Limited Reliability of Radiographic Assessment of Sacroiliac Joints in Patients with Suspected Early Spondyloarthritis

Alice Ashouri Christiansen, Oliver Hendricks, Dorota Kuettel, Kim Hørslev-Petersen, Anne Grethe Jurik, Steen Nielsen, Kaspar Rufibach, Anne Gitte Loft, Susanne Juhl Pedersen, Louise Thuesen Hermansen, Mikkel Østergaard, Bodil Arnbak, Claus Manniche, and Ulrich Weber

ABSTRACT. ***Objective.*** To determine the reproducibility of evaluation of sacroiliac joint (SIJ) radiographs among readers with varying levels of experience, and to identify potential drivers of disagreement in classification among 5 predefined radiographic lesion types.

Methods. The study sample consisted of 104 consecutive patients aged 18–40 with low back pain ≥ 3 months of duration who met the Assessment of SpondyloArthritis international Society (ASAS) definition for a positive SIJ magnetic resonance image, or were HLA-B27–positive and had ≥ 1 spondyloarthritis (SpA)-related clinical/laboratory feature according to the ASAS classification criteria for axial SpA. Seven blinded readers (2 musculoskeletal radiologists, 5 rheumatologists) classified pelvic radiographs according to the modified New York criteria (mNY) and recorded presence/absence of 5 lesion types in both SIJ: erosion, sclerosis, ankylosis, joint space widening, and joint space narrowing. Reproducibility of mNY classification among 21 reader pairs was assessed and potential drivers of disagreement were identified among 5 lesion types. A generalized linear mixed logistic regression model served to analyze to what extent discordance in lesion type was associated with discrepant mNY classification.

Results. Mean κ values (percent concordance) were 0.39 (84.1%) for mNY classification over 21 reader pairs, 0.46 (79.8%) between 2 musculoskeletal radiologists, and 0.55 (86.5%) and 0.36 (77.9%) between the most experienced rheumatologist and the 2 radiologists. Erosion showed the lowest agreement (25%) among patients with discordant classification and gave the highest OR of 13.5 for disagreement.

Conclusion. Reproducibility of radiographic SIJ classification in an SpA inception cohort was only fair to at best moderate among 7 readers with varying levels of experience, questioning the applicability of mNY in early SpA. (J Rheumatol First Release October 15 2016; doi:10.3899/jrheum.160079)

Key Indexing Terms:
SPONDYLOARTHRITIS
INTERREADER AGREEMENT
RADIOGRAPHIC SACROILIITIS
MODIFIED NEW YORK CRITERIA

From the King Christian 10th Hospital for Rheumatic Diseases, Gråsten; Hospital of Southern Jutland, Aabenraa; Institute of Regional Health Research, University of Southern Denmark, Odense; Department of Radiology, and Department of Rheumatology, Aarhus University Hospital, Aarhus; Research Department, Spine Centre of Southern Denmark, Hospital Lillebaelt Middelfart, Middelfart; Department of Internal Medicine, Hospital Lillebaelt Vejle, Vejle; Copenhagen Center for Arthritis Research (COPECARE), Center for Rheumatology and Spine Diseases, Rigshospitalet – Glostrup, Glostrup; Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; Rufibach rePROstat, Biostatistical Consulting and Training, Basel, Switzerland.

Dr. Rufibach is founder and owner of Rufibach rePROstat and is an employee of F. Hoffmann-La Roche, Basel, Switzerland. The Hospital of Southern Jutland, University of Southern Denmark, Hospital Lillebaelt, Vejle, and Knud og Edith Eriksens Mindefond funded Dr. Christiansen's salary during the course of a PhD program, including this study.

A.A. Christiansen, MD, King Christian 10th Hospital for Rheumatic Diseases, and Hospital of Southern Jutland, and Institute of Regional Health Research, University of Southern Denmark; O. Hendricks, MD, PhD, King Christian 10th Hospital for Rheumatic Diseases, and Hospital of Southern Jutland, and Institute of Regional Health Research, University of Southern Denmark; D. Kuettel, MD, King Christian 10th Hospital for Rheumatic Diseases, and Hospital of Southern Jutland; K. Hørslev-Petersen, MD, DMSc, King Christian 10th Hospital for Rheumatic Diseases, and Hospital of Southern Jutland, and Institute of Regional Health Research, University of Southern Denmark; A.G. Jurik, MD, DMSc, Institute of Regional Health Research, University of Southern Denmark, and Department of Radiology, Aarhus University Hospital, and Research Department, Spine Centre of Southern Denmark, Hospital Lillebaelt Middelfart; S. Nielsen, MD, King Christian 10th Hospital for Rheumatic Diseases, and Hospital of Southern Jutland; K. Rufibach, PhD, Rufibach rePROstat, Biostatistical Consulting and Training; A.G. Loft, MD, DMSc, Department of Internal Medicine, Hospital Lillebaelt Vejle, and Department of Rheumatology, Aarhus University Hospital; S.J. Pedersen, MD, PhD, COPECARE, Center for Rheumatology and Spine Diseases, Rigshospitalet – Glostrup, and Department of Clinical Medicine, University of Copenhagen; L.T. Hermansen, MSc, King Christian 10th Hospital for Rheumatic Diseases, and Hospital of Southern Jutland; M. Østergaard, MD, PhD, DMSc, COPECARE, Center for

Rheumatology and Spine Diseases, Rigshospitalet – Glostrup, and Department of Clinical Medicine, University of Copenhagen; B. Arnbak, MSc, PhD, Institute of Regional Health Research, University of Southern Denmark, and Research Department, Spine Centre of Southern Denmark, Hospital Lillebaelt Middelfart; C. Manniche, MD, DMSc, Institute of Regional Health Research, University of Southern Denmark, and Research Department, Spine Centre of Southern Denmark, Hospital Lillebaelt Middelfart; U. Weber, MD, King Christian 10th Hospital for Rheumatic Diseases, and Hospital of Southern Jutland, and Institute of Regional Health Research, University of Southern Denmark.

