OP0253

STRUCTURAL DAMAGE IN AXIAL SPONDYLOARTHRITIS: IS THERE A PREFERRED WAY TO ASSESS PROGRESSION OVER TIME?

X. Baraliakos¹, V. Rios Rodriguez², M. Torgutalp², A. Dilbaryan², H. Haibel², M. Verba², J. Sieper², J. Braun¹, M. Rudwaleit³, D. Poddubnyy^{2,4}. ¹Ruhr-University Bochum, Rheumazentrum Ruhrgebiet, Herne, Germany; ²Charité-Universitätsmedizin, Gastroenterology, Infectious Diseases and Rheumatology, Berlin, Germany; ³Klinikum Rosenhöhe, Universitätsklinikum OWL, Bielefeld, Germany; ⁴German Rheumatism Research Centre, Epidemiology Unit, Berlin, Germany

Background: Axial spondyloarthritis (axSpA) is characterized by development of structural damage in the spine, assessed by conventional radiographs (CR). The modified Stokes Ankylosing Spondylitis Scoring System (mSASSS) includes all the SpA-relevant CR findings and is the most frequently used scoring system for quantification of spinal structural damage in patients (pts) with axSpA. Radiographic progression over 2 years using mean mSASSS progression, scored by a blinded 'paired' (sets of CRs of the same patient) approach, is considered as a key treatment outcome reflecting structural progression in clinical trials. However, this approach has been accompanied by questions about the influence of pts showing mSASSS 'improvement', indicating possible disappearance of syndesmophytes over time, something 'naturally unexpected'.

Objectives: To investigate the performance of mSASSS in assessing spinal radiographic damage and progression in axSpA pts using different approaches of CR evaluations.

Methods: Complete sets (cervical and lumbar CRs) from the German SpA Inception Cohort (GESPIC) at baseline and after 2 years were blinded to all clinical and demographic characteristics and then scored using the mSASSS by 5 different experienced readers, 2 blinded and 3 unblinded to the timepoint of CR performance. The final mSASSS score was calculated as a mean of 2 (blinded) or 3 (unblinded) readers. Descriptive statistics, cumulative probability plots and shift analyses (agreement of 2/2 readers in the blinded and 2/3 readers in the unblinded group) were performed for each reader group.

Results: A total of 210 pts (mean age 37.3y, 51% male, 79% HLA-B27+) with non-radiographic (n=115) and radiographic (n=95) axSpA at baseline were included. The mean mSASSS at baseline was 4.3±8.3 vs. 3.4±7.9, while the mean progression was 0.7±2.3 vs. 1.0±1.9 mSASSS units for the blinded vs. the unblinded group, respectively (Figure). On the patient level, progression of ≥2 mSASSS units was found in 30 (14.3%) vs. 37 (17.6%) pts in the blinded vs. the unblinded group, while agreement between groups was seen in 179 (85.2%) pts, 18 (8.9%) pts for progression and 161 (76.7%) for no progression.

Figure 1. Cumulative probability plot for mean mSASSS scores (all scores) in blinded (A) and unblinded (B) and 'definite' scores in blinded (C) and unblinded (D) scoring approach

In the analysis of 'definite' CR findings (only scores of 2=syndesmophytes or 3=ankylosis), the mean mSASSS score at baseline was 3.3±8.0 vs. 2.6±7.2 and the mean radiographic progression was 0.6±2.4 vs. 0.8±2.1 mSASSS units for the blinded vs. the unblinded group, respectively (Figure). On the patient level, progression was found in 37 (17.6%) vs. 33 (15.7%) pts in the blinded vs. the unblinded group, while agreement between groups was seen in 188 (89.5%) pts, 24 (11.3%) for progression and in 164 (78.1%) pts for no progression.

In the shift analysis, mSASSS worsening was found in 35 (0.8%) and mSASSS 'improvement' in only 4 (0.1%) out of a total of 4.373 vertebral edges (VE) analyzed in the blinded and in 109 (2.2%) and 2 (0.04%), respectively, out of a total of 4.914 VE analyzed in the unblinded group (Table). The majority of progression was found for the development of 'definite' signs of progression in both the blinded (25/35, 71.4%) and the unblinded (61/109, 56%) group, while more VE showing 'minor' (only scores of 1=sclerosis, squaring, erosion) signs of progression were found in the unblinded (48/109, 44%) vs. to the blinded (10/25, 28.6%) group.

