

RADIOLOGICAL SCORING METHODS IN ANKYLOSING SPONDYLITIS: A COMPARISON OF THE RELIABILITY OF AVAILABLE METHODS

Hasan Ulusoy*, Arzu Kaya*, Ayhan Kamanlı*, Gurkan Akgol*, Salih Ozgocmen**

Abstract

Objective: To assess intra and inter-rater reliability of available radiological scoring methods in ankylosing spondylitis (AS).

Patients and Methods: Two trained raters evaluated 44 complete sets of AS radiographs. The cervical and lumbar spine was graded from zero to 4 according to the Bath Ankylosing Spondylitis Radiology Index (BASRI). Hip joints were graded according to the BASRI-hip method. Sacroiliac (SI) joints were scored according to the New York method (0-4). The anterior and posterior sites of the lumbar spine were scored according to the Stoke Ankylosing Spondylitis Spinal Score (SASSS) method (0-72). Modified-SASSS was assessed by using the anterior sites of both the cervical and lumbar spine (0-72).

Results: Both intra and inter-rater reliability were almost perfect for all the methods and intra-class correlation coefficient (ICC) for all the methods was relatively similar to each other. The BASRI-spine and BASRI-total showed intra and inter-rater ICC between 0.78 and 0.98. Both SASSS and modified-SASSS reached perfect intra and inter-rater reliability with ICC between 0.86 and 0.99. The ICC of the BASRI-hip was substantial to perfect, ranging from 0.77 to 0.88. Time spent to score a set of radiographs using the BASRI-spine was <45 seconds, whereas >60 seconds for both SASSS and mSASSS methods.

Conclusion: After training, all of these methods have demonstrated almost perfect intra and inter-rater reliability. The BASRI was easier to perform and less time consuming than SASSS methods.

Keywords: Ankylosing Spondylitis; Radiological Scoring; BASRI; SASSS.

Introduction

Ankylosing spondylitis (AS) is a chronic inflammatory disease affecting the sacroiliac joints, spine, peripheral joints and entheses¹. Radiological signs have diagnostic value in AS. Characteristic radiological findings in AS are the growth of syndesmophytes and other features of new bone formation, possibly leading to ankylosis and spinal fusion². The evaluation of disease outcome in AS is a complex and multifactorial issue. Traditionally, various clinical variables like pain and stiffness, spinal mobility and anthropometric measurements have been used to assess outcome in patients with AS^{3,4}. However, these clinical variables are prone to change throughout the day even in the same patient and interfere with momentary disease activity, the use of non-steroidal anti-inflammatory drugs and physiotherapy interventions. Moreover, clinical measurements may show inter- and intra-observer variation^{5,6}. Radiographic scoring methods have been considered to assess the disease outcome in patients with AS. Radiographs have major advantage of providing an objective result of a cumulative process of destruction over time that is not complicated by factors such as diurnal variation and momentary disease activity. Additionally, radiographs provide a permanent record necessary for serial evaluation of the disease course, and they can be randomized and blinded for an objective evaluation⁷⁻¹⁰.

The Stoke Ankylosing Spondylitis Spinal Score (SASSS) and Bath Ankylosing Spondylitis Radiology Index (BASRI) are standardized, reproducible and validated measures of outcome in AS^{7,10-12}. In the early 1990s, Taylor et al¹¹ developed a new detailed radiological scoring system known as the SASSS. The SASSS evaluates lateral view of the lumbar spi-

*Division of Rheumatology, Department of Physical Medicine and Rehabilitation, Firat University, Faculty of Medicine, Elazig, Turkey

**Division of Rheumatology, Department of Physical Medicine and Rehabilitation, Erciyes University, School of Medicine, Gevher Nesibe Hospital, Kayseri, Turkey

ne, with a total score ranging from 0 to 72. Later, SASSS was modified by Creemers et al¹³ by adding the cervical spine to the scoring system. This new version was called the modified-SASSS. Unlike SASSS, BASRI is a global grading method to assess radiological outcome in AS. BASRI score ranges from 0 to 4 in each evaluated part of the spine. Two versions of the BASRI have been developed. The first one was published in 1995, and the second partially modified later^{12,14}. BASRI-total consists BASRI-hip and BASRI-spine, which combines the scores of the lumbar spine, cervical spine and sacroiliac joints^{10,15}. For clinicians, it is important to be aware of the reliability of the measurement and scoring methods which are widely used in clinical studies and daily clinical practice.