Address correspondence to Dr. A.A. Christiansen, Department of Research, King Christian 10th Hospital for Rheumatic Diseases, Toldbodgade 3, 6300 Gråsten, Denmark.
E-mail: achristiansen@gigtforeningen.dk

Accepted for publication August 31, 2016.

Radiographic evaluation of sacroiliac joints (SIJ) according to the modified New York criteria (mNY)[1] is the gold standard in the classification of axial spondyloarthritis (axSpA) and may affect treatment decisions in this chronic inflammatory condition. However, several studies have consistently shown limited agreement among trained readers in radiographic classification of SIJ, with κ values around 0.5[2,3,4]. The limited reproducibility of SIJ evaluation on pelvic radiographs of patients suspected of having SpA was also featured at a public hearing of the US Food and Drug Administration[5]. Two interventional trials in patients with nonradiographic axSpA used the mNY, assessed by local rheumatology and radiology readers from different sites as inclusion criterion. A posthoc analysis by trained central readers resulted in the reclassification of 36% and 37% of the patients regarding fulfillment of the radiographic mNY[6,7].

These concerns about low reliability of radiographic mNY were confirmed by a report highlighting at best moderate reproducibility of SIJ evaluation on pelvic radiographs by rheumatologist and radiologist readers, which even put the role of radiographic sacroiliitis for classification of axSpA into question[3]. However, possible data-driven explanations for the marked variability in interpretation of SIJ radiographs are scarce. We therefore hypothesized that certain radiographic lesion types contained in the radiographic mNY such as erosion, sclerosis, or joint space variation may contribute more to interreader disagreement than others.

The objectives of our study in an SpA inception cohort recruited from primary care were (1) to determine the reproducibility of radiographic SIJ classification according to the mNY among 7 rheumatology and radiology readers with varying levels of experience in imaging in SpA, and (2) to identify potential drivers of disagreement in classification among 5 predefined radiographic lesion types according to the mNY.

MATERIALS AND METHODS

Patients. Our study sample was recruited from the cohort Spines of Southern Denmark, which has been described in detail elsewhere[8,9,10]. Briefly, the cohort consisted of 1037 patients aged 18–40 years referred to the Spine Centre of Southern Denmark, Middelfart, for evaluation of low back pain of 2–12 months' duration that was refractory to treatment in primary care.

All referred patients were screened according to a standardized protocol, which included a clinical visit, back pain questionnaires, laboratory testing [HLA-B27, high-sensitivity C-reactive protein (CRP)], and magnetic resonance imaging (MRI) of the SIJ and the entire spine. Patients with back pain of ≥ 3 months' duration, who either fulfilled the Assessment of SpondyloArthritis international Society (ASAS) criteria for a positive SIJ MRI[11] or were HLA-B27–positive with at least 1 concomitant clinical or laboratory feature suggestive of SpA according to ASAS classification criteria for axSpA[12] were referred for clinical evaluation by 1 of 3 specialists in rheumatology (AGL, LHH, or OH). ASAS concomitant clinical or laboratory features suggestive of SpA were inflammatory back pain according to ASAS criteria[13], arthritis, heel enthesitis, uveitis, dactylitis, psoriasis, inflammatory bowel disease, good response to nonsteroidal anti-inflammatory drugs, family history of SpA, and elevated CRP.

Our study sample consisted of 104 patients in whom a diagnosis of axSpA was considered possible by the clinical rheumatologic assessment, and in whom pelvic radiographs of sufficient technical quality were available. Among the 104 patients, 92 met the ASAS criteria for a positive SIJ MRI and 12 were HLA-B27–positive showing ≥ 1 clinical or laboratory SpA feature. Eighty-one patients (77.9%) met the ASAS criteria for axSpA: 56 (53.8%) through the imaging arm only (MRI-only), 8 (7.7%) through the clinical arm only, and 17 (16.3%) through both arms. Twenty-three patients (22.1%) did not meet the ASAS criteria for axSpA: 19 (18.3%) with a positive SIJ MRI only and 4 (3.8%) being HLA-B27–positive with only 1 SpA feature.

The study was approved by the Danish Data Protection Agency and by the Ethics Committee of the Region of Southern Denmark (project ID S-20110029). All participating patients gave written informed consent.

Evaluation of SIJ radiographs. SIJ radiographs were obtained according to local protocols used in daily routine in 6 radiology departments in Denmark. Among the 104 SIJ radiographs, 88 (84.6%) were standard anteroposterior pelvic radiographs, 14 (13.5%) were radiographs of the lumbar spine including the SIJ, and 2 examinations (2.0%) consisted of oblique SIJ projections. All 104 digital SIJ radiographs were centrally anonymized and randomized. Seven readers (2 musculoskeletal radiologists, 5 rheumatologists) blinded to clinical, biochemical, and MRI data independently assessed the SIJ radiographs in random order on electronic workstations. First, the readers classified the SIJ radiographs according to the mNY that was considered met if there was at least bilateral grade 2 or unilateral grade 3 sacroiliitis[1]. Second, the readers recorded the presence/absence of 5 radiographic lesion types in both SIJ as described in the mNY: erosion, sclerosis, ankylosis, joint space widening (JSW), and joint space narrowing (JSN). Erosion and sclerosis were recorded per 4 joint surfaces, i.e., on the sacral and the iliac side of the right and left SIJ, respectively, whereas ankylosis, JSW, and JSN were reported separately per right and left SIJ, respectively. We followed the definitions of SIJ grades and radiographic lesion types as stated in the mNY[1]: grade 0 = normal, grade 1 = suspicious changes, grade 2 = minimum abnormality (small localized areas with erosion or sclerosis, without alteration in the joint width), grade 3 = unequivocal abnormality (moderate or advanced sacroiliitis with erosion, evidence of sclerosis, widening, narrowing, or partial ankylosis), and grade 4 = severe abnormality or total ankylosis. SIJ scores and radiographic lesions were entered into a standardized electronic data sheet identical to the one used during reader calibration.

Reader calibration. The 7 readers consisted of 2 senior musculoskeletal radiologists having more than 20 years each of experience in interpretation of pelvic radiographs (AGJ, SN), and of 3 senior and 2 junior staff rheumatologists from 1 institution (King Christian 10th Hospital for Rheumatic Diseases, Gråsten, Denmark). The 2 radiologists came from different institutions and were not involved previously in shared imaging research. One of the rheumatologist readers (UW), who had more than 10 years of research experience in conventional and tomographic imaging in SpA, was responsible for calibration of the reader team.

All 7 readers were calibrated by reference images of pelvic radiographs covering all mNY grades. The reference images were derived from clinical practice in patients with various stages of SpA to best match the original grading description, which lacks standardized and validated lesion definitions. The definitions of the 5 grades were adopted from the original description of the mNY[1], which was based on the Atlas of Standard Radiographs in Arthritis[14]. Because of their longstanding experience in scoring SIJ on pelvic radiographs, the 2 musculoskeletal radiologists did not participate in the additional calibration for the rheumatologists. The 5 rheumatologists had three 2-h calibration sessions and independently performed a training readout. The first session consisted of an introduction to the scoring method, a review of the relevant literature, and a group discussion of 10 pelvic radiographs. This was followed by an independent evaluation of 15 pelvic radiographs by each rheumatologist according to the same scientific protocol that was later used in the main study. SIJ scores and radiographic lesion types reported in this training readout were evaluated in a second calibration session. A third calibration session with group discussion of another 10 pelvic radiographs served to refine the reference images set. All pelvic radiographs used in the training sessions were unrelated to the main study.

Descriptive analysis. Categorical demographic, clinical, and laboratory variables were described as proportion of subjects showing these features, and continuous variables as median [interquartile range (IQR)]. We expressed the presence of single radiographic features and fulfillment of the mNY as mean proportion of study subjects over 7 readers, and as mean proportions stratified according to level of reader experience. To determine the frequency of advanced sacroiliitis in our sample, we calculated the proportion of study subjects showing SIJ scores > 2 in the right and left SIJ separately. Presence of erosion and sclerosis was defined as ≥ 1 lesion on ≥ 1 of the 4 joint surfaces on both sides, while ankylosis, JSW, and JSN were defined as ≥ 1 lesion in ≥ 1 of the 2 joints, respectively. The frequency of the 5 radiographic lesions was calculated as mean proportion of patients having each lesion type over 7 readers, and as proportion of each lesion type among mNY-positive and mNY-negative study subjects for all 7 readers individually. Finally, we calculated the frequency of ≥ 2 concomitant lesion types per patient.

Interreader agreement. Interreader agreement for classification according to the mNY and for the 5 radiographic features was assessed by means of 2 × 2 tables and calculating percent agreement (total; positive/negative) and by Cohen's κ[15]. Interreader agreement for the ordinal SIJ grades for both sides separately was evaluated by weighted Cohen's κ. Agreement was interpreted according to Landis and Koch[16] as slight ($\kappa < 0.2$), fair ($0.2 \le \kappa < 0.4$), moderate ($0.4 \le \kappa < 0.6$), substantial ($0.6 \le \kappa < 0.8$), and almost perfect ($0.8 \le \kappa < 1.0$). The computations were made for each reader pair and for all readers jointly as mean value over all 21 reader pairs. For the pairwise κ values, a bootstrap CI based on 1000 bootstrap replications and computed at a CI of 95% was provided. We additionally compared 5 selected reader pairs regarding agreement: the 2 musculoskeletal radiologists, the most experienced rheumatologist versus each of the 2 musculoskeletal radiologists, and the 2 senior and the 2 junior rheumatologists. The proportion of concordant single grades according to mNY among ≥ 2 readers (any reader pair) and ≥ 4 readers (majority of readers) was described for the right and left SIJ separately.

Candidate lesion types driving discrepancies in mNY classification. To assess the relative contribution of each of the 5 lesion types to disagreement in mNY classification, we first identified patients with discrepant mNY classification for each reader pair. Among these, we computed the proportion of patients with concordance for each radiographic lesion type for all reader pairs.

Finally, a generalized linear mixed logistic regression model was computed to estimate the relative effect size of each individual radiographic lesion type. Results were expressed as OR for disagreement in mNY classification with 95% CI. P values ≤ 0.05 were considered significant.

All computations were done with R (R Core Team, version 3.1.1.).

