Rheumatology 2004;43:750–757 doi:10.1093/rheumatology/keh169
Advance Access publication 7 April 2004

Spinal mobility in ankylosing spondylitis: reliability, validity and responsiveness

K. L. Haywood, A. M. Garratt, K. Jordan[1], K. Dziedzic[2] and P. T. Dawes[3]

***Objective*. To evaluate the measurement properties of an evidence-based selection of measures of spinal mobility in patients with ankylosing spondylitis (AS).**

***Methods*. Measurements of spinal mobility were taken by trained observers within a UK rheumatology centre. Inter-observer reliability was assessed. Intra-observer reliability was assessed in patients reporting no change in AS-specific health at 2 weeks. Validity was assessed and scores were correlated with responses to health transition questions. Responsiveness was evaluated for patients reporting change in health at 6 months.**

***Results*. Reliability estimates support the use of all measures in individual evaluation (intraclass correlation > 0.90). Correlations between measures of spinal mobility were in the hypothesized direction; the largest was between the modified Schober index (15 cm) (MSI) and the other measures. As hypothesized, small to moderate levels of correlation were found between mobility measures and patient-assessed health status. There was no significant linear relationship between mobility measures and self-reported health transition. Fingertip-to-floor distance following trunk forward flexion (FFD) was the most responsive mobility measure but was not as responsive as two AS-specific patient-assessed instruments, the Ankylosing Spondylitis Quality of Life Questionnaire (ASQoL) and the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI). The MSI and cervical rotation (Crot) also had evidence of responsiveness. Low levels of responsiveness were found for the remaining measures.**

***Conclusion*. All mobility measures had adequate levels of reliability and validity. The MSI had a strong relationship with all mobility measures, and the FFD and Crot were the most responsive to self-perceived changes in health at 6 months. The MSI, FFD and Crot are recommended for clinical practice and research.**

KEY WORDS: Ankylosing spondylitis, Spinal mobility, Reliability, Validity, Responsiveness, Patient-assessed health.

Limited spinal mobility is a cardinal sign of ankylosing spondylitis (AS), featuring strongly in the AS diagnostic criteria (Modified New York Criteria) [1]. Providing insight into the natural history of disease, the assessment of spinal mobility assists in the identification of patient sub-groups and informs clinical outcome in AS [2–4]. Consequently, the Assessment in AS international group (ASAS) has recommended spinal mobility as a core domain in the evaluation of patients in both clinical practice and trials [5, 6], confirming its use in routine practice [7–9]. However, its inclusion in Ankylosing Spondylitis diagnostic criteria [7, 8] may have influenced measurement selection while neglecting evidence of necessary measurement properties. Whereas measures included in diagnostic criteria are not required to be responsive to change in clinical status over time, this is an essential feature of measures used in the evaluation of health-care [10].

Following a structured literature review that synthesized published evidence up to 2000 for measurement reliability, validity and responsiveness of spinal mobility measures in AS assessment, 41 measures met the review inclusion criteria [11]. Although most measures had similar levels of reliability, several had consistently higher levels, supporting application in individual assessment, including tape measure [12, 13] or inclinometer [12–17] assessment of cervical rotation and the measurement of occiput [12, 13, 15, 16, 18] or tragus-to-wall distance [13, 14, 18, 19] (OWD and TWD respectively). Good evidence of validity was found for the same assessments [12, 13, 15, 16, 20–26] and for cervical lateral flexion [12, 13], chest expansion (4th intercostal space) [13, 15, 26–30], the Modified Schober Index (MSI) (15 cm) [3, 13, 21–23, 25, 27, 29–34] and lateral lumber flexion assessment [13, 21, 22, 25, 27, 32, 35]. There was limited evidence of responsiveness for all measures. The most extensive evidence was found for the MSI, which indicated that it was not responsive to short-term change following either drug [for example, 36, 37] or physical therapy [for example, 21, 38]. However, fingertip-to-floor distance (trunk forward flexion; FFD) had good evidence of responsiveness following physical therapy [for example, 16, 26, 39, 40].