In this study we aimed to assess intra- and inter-rater reliability of available radiological scoring methods in AS. Additionally, we aimed to assess the reliability of SASSS with particular reference to score posterior and anterior sites separately and the reliability of the BASRI-hip, which has not been completely clarified.

Patients and Methods

Forty-four complete sets of radiographs were randomly selected from archives of our rheumatology division. These complete sets of radiographs were taken from patients who met modified New York criteria for the classification of AS¹⁶. Patients (33 male and 11 female) had a mean age of 32.8 ± 8.0 years and symptom duration of 10.5 ± 7.8 years. One set of radiographs consisted of a posteroanterior view of the pelvis to score the sacroiliac joints and the hips, an anteroposterior and lateral view of the lumbar spine, and a lateral view of the cervical spine. All radiographs were blinded for patients' identity and numbered.

Scoring procedure

An experienced rheumatologist (SO) trained two physicians (HU and AK, each had ten years of experience in the field of musculoskeletal medicine) for the scoring methods. This training was done in three separate sessions by using the teaching radiographs of our division. Four weeks later, raters (HU and AK) scored radiographs on four sessions including eleven sets of radiographs in each. In every occasion, raters scored all sets according to one method only. Raters noted the time spend to

score all sets of radiographs. For intra-rater reliability, one of the raters (HU) scored the same radiographs 6 weeks later.

For the SASSS scoring, all anterior and posterior corners of each vertebra between the lower border of T12 and the upper border of S1 were examined and scored as: 0 for normal; 1 for sclerosis, erosion or squaring; 2 for syndesmophyte formation; and 3 for total bony bridging, giving a maximum possible score of 36 for each lumbar site (36 for anterior and 36 for posterior)¹¹. A similar process was used for the assessment of only anterior site of the cervical spine between the lower border of C2 and the upper border of T1, giving a maximum possible score of 36¹³. The SASSS-spine is a composite score of the anterior and posterior lumbar SASSS scores and ranges from 0 to 72. The modified-SASSS is also a composite score of the anterior lumbar and anterior cervical SASSS, ranging from 0 to 72^{11,13}. If 3 or fewer scoring sites were missing (lumbar anterior, lumbar posterior and cervical anterior) the mean of the other scoring sites was used as a substitute for the missing sites¹⁷. If more than 3 scoring sites were missing, these radiographs were excluded from the study.

For the BASRI, the lumbar spine was defined extending from the lower border of T12 to the upper border of S1, and the cervical spine extending from the lower border of C1 to the upper border of C7. At the lumbar spine, both anteroposterior and lateral lumbar radiographs were scored and the highest result was accepted as the lumbar spine BASRI. Only lateral cervical radiographs were assessed to obtain cervical spine BASRI. BASRI was scored as 0 for normal; 1 for suspicious changes; 2 for erosions, squaring, sclerosis or syndesmophytes on ≤ 2 vertebrae; 3 for syndesmophytes on ≥ 3 vertebrae and/or fusion involving ≤ 2 vertebrae; 4 for fusion involving ≥ 3 vertebrae on both the lumbar and cervical spine^{10,12}.

Both hips were scored separately according to the BASRI-hip; 0 = normal; 1 = suspicious (possible focal joint space narrowing); 2 = minimal (definite narrowing, leaving a circumferential joint space > 2 mm); 3 = moderate (narrowing with circumferential joint space ≤ 2 mm or bone-on-bone apposition of < 2 cm); 4 = severe (bone deformity or bone-on-bone apposition ≥ 2 cm or total hip replacement)¹⁵.

The New York method, ranging from 0 to 4, was used to score the sacroiliac joints. In this method,

each sacroiliac joint was scored as; 0 = no disease; 1 = suspicious change; 2 = loss of definition at the edge of the joint, there was some sclerosis and minimal erosions, there might be some joint space narrowing; 3 = definite sclerosis on both sides, larger erosions with loss of joint space; and 4 = complete fusion or ankylosis of the joint (without some residual sclerosis)¹⁸.

