funding grants. Charles Peterfy reports a relationship with Hutchmed that includes: funding grants. Charles Peterfy reports a relationship with Labcorp that includes: funding grants. Charles Peterfy reports a relationship with OrthoTrophix Inc that includes: funding grants. Charles Peterfy reports a relationship with Pacira that includes: funding grants. Charles Peterfy reports a relationship with Paradigm that includes: funding grants. Charles Peterfy reports a relationship with Sarepta that includes: funding grants. Charles Peterfy reports a relationship with SynOx that includes: funding grants. Charles Peterfy reports a relationship with Vivex that includes: funding grants. Maria Simona Stoenoiu reports a relationship with Novartis that includes: funding grants. Maria Simona Stoenoiu reports a relationship with UCB that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We want to thank the Danish Rheumatism Association for the salary of STW.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2026.153013.

References

[1] Navarro-Compán V, Sepriano A, Capelusnik D, et al. Axial spondyloarthritis. Lancet 2025;405:159–72.
[2] Navarro-Compán V, Boel A, Boonen A, et al. The ASAS-OMERACT core domain set for axial spondyloarthritis. Semin Arthritis Rheum 2021;51:1342–9.
[3] Silva ABD, van der Heijde D, Gaalen FV, et al. Transforming myth into reality: a narrative review on the effect of therapies in slowing structural damage progression in axial spondyloarthritis. Jt Bone Spine 2025:105963.
[4] Maxwell LJ, Beaton DE, Boers M, et al. The evolution of instrument selection for inclusion in core outcome sets at OMERACT: filter 2.2. Semin Arthritis Rheum 2021;51:1320–30.
[5] Baraliakos X, Listing J, Rudwaleit M, et al. Progression of radiographic damage in patients with ankylosing spondylitis: defining the central role of syndesmophytes. Ann Rheum Dis 2007;66:910–5.
[6] Creemers MC, Franssen MJ, van't Hof MA, et al. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. Ann Rheum Dis 2005;64:127–9.
[7] van der Heijde D, Braun J, Deodhar A, et al. Modified stoke ankylosing spondylitis spinal score as an outcome measure to assess the impact of treatment on structural progression in ankylosing spondylitis. Rheumatology 2019;58:388–400.
[8] Navarro-Compán V, Boel A, Boonen A, et al. Instrument selection for the ASAS core outcome set for axial spondyloarthritis. Ann Rheum Dis 2023;82:763–72.
[9] van der Heijde D, Landewé R. Selection of a method for scoring radiographs for ankylosing spondylitis clinical trials, by the assessment in ankylosing spondylitis working group and OMERACT. J Rheumatol 2005;32:2048–9.
[10] Spoorenberg A, de Vlam K, van der Linden S, et al. Radiological scoring methods in ankylosing spondylitis. Reliability and change over 1 and 2 years. J Rheumatol 2004;31:125–32.
[11] Baraliakos X, Østergaard M, Lambert RG, et al. MRI lesions of the spine in patients with axial spondyloarthritis: an update of lesion definitions and validation by the ASAS MRI working group. Ann Rheum Dis 2022;81:1243–51.
[12] Diekhoff T, Hermann KGA, Lambert RG. Future of low-dose computed tomography and dual-energy computed tomography in axial spondyloarthritis. Curr Rheumatol Rep 2022;24:198–205.
[13] Lambert RGW, Hermann KGA, Diekhoff T. Low-dose computed tomography for axial spondyloarthritis: update on use and limitations. Curr Opin Rheumatol 2021;33:326–32.
[14] Maksymowych WP, Learch T, Lambert RG, et al. Development and validation of the spondyloarthritis radiography module for calibration of readers using the modified Stoke Ankylosing Spondylitis Spine Score. Arthritis Care Res 2014;66:55–62.
[15] Gazel U, Ayan G, Hryciw N, et al. Disease-specific definitions of new bone formation on spine radiographs: a systematic literature review. Rheumatol Adv Pr 2024;8:rkae061.
[16] Florkow MC, Zijlstra F, Willemsen K, et al. Deep learning-based MR-to-CT synthesis: the influence of varying gradient echo-based MR images as input channels. Magn Reson Med 2020;83:1429–41.
[17] Morbée L, Chen M, Herregods N, et al. MRI-based synthetic CT of the lumbar spine: geometric measurements for surgery planning in comparison with CT. Eur J Radiol 2021;144:109999.
[18] Staartjes VE, Seevinck PR, Vandertop WP, et al. Magnetic resonance imaging-based synthetic computed tomography of the lumbar spine for surgical planning: a clinical proof-of-concept. Neurosurg Focus 2021;50:e13.
[19] van der Kolk B, Slotman DJJ, Nijholt IM, et al. Bone visualization of the cervical spine with deep learning-based synthetic CT compared to conventional CT: a single-center noninferiority study on image quality. Eur J Radiol 2022;154:110414.
[20] Florkow MC, Willemsen K, Zijlstra F, et al. MRI-based synthetic CT shows equivalence to conventional CT for the morphological assessment of the hip joint. J Orthop Res 2022;40:954–64.
[21] Jans LBO, Chen M, Elewaut D, et al. MRI-based synthetic CT in the detection of structural lesions in patients with suspected sacroiliitis: comparison with MRI. Radiology 2021;298:343–9.
[22] Morbée L, Chen M, Van Den Berghe T, et al. MRI-based synthetic CT of the hip: can it be an alternative to conventional CT in the evaluation of osseous morphology? Eur Radiol 2022;32:3112–20.
[23] Morbée L, Vereecke E, Laloo F, et al. Common incidental findings on sacroiliac joint MRI: added value of MRI-based synthetic CT. Eur J Radiol 2023;158:110651.

[24] Krabbe S, Møller JM, Hadsbjerg AEF, et al. Detection of structural lesions of the sacroiliac joints in patients with spondyloarthritis: a comparison of T1-weighted 3D spoiled gradient echo MRI and MRI-based synthetic CT versus T1-weighted turbo spin echo MRI. Skelet Radiol 2024;53:2459–68.
[25] Willesen ST, Hadsbjerg AE, Møller JM, et al. MRI-based synthetic CT: a new method for structural damage assessment in the spine in patients with axial spondyloarthritis - a comparison with low-dose CT and radiography. Ann Rheum Dis 2024;83:807–15.
[26] Li Y, Xiong Y, Hou B, et al. Comparison of zero echo time MRI with T1-weighted fast spin echo for the recognition of sacroiliac joint structural lesions using CT as the reference standard. Eur Radiol 2022;32:3963–73.
[27] Deppe D, Hermann KG, Proft F, et al. CT-like images of the sacroiliac joint generated from MRI using susceptibility-weighted imaging (SWI) in patients with axial spondyloarthritis. RMD Open 2021;7.
[28] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.
[29] Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6:284–90.
[30] Beaton et al. Instrument selection part B (OMERACT handbook chapter 5B) 2021.
[31] D'Agostino MA, Beaton DE, Maxwell LJ, et al. Improving domain definition and outcome instrument selection: lessons learned for OMERACT from imaging. Semin Arthritis Rheum 2021;51:1125–33.