Eleven measures were initially selected for further evaluation of feasibility and reliability in a pre-pilot study [11, 41]. Six approaches were excluded due to poor clinical feasibility: cervical rotation (inclinometer) [39, 42], FFD (vertically mounted ruler) [19], lateral lumbar flexion (LLF) (skin distraction [35] and vertically mounted ruler [43]), OWD (tape measure) [44] and TWD (T-square) [19]. Five alternative measures of spinal mobility were considered suitable for further evaluation: cervical rotation (Crot) (tape measure) [12], FFD (trunk forward flexion) [43], LLF [14, 43], MSI [27, 31] and TWD [14, 19] (Table 1). Informed by the pre-pilot evaluation, the mounted ruler (FFD and LLF) [43] and T-square (TWD) [19] were replaced with a retractable steel ruler

Unit of Health-Care Epidemiology, Department of Public Health, University of Oxford, Oxford OX3 7LF, [1]Primary Care Sciences Research Centre and [2]Department of Physiotherapy Studies and Primary Care Sciences Research Centre, Keele University, Keele, Staffordshire ST5 5BG and [3]Staffordshire Rheumatology Centre, University Hospital of North Staffordshire NHS Trust, Staffordshire ST6 7AG, UK.

Submitted 27 October 2003; revised version accepted 4 February 2004.

Correspondence to: K. L. Haywood, Unit of Health-Care Epidemiology, Department of Public Health, University of Oxford, Oxford OX3 7LF, UK. E-mail: kirstie.haywood@uhce.ox.ac.uk

TABLE 1. Measures of spinal mobility and patient-assessed health status

Measure	Description[a]
Spinal mobility	
Cervical rotation (Crot)[c]	[b]Distance between tip of nose and ACJ in neutral and maximal ipsilateral rotation. Difference between two positions calculated for right/left rotation. Smaller difference indicates a more restricted range. Measured with plastic tape measure
Fingertip-to-floor distance (lumbar forward flexion; FFD)	[b]Distance between tip of right middle finger and the floor following maximal lumbar flexion, whilst maintaining knee extension; smaller distance indicates greater movement Measured with a retractable steel tape measure
Lumbar lateral flexion (LLF)[c]	[b]Distance between tip of ipsilateral middle finger and floor following maximal LLF, maintaining heel contact with floor and without trunk rotation. Smaller distance indicates greater movement. Measured with a retractable steel tape measure
Modified Schober index (15 cm) (MSI) [27, 31][c]	Distance between two marks placed 15 cm apart in standing (10 cm proximal and 5 cm distal to the PSIS) following maximal forward flexion of the spine. Larger difference indicates greater lumbar movement. Measured with a plastic tape measure
Tragus-to-wall distance (TWD)[c]	[b]Horizontal distance between right tragus and wall, standing with heels and buttocks against the wall (to prevent pivoting), knees extended and chin drawn in. Larger distance indicates worse spinal/upper cervical posture. Measured with a retractable steel tape measure
Patient-assessed health instruments	
Ankylosing Spondylitis Quality of Life (ASQoL) Questionnaire [58]	18-item, AS-specific HRQL. Response 'Yes' or 'No'. Score 0–18; 0 is better HRQL
Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) [59]	6-item, AS-specific disease activity. Response on a 6×10 cm horizontal VAS. Score 0–10; 0 is less disease activity
Revised Leeds Disability Index (RLDQ) [21]	16-item, AS-specific functional disability. Four-point ordinal response scale. Perceived activity completion. Response 'Yes, with no difficulty' (0) to 'Unable to do' (3). Score 0–48; 0 is better functional ability

ACJ, acromioclavicular joint; HRQL, health-related quality of life; PSIS, posterior superior iliac spine; VAS, visual analogue scale.
[a]All mobility measures were practised once before the movement was recorded.
[b]Instrumentation (and landmarks for cervical rotation) were modified following pre-pilot evaluation [11, 41].
[c]Similarity between ROM selection for present study and the Bath Ankylosing Spondylitis Metrology Index (BASMI) [14]: BASMI Crot was assessed with an inclinometer (patient supine); BASMI FFD and LLF use a mounted ruler [43]; BASMI TWD uses a T-square [19].

which combined both the portability and adaptability of a tape measure with the rigidity of a mounted ruler. The landmarks for cervical rotation were also modified, identifying the tip of the nose as a more fixed facial landmark than the chin (Table 1). High levels of reliability were found for these selected measures [intraclass correlation coefficient (ICC) > 0.85] [11, 41].

The aim of this study was to examine the measurement properties of this evidence-based selection of spinal mobility measures in AS patients recruited from a UK rheumatology centre. The results of the study will inform recommendation of which measures of spinal mobility to include in routine practice and clinical research.

Methods

Data collection

Following published clinic-based studies [26], a sample of 150 patients was considered appropriate for the evaluation of measures of spinal range of movement (ROM). A random sample of 269 patients with a confirmed clinical diagnosis of AS (Modified New York Criteria) [1], registered with a specialist centre of rheumatology in England and aged between 18 and 75 yr, were invited to participate. Pregnancy was an exclusion criterion. The study was approved by the North Staffordshire Local Research Ethics Committee. Written consent was obtained from all patients. For the duration of the study patients received their usual care.