BASRI-spine is the total of lumbar spine, cervical spine, and sacroiliac scores (New York method, mean of both sites to one decimal place), ranging from 2 to 12¹⁰. BASRI-total is the totted up BASRI-hip (mean of both hips to one decimal place) and BASRI-spine, with a range of 2-16¹⁵.

Statistics

In order to compute inter-rater reliability we used intraclass correlation coefficient (ICC)¹⁹. Analysis were performed by using SPSS version 15.0 (SPSS Inc. Chicago, USA). We chose two-way random model ICC with measures of absolute agreement to calculate inter-rater agreement. Analysis of variance (ANOVA) statistics was applied to the data to get ICC estimate. A two-way random model treats both the differences among objects measured and the variability among raters as random factors. We chose the single measure estimate, which was more conservative and show the reliability of a single rating. Intra-rater reliability was computed using the one-way random model with measures of absolute agreement. Intraclass correlation coefficient value ranges from 0 to 1. According to the criteria of Landis and Koch, ICC estimate of 0.00–0.20 is slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, 0.81 or above is almost perfect agreement²⁰. Sample size requirements for both inter-rater and intra-rater correlation coefficients at 0.80 and 0.90

(inter- and intra-rater reliability, respectively) were estimated using tables by Eliasziw et al (at 0.95 confidence interval and 0.80 power level, two raters and for two ratings per subject)²¹. Sample size required was estimated as >40.

Results

The mean time elapsed to score one set of radiographs was 68 seconds for the SASSS method, 74 seconds for the modified-SASSS method and 42 seconds for the BASRI-spine method. Scoring values of the raters are shown in Table I.

Both intra and inter-rater reliability were almost good to perfect for all the methods and ICC for all the methods was relatively similar to each other (Table II). The intra-rater reliability was relatively higher than the inter-rater reliability for all the methods (Table II).

We also compared the radiological scoring of the anterior site of the lumbar spine with that on the posterior site on the SASSS method. The inter-rater reliability for posterior site of the lumbar spine (ICC 0.81, 95% CI 0.35-0.93) was relatively poorer than anterior site (ICC 0.92, 95% CI 0.35-0.98) of the lumbar spine. Intra-rater reliability of the anterior site was also slightly better with respect to posterior site (ICC 0.98, 96% CI 0.97-0.99 vs ICC 0.94, 95% CI 0.90-0.99, respectively). This finding enhances the notion of using anterior sites of the lumbar and cervical spine in the modified-SASSS method. Regarding the BASRI-hip, inter-rater reliability was similar for right (ICC 0.79, 95% CI 0.65-0.88) and left hips (ICC 0.84, 95% CI 0.72-0.91). The intra-rater reliability was also perfect for both hips (right hip ICC 0.88, 95% CI 0.79-0.93, left hip ICC 0.77, 95% CI 0.61-0.87).

Table I. Scored values of the raters for each method

	Rater 1			Rater 2			Rater 2 second assessment		
	Mean	SD	25-75th Percentile	Mean	SD	25-75th Percentile	Mean	SD	25-75th Percentile
BASRI-spine	8.5	1.56	7.00-10.0	8.09	1.46	7.0-9.0	7.62	1.26	6.75-8.25
BASRI-total	9.22	1.93	7.75-10.0	8.63	2.2	7.0-9.25	7.96	1.88	6.75-8.37
SASSS	23.07	17.25	10.0-32.0	16.7	17.85	7.0-19.75	14.88	16.69	3.25-21.0
Modified-SASSS	27.44	11.93	19.0-32.0	21.56	13.12	12.0-29.0	17.79	12.1	9.0-24.0

BASRI: Bath Ankylosing Spondylitis Radiology Index; SASSS: Stoke Ankylosing Spondylitis Spine Score; SD: Standard Deviation.

