Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system

M C W Creemers, M J A M Franssen, M A van ‘t Hof, F W J Gribnau, L B A van de Putte, P L C M van Riel

Objective: To develop and validate an extensive radiographic scoring system for ankylosing spondylitis (AS).

Methods: The Stoke Ankylosing Spondylitis Spinal Score (SASSS) was modified by adding a score for the cervical spine and defining squaring. This modified SASSS (mSASSS) is the sum of the lumbar and cervical spine score (range 0–72). 370 lateral views of the lumbar and cervical spine were used for development of the mSASSS, standardisation of observers, and for studying reliability. In a 48 week NSAID study of 57 patients, change over time and construct validity were studied.

Results: Interoobserver correlations of the lumbar and cervical spine scores were good (r>0.95). The interobserver duplicate error was 0.55 in a range from 0 to 36. The mean change in the cervical and lumbar spine scores between weeks 0 and 48 of all patients was 1.45 (range 0–6.0) and 1.06 (0–5.0), respectively (paired t testing, p<0.001). Change in radiological score was seen in 36/57 (63%) patients (lumbar and cervical spine 11, cervical spine 12, lumbar spine 13 patients).

Conclusion: The mSASSS is useful for assessing extensive radiographic damage in AS. It is reliable, detects changes over 48 weeks, and shows a satisfactory face and construct validity.

We here present the original study of this extended radiographic scoring method—called the modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS)—briefly, because recent studies with the mSASSS have shown promising results compared with the other two published radiographic scoring systems.

METHODS

The study was conducted with radiographs of patients with AS, diagnosed according to the modified New York criteria. Three hundred and seventy lateral views of the lumbar and cervical spine were used for development of the scoring system and standardisation of the two observers (MF, MC). Radiographs were scored blinded for the patient and the previous score.

Reliability

Radiographs, which had not been used previously, were scored twice by two observers, after standardisation of the observers who were unaware of the patient and the previous score. The time between the duplicate scores was 8 weeks in order to guarantee independency of the scores. Interobserver correlations and interobserver duplicate error (Σd2/n) of the different scores were calculated. Intraobserver correlation and intraobserver error were only studied if the interobserver correlation was <0.95.

To study whether this method could detect changes over time of radiographs of 57 patients with AS (43 male, 13 female; mean disease duration 10 years; 54 HLA-B27 positive) participating in a 48 week NSAID trial were used. Radiographs were obtained at weeks 0 and 48 of the trial and scored sequentially blinded for the patient and treatment.

Construct validity was studied in this NSAID study by calculating Pearson correlations at weeks 0, 12, and 48 between the mSASSS and assessed clinical variables, if necessary transformed to obtain normality. The following variables were assessed at weeks 0, 2, 4, 8, 12, 18, 24, 30, 36, 42, and 48: physicians’ assessed variables (performed by a single observer): occiput to wall distance, chest expansion, 10 cm Schober test, lumbar flexion index, lumbar lateral flexion, 53 swollen joint count, enthesis index, and mobility of the cervical spine in all three planes; patients’ assessed variables: duration of morning stiffness, 100 mm visual analogue scale (VAS) score for spinal pain during the day, VAS spinal pain during the night, and VAS general wellbeing; and laboratory variables: erythrocyte sedimentation rate, C reactive protein, haemoglobin, leucocytes, and immunoglobulin A. In addition, every 12 weeks the Dutch Functional Index was completed.

Abbreviations: AS, ankylosing spondylitis; mSASSS, modified Stoke Ankylosing Spondylitis Spinal Score; NSAID, non-steroidal anti-inflammatory drug; VAS, visual analogue scale
The mSASSS system contained a score for the lumbar spine and a score for the cervical spine. The total score is the sum of both scores (range 0–72):

- **Lumbar spine:** the scoring system, developed by Taylor et al., was applied to the lower border of the 12th thoracic vertebra, all five lumbar vertebrae, and the upper border of the sacrum, which were viewed on the lateral radiograph. The corresponding nominal scoring system was used: 0 = no abnormality; 1 = erosion, sclerosis or squaring; 2 = syndesmophyte; and 3 = total bony bridging at each site. Radiological abnormalities, not related to AS, such as osteophytes, and sites not clearly visible on the radiograph were not considered for scoring. Radiographs were only taken into account if no more than three scoring sites were missing. Squaring was defined as present if a line, fictively drawn with a transparent ruler, from the upper and lower border of each vertebral body overlayed 50% or more with the surface of the vertebra, starting at either the upper or lower border; or if the surface of the vertebra was convex, approaching the method of Ralston et al. To deal efficiently with missing observations the total score (range 0–36) was calculated as 12 times the mean score of all scoring sites.

- **Cervical spine:** The scoring system was identical to that of the lumbar spine. The lower border of the 2nd cervical vertebra up to and including the upper border of the 1st thoracic vertebra were viewed on the lateral radiograph. Owing to the original straight shape of the lateral surface of the 3rd cervical vertebra it was decided that this vertebra should not be scored for squaring, although erosions and sclerosis were scored.