Assessment of spinal mobility

Participants were assessed in a clinic-based setting by a trained observer (KH, KJ or JW). Spinal mobility was assessed following patient self-completion of a questionnaire, which included patient-assessed health instruments (Table 1), health transition items and socio-demographic questions [45, 46]. Each movement was practised once before the range of movement was recorded. All baseline and repeat appointments were arranged for the same time of day, thereby limiting the impact of diurnal change in spinal stiffness on measurement [47].

Reliability

Participants regularly attending the rheumatology centre for physiotherapy and participation in a self-help group were invited to participate in the study at baseline and 2 weeks for the reliability assessment. Participants were advised to maintain their usual activities. For the assessment of inter-observer reliability, two observers (KH, KJ) recorded the baseline measures for all participants, following a randomized order of assessment. For the assessment of intra-observer reliability, one observer (KH) repeated the measurements at 2 weeks. This is an appropriate period for reliability assessment in patients with a stable condition [48]. Intra-observer reliability was assessed for those patients indicating that their AS-specific health had remained the same at 2 weeks on a patient-reported health transition question [49, 50].

The ICC 2, 1 [51] was used to measure the agreement for both inter- and intra-observer assessment [48, 52]. Reliability levels over 0.70 are required for the comparison of groups [10, 48], and levels above 0.90 have been recommended for individual evaluation [10, 52, 53]. The 95% limits of agreement test–retest reliability estimate gives a score range that is expected to describe the range of agreement between repeat administrations of a measurement, and acknowledges that few repeat observations will be identical due to random error [54, 55].

Validity

Validity was assessed by correlating the ROM measures with the patient-assessed health instruments. Hypothesized associations between all measures and instruments were considered *a priori* (Table 2). A large correlation between the MSI and LLF and progressive AS-specific radiographic change in the lumbar spine and between TWD and change in the cervical spine has been reported [3, 22], indicating structural and irreversible change in AS. It was hypothesized that the MSI would have a moderate to strong correlation with LLF and TWD (0.5–0.8). Changes in cervical rotation and FFD may have a reversible component [56] and a moderate correlation between cervical rotation and FFD and TWD was hypothesized (0.5–0.7). In addition, cervical rotation and FFD would have small levels of correlation with the other ROM measures (<0.3).

Limited mobility leads to reduced levels of functioning and psychological well-being [47, 57]. The Ankylosing Spondylitis Quality of Life Questionnaire (ASQoL) [58], the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) [59] and Revised Leeds Disability Questionnaire (RLDQ) [21] are AS-specific, patient-assessed health instruments that respectively purport to assess quality of life, disease activity and functional disability. Measures of spinal mobility reflecting irreversible damage are associated with disease progression, with change occurring over time [60, 61]. Patients may compensate for movement limitation with a consequent change in expectation, and a small relationship between the MSI and TWD and patient-assessed health was expected (Table 2). It was hypothesized that ROM measures which reflect reversible change would have small to moderate levels of correlation with patient-assessed health. More specifically, limited mobility affects quality of life, as assessed by items in the ASQoL, and a small to moderate relationship with the ROM measures was hypothesized. Pain can affect mobility and a small to moderate correlation between the ROM measures and the BASDAI was hypothesized. Functional activities measured by the RLDQ are adversely affected by limited mobility and a moderate association with the ROM measures was hypothesized. Several activities are closely related to cervical mobility, and the larger association was hypothesized with Crot and TWD. The smallest association was expected for the correlation between the RLDQ and LLF.

Changes in spinal mobility and patient response to the AS-specific health transition question at 6 months was assessed for a linear trend [62]. To the extent that ROM measures are valid measures of health capable of measuring change, a statistically significant moderate level of association with the patient reported health transition item was expected [62].

Responsiveness

Patients were re-assessed at 6 months. Measures of spinal mobility were assessed for responsiveness to change by calculating the modified standardized response mean (MSRM) which is equal to the mean change in scores divided by the standard deviation of change scores in patients defined as stable [10]. MSRMs were calculated for patients reporting an improvement or deterioration in health on generic or AS-specific health transition. MSRMs are presented for three patient-assessed health instruments for purposes of comparison.

Results

Data collection

Of the 269 patients who were invited to take part in the study, 159 (59.9%) participated in the baseline assessment; 51 of these patