Radiological scoring methods for ankylosing spondylitis: a comparison between the Bath Ankylosing Spondylitis Radiology Index and the modified Stoke Ankylosing Spondylitis Spine Score

F. Salaffi[1], M. Carotti[2], G. Garofalo[1], G.M. Giuseppetti[2], W. Grassi[1]

[1]Department of Rheumatology and [2]Department of Radiology, Università Politecnica delle Marche, Ancona, Italy.

Abstract

Objective

The main objective of the present study was to test the interobserver reliability, truth, discrimination and feasibility of two scoring methods available in ankylosing spondylitis (AS) over a follow-up period of 3 years.

Methods

Two blinded trained observers scored 95 AS radiographs from a cohort of AS patients. Each radiograph was scored by two scoring methods, the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS), and the Bath Ankylosing Spondylitis Radiology Index - spine (BASRI-spine). Interobserver agreement was analyzed by intraclass correlation coefficients (ICC). The construct validity was assessed by examining the correlation of the scoring methods with measures of spinal mobility (Bath Ankylosing Spondylitis Metronomy Index - BASMI), functional limitation (Bath Ankylosing Spondylitis Functional Index - BASFI) and disease duration. Bland and Altman's 95% limits of agreement method and effect size (ES) analysis were used to estimate the smallest detectable difference (SDD) of radiological progression and responsiveness.

Results

The BASRI-spine reached intra- and interobserver ICC of 0.755 and 0.831, respectively. The mSASSS scores were more reliable, with ICC of 0.874 and 0.941, respectively. Both scoring systems correlated significantly with BASMI ($p = 0.01$), while only the mSASSS showed a significant correlation ($p = 0.02$) with BASFI. With regards to sensitivity to change, it was found that mSASSS classified the highest percentage of patients with more changes than the BASRI-spine (mSASSS: 35.8% vs. BASRI-spine: 15.8%). The ES analysis also suggested that the mSASSS was more responsive than BASRI-spine. Concerning feasibility, the BASRI-spine takes less time for scoring.

Conclusion

We have shown that the mSASSS offers advantages in measurement properties and is the most appropriate method by which to assess progression of structural damage in AS.

Key words

Ankylosing spondylitis, radiographic scoring methods, intra- and interobserver reliability, BASRI, mSASSS.

Fausto Salaffi, MD, PhD, Associate Professor of Rheumatology; Marina Carotti, MD, Assistant Professor of Radiology; Giovanna Garofalo, MD, Researcher; Gian Marco Giuseppetti, MD, Head, Professor of Radiology; Walter Grassi, MD, Head, Professor of Rheumatology.

Please address correspondence to: Fausto Salaffi, MD, PhD, Department of Rheumatology – Università Politecnica delle Marche, Ospedale A. Murri, Via dei Colli 52, 60035 - Jesi (AN), Italy. E-mail: fsalaff@tin.it

Received on April 10, 2006; accepted in revised form on September 12, 2006.

Introduction

Ankylosing spondylitis (AS) is an important chronic inflammatory disease with an onset before the age of 40 in about 80% of patients and a prevalence between 0.1% and 0.9% (1-4). It has been estimated that at least 30% of patients develop severe spinal restriction during the natural course of the disease (5, 6). The assessment of the outcome in AS relies on a number of measures although few endpoints are clearly defined (7). Radiological changes are important hallmarks in AS because they reflect the cumulative process of destruction over time (8-10). A number of radiological scoring methods are available for this purpose: the Bath Ankylosing Spondylitis Radiology Index - spine (BASRI-spine) (11), the Stoke Ankylosing Spondylitis Spine Score (SASSS) (12), and a modification of the SASSS (mSASSS) (13). In particular, SASSS assesses the lumbar spine region, the mSASSS the cervical and lumbar spine, and the BASRI-spine, the lumbar, cervical spine and the sacroiliac (SI) joints. All methods have been validated by their developers (11-13). In these studies, some aspects of reliability (intra- and interobserver reliability of status scores) in all 3 methods were established. Moreover, agreements between 2 observers on progression in individual patients was assessed, but only with a strict definition of "agreement." In clinical trials, however, the subject of interest is change in radiographic damage, primarily on the group level, and not the absolute level of damage itself. Apart from that, and according to the Outcome Measures in Rheumatology Clinical Trials (OMERACT), truth (construct validity), discrimination (sensitivity to change of scoring methods), and feasibility should be investigated before a preference is made (7). The main objective of the present study was therefore to test two (mSASSS and BASRI-spine) of the radiographic scoring methods covering all aspects of the OMERACT filter over a follow-up period of 3 years.