### RESULTS AND DISCUSSION

**Reliability**

In total 46 radiographs of the lumbar and 26 of the cervical spine were scored (table 1). The interobserver correlations of the lumbar and cervical spine scores were good ($r>0.95$). Although the cervical spine score showed a statistically significant difference between the scores of the two observers, the interobserver duplicate error of 0.55 in a range from 0 to 36 can be considered as relatively small. Intraobserver correlation and intraobserver duplicate error were not calculated because the interobserver correlation was 0.99.

Radiographs of patients in the NSAID study showed total bony bridging from C7 to Th1, C2 to C3, C3 to C4, and from Th12 to L1. The 12th thoracic, the 1st lumbar, and the 4th cervical vertebrae showed most frequently radiological involvement. In total 51/57 (89%) patients showed involvement of the cervical spine and 37/57 (65%) patients of the lumbar spine at baseline. Differences between men and women were not statistically significant if adjusted for disease duration, which was 11 (SD 8) years and 6 (4) years for men and women, respectively.

### Change over time

The mean change in scores between weeks 0 and 48 of all patients was 1.45 (range 0–6.0) in the cervical spine score and 1.06 (0–5.0) in the lumbar spine score; all changes were significant (paired t testing, $p<0.001$). Change in score was seen in 36 (63%) patients—in 11 patients in both the lumbar and cervical spine, in the cervical spine only in 12 patients, and in the lumbar spine only in 13 patients (fig 1).

The radiographic scores changed statistically significantly. However, it should be noted that the radiographs were scored in sequential order in this longitudinal study, allowing for a positive difference only, for reduction of measurement error.

### Construct validity

Chest expansion, occiput to wall distance, cervical anteflexion, cervical retroflexion, cervical rotation, cervical lateral flexion, lumbar flexion index, and lumbar lateral flexion showed significant Pearson correlations with the mSASSS at weeks 0, 12, and 48 ($p$ value varied from $<0.05$ to $<0.0005$).

In conclusion, this mSASSS is a useful method for assessing radiographic damage in AS: it is reliable, can detect change over 1 year, and has an acceptable construct validity. It shows that radiological involvement is present in the cervical spine when lumbar spine involvement is not found.

---

**Table 1** Means (SD), inter- and intraobserver correlations, duplicate errors, and p values of t testing

<table>
<thead>
<tr>
<th>Observer</th>
<th>Score (systematic differences)</th>
<th>p Value</th>
<th>Interobserver Correlation</th>
<th>Duplicate error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lumbar spine (n = 46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11.78 (13.32)</td>
<td>0.03</td>
<td>0.99</td>
<td>0.55</td>
</tr>
<tr>
<td>2</td>
<td>11.88 (13.48)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cervical spine (n = 26)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6.12 (11.35)</td>
<td>0.70</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>2</td>
<td>6.36 (11.32)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Thus, it should be emphasised that inclusion of the cervical spine score in the mSASSS \(^7\) adds important information, which should be used in the evaluation of treatments such as anti-tumour necrosis factor agents. The mSASSS should be validated further in long term studies, paying particular attention to other aspects of validity such as predictive validity and validation in respect of clinical variables.

\[\text{Authors’ affiliations}\]
M C W Creemers, L B A van de Putte, P L C M van Riel, Department of Rheumatology, University Medical Centre St Radboud, Nijmegen, The Netherlands
M J A M Franssen, Department of Rheumatology, St Maartenshospital, Nijmegen, The Netherlands
M A van ‘t Hof, Department of Medical Statistics, Catholic University, Nijmegen, The Netherlands
F W J Gribnau, Department of Clinical Pharmacology, University Medical Centre St Radboud, Nijmegen, The Netherlands

Correspondence to: Dr M C W Creemers, Department of Rheumatology, University Medical Centre St Radboud, PO Box 9101, 6500 HB Nijmegen, The Netherlands; m.creemers@reuma.umcn.nl

Accepted 5 March 2004
Published Online First 29 March 2003

\[\text{REFERENCES}\]
6. **Creemers M CW.** Clinical and methodological studies in ankylosing spondylitis. The Netherlands: Catholic University Nijmegen, 1994 (thesis.)
Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system

M C W Creemers, M J A M Franssen, M A van 't Hof, F W J Gribnau, L B A van de Putte and P L C M van Riel

Ann Rheum Dis 2005 64: 127-129 originally published online March 29, 2004
doi: 10.1136/ard.2004.020503

Updated information and services can be found at: http://ard.bmj.com/content/64/1/127

These include:

References
This article cites 15 articles, 2 of which you can access for free at: http://ard.bmj.com/content/64/1/127#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

Ankylosing spondylitis (417)
Calcium and bone (725)
Connective tissue disease (4253)
Degenerative joint disease (4641)
Immunology (including allergy) (5144)
Musculoskeletal syndromes (4951)
Rheumatoid arthritis (3258)
Biological agents (545)
Drugs: musculoskeletal and joint diseases (700)

Notes

To request permissions go to: http://group.bmj.com/group/rights-licensing/permissions
To order reprints go to: http://journals.bmj.com/cgi/reprintform
To subscribe to BMJ go to: http://group.bmj.com/subscribe/