EXTENDED REPORT

Reliability of self assessed joint counts in ankylosing spondylitis

A Spoorenberg, D van der Heijde, M Dougados, K de Vlam, H Mielants, H van de Tempel, S van der Linden

....

See end of article for authors' affiliations
....

Correspondence to:
Dr D van der Heijde,
Department of Internal
Medicine, Division of
Rheumatology, University
Hospital Maastricht, PO
Box 5800, 6202 AZ,
Maastricht, The
Netherlands;
dhe@sint.azm.nl

Accepted 27 March 2002
....

Objective: To determine the reliability of self reported joint counts to assess pain or swelling in ankylosing spondylitis (AS).

Methods: 217 outpatients fulfilling the modified New York criteria for AS were asked to mark painful joints and swollen joints on two mannequins presenting 44 and 40 joints respectively. A doctor or research nurse assessed the same joints for pain and swelling on the same day, after completion by the patient, without information on the results of the patient's assessment.

Results: Forty six (21%) patients reported one or more swollen joints (mean number of swollen joints 0.5, range 0–8); the doctor found one or more swollen joints in 54 (25%) of the patients (mean number of swollen joints 0.8, range 0–31). The overall agreement on the number of swollen joints between patients and doctor was moderate (intraclass correlation coefficient (ICC) 0.53). Agreement on individual swollen joints was poor to moderate (κ 0.1–0.64). 128 (60%) patients reported tender joints (mean number of joints 2.4, range 0–26). The doctors reported one of more tender joints in 50% of the patients (mean number of tender joints 2.2, range 0–34). The overall agreement was also moderate (ICC 0.71). The agreement on individual tender joints was again poor to moderate (κ 0.19–0.43). There was only high concordance between doctors and patients on the absence of swollen joints (82%). The concordance on the presence of monoarthritis, oligoarthritis, or polyarthritis was low (17–22%).

Conclusion: Owing to these discrepancies in assessment of individual joints and total number of affected joints, joint counts in AS assessed by doctors cannot be replaced by joint counts reported by the patients. Patients are only able to judge if their joints are not swollen.

In patients with ankylosing spondylitis (AS) a minor part of the population has peripheral arthritis (±20%). Traditionally, either a doctor or a well trained health care professional is involved in the clinical assessment of arthritis. The reliability of joint counts for swelling and pain reported by patients is often studied in the assessment of disease activity in rheumatoid arthritis (RA). To evaluate these joint counts different methods have been used. Some authors used mannequins to mark painful or swollen joints[1–3]; Hanly *et al* used a questionnaire (a modified version of the rapid assessment of disease activity in rheumatology (RADAR) questionnaire)[4 5]; and some authors evaluated both.[6–11] The reported results were ambiguous but there were no differences in the results related to the method used. Some authors reported good reliability and suggested that patients' self reported joint counts can be used to measure disease activity in RA.[1 2 7 10 11] Others found moderate to poor reliability and suggested that joint counts derived by patients could be used but were not interchangeable with joint counts derived by doctors.[3 4 6 8] As far as we know reliability of joint counts reported by patients with AS has not been studied. Our aim was to determine the reliability of self reported swollen joint counts and tender joint counts marked on mannequins by patients with AS.

PATIENTS AND METHODS

The study sample comprised consecutive outpatients with AS at the University Hospital Maastricht, The Netherlands, the Maasland Ziekenhuis Sittard, The Netherlands, the University Hospital Gent, Belgium, and the Hôpital Cochin, Paris, France. These hospitals are secondary and tertiary referral centres. All patients fulfilled the modified New York criteria for AS[12] and are participating in a longitudinal, observational study with follow up visits according to a fixed protocol. The patients were asked to mark their painful joints on a mannequin presenting 44 joints and their swollen joints on a mannequin presenting 40 joints (fig 1). The mannequin diagram was designed after the method of Stewart *et al*.[2] Shoulder and hip joints were not represented on the swollen joint mannequin because it is very difficult to see swelling of these joints, especially by untrained people. At the same day but after the patient assessment, two doctors and one research nurse, one person for each participating centre, assessed the joints for pain and swelling. These results were reported on similar mannequins without knowledge of the patients' assessments. Data were collected at yearly intervals. In this paper the results of the baseline and one year data are presented.