Conclusion: Despite lower mean baseline mSASSS, higher mean mSASSS progression was found using the unblinded approach, while in the shift analysis the unblinded approach was also more specific, confirming the absence of 'improvement' of radiographic damage over time. These results were similar in the analysis on the patient's level. The scoring approach and the influence of 'minor' radiographic changes should be taken into account for avoiding the so-called 'background noise' in the evaluation of mSASSS.

Acknowledgements: GESPIC was initially supported by the BMBF. As a consequence of the funding reduction by BMBF according to schedule in 2005 and stopped in 2007, complementary financial support has been obtained also from Abbott, Amgen, Centocor, Schering-Plough, and Wyeth. Starting from 2010, the core GESPIC cohort was supported by AbbVie.

Table 1. Shift analysis of the blinded and unblinded reading approach on the basis of VE.

Blinded reading						
mSASS Scores		2-year follow-up				Total
		0	1	2	3	
Baseline	0	4.166	10	8	9	4.193
	1	0	29	4	0	33
	2	3	1	48	4	56
	3	0	0	0	92	92
Total		4.169	40	60	105	4.374

Unblinded reading						
mSASS Scores		2-year follow-up				Total
		0	1	2	3	
Baseline	0	4.575	48	14	3	4.640
	1	1	75	16	8	100
	2	0	0	69	20	89
	3	0	1	0	84	85
Total		4.576	124	99	115	4.914

Disclosure of Interests: None declared

DOI: 10.1136/annrheumdis-2021-eular.2532

OP0254

AN ARTIFICIAL NEURAL NETWORK FOR THE DETECTION OF DEFINITE RADIOGRAPHIC SACROILIITIS WITH HIGH SPECIFICITY IN THE DIAGNOSTIC SETTING

J. L. Vahldiek¹, K. K. Bressen¹, S. Niehues¹, L. Adams¹, L. Spiller², M. Protopopov², V. Rios Rodriguez², B. Mueche², J. Rademacher², H. Haibel², M. Torgutalp², F. Proft², D. Poddubnyy^{2,3}. ¹Charité – Universitätsmedizin Berlin, Division of Radiology, Berlin, Germany; ²Charité – Universitätsmedizin Berlin, Division of Gastroenterology, Infectious Diseases and Rheumatology, Berlin, Germany; ³Deutsches Rheuma-Forschungszentrum Berlin (DRFZ), ein Institut der Leibniz-Gemeinschaft, Epidemiology Unit, Berlin, Germany

Background: Radiographs of the sacroiliac joints are commonly used as the first imaging method for the diagnosis of axial spondyloarthritis (axSpA), but the reliability of sacroiliitis detection is usually low. Recently we developed a deep convolutional neural network (CNN) that can detect radiographic sacroiliitis with expert-like accuracy in patients with axSpA, i.e., classification as radiographic or non-radiographic [1]. There is frequent criticism that many artificial intelligence algorithms for diagnostic analysis of medical images lack robust validation in real-world clinical applications.

Objectives: The aim of this study was to evaluate the performance of our deep CNN in detecting definite radiographic sacroiliitis in the diagnostic setting.

Methods: In this study, we included a total of 361 patients with chronic back pain who presented to a rheumatologist in a specialized SpA center with a suspicion of axSpA within the OptiRef project [2]. All patients received a structured rheumatologic diagnostic work-up that resulted in the final diagnosis of axial SpA/no axial SpA. Radiographs of sacroiliac joints were evaluated by a rheumatologist and radiologist according to the modified New York criteria; the consensus judgement of the presence of definite radiographic sacroiliitis (≥grade 2 bilaterally or ≥grade 3 unilaterally) was used as a reference. The predictions of the deep CNN's inference (with a balanced cutoff of 0.724 for the predictions of the model that was derived from the training and validation steps [1]) on all available pelvic radiographs was compared to this reference judgement.