RESULTS

Descriptive analysis. Of the 104 patients, 38.5% were men and 33.7% were HLA-B27–positive (Table 1). Median age was 33.0 years. Over all 7 readers, a mean proportion of 15.7% of the patients met the mNY, and 8.1% showed mNY grades 3 or 4 (Table 1). Sclerosis and erosion were the 2 most frequent lesions reported in 50.1% and in 25.7% of the patients, respectively. The 3 more experienced readers scored more lesions of all types than the 4 less experienced readers, and they also considered more patients to be mNY-positive (21.5% vs 11.3%). Patients with erosion concomitantly showed sclerosis in 93.5%, JSN in 48.5%, JSW in 27.8%, and ankylosis in 19.5%. The distribution of the 5 lesion types among mNY-positive and -negative patients for all 7 readers individually is shown in Figure 1. Among the 5 radiographic lesion types, erosion and sclerosis showed the largest variation between individual readers. The most frequent constellation when reporting erosion in mNY-negative patients was unilateral grade 2 sacroiliitis (data not shown). Both more and less experienced readers reported joint space alterations in a small minority of subjects classified as mNY-negative.

Interreader agreement. Kappa (percent) agreement for mNY classification was 0.39 (84.1%) over 7 readers, 0.46 (79.8%) between 2 musculoskeletal radiologists, and 0.55 (86.5%) and 0.36 (77.9%) among the most experienced rheumatologist and each of the 2 musculoskeletal radiologists, respec-

Table 1. Patient characteristics and distribution of radiographic features. Values are n (%) unless otherwise specified.

Variables	All Readers	More Experienced Readers*	Less Experienced Readers**
Patient characteristics			
Age, yrs, median (IQR)	33.0 (8.0)	N/A	N/A
Male	40 (38.5)	N/A	N/A
HLA-B27–positive	35 (33.7)	N/A	N/A
Elevated hsCRP[1]	15 (16.0)	N/A	N/A
Current smoker[2]	41 (39.8)	N/A	N/A
Radiographic features[3]			
mNY-positive	16.3 (15.7)	22.3 (21.5)	11.8 (11.3)
SIJ right score > 2	8.6 (8.2)	12.7 (12.2)	5.5 (5.3)
SIJ left score > 2	8.3 (8.0)	12.7 (12.2)	5.0 (4.8)
Erosion	26.7 (25.7)	34.0 (32.7)	21.3 (20.4)
Sclerosis	52.1 (50.1)	62.0 (59.6)	44.8 (43.0)
Ankylosis	6.4 (6.2)	7.7 (7.3)	5.5 (5.3)
Joint space widening	8.7 (8.4)	11.0 (10.6)	7.0 (6.7)
Joint space narrowing	20.7 (19.9)	32.0 (30.8)	12.3 (11.8)

* More experienced readers: 2 musculoskeletal radiologists and 1 rheumatologist experienced in imaging in SpA. ** Less experienced readers: 4 rheumatologists less experienced in imaging in SpA. [1] Reference range ≤ 6 mg/l; missing values in 10 patients. [2] Missing value in 1 patient. [3] Mean no. study subjects over all/more/less experienced readers showing predefined radiographic features according to the mNY. hsCRP: high-sensitivity C-reactive protein; mNY: modified New York criteria; N/A: not applicable; SIJ: sacroiliac joint; SpA: spondyloarthritis; IQR: interquartile range.

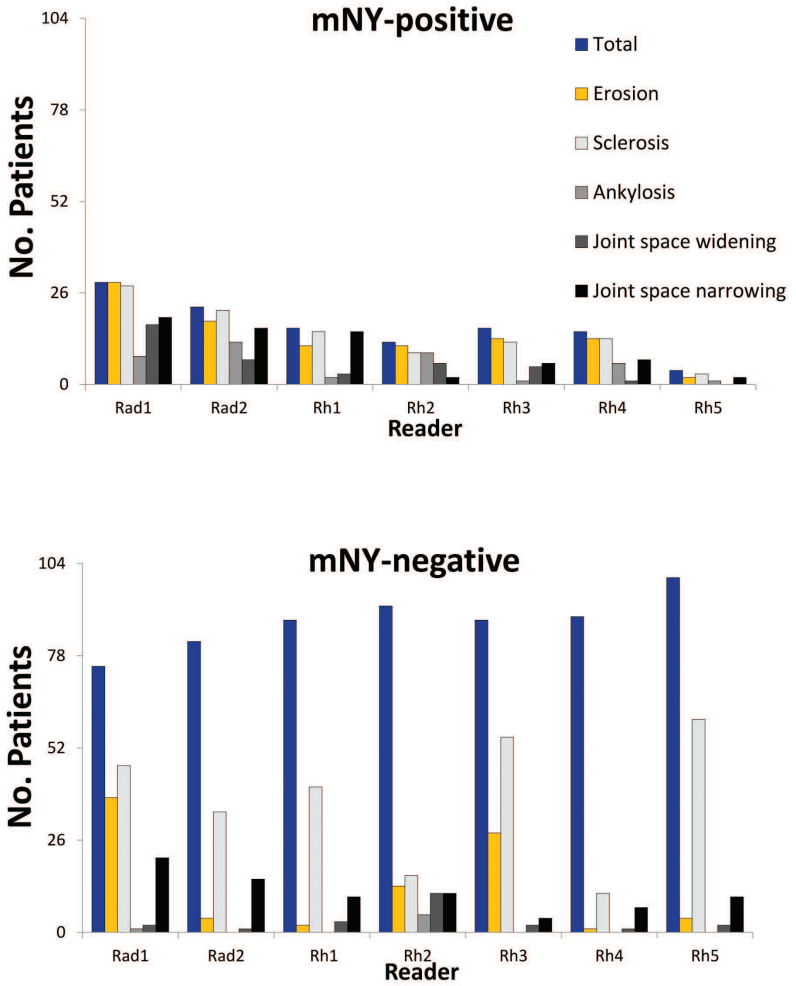

Figure 1. Distribution of 5 lesion types among mNY-positive and -negative patients for all 7 readers individually. mNY: modified New York criteria; Rad1: musculoskeletal radiologist 1; Rad2: musculoskeletal radiologist 2; Rh1: most experienced rheumatologist; Rh2, Rh3: 2 senior rheumatologists; Rh4, Rh5: 2 junior rheumatologists.

tively (Table 2). Among the rheumatologists less experienced in radiographic SIJ assessment, agreement between the 2 senior and the 2 junior rheumatologists was 0.34 (84.6%) and 0.27 (87.5%), respectively. Among the 5 radiographic lesion types, ankylosis showed the highest (κ 0.34) and JSW the lowest agreement (κ 0.12) over 21 reader pairs. Reliability among all 21 reader pairs for standard pelvic versus lumbar spine radiographs was in the agreement category "fair" as defined above with mean κ values of 0.39 and 0.33, respectively.

Candidate lesion types driving discrepancies in mNY classification. Among the 21 reader pairs, 15.9% of the patients had discrepant mNY classification. Among patients with discordant mNY classification, erosion was the lesion with the lowest interreader agreement: the proportion of mNY-discrepant patients with concordance for erosion was only 0.25 (IQR 0.09; Figure 2). The assessment of the effect size of each of the 5 lesion types showed that erosion was the strongest driver of discordance in mNY classification. Erosion was associated with statistically significant 13.5× higher odds (95% CI 9.1–20.1) for discrepant mNY classification (Table 3).

Figure 3 shows a pelvic radiograph in which 4 of 7 readers considered the mNY as being met and 5 of 7 readers scored erosion.

DISCUSSION

Our study on the reliability of radiographic SIJ classification according to the mNY in an SpA inception cohort suggests that SIJ erosion may be the primary driver of interreader disagreement. The only fair to at best moderate level of

Table 2. Agreement of all 21 reader pairs and among selected reader pairs: Cohen's κ values (95% CI; upper row) and percent agreement (positive/negative; lower row).

Reader Pairs	mNY	SIJ Right Grades 0–4	SIJ Left Grades 0–4	Erosion	Sclerosis	Ankylosis	Joint Space Widening	Joint Space Narrowing
Mean of 21	0.39	0.33	0.33	0.22	0.28	0.34	0.12	0.21
reader pairs	84.1 (7.7/76.4)	N/A	N/A	67.9 (9.6/58.2)	62.9 (31.6/31.3)	91.8 (2.1/89.7)	86.8 (1.8/85.0)	74.4 (7.1/67.3)
2 MSK radiologists	0.46 (0.26–0.64)	0.42 (0.29–0.53)	0.33 (0.17–0.46)	0.16 (0.04–0.29)	0.41 (0.24–0.58)	0.42 (0.12–0.67)	0.21 (–0.02 to 0.44)	0.26 (0.07–0.45)
	79.8 (14.4/65.4)	N/A	N/A	51.0 (18.3/32.7)	71.2 (48.1/23.1)	89.4 (4.8/84.6)	81.7 (3.8/77.9)	66.3 (17.3/49.0)
MSK radiologist 1 vs most								
experienced	0.36 (0.14–0.56)	0.28 (0.14–0.41)	0.23 (0.10–0.35)	0.05 (–0.05–0.14)	0.26 (0.08–0.42)	0.34 (0.00–0.66)	–0.01 (–0.12 to 0.16)	0.37 (0.16–0.53)
rheumatologist	77.9 (10.6/67.3)	N/A	N/A	42.3 (9.6/32.7)	64.4 (45.2/19.2)	93.3 (1.9/91.3)	77.9 (1.0/76.9)	72.1 (17.3/54.8)
MSK radiologist 2 vs most								
experienced	0.55 (0.33–0.75)	0.49 (0.33–0.62)	0.47 (0.30–0.61)	0.42 (0.19–0.64)	0.44 (0.26–0.62)	0.26 (0.00–0.57)	0.08 (–0.08 to 0.36)	0.37 (0.16–0.55)
rheumatologist	86.5 (11.5/75.0)	N/A	N/A	83.7 (8.7/75.0)	72.1 (39.4/32.7)	90.4 (1.9/88.5)	88.5 (1.0/87.5)	75.0 (14.4/60.6)
2 senior	0.34 (0.08–0.59)	0.23 (0.08–0.37)	0.28 (0.13–0.41)	0.07 (–0.11 to 0.25)	0.13 (0.01–0.26)	0.12 (0.00–0.32)	0.36 (0.09–0.59)	0.37 (0.06–0.64)
rheumatologists	84.6 (5.8/78.8)	N/A	N/A	58.7 (10.6/48.1)	50.0 (19.2/30.8)	87.5 (1.0/86.5)	86.5 (4.8/81.7)	87.5 (4.8/82.7)
2 junior	0.27 (–0.02 to 0.56)	0.30 (0.15–0.46)	0.26 (0.11–0.41)	0.35 (0.06–0.60)	0.22 (0.10–0.35)	0.27 (0.00–0.67)	–0.02 (–0.04 to 0.00)	0.21 (–0.03 to 0.45)
rheumatologists	87.5 (2.9/84.6)	N/A	N/A	88.5 (3.8/84.6)	56.7 (20.2/36.5)	95.2 (1.0/94.2)	96.2 (0.0/96.2)	82.7 (3.8/78.8)

mNY: modified New York criteria; SIJ: sacroiliac joint; MSK: musculoskeletal; N/A: not applicable.

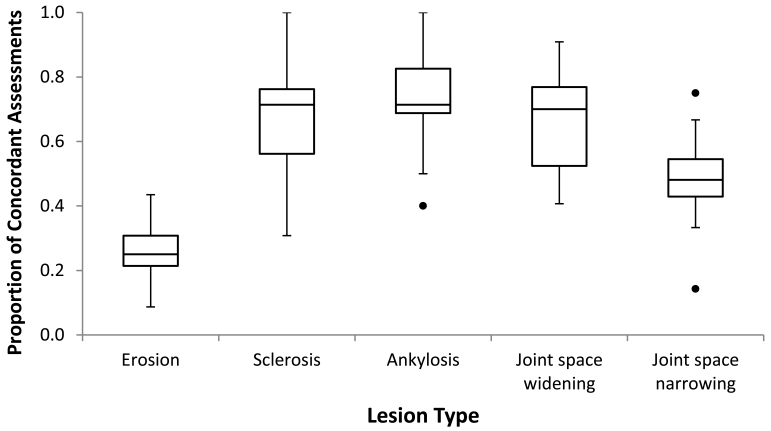

Figure 2. Proportion of patients with concordant lesion types among patients with discordant mNY classification for all 21 reader pairs. The horizontal bands within the boxes represent the medians. mNY: modified New York criteria.

Table 3. Relative contribution of 5 predefined lesion types to disagreement in modified New York criteria classification (generalized linear mixed model).

Lesion Type	OR	95% CI	p
Erosion	13.47	9.05–20.06	< 0.0001
Sclerosis	0.86	0.57–1.31	0.49
Ankylosis	4.75	2.73–8.28	< 0.0001
Joint space widening	5.59	3.41–9.17	< 0.0001
Joint space narrowing	3.04	2.02–4.56	< 0.0001

concordance for mNY (κ 0.39) among 7 radiology and rheumatology readers with varying experience in imaging in SpA was even slightly lower than a reported only moderate agreement (κ 0.54) between 2 central readers in another axSpA inception cohort[3].

The limited reproducibility of radiographic SIJ classification according to the mNY is well documented. However, the characteristics of a given study sample may affect the level of interreader agreement. Previous reports suggest that the higher the proportion of patients with ankylosing spondylitis (AS) in a given study sample, the better the concordance in radiographic mNY. A study from Turkey applying the radiographic mNY in patients with Behçet disease recorded pre/post-training κ agreement of 0.32/0.19, 0.32/0.36, and 0.44/0.41 for 3 reader pairs (1 radiologist and 2 rheumatologists, respectively)[2]. Our study with 15.7% mNY-positive patients showed a κ concordance of 0.39 among 7 radiology and rheumatology readers. In a report on patients with inflammatory back pain suggestive of axSpA, 21.1%/26.6% (central/local reading) had obvious sacroiliitis; concordance for radiographic mNY by κ values was 0.54

Figure 3. Pelvic radiograph illustrating disagreement in modified New York criteria (mNY) and erosion among 7 radiology and rheumatology readers with varying experience in imaging in spondyloarthritis (SpA). Radiograph is of a 36-year-old man suspected of having early SpA. The ASAS criteria for axSpA were met, as well as the ASAS criteria for a positive sacroiliac joint magnetic resonance image. Four of 7 readers (2 radiologists and 2 rheumatologists) considered the mNY met, and 5 of 7 readers (2 radiologists and 3 rheumatologists) reported presence of erosion. Sclerosis and joint space irregularities show mainly in the middle portion of both SIJ, while the lowest third, which represents mostly the cartilaginous joint compartment and is usually involved first by inflammation, seems to be less affeccted. The right joint displays additionally a lumbosacral transitional anomaly with an accessory joint between the transverse process of the fifth lumbar vertebra and the basis of the sacrum. ASAS: Assessment of SpondyloArthritis international Society; axSpA: axial SpA; SIJ: sacroiliac joint.

among 2 trained central rheumatologist readers and 0.55 for central versus local radiologist and rheumatologist reading[3]. A study assessing the rate of radiographic sacroiliitis progression over 2 years consisted of a high proportion of 54.8% patients with AS[4]. Interreader agreement between 2 trained rheumatology readers blinded to time sequence was moderate at baseline with a κ value of 0.57, but increased to a substantial κ of 0.67 at followup together with a progression rate from nonradiographic SpA to AS of 11.6% over 2 years. The highest interobserver concordance was reported in a study of 217 patients with AS who all met the mNY[17]. Kappa values between 2 trained rheumatologist readers were 0.68, 0.69, and 0.66 at baseline, 1-year followup, and 2-year followup.

Our agreement for classification according to the radiographic mNY (κ 0.39) was higher than for each single of the 5 radiographic lesion types (κ values 0.12–0.34). This is in line with the above-mentioned report on inflammatory back pain patients with κ values for the single lesion types between 0.12–0.44 as opposed to agreement of 0.54 for mNY[3].

A potential source of disagreement is the lack of standardized and validated definitions for each of the radiographic lesion types contained in the original description[1,14], which were used in our study. However, it remains to be shown whether an attempt to standardize and validate lesion definitions might facilitate agreement in view of the broad morphologic spectrum of the radiographic lesion types.

All lesions except sclerosis contributed to discordant mNY classification, but erosion was the main driver of disagreement. Technical issues such as bowel overlapping the SIJ or various radiographic SIJ projections can only partially explain this finding because they also affect recognition of other radiographic lesions such as sclerosis or joint space variation. Our results need to be confirmed in other cohorts of patients with clinically suspected axSpA because erosion is widely regarded as a key lesion indicating radiographic sacroiliitis.

Our SpA inception cohort recruited from primary care with low back pain of ≥ 3 months' duration showed a low frequency of HLA-B27 and male sex. Multiple studies in other early axSpA cohorts have shown lower proportions of male sex and HLA-B27 positivity when compared with AS[18,19,20,21,22]. However, these cohorts were not or not entirely based on recruitment from primary care, and usually excluded patients with just suspected SpA, which may explain the higher prevalence of male sex and HLA-B27 positivity, when compared to ours. A Dutch cohort of patients with suspected axSpA similar to ours and also recruited from primary care[23] showed an even lower proportion of

HLA-B27 positivity of 20% among patients meeting the ASAS criteria for axSpA. Our cohort reflects daily routine in which young patients with treatment-refractory back pain referred from primary care with suspected early SpA often need to be followed over time before a final diagnosis can be made. However, pelvic radiographs are often performed as 1 element of the rheumatologic evaluation in such a clinical setting of suspected early SpA, despite the limited evidence of whether they may enhance confidence in a diagnosis of early SpA.

The mNY derived from a cohort of 183 HLA-B27–positive patients with AS, their HLA-B27–positive or –negative first-degree relatives, and population controls[1] may not be directly applicable to chronic back pain patients clinically suspected of having axSpA. Further, there are no normative data regarding frequency and morphology of the 5 radiographic mNY lesion types in healthy controls, mechanical back pain patients, subjects with increased physical activity, or multiparous women. A back pain cohort from chiropractic practices in Canada with a recruitment mode similar to ours but with older patients showed degenerative SIJ changes in 35.2% of 142 women ages 18–60 years, which might be a factor leading to reader disagreement in low grade sacroiliitis in women[24]. In our study, sclerosis was the most frequently reported lesion type by all readers among patients classified as not meeting the radiographic mNY.

A Dutch report on radiographic assessment of sacroiliitis by 100 rheumatologists and 23 radiologists showed only modest sensitivity and specificity for sacroiliitis and sizable intraobserver variation[25]. Evaluation of the same image set after 3–6 months upon individual training and workshops did not improve performance. However, no pairwise analysis among all possible reader pairs was performed as in our study, but the scores of an expert panel (2 rheumatologists, 1 epidemiologist, and 1 radiologist) served as gold standard. Future studies with pairwise analysis of all possible reader pairs involving both radiologists and rheumatologists are needed to determine whether training and calibration in recognition of various radiographic SIJ lesion types, especially erosion, might improve agreement in classification according to the radiographic mNY.

Our SIJ radiographs were acquired according to local protocols in 6 radiology centers resulting in 84.6% standard anteroposterior pelvic radiographs, the remaining being lumbar spine radiographs including the SIJ and 2 oblique SIJ projections. The different visualizations of the SIJ might have had an effect on reproducibility. The lack of a full calibration of the 2 musculoskeletal radiologists may have affected inter-reader agreement as well. However, both limitations regarding imaging protocols and reader calibration reflect the conditions in daily routine. Another potential limitation is that κ statistics inherently perform less well in cases of skewed distribution of the variables under observation[26,27,28,29], as with our relatively low prevalence of mNY grades 3–4 of only 8.0%.

Reproducibility of SIJ classification according to the mNY in a SpA inception cohort was only fair to at best moderate among 7 radiology and rheumatology readers with varying experience in imaging in SpA. Erosion was the main driver of discordant classification. These findings question the applicability of the radiographic mNY in back pain patients clinically suspected of having early axSpA, particularly in healthcare settings where access to SIJ MRI is readily available.

ACKNOWLEDGMENT

The authors thank Laila Dungart and Henning Jakobsen from the Radiology Department at King Christian 10th Hospital for Rheumatic Diseases, Gråsten, Denmark, for anonymization and randomization of the pelvic radiographs; Lone Holm Hansen (LHH) for clinical evaluation of patients at Hospital Lillebaelt, Vejle, Denmark; Charlotte Drachmann and Lis Schubert at King Christian 10th Hospital for Rheumatic Diseases, Gråsten, Denmark, for high-sensitivity C-reactive protein and HLA-B27 analysis; Tue Secher Jensen at the Spine Centre of Southern Denmark, Denmark, for his role in the conception and design of the Spines of Southern Denmark Cohort; and the radiologic departments at these Danish hospitals for kindly providing the radiographs used in this study: Hospital Lillebaelt, Vejle; Odense University Hospital; Odense University Hospital at Svendborg Hospital; Hospital South West Jutland; Hospital of Nykøbing Falster; and King Christian 10th Hospital for Rheumatic Diseases, Gråsten.