Statistics

Reliability was determined by the intraclass correlation coefficient (I CC, type 3.1) and kappa (κ) statistics. The κ statistic was used for between rater agreement on categorical data such as the individual joint scores. The ICC was used for overall agreement on linear data such as the total number of tender and swollen joints. To visualise this overall agreement we plotted the data using the method of Bland and Altman[13] and calculated the 95% limits of agreement. This method is designed as an absolute measure of agreement between two

....

1. Please indicate with a mark on the picture below all joints that are painful at present

RIGHT   LEFT

Sternal clavicular joints

Shoulder joint

Elbow joint

Hip joint

Wrist

Finger joints

Knee joint

Ankle joint

Toe joints

2. Please indicate with a mark on the picture below all joints that are swollen at present

RIGHT   LEFT

Sternal clavicular joints

Elbow joint

Wrist

Knee joint

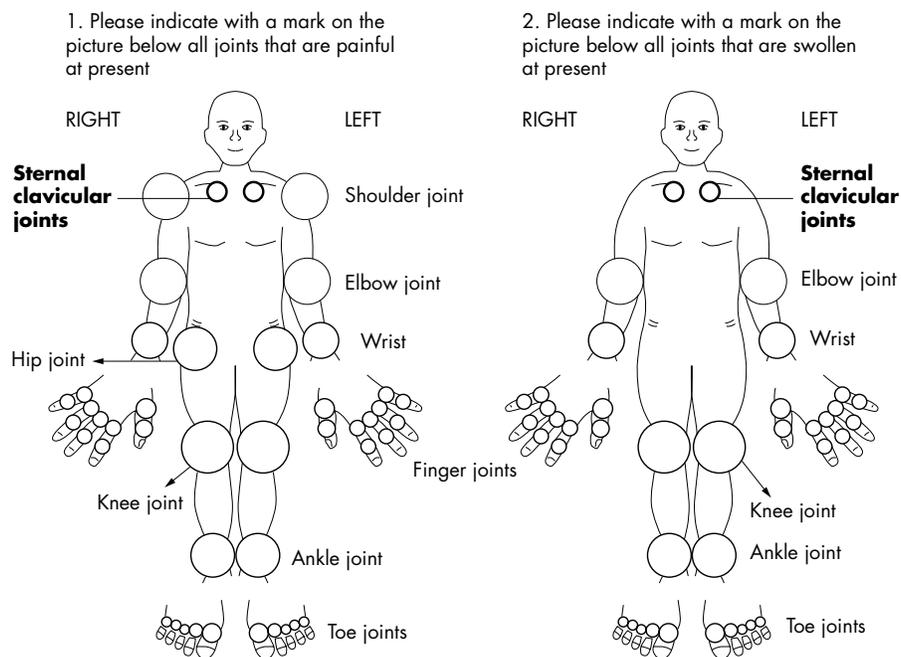Ankle joint

Toe joints

Figure 1 Joint mannequins

RESULTS

instruments which are on the same scale of measurement. To visualise this, the difference between two observations was plotted against the mean of the pairs of observations. Furthermore, Spearman's correlation coefficients were computed for data without a normal distribution. All analyses were done with SPSS 10.0 for Windows.

RESULTS

There were 217 outpatients in our study, with a male to female ratio of 2:1. Table 1 describes the demographic and clinical features of all patients. Sixty one (28%) patients had finished more than secondary school and 39 (18%) attended elementary school only. In 58 (27%) patients peripheral arthritis was diagnosed by the treating rheumatologist. Psoriasis was diagnosed in 10 (4.6%) of the patients and dactylitis in 20 (9.2%) during the whole course of the disease. The mean score of the Bath ankylosing spondylitis disease activity index, BASDAI[14] and the Bath ankylosing spondylitis functional index (BASFI)[15] indicated an overall mild disease activity and mild functional impairment for this group of patients with AS. The mean score for the Bath ankylosing spondylitis radiology index (BASRI)[16] of the lumbar and cervical spine was 1.9 (range 0–4) with 93 (43%) patients having at least three syndesmophytes at the lumbar and/or cervical spine, indicating moderate to severe damage. The enthesis index according to Mander *et al* was also computed with a mean score of 7.7 (range 0–90) at baseline (table 1).[17]

For baseline and one year data, both patients and doctors reported more tender than swollen joints (table 2). The average tender joint count and the swollen joint count were comparable between doctors and patients; however, there was a striking difference in the maximum number of swollen joints scored by the doctors (31) compared with the number scored by the patients (eight). The overall between observer agreement (ICC) between the patients and doctors for the total number of tender joints was moderate (0.71 and 0.54 for baseline and one year respectively) and was slightly worse for the total number of swollen joints (0.53 and 0.51 for baseline and one year respectively). There was no difference in agreement between patients and the two doctors and patients and research nurse for both swollen joint counts (0.53 and 0.54 respectively) and tender joint counts (0.71 and 0.70

respectively). In general, the baseline and one year data were very similar. For the remaining analyses we present baseline data only. By contrast with RA, most patients with AS have just a few inflamed joints. Therefore we also analysed the data in a different way: the number of patients with none or one, two, three, or more than three swollen joints (table 3). The doctors found one or more swollen joints in 54 of 217 (25%) patients whereas 44 of 214 (21%) of the patients reported one or more swollen joints. These percentages were very similar. However, this is misleading as there was low concordance on one or more swollen joints between the patients and doctors (51%). So the patients who judged that they had one or more swollen joints were often different from those judged by the doctor to have swollen joints. When the doctors' assessment was used as the gold standard, our results indicated that patients with AS can judge whether their joints are not swollen (specificity 93%) but have difficulty judging one or more swollen joints (sensitivity 61%).

According to the doctors 50% of the patients had one or more tender joints as opposed to 60% according to the patients. Again the concordance on assessing one or more

Table 1 Characteristics of the patients (n=217) presented as mean (SD, min–max) or percentage

Age (years)	43 (12, 18–77)
Male	67%
Disease duration since diagnosis (years)	9.2 (8.6, 0–42)
Reported peripheral arthritis*	27%
Reported psoriasis*	4.6%
Reported dactylitis*	9.2%
Formal education	
>12 years	28%
≤6 years	18%
BASRI† (patients with ≥3 syndesmophytes at lumbar and/or cervical spine)	43%
BASDAI‡ (range 0–10)	3.5 (2.1, 0–9.7)
BASFI§ (range 0–10)	3.4 (2.6, 0–10)
Mander enthesis index (range 0–90)	7.7 (11.1, 0–56)

*Diagnosed by treating rheumatologist, ever during the disease course; †Bath ankylosing spondylitis radiology index; ‡Bath ankylosing spondyitis disease activity index; §Bath ankylosing spondylitis functional index.

Table 2 Summary statistics of total number of tender and swollen joints (n=217)

	Physician mean (median, centiles 10 and 90)	Patient mean (median, centiles 10 and 90)	Between observer agreement (ICC)
Tender joints:			
Baseline	2.2 (1.0, 0.0–7.0)	2.4 (1.0, 0.0–6.0)	0.71
1Year follow up	2.1 (0.0, 0.0–7.0)	3.1 (2.0, 0.0–8.7)	0.54
Swollen joints:			
Baseline	0.8 (0.0, 0.0–2.0)	0.5 (0.0, 0.0–2.0)	0.53
1 Year follow up	0.5 (0.0, 0.0–1.0)	0.5 (0.0, 0.0–2.0)	0.51

Table 3 Number of painful and swollen joints, concordance, and distribution of root joints*

	Painful				Swollen			
Affected joints	Physician	Patient	Perfect concordance (%)	Root joints (%)	Physician	Patient	Perfect concordance (%)	Root joints (%)
0	107	86	53		162	170	82	
1 Monoarthritis	30	26	12	80	22	20	17	50
2 or 3 Oligoarthritis:			23				17	
Patients	30	49			20	14		
Joints	70	115		63	48	32		50
>3 Polyarthritis:			47				22	
Patients	47	53			12	10		
Joints	320	372		27	97	54		0

*Root joints are shoulders, hips, and sternoclavicular joints. Only sternoclavicular joints were scored at the swollen joint mannequin. The percentages of root joints are given for the concordant pairs: agreement between patient and physician.