Table II. Inter-and intra-rater reliability of the scoring methods ($p < 0.0001$ for all values)

	Inter-rater		Intra-rater	
	ICC	95% CI	ICC	95% CI
BASRI-spine	0.89	0.59-0.96	0.94	0.82-0.98
BASRI-total	0.83	0.56-0.93	0.93	0.78-0.98
SASSS	0.89	0.19-0.97	0.98	0.95-0.99
modified-SASSS	0.86	0.03-0.97	0.97	0.91-0.99

BASRI: Bath Ankylosing Spondylitis Radiology Index; SASSS: Stoke Ankylosing Spondylitis Spine Score; ICC: Intraclass Correlation Coefficient.

Discussion

Structural damage of the spine, such as vertebral squaring, bony erosions, sclerosis, syndesmophytes and ankylosis, is a hallmark of AS. The use of a standardized scoring system for spinal radiographs allows quantification of damage as a measurable outcome for both natural course of the disease and clinical trials. Currently various radiological scoring methods have been validated and used as outcome measures in AS²²⁻²⁴. Any measure that documents disease outcome needs to be reliable and sensitive to change. Additionally, it must be valid, easy and practical to use in daily clinical practice. This study results confirmed that available radiological scoring methods, including SASSS, modified-SASSS, BASRI-spine, BASRI-hip and BASRI-total, have substantial to perfect intra and inter-rater reliability.

The developers of the SASSS method found a good inter-rater reliability but, unexpectedly, poorer intra-rater reliability¹¹. Spoorenberg et al²⁵ reported perfect intra and inter-rater reliability for both the SASSS and modified-SASSS scores, with ICC between 0.92 and 0.98. Similarly, both intra and inter-rater reliability of the SASSS scores were almost perfect in our study. The main problem for SASSS method is the difficulty for scoring the posterior site of the lumbar spine. We also found relatively poorer reliability scores for the posterior site than anterior site of the lumbar spine. Additionally, it should be emphasized that inclusion of the cervical spine in the modified-SASSS gathers important information. MacKay et al²⁶ presented data on the involvement of the cervical spine in a group of 470 patients with AS, 43% of whom showed involvement of both cervical and lumbar spine, and 8% showed changes only at the cervical, but not at the

lumbar spine.

In accordance with previous findings^{10,25} both intra and inter-rater reliability for the composite BASRI scores (BASRI-spine and BASRI-total) were good to perfect in our study. Additionally, the reliability of the BASRI-hip was also good to perfect in agreement with the previous researches^{10,15,25}.

In this study, the mean time spend to score a set of radiographs according to BASRI-spine was <45 seconds, whereas this time was >60 seconds for both the SASSS and modified-SASSS. Similarly, previous studies have reported that SASSS takes more time than the BASRI method²⁷. Calin et al¹⁰ reported that the mean time taken to score the BASRI-total was less than 30 seconds. To maintain simplicity, BASRI does not pick up minor radiologic change including erosion, squaring and sclerosis. The score remains at grade 2 until fusion between 2 vertebrae or the presence of more than 3 syndesmophytes is identified¹². Another advantage of the BASRI method may be the inclusion of hip joints. Hip disease affects 18-37% of the AS population²⁸ and is associated with a more severe outcome²⁹. Using the SASSS or modified-SASSS, which are more detailed methods, is time-consuming with respect to BASRI. However, the modified-SASSS has been identified as the most sensitive radiological method for the evaluation of chronic spinal changes in AS^{23,30,31}. These scoring methods were compared by a group of international experts using criteria developed by the outcome measures in rheumatology clinical trials (OMERACT). They found that SASSS and modified-SASSS were more sensitive than BASRI to detect changes at the spine over a 4-year follow-up. The changes were detected at the lumbar spine in 18% of patients according to BASRI method, in 46% according to SASSS and in 43% according to modified-SASSS. Change were detected at the cervical spine in 23% of patients according to BASRI method and in 41% according to modified-SASSS. SASSS method, which does not include cervical spinal changes, omits a high percentage of patients who show progression in the cervical spine over a 4-year period. Therefore, the modified-SASSS is more preferable than the SASSS^{31,32}.