REFERENCES

1. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984;27:361-8.
2. Yazici H, Turunç M, Ozdoğan H, Yurdakul S, Akinci A, Barnes CG. Observer variation in grading sacroiliac radiographs might be a cause of 'sacroiliitis' reported in certain disease states. Ann Rheum Dis 1987;46:139-45.
3. van den Berg R, Lenczner G, Feydy A, van der Heijde D, Reijnierse M, Saraux A, et al. Agreement between clinical practice and trained central reading in reading of sacroiliac joints on plain pelvic radiographs. Results from the DESIR cohort. Arthritis Rheumatol 2014;66:2403-11.
4. Poddubnyy D, Rudwaleit M, Haibel H, Listing J, Märker-Hermann E, Zeidler H, et al. Rates and predictors of radiographic sacroiliitis progression over 2 years in patients with axial spondyloarthritis. Ann Rheum Dis 2011;70:1369-74.
5. Deodhar A, Reveille JD, van den Bosch F, Braun J, Burgos-Vargas R, Caplan L, et al. The concept of axial spondyloarthritis: joint statement of the spondyloarthritis research and treatment network and the Assessment of SpondyloArthritis international Society in response to the US Food and Drug Administration's comments and concerns. Arthritis Rheumatol 2014;66:2649-56.
6. U.S. Food and Drug Administration, Department of Health & Human Services. Arthritis Advisory Committee Meeting: sBLA 125057/323: adalimumab for the treatment of active non-radiographic axial spondyloarthritis in adults with objective signs of inflammation by elevated C-reactive protein (CRP) or magnetic resonance imaging (MRI), who have had an inadequate response to, or are intolerant to, a nonsteroidal anti-inflammatory drug [Internet. Accessed August 31, 2016.] Available from: www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisAdvisoryCommittee/UCM361563.pdf
7. U.S. Food and Drug Administration, Department of Health & Human Services. Arthritis Advisory Committee Meeting: sBLA 125160/215: Cimzia (certolizumab) for the treatment of active axial spondyloarthritis, including patients with ankylosing spondylitis

[Internet. Accessed August 31, 2016.] Available from: www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisAdvisoryCommittee/UCM361565.pdf
8. Arnbak B, Jensen TS, Egund N, Zejden A, Hørslev-Petersen K, Manniche C, et al. Prevalence of degenerative and spondyloarthritis-related magnetic resonance imaging findings in the spine and sacroiliac joints in patients with persistent low back pain. Eur Radiol 2016;26:1191-203.
9. Arnbak B, Hendricks O, Hørslev-Petersen K, Jurik AG, Pedersen SJ, Østergaard M, et al. The discriminative value of inflammatory back pain in patients with persistent low back pain. Scand J Rheumatol 2016;45:321-8.
10. Arnbak B, Grethe Jurik A, Hørslev-Petersen K, Hendricks O, Hermansen LT, Loft AG, et al. Associations between spondyloarthritis features and magnetic resonance imaging findings: a cross-sectional analysis of 1,020 patients with persistent low back pain. Arthritis Rheumatol 2016;68:892-900.
11. Rudwaleit M, Jurik AG, Hermann KG, Landewé R, van der Heijde D, Baraliakos X, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. Ann Rheum Dis 2009;68:1520-7.
12. Rudwaleit M, van der Heijde D, Landewé R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009;68:777-83.
13. Sieper J, van der Heijde D, Landewé R, Brandt J, Burgos-Vagas R, Collantes-Estevez E, et al. New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the Assessment of SpondyloArthritis international Society (ASAS). Ann Rheum Dis 2009;68:784-8.
14. Kellgren JH, Jeffrey MR. The epidemiology of chronic rheumatism; volume 2: atlas of standard radiographs of arthritis. Oxford: Blackwell Scientific Publications; 1963:36-40.
15. Conger AJ. Integration and generalization of kappas for multiple raters. Psychol Bull 1980;88:322-8.
16. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics 1977;33:363-74.
17. Spoorenberg A, de Vlam K, van der Linden S, Dougados M, Mielants H, van de Tempel H, et al. Radiological scoring methods in ankylosing spondylitis. Reliability and change over 1 and 2 years. J Rheumatol 2004;31:125-32.
18. Rudwaleit M, Haibel H, Baraliakos X, Listing J, Märker-Hermann E, Zeidler H, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. Arthritis Rheum 2009;60:717-27.
19. Ciurea A, Scherer A, Exer P, Bernhard J, Dudler J, Beyeler B, et al; Rheumatologists of the Swiss Clinical Quality Management Program for Axial Spondyloarthritis. Tumor necrosis factor alpha inhibition in radiographic and nonradiographic axial spondyloarthritis: results from a large observational cohort. Arthritis Rheum 2013;65:3096-106.
20. van den Berg R, de Hooge M, van Gaalen F, Reijnierse M, Huizinga T, van der Heijde D. Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. Rheumatology 2013;52:1492-9.
21. Moltó A, Paternotte S, van der Heijde D, Claudepierre P, Rudwaleit M, Dougados M. Evaluation of the validity of the different arms of the ASAS set of criteria for axial spondyloarthritis and description of the different imaging abnormalities suggestive of spondyloarthritis: data from the DESIR cohort. Ann Rheum Dis 2015;74:746-51.
22. Kiltz U, Baraliakos X, Karakostas P, Igelmann M, Kalthoff L, Klink C, et al. The degree of spinal inflammation is similar in patients with axial spondyloarthritis who report high or low levels of disease activity: a cohort study. Ann Rheum Dis 2012;71:1207-11.
23. van Hoeven L, Luime J, Han H, Vergouwe Y, Weel A. Identifying axial spondyloarthritis in Dutch primary care patients, ages 20-45 years, with chronic low back pain. Arthritis Care Res 2014; 66:446-53.
24. O'Shea FD, Boyle E, Salonen DC, Ammendolia C, Peterson C, Hsu W, et al. Inflammatory and degenerative sacroiliac joint disease in a primary back pain cohort. Arthritis Care Res 2010;62:447-54.
25. van Tubergen A, Heuft-Dorenbosch L, Schulpen G, Landewé R, Wijers R, van der Heijde D, et al. Radiographic assessment of sacroiliitis by radiologists and rheumatologists: does training improve quality? Ann Rheum Dis 2003;62:519-25.
26. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43:543-9.
27. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43:551-8.
28. Vach W. The dependence of Cohen's kappa on the prevalence does not matter. J Clin Epidemiol 2005;58:655-61.
29. Flight L, Julious SA. The disagreeable behaviour of the kappa statistic. Pharm Stat 2015;14:74-8.