Materials and methods

Study population

Our study population included a cross sectional cohort of 129 patients who fulfilled the modified New York Criteria for AS diagnosis (14) and attended the Department of Rheumatology of the Università Politecnica delle Marche. Only 95/129 sets of patients radiographs were available and included in the 3-year follow-up study. Seventy-seven patients were males (mean age ± SD at baseline was 47.9 ± 9.3 years, with a disease duration of 12.4 ± 6.6 years), and 18 patients were females (mean age at baseline 45.9 ± 8.7 years, with a disease duration of 11.3 ± 8.2 years). The disease duration was obtained retrospectively from medical records at the time the radiograph films were performed and from the time of the disease diagnosis. The data was then corroborated by questionnaires administered to patients. All patients completed the Italian version of Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (15) for the activity disease assessment, on the same day as the visit.

Clinical and functional assessment

For each patient, a clinical evaluation was carried out by the same investigator during the initial visit and at the follow up visit, 3 years later. A measure of spinal mobility (Bath Ankylosing Spondylitis Metronomy Index - BASMI) (16), and of functional limitation (Bath Ankylosing Spondylitis Functional Index - BASFI) (15) were used. The BASMI is a combined measure that assesses spine mobility and hip function, it consists of the following items: cervical rotation, tragus to wall distance, lateral spinal flexion, modified Schöber, and intermalleolar distance. Each of these clinical measures is rated on a scale from 0-2, reflecting mild to severe disease involvement. The total score ranges from 0 to 10 (16). The BASFI includes 8 items of daily activities, while 2 assess the patient's ability to cope with everyday life. Each item is answered on a 10 cm horizontal numerical rating scale (NRS). The final score (ranging from 0 to 100) is the average of the scores of the 10 items. Higher scores indicate more severe impairment (15, 17). The BASDAI is a composite index that includes questions on fatigue,

axial pain, peripheral pain, stiffness and discomfort. Responses are reported on a 10 cm NRS (range 0-100 from none to severe) and the index is calculated as a simple sum of its components (15, 18). In addition, the erythrocyte sedimentation rate (ESR) was assessed by the Westergren method (mm/1st h). C reactive protein (CRP) by the turbinimetric method (mg/l), and HLA-B27 by flow cytometry.

Conventional x-rays

All patients had x-ray films those that were taken up to three months from the initial visit were accepted the others had to undergo a new x-ray procedure that consisted of taking an anteroposterior and lateral view of the spine and an anteroposterior view of the pelvis. A well known problem of imaging and scoring procedures occurs when not all images are of perfect quality (over- or underexposure of the radiograph) and the spinal segments are not always completely captured on the film, therefore some sites are missed. In this study we excluded patients who had more than three vertebral sites missed. In 9 of the 95 cases (9.5%) in which ≤ 3 vertebral sites were missed, the missing scores were substituted by the mean value score of the vertebra of the same spinal segment of the patient. The films were scored by two readers (FS, MC) in a random time order sequence (paired films for each patients).

Radiologic scoring methods

Films were available twice: at the initial visit and at the 3 year follow up visit. The BASRI-spine (11) and the mSASSS (13) scoring system were used to analyse the conventional x-ray findings in all patients (Table I). The BASRI-spine score includes the SI joints (still scored according to the New York criteria) and the lumbar and cervical spine. To assess the cervical spine, a lateral view is obtained, and is defined extending from the lower border of C1 to the upper border of C7 (11). To assess the lumbar spine, anterior-posterior (AP) and lateral view are obtained. The lumbar spine is defined extending from the lower border of T12 to the upper border of S1. The highest score of the 2 views of the lumbar spine is applied in the BASRI-spine. The lumbar spine and cervical spine are graded separately on a scale of 0-4. The New York scoring method for the SI joints (19) was followed: 0 = no abnormalities; 1 = suspicious changes (no specific abnormalities); 2 = minimal sacroiliitis (loss of definition at the edge of the SI joints, there is some sclerosis and perhaps minimal erosions, there may be some joint space narrowing); 3 = moderate sacroiliitis (definite sclerosis on both sides of the joint, blurring and indistinct margins, and erosive changes with loss of joint space); 4 = complete fusion or ankylosis of the SI joints (without any residual sclerosis). According to the modified New York criteria (19), at least grade 2 bilaterally or grade 3 or 4 unilaterally is necessary for the diagnosis of AS. The BASRI spine is a composite score and corresponds to the sum of the mean score of the right and left SI joints plus the scores of the lumbar spine and cervical spine. According to the New York criteria, AS patients are supposed to have radiographic sacroiliitis, so the range of the BASRI-spine in patients fulfilling the criteria is 2-12 (19, 20).