Unlike rheumatoid arthritis, radiologic progression in AS appears to proceed quite slowly. At least 2 years of follow up is required to sensitively detect spinal changes according to BASRI and SASSS^{10,12,15}.

Evaluation of radiographs with an interval of one year does not seem to be useful^{10,17}. The radiation exposure is an important disadvantage of the all radiological scoring methods. BASRI method requires an anteroposterior view of the lumbar spine which increases the radiation exposure. Taking radiographs with an interval more than 2 years seems to be reasonable and may reduce the risk of radiation. Lack of facet joint examination may be another limitation of BASRI and SASSS methods. According to these methods, a patient with only fusion of posterior elements would have low radiologic scores despite severe spinal involvement. On the other hand, it must be emphasized that the posterior spinal elements are difficult to evaluate even by an experienced musculoskeletal radiologist. Additionally, none of these scoring systems evaluates the thoracic spine, where specific anatomic features and superimpositions complicate radiologic evaluation.

Both BASRI and SASSS methods are significantly correlated with other outcome measures such as the Bath Ankylosing Spondylitis Metrology Index^{3,23,33}. Additionally, the modified-SASSS and the BASRI-spine scores were shown to be correlated well with each other²³.

Conclusions

Radiological scoring methods remain important outcome measures in AS. There are essentially three different scoring methods: BASRI, SASSS and the modified-SASSS. With a training period, all of these methods have substantial to perfect intra and inter-rater reliability. BASRI is a global scoring method that is easy to perform and time conservative. The modified-SASSS is more preferable than SASSS, because it includes the cervical spine. Using SASSS or modified-SASSS is time consuming; however, these detailed methods are more sensitive to change.

Correspondence to

Dr. Hasan Ulusoy
Firat Universitesi, Tip Fakultesi, Fiziksel Tip ve Rehabilitasyon AD
23119 Elazığ, Turkey
E-mail: ulusoyh@mynet.com, drulusoyh@gmail.com

References

1. Sieper J, Braun J, Rudwaleit M, Bonen A, Zink A. Ankylosing spondylitis: an overview. *Ann Rheum Dis* 2002;61:8-18.
2. Braun J, BM, Sieper J. Radiology and pathology of the spondyloarthropathies. *Rheum Dis Clin North Am* 1998;697-735.
3. Jenkinson T, Mollorie P, Whitelock H, Kennedy G, Garrett S, Calin A. Defining spinal mobility in ankylosing spondylitis. The Bath Metrology Index. *J Rheumatol* 1994;21:1694-1698.
4. Creemers MCW, van't Hof MA, Franssen MJAM, van de Putte LBA, Gribnau FWJ, van Riel PLCM. Disease activity in ankylosing spondylitis: selection of a core set of variables and a first step in the development of a disease activity score. *Br J Rheumatol* 1996;35:867-873.
5. Goodacre JA, Mander M, Carson DW. Patients with ankylosing spondylitis show individual patterns of variation in disease activity. *Br J Rheumatol* 1991;30:336-338.
6. Pile KD, Laurent MR, Salmond CE, Best MJ, Pyle EA, Moloney RO. Clinical assessment of ankylosing spondylitis: a study of observer variation in spinal measurements. *Br J Rheumatol* 1991;30:29-34.
7. Aaverns HL, Oxtoby J, Taylor HG, Jones PW, Dziedzic K, Daves PT. Radiological outcome in ankylosing spondylitis: use of the Stoke Ankylosing Spondylitis Spine Score (SASSS). *Br J Rheumatol* 1996;35:373-376.
8. van der Heijde D, Bellamy N, Calin A, Dougados M, Khan MA, van der Linden S, for the Assessment in Ankylosing Spondylitis Working Group. Preliminary core sets for end points in ankylosing spondylitis. *J Rheumatol* 1997;24:2225-2229.
9. van der Heijde D, Spoorenberg A. Plain radiographs as an outcome measure in ankylosing spondylitis. *J Rheumatol* 1999;26:985-987.
10. Calin A, Makay K, Santos H, Brophy S. A new dimension to outcome. Application of the Bath Ankylosing Spondylitis Radiology Index. *J Rheumatol* 1999;26:988-992.
11. Taylor HG, Wardle T, Beswick EJ, Dawes PT. The relationship of clinical and laboratory measurements to radiological change in ankylosing spondylitis. *Br J Rheumatol* 1991;30:330-335.
12. MacKay K, Mack C, Brophy S, Calin A. The Bath Ankylosing Spondylitis Radiology Index: a new validated approach to disease assessment. *Arthritis Rheum* 1998;41:2263-2270.
13. Creemers MCW, Franssen MJAM, van't Hof MA, Gribnau FWJ, van de Putte LBA, van Riel PLCM. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. *Ann Rheum Dis* 2005;64:127-129.
14. Kennedy LG, Jenkinson TR, Mallorie PA, Whitelock HC, Garrett SL, Calin A. Ankylosing spondylitis: the correlation between a new metrology score and radiology. *Br J Rheumatol* 1995;34:767-770.
15. MacKay K, Brophy S, Mack C, Doran M, Calin A. The development and validation of a radiographic grading system for the hip in ankylosing spondylitis: the