tender joints was rather low (60%). Sensitivity of the patients' judgement on tender joints was rather high (82%) but the specificity was low (62%).
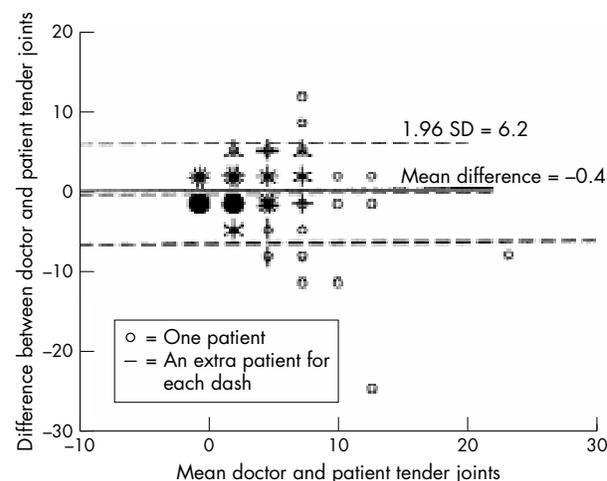
Table 3 shows the number of tender joints and swollen joints and concordance rate of patients and doctors if the assessed joints were split into four categories: no arthritis, monoarthritis, oligoarthritis, or polyarthritis. The only high concordance rate found was 82% (category, non-affected swollen joints) again suggesting that patients could only judge whether their joints were not swollen. The other concordance rates were at best moderate but overall they were low.

The distribution of monoarthritis and oligoarthritis according to the doctors gave more or less the expected distribution in patients with AS (table 3). In tender joints we found 53% and 34% involvement of root joints (shoulders and hips) and mostly large joints were affected instead of small joints in the hands and feet.

Analysis by κ statistics showed moderate to poor and non-consistent agreement between doctor and patients on individual joint counts for either pain or swelling (table 4). Because there were only very few affected small joints in the

hands and feet we clustered these for statistical analysis using the ICC instead of κ statistics.

Figure 2 shows the Bland and Altman plot of the total number of tender joints. There was a maximum difference of 25 joints between the doctors and the patients on the scoring range of 0 to 40; the 95% limit of agreement of the difference was 6.2 (1.96*SD). It also showed that the doctors consistently scored somewhat lower than the patients (mean difference −0.4). The Bland and Altman plot of the total number of swollen joints showed similar results (fig 3). However, now the doctors consistently scored somewhat higher than the patients. There was one outlier, which showed a difference of 26 swollen joints between doctor and patient. Both Bland and Altman plots showed the influence of the number of affected joints: by increasing the number of affected joints, the disagreement between patients and doctors was slightly larger although this was based on few patients.

Table 4 Levels of agreement on individual joints (κ)

	Tender joints		Swollen joints	
	Left	Right	Left	Right
Sternoclavicular	0.38	0.23	0.53	0.36
Shoulder	0.42	0.27	–	–
Elbow	0.43	0.30	0.10	0.10
Wrist	0.37	0.32	0.48	0.48
Hand (MCP,PIP)	0.73*	0.62*	0.05*	0.25*
Hip	0.26	0.19	–	–
Knee	0.42	0.34	0.52	0.50
Ankle	0.40	0.27	0.48	0.49
Foot (MTP)	0.72*	0.57*	0.64	0.23

*Intraclass correlation coefficient (because of linear data).
MCP, metacarpophalangeal; MTP, metatarsophalangeal; PIP, proximal interphalangeal.

Figure 2 Bland and Altman plot: mean versus difference between patient and doctor for total number of tender joints.
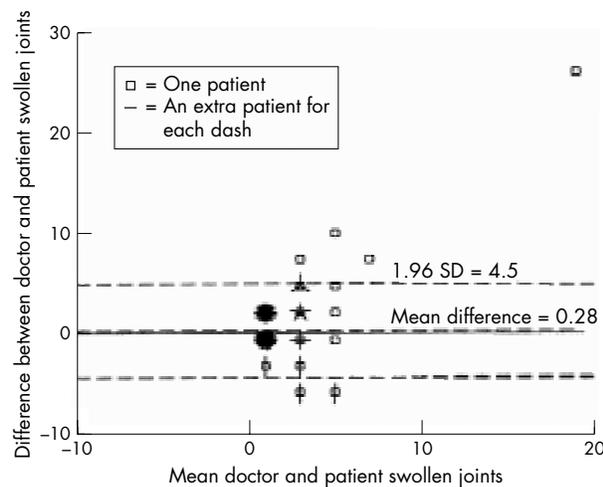
Figure 3 Bland and Altman: mean versus difference between patient and doctor for total number of swollen joints.

We also computed Spearman's correlations of the total number of tender and swollen joints judged by doctors and the patients with the enthesis index according to Mander *et al* at baseline and dactylitis.The highest correlation found was 0.66 for tender joints assessed by doctors with the enthesis index. The correlations of swollen joints assessed by the doctors, and tender and swollen joints assessed by the patient with the index of Mander *et al* were 0.38, 0.41, and 0.30 respectively. Correlations with dactylitis were low; 0.23 for both tender joints and swollen joints assessed by the doctors and 0.27 and 0.23 for tender joints and swollen joints assessed by the patients.

DISCUSSION
Collecting accurate and reproducible information from patients in routine rheumatology practice, epidemiological surveys, and clinical trials is often labour intensive and time consuming. This is one of the reasons for an increased use of self administration forms such as questionnaires on function, disease activity, and quality of life to assess the course and outcome of the disease. If joint counts assessed by doctors could be replaced by joint counts assessed by the patients, this would again lighten the task for rheumatologists in practice and researchers in particular. It would also make it easier to collect these data more often and even with postal questionnaires. However, to be able to replace joint counts derived by doctors with joint counts derived by patients, the validity of the second needs to be assessed. So far, studies comparing results from patients with those from doctors have only been carried out on patients with RA. Five of these studies showed good reliability,[1 2 7 10 11] by contrast with four that showed only poor to moderate reliability.[3 4 6 8] The authors of the second group of studies concluded that joint counts derived by patients could be used, but were not interchangeable with those derived by doctors. Furthermore, reliability for swollen and tender joint counts was the same[3 11] or slightly better for assessing tender joints.[6 9]

As far as we know this is the first study in AS to compare the patient's assessment of tender joint and swollen joint counts with that of the doctor as a gold standard. A major difference between patients with RA and those with AS is the fact that in RA all patients' peripheral joints are affected, although to a different degree during the course of the disease. In patients with AS only about 20 to 30% of the patients have involvement of peripheral joints. Root joints are involved in about 30% of patients and this is relatively more common in those with juvenile onset AS.[18] Moreover, if patients with AS have involvement of peripheral joints, this is often to a lesser extent than patients with RA and often a different pattern of joints is involved.

Our results show that, on a group level, there is a consistent difference between the number of tender and swollen joints assessed by the patients or by the doctors. The patients score consistently more tender joints, and the doctors more swollen joints. An explanation for the first finding could be that it is difficult for patients to differentiate between a tender joint and pain caused by enthesitis, although according to our data the enthesis index of Mander *et al* was only significantly correlated with the total number of tender joints assessed by the doctor. This was possibly because both assessments used the same methodology and most peripheral entheses are located near the joint. The second finding could be caused by the fact that patients with AS are not educated as to what a swollen joint means. From the group level results it could be concluded that absolute scores assessed by patients cannot replace those assessed by doctors. If we look on a patient level, the results are even worse. The 95% limits of agreement were ±6.2 for tender joints, indicating that there may be a difference of 12 joints between the patient and doctor assessment, whereas there is no real difference. For the swollen joints the 95% limits of agreement were ±4.5.

Although actual joint counts assessed by a doctor cannot be replaced by self assessed joint counts, self assessment could still be valuable if the patients could differentiate between the absence of arthritis and the presence of monoarthritis, oligoarthritis, or polyarthritis. Again the concordance rates were very low for all groups of tender joints, and for the various levels of swollen joints. The only good concordance rate was in the absence of swollen joints. Consequently, patients are able to tell if they do not have swollen joints. However, if they have swollen joints, they are unable to judge the extent of the swelling, even within rough categories of monoarthritis, oligoarthritis, and polyarthritis. Perhaps further studies could investigate if training of the patients would make a difference.

A limitation of our study is that we did not assess test-retest reliability formally. However, the results obtained at baseline and after one year of follow up showed very similar results, indicating good reliability.

Our results show a major discrepancy between the number of tender joints and swollen joints assessed by a doctor or the patient. Therefore, joint scores derived by doctors cannot be replaced by self assessed joint scores in AS. The only reliable result is the judgment of the patient that no joints are swollen.

. . . .

Authors' affiliations
A Spoorenberg, D van der Heijde, S van der Linden, University Hospital Maastricht, Maastricht, The Netherlands
D van der Heijde, Limburg University Centre, Diepenbeek, Belgium
M Dougados, Hôpital Cochin, Paris, France
K de Vlam, H Mielants, University Hospital Gent, Gent, Belgium
H van de Tempel, Maasland Hospital, Sittard, The Netherlands
A Spoorenberg, Leeuworden Medical Centre, The Netherlands

REFERENCES
1 **Escalante A**. What do self-administered joint counts tell us about patients with rheumatoid arthritis? Arthritis Care Res 1998;11:280–90.
2 **Stewart MW**, Palmer DG, Knight RG. A self-report articular index measure of arthritic activity: investigations of reliability, validity and sensitivity. J Rheumatol 1990;17:1011–5.
3 **Prevoo ML**, Kuper IH, van't Hof MA, van Leeuwen MA, van de Putte LB, van Riel PL. Validity and reproducibility of self-administered joint counts. A prospective longitudinal follow up study in patients with rheumatoid arthritis. J Rheumatol 1996;23:841–5.
4 **Hanly JG**, Mosher D, Sutton E, Weerasinghe S, Theriault D. Self-assessment of disease activity by patients with rheumatoid arthritis. J Rheumatol 1996;23:1531–8.
5 **Mason JH**, Anderson JJ, Meenan RF, Haralson KM, Lewis Stevens D, Kaine JL. The rapid assessment of disease activity in rheumatology (radar) questionnaire. Validity and sensitivity to change of a patient self-report measure of joint count and clinical status. Arthritis Rheum 1992;35:156–62.
6 **Alarcon GS**, Tilley BC, Li SH, Fowler SE, Pillemer SR. Self-administered joint counts and standard joint counts in the assessment of rheumatoid arthritis. J Rheumatol 1999;26:1065–67.

7 **Abraham N**, Blackmon D, Jackson JR, Bradley LA, Lorish CD, Alarcon GS. Use of self-administered joint counts in the evaluation of rheumatoid arthritis patients. Arthritis Care Res 1993;6:78–81.

8 **Calvo FA**, Calvo A, Berrocal A, Pevez C, Romero F, Vega E, *et al.* Self-administered joint counts in rheumatoid arthritis: comparison with standard joint counts. J Rheumatol 1999;26:536–39.

9 **Houssien DA**, Stucki G, Scott DL. A patient-derived disease activity score can substitute for a physician-derived disease activity score in clinical research. Rheumatology 1999;38:48–52.

10 **Stucki G**, Stucki S, Bruhlmann P, Maus S, Michel BA. Comparison of the validity and reliability of self-reported articular indices. Br J Rheumatol 1995;34:760–6.

11 **Wong AL**, Wong WK, Harker J, Sterz M, Bulpitt K, Park G, *et al.* Patient self-report tender and swollen joint counts in early rheumatoid arthritis. J WC. 1999;26:2551–61.

12 **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.

13 **Bland JM**, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;i:307–10.

14 **Garrett S**, Jenkinson 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 index. J Rheumatol 1994;21:2286–91.

15 **Calin A**, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, *et al.* A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath ankylosing spondylitis functional index. J Rheumatol 1994;21:2281–5.

16 **Calin A**, MacKay K, Santos H, Brophy S. A new dimension to outcome: application of the Bath ankylosing spondylitis radiology index. J Rheumatol 1999;26:988–92.

17 **Mander M**, Simpson JM, Mc Lellan A, Wahler D, Goodacre JA, Dick WC. Studies with an enthesis index as a method for patients with ankylosing spondylitis. Ann Rheum Dis 1987;46:197–202.

18 **Khan MA**, Ankylosing spondylitis: clinical features. In: Klippel JH, Dieppe PA, eds. *Rheumatology*. St Louis: Mosby, 1998;6:16.2–3.