The SASSS for the spine used in the modified form as proposed by Creemers *et al*. (13) is different from the original SASSS (12) where the anterior and the posterior border of only the lumbar spine are scored. The mSASSS system contained a score for the lumbar spine and a score for the cervical spine. The mSASSS method (13) scores every corner of the anterior site of the lumbar and cervical vertebrae on a scale fro 0 to 3, in which 0 indicates no abnormality; 1 indicates erosion, sclerosis or squaring; 2 indicates a syndesmophyte; and 3 a bridging syndesmophyte. The total score is the sum of both scores and ranges from 0 to 72 (from 0 to 36 for the cervical spine and from 0 to 36 for the lumbar spine). The cervical spine is scored from the lower border of the second cervical vertebra to the upper border of the first thoracic vertebra, and the lumbar spine is scored from the lower border of the 12th thoracic vertebra to the upper border of the sacrum. The thoracic spine and zygapophyseal joints are not part of this x-ray based system because of the limited ability of two dimensional imaging at these sites. Radiological abnormalities, not related to AS, such as osteophytes, and sites not clearly visible on the radiograph were not considered for scoring (21).

Table I. Radiologic scoring methods for ankylosing spondylitis*.

BASRI-spine (range 2-12) (for the lumbar spine, AP and lateral views are scored, and the view with the highest score is taken; for the cervical spine, lateral view is scored)
0 = normal (no change)
1 = suspicious (no definite change)
2 = mild (any number of erosions, squaring, or sclerosis, with or without syndesmophytes, on ≤ 2 vertebrae)
3 = moderate (syndesmophytes on ≥ 3 vertebrae, with or without fusion involving 2 vertebrae)
4 = severe (fusion involving ≥ 3 vertebrae)

Modified SASSS (range 0–72) (the anterior site of the lumbar spine and the anterior site of the cervical spine from the lower border of C2 to the upper border of T1 are scored on a lateral view)
0 = normal
1 = erosion, sclerosis, or squaring
2 = syndesmophyte
3 = bridging syndesmophyte

New York criteria for sacroiliitis (mean score of both SI joints is used in the BASRI)
0 = normal
1 = suspicious (no definite change)
2 = minimal (minimal sacroiliitis, defined as the loss of definition at the edge of the SI joints, some juxtaarticular sclerosis, minimal erosions, and possible joint space narrowing)
3 = moderate (moderate sacroiliitis, defined as definite sclerosis on both sides, blurring and indistinct margins, and erosive changes, with loss of joint space)
4 = severe (complete fusion or ankylosis of the joints)

*SI: sacroiliac; BASRI: Bath Ankylosing Spondylitis Radiology Index; AP: anteroposterior; SASSS: Stoke Ankylosing Spondylitis Spine Score.

Statistical analysis

Descriptive statistics (mean, standard deviation, median, 95% coefficient interval – 95% CI for the median) are available for BASRI-spine and mSASSS for baseline scores, as well as the progression scores. Intra- and interobserver reliability of the two different scoring methods were analyzed by intraclass correlation coefficient (ICC). The ICC is regarded as excellent if above 0.75, if between 0.4 and 0.75 reliability is defined as fair to good, and below 0.4 reliability is poor. The construct validity of the methods was assessed by examining the correlation of the mean combined radiological scores of the two observers with measures of spinal mobility, functional limitation and disease duration. The correlation was expressed as Spearman's rho. Change over time of the scoring methods was computed by Wilcoxon signed-rank test for nonparametric data. To evaluate responsiveness we applied the effect size (ES) statistic. The ES is calculated as the mean change in score from baseline divided by standard deviation of the baseline scores. The variation in baseline score is a reference against which to judge change. Absolute values of 0.2 are considered small, values of 0.5 are moderate, and those of 0.80 or more represent large effects. Furthermore, in order to assess whether an individual difference between 2 scores in a patient is a real change or whether it is a change that cannot be separated reliably from measurement error, a smallest detectable difference (SDD) was estimated (22). The SDD is a statistical method for defining measurement error and is based on the 95% limits of agreement as described by Bland and Altman (23). So, progression scores smaller than the SDD cannot be distinguished from measurement errors. Differences in the sensitivity for detecting clinically relevant changes by using the interobserver SDD as the threshold level between scoring methods was analyzed with McNemar chi-square tests for paired proportions, with a P value of 0.05 as the significance level. Also, 95% CI of these differences were assessed. Descriptive analyses, kappa statistics, and the Spearman's rho were performed by SPSS 11.0 for Windows. Bland and Altman's plots were analyzed by MedCalc statistical (version 8.2) software.

Table II. Mean values (standard deviation-SD), median, and 95% confidence interval (95% CI) for the median of the demographic, clinical and radiological variables in 95 patients with AS included into study.

	mean ± SD	Median	CI 95% for the median
Age (years)	47.7 ± 9.2	47	44 – 49
Disease duration (years)	12.2 ± 6.8	12	10 – 14
VAS-pain (range 0-100)	50.9 ± 22.9	55	45 – 62
VAS-stiffness (range 0-100	53.2 ± 24.5	55	45.8 – 65.0
BASMI (range 0-10)	5.7 ± 1.8	6	5 – 6
BASDAI (range 0-100)	50.9 ± 23.6	53	39.0 - 64.1
BASFI (range 0-100)	52.1 ± 27.6	55	45.9 - 66.1
CRP (mg /l)	18.2 ± 23.1	16	6.6 – 22.4
ESR (mm/1st h)	29.4 ± 16.1	26	23 - 30
BASRI-spine (range 2–12)	6.0 ± 3.1	5	5 – 6.5
mSASSS (range 0-72)	14.4 ± 4.7	13	11-17

VAS: Visual Analogic Scale; BASMI: Bath Ankylosing Spondylitis Metrology Index; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; CRP: C reactive protein; ESR: erythrocyte sedimentation rate; BASFI: Bath Ankylosing Spondylitis Functional Index); BASRI-spine: Bath Ankylosing Spondylitis Radiology Index-spine; mSASSS: modified Stoke Ankylosing Spondylitis Spine score.

Results

Demographic and clinical data

Demographic data and distribution of the variables tested among the 95 patients are shown in Table II. The patients had a BASDAI of (mean ± SD) 50.9 ± 23.6, a BASMI of 5.7 ± 1.8, and a BASFI of 52.1 ± 27.6. The mSASSS scores had an average baseline score of 14.4 (± 4.7), with a median of 13 (95% CI for the median, 11-17). The BASRI-spine scores had an average baseline score of 6.0 (± 3.1), with a median of 5 (95% CI for the median, 5 - 6.5). The radiographic total scores of both methods were not normally distributed (Kolmogorov-Smirnov test for normal distribution). The distributions of the baseline scores are shown in Figs. 1 and 2. The bar on the left of each graph represents the number of subjects with a score of 0; the bar on the right represents the number of subjects with a maximum possible score. The mSASSS had negligible floor and ceiling effects while with BASRI-spine method the highest possible score was found in 8% of the patients for the cervical spine and in 12% of the patients for the lumbar spine (overall, the 10% of baseline radiographs had reached ceiling effect). Seventy three of the 95 patients (76.8%) with AS were HLA-B27 positive. The mean (SD) CRP was 18.3 (23.1) mg/l and the mean ESR 29.4 (16.1) mm/1st h, respectively (Table II).

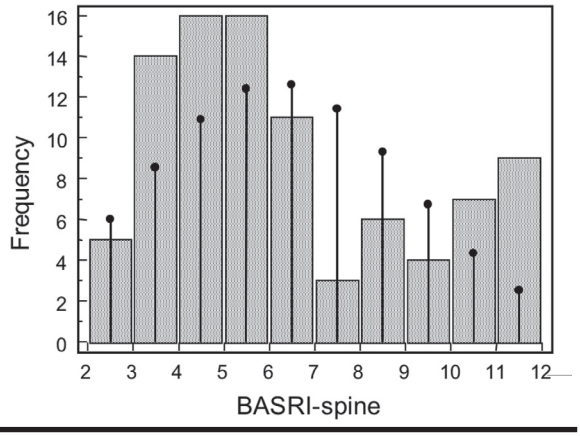


Fig. 1. Distribution of the scores in the Bath Ankylosing Spondylitis Radiology Index - spine (BASRI-spine).

Intra- and interobserver reliability

The intraobserver reliability was good for the BASRI-spine with ICC of 0.755 (95% CI 0.694 to 0.796) and excellent for the mSASSS with ICC of 0.874 (95% CI 0.806 to 0.898). Concerning the interobserver reliability, the BASRI-spine scores of 95 baseline radio-graphs showed excellent reliability with ICC of 0.831 (95% CI 0.798 to 0.881). The mSASSS scored on the anterior sites of both the lateral view of the lumbar and cervical spine (0-72) also showed excellent reliability with interobserver ICC of 0.941 (95% CI 0.891 to 0.972). To illustrate observer agreement over the complete range of observed scores, Figures 3 and 4 show Bland and Altman plots of baseline data of the BASRI-spine and mSASSS. Differences in radiographic scores plotted against the average scores. The mean difference in BASRI-spine scores is -0.2 and the limits of agreement are -2.4 and 2.6 (mean ± 1.96 SD) (Fig. 3). For the mSASSS method, the mean difference in score is 1.1 and the limits of agreement are -2.3 and 4.5 (mean ± 1.96 SD) (Fig. 4). In each case, all observed differences mSASSS scores are within 1.96 SD of the mean difference (dotted line). The SDDs with the BASRI-spine radiographic scoring method, and mSASSS were 2.5, and 3.4, respectively.

Construct validity

We proved the construct validity of the two scoring systems for their correlation with functional findings, as assessed by the BASMI and the BASFI, and with disease duration. The BASMI correlated significantly with the two scoring systems: *rho* = 0.41 compared with the mSASSS (p = 0.001), and *rho* = 0.34 compared with the BASRI-spine (p = 0.01). When the BASFI was compared to the two scoring methods, only the mSASSS, showed a significant correlation of *rho* = 0.31 (p = 0.02). For the two methods, the correlation with disease duration showed the same magnitude (p = 0.0001). No significant correlation has been observed compared to the BASDAI.

Change over time

Figures 5 and 6 show the mSASSS and BASRI-spine scores at baseline and at the 3-year follow-up. After 3 years, the mSASSS showed changes of 3.7 (± 2.6), with a median of 2.5 (95% CI for the median, 1-5) (significant at p = 0.03, by Wilcoxon signed-rank test). The corresponding figures for the BASRI-spinal progression scores were 0.7 points (± 1.1), with a median of 0 (95% CI for the median, 0-1). This difference did not reach statistical significance. Concerning sensitivity to change, if a patient deteriorated or improved more than the SDD, the change was judged as real. Our results demonstrated that the mSASSS quantifies a higher proportion (35.8%) of patients having radiographic progression compared to the BASRI-spine (15.8%). This difference reaches statistical significance (p = 0.02). The results of the ES analysis confirmed that the mSASSS to improvement appeared to be more responsive than BASRI-spine (ES = 0.83 vs 0.24).

Feasibility

This last aspect of the OMERACT fil-

Fig. 2. Distribution of the scores in the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS).

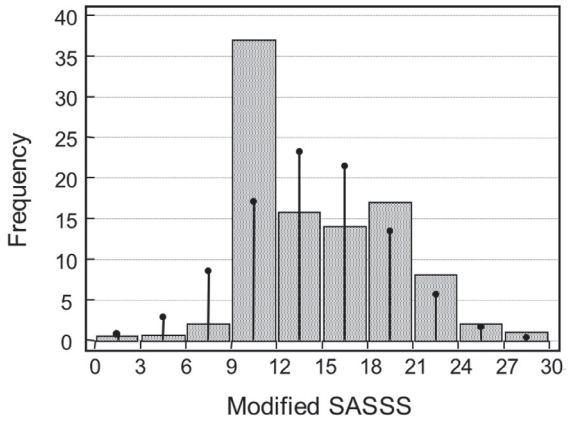


Fig. 3. Bland and Altman's plot of baseline data of the BASRI-spine. Differences in radiographic scores plotted against the average scores. The mean difference in BASRI-spine scores is -0.2 and the limits of agreement are -2.4 and 2.6 (mean 1.96 SD). All observed differences in scores are within 1.96 SD of the mean difference (dotted line).

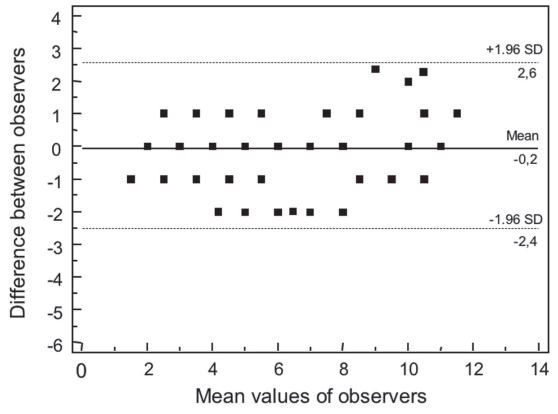
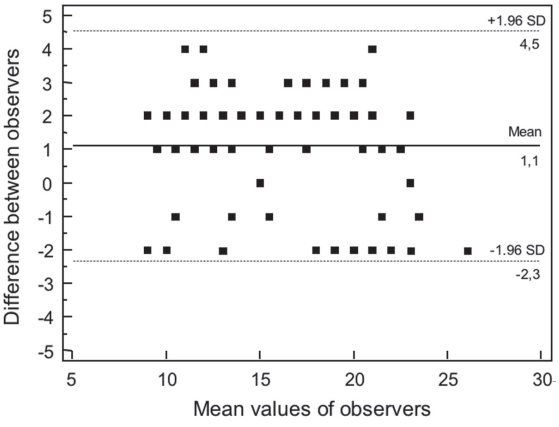


Fig. 4. Bland and Altman's plot of baseline data of the mSASSS. Differences in radiographic scores plotted against the average scores. The mean difference in mSASSS score is 1.1 and the limits of agreement are -2.3 and 4.5 (mean 1.96 SD). All observed differences in scores are within 1.96 SD of the mean difference (dotted line).



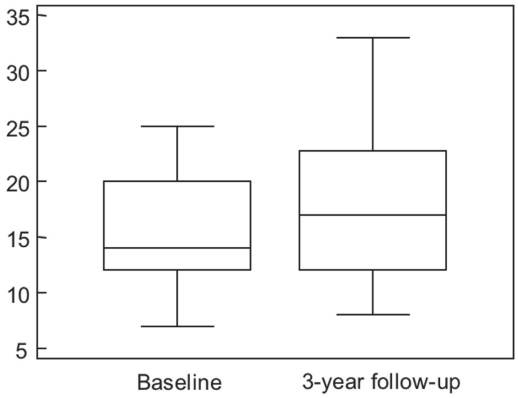


Fig. 5. Baseline and 3-year follow-up of mSASSS scores. The box plots provide information on distribution symmetry, on the numerical measures of central tendency, and on the variability and spread of data in the distribution tails. The box contains the median values (represented by a horizontal line within the box), 25th and 75th percentiles, and whiskers representing the 10th and 90th percentiles.

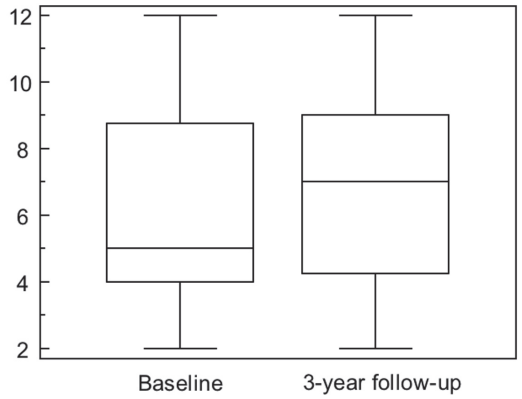


Fig. 6. Baseline and 3-year follow-up of BASRI-spine scores. The box plots provide information on distribution symmetry, on the numerical measures of central tendency, and on the variability and spread of data in the distribution tails. The box contains the median values (represented by a horizontal line within the box), 25th and 75th percentiles, and whiskers representing the 10^{th} and 90^{th} percentiles.

ter focuses on the question of whether the measure can be applied easily given the constraints of time (7, 24). In order to provide insight into these matters, we have given information regarding the time required for scoring the two methods as well as the radiation exposure of the patients. The time needed for scoring the BASRI-spine was less (mean time 3 minutes, range 2 to 6 minutes) than the time needed for scoring the mSASSS (mean time 5 minutes, range 4 to 9 minutes). Concerning the radiation exposure for the patients was as follows (based on data provided by the Radiology Department of the Università Politecnica delle Marche): anteroposterior view of the pelvis = 0.85 mSv, anteroposterior view of the lumbar spine = 0.77 mSv, lateral view of the lumbar spine = 0.95 mSv, and lateral view of the cervical spine = 0.1 mSv. The total exposure for the two scoring methods was 2.67 mSv for the BASRI-spine and 1.05 mSv for the mSASSS.

Discussion

Radiographs of the spine have been selected by the international ASsessment in Ankylosing Spondylitis (ASAS) Working Group as an important outcome domain in clinical trials (25). Several new antirheumatic drugs, such as biologic agents have been recently introduced in the treatment of early aggressive AS, as a matter of fact the recognition of early bone changes is of great benefit in a patient's early treatment response to disease progression (26-28).

Plain radiographs are inexpensive, easy to generate, and widely available and accepted. They also give rapid results and provide a permanent record that can be easily studied in a randomised and blinded fashion. Radiographs are reproducible, allow measurement of severity, and can identify single point damage and progression with fairly high precision and accuracy. In addition, the overwhelming majority of reported data related to radiological progression of AS come from plain film radiography (20, 21, 32, 33).

However, there are several disadvantages using plain film radiography. At first, technical limitations include the need for proper technique and positioning, which can falsely obscure or enhance various findings on an initial film or follow up radiographs. Secondly, there are "floor" and "ceiling" effects related to the available radiological scoring methods for detection of AS induced disease seen on conventional radiographs. The floor effect stems from the fact that the hallmark radiographic findings of syndesmophyte or a bridging syndesmophyte may occur late in the pathophysiology of the disease. The ceiling effect refers to the fact that radiographic progression of disease can continue even after the highest damage score has been assigned by the scoring system. In this case, the BASRI-spine method suffered from a substantial ceiling effect (the highest possible score was found in 10% of our patients, as compared to 1% if assessed by the mSASSS).