- bath ankylosing spondylitis radiology hip index. *J Rheumatol* 2000;27:2866-2872.
16. 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;27:361-368.
 17. Spoorenberg A, de Vlam K, van der Heijde D et al. Radiological scoring methods in ankylosing spondylitis: Reliability and sensitivity to change over one year. *J Rheumatol* 1999;26:997-1002.
 18. Moll JMH, Wright V. New York clinical criteria for ankylosing spondylitis. *Ann Rheum Dis* 1973;32:354-363.
 19. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86:420-428.
 20. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159-174.
 21. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther* 1994;74:777-788.
 22. Brophy S, MacKay K, Al-Saidi A, Taylor G, Calin A. The natural history of ankylosing spondylitis as defined by radiological progression. *J Rheumatol* 2002;29:1236-1243.
 23. Salaffi F, Carotti M, Garofalo G, Giuseppetti GM, Grassi W. Radiological scoring methods for ankylosing spondylitis: a comparison between the Bath Ankylosing Spondylitis Radiology Index and the modified Stoke Ankylosing Spondylitis Spine Score. *Clin Exp Rheumatol* 2007;25:67-74.
 24. Zochling J. Assessment and treatment of ankylosing spondylitis: current status and future directions. *Curr Opin Rheumatol* 2008;20:398-403.
 25. 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; 31:125-132.
 26. MacKay K, Brophy S, Mack C, Calin A. Patterns of radiological axial involvement in 470 ankylosing spondylitis patients [abstract]. *Arthritis Rheum* 1997;40:61.
 27. Mack CS, MacKay KR, Calin A. Radiology in ankylosing spondylitis: a comparison of two scoring systems. *Br J Rheumatol* 1996;35:23.
 28. Khan MA. Ankylosing Spondylitis: clinical features. In: Klippel J, Dieppe P, editors. *Rheumatology. Mosby-Year Book Europe Limited*; 1994. p. 3:25.1-25.10.
 29. Amor B, Santos RS, Nahal R, Listrat V, Dougados M. Predictive factors for the long term outcome of spondyloarthropathies. *J Rheumatol* 1994;21:1883-1887.
 30. Wanders A, van der Heijde D, Landewe R et al. Inhibition of radiographic progression in ankylosing spondylitis (AS) by continuous use of NSAIDs. *Arthritis Rheum* 2005;52:1756-1765.
 31. Wanders AJB, Landewe RBM, Spoorenberg A et al. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the available methods based on the outcome measures in rheumatology clinical trials. *Arth Rheum* 2004;50:2622-2632.
 32. van der Heijde D, Landewe R. Imaging in spondylitis. *Curr Opin Rheumatol* 2005;17:413-417.
 33. Baraliakos X, Listing J, Rudwaleit M, Brandt J, Sieper J, Braun J. Radiographic progression in patients with ankylosing spondylitis after 2 years of treatment with the tumour necrosis factor a antibody infliximab. *Ann Rheum Dis* 2005; 64: 1462-1466.

3e Initiative 2010

Cascais, Portugal
8 a 9 de Outubro 2010