Several studies related to the scoring of radiographs of AS patients have been published. Wander *et al.* (29), compared the existing radiographic scoring methods for various aspects of validity. They concluded that mSASSS is the most appropriate method for use in clinical trials. In a recent study the SASSS and the BASRI were found to be reproducible, but both had a rather low sensitivity to change (21). Furthermore, the mSASSS was found to be the most reliable in comparison with the original SASSS and the BASRI for scoring chronic spinal lesions in AS (20). All three *x* ray scoring systems assess only parts of the spine – the SASSS only the lumbar spine, the modified SASSS the cervical and the lumbar spine, and the BASRI-spine the lumbar and cervical spine and the sacroiliac joints. An important disadvantage of the BASRI in comparison with the mSASSS method is that it does not pick up minor radiological changes (21).

It appears that our results are consistent with the results of the above mentioned studies. In fact, the mSASSS seems to be the most appropriate method for scoring radiological progression in AS patients. This conclusion is based on the following aspects of the OMERACT filter: truth, discrimination, and feasibility (7, 24).

With regard to truth, a valid scoring

system requires assessments of relationship of x-ray findings with clinical disease course. Multiple studies have been performed in an attempt to correlate commonly used clinical indices, such as the BASFI, BASMI or other anthropometric and clinical measures with radiographic findings (15, 30, 31). The results have been variable. More recent data from clinical trials, however, indicate a clear relation between functional limitation (BASFI) and structural damage. We found that both scoring methods correlated significantly with BASMI (mSASSS, p = 0.001; BASRI-spine, p = 0.01), while only the mSASSS showed a significant correlation (p = 0.02) with BASFI. For two methods, the correlation with disease duration showed the same magnitude (p = 0.0001). Although, the natural history of the disease is poorly understood, spinal involvement in AS is largely an expression of disease duration. In 2002, Brophy *et al*. (32) showed that AS is a linearly progressive disease with about 35% change or an increase of 2.5 point on the BASRI scale every 10 years. In our study we observed a comparable rate of annual progression on BASRI method in a 3-year follow up.

With regard to discrimination, the mSASSS demonstrated superior intra - and interobserver reliability. In terms of sensitivity to change, this method quantifies a higher proportion of patients as having progression as compared with the BASRI. To determine whether an individual difference between 2 scores in a patient is a real change or whether it is a change that cannot be separated reliably from measurement error, a SDD was estimated. It seems logical that such a cut off value should at least be greater than the measurement error of the instrument used to quantify the response. The SDD expresses the smallest difference between two independently obtained measures that can be interpreted as ''real''– that is, a difference greater than the measurement error (22). In addition, the results of the ES analysis confirmed that the mSASSS appeared to be more responsive than BASRI-spine.

In conclusion, comparing the BASRI-spine and the mSASSS with respect to their use in clinical trials, we have showed that the mSASSS offers advantages in measurement properties. However, the BASRI-spine is a feasible method that takes somewhat less time to perform, but the radiation exposure is more higher because of the additional anteroposterior radiographs at lumbar spine and pelvis.

References

1. SALAFFI F, DE ANGELIS R, GRASSI W: Prevalence of musculoskeletal conditions in an Italian population sample: results of a regional community-based study. I. The MAPPING study. *Clin Exp Rheumatol* 2005; 23: 819-28.
2. 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.
3. BRAUN J, BOLLOW M, REMLINGER G *et al.:* Prevalence of spondyloarthropathies in HLA-B27 positive and negative blood donors. *Arthritis Rheum* 1998; 41: 58-67.
4. SARAUX A, GUILLEMIN F, GUGGENBUHL P *et al.*: Prevalence of spondyloarthropathies in France: 2001. *Ann Rheum Dis* 2005; 64: 1431-5.
5. OLIVIERI I, VAN TUBERGEN A, SALVARANI C, VAN DER LINDEN S: Seronegative spondyloarthritides. *Best Pract Res Clin Rheumatol* 2002; 16: 723-39.
6. ZOCHLING J, BRANDT J, BRAUN J: The current concept of spondyloarthritis with special emphasis on undifferentiated spondyloarthritis. *Rheumatology* 2005; 44: 1483-91.
7. BOERS M, BROOKS P, STRAND CV, TUGWELL P: The OMERACT filter for Outcome Measures in Rheumatology (editorial). *J Rheumatol* 1998; 25: 198-9.
8. BRAUN J, SIEPER J: The sacroiliac joint in the spondyloarthropathies. *Curr Opin Rheumatol* 1996; 8: 275-87.
9. BRAUN J, BOLLOW M, SIEPER J:Radiologic diagnosis and pathology of the spondyloarthropathies. *Rheum Dis Clin North Am* 1998; 4: 697-735.
10. VAN DER HEIJDE D, DOUGADOS M, DAVIS J *et al.*: Assessment in Ankylosing Spondylitis International Working Group/Spondylitis Association of America Recommendations for Conducting Clinical Trials in Ankylosing Spondylitis. *Arthritis Rheum* 2005; 52: 386-94.
11. 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: 2866-72.
12. AVERNS HL, OXTOBY J, TAYLOR HG, JONES PW, DZIEDZIC K, DAWES PT: Radiological outcome in ankylosing spondylitis: use of the Stoke Ankylosing Spondylitis Spine Score (SASSS). *Br J Rheumatol* 1996; 35: 373-6.
13. CREEMERS MC, FRANSSEN MJ, VAN 'T HOF MA, GRIBNAU FW, VAN DE PUTTE LB, VAN RIEL PL: A radiographic scoring system and identification of variables measuring structural damage in ankylosing spondylitis (thesis). Nijmegen (The Netherlands): University of Nijmegen; 1993.
14. 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.
15. SALAFFI F, STANCATI A, SILVESTRI A, CAROTTI M, GRASSI W: Validation of the Italian versions of the Bath Ankylosing Spondylitis Functional Index (BASFI) and the Dougados Functional Index (DFI) in patients with ankylosing spondylitis. *Reumatismo* 2005; 57: 161-73.
16. JENKINSON TR, MALLORIE PA, WHITELOCK HC, KENNEDY LG, GARRETT SL, CALIN A: Defining spinal mobility in ankylosing spondylitis (AS): the Bath AS Metrology Index. *J Rheumatol* 1994; 21: 1694-8.
17. CALIN A, GARRETT S, WHITELOCK H *et al.:* A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Ankylosing Spondylitis Functional Index. *J Rheumatol* 1994; 21: 2281-5.
18. GARRET S, JENKISON T, KENNEDY LG, WHITELOCK H, GAISFORD P, CALIN A: A new approach to defining disease status in ankylosing spondylitis. The Bath Ankylosing Spondylitis Disease Activity. *J Rheumatol* 1994; 21: 2286-91.
19. DALE K: Radiographic gradings of sacroiliitis in Bechterew's syndrome and allied disorders. *Scand J Rheumatol* 1979; 32 (Suppl. 32): 92-7.
20. WANDERS A , LANDEWÉ R, SPOORENBERG A *et al.*: What is the most appropriate scoring method for ankylosing spondylitis? A comparison of the available methods based on the Outcome Measures in Rheumatology Clinical Trials Filter. *Arthritis Rheum* 2004; 50: 2622-32.
21. 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.
22. LASSERE M, BOERS M, VAN DER HEIJDE D *et al.*: Smallest detectable difference in radiological progression. *J Rheumatol* 1999; 26: 731-9.
23. BLAND JM, ALTMAN DG: Comparing methods of measurement: why plotting differences against standard methods is misleading. *Lancet* 1995; 346: 1085-7.
24. VAN DER HEIJDE D, LANDEVÉ R, AND THE ASAS WORKING GROUP: Selection of a method for scoring radiographs for ankylosing spon-dylitis clinical trials, by the Assessment in Ankylosing Spondylitis Working Group and OMERACT. *J Rheumatol* 2005; 32: 2048-9.
25. 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 endpoints in ankylosing spondylitis. *J*

Rheumatol 1997; 24: 2225-9.

26. ZOCHLING J, BRAUN J: Assessment of ankylosing spondylitis. *Clin Exp Rheumatol* 2005; 23 (5 Suppl. 39): S133-41.
27. SALVARANI C, OLIVIERI I, CANTINI F *et al.*: Recommendations for the appropriate use of anti-TNF-alpha therapy in patients with psoriatic arthritis. Italian Rheumatology Society. *Reumatismo* 2004; 56: 133-8.
28. BRAUN J, DAVIS J, DOUGADOS M, SIEPER J, VAN DER LINDEN S, VAN DER HEIJDE D AND FOR THE ASAS WORKING GROUP: First update of the international ASAS consensus statement for the use of anti-TNF agents in patients with ankylosing spondylitis. *Ann Rheum Dis* 2005; 65: 316-20.
29. WANDERS A, LANDEWÉ R, SPOORENBEG A *et al.:* Scoring of radiographic progression in randomised clinical trials in ankylosing spondylitis: a preference for paired reading order. *Ann Rheum Dis* 2004; 63: 1601-4.
30. WANDERS A, LANDEWÉ R, DOUGADOS M, MIELANTS H, VAN DER LINDEN SJ, VAN DER HEIJDE D: Association between radiographic damage of the spine and spinal mobility for individual patients with ankylosing spondylitis: can assessment of spinal mobility be a proxy for radiographic evaluation? *Ann Rheum Dis* 2005; 64: 988-94.
31. 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: 767-70.
32. 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: 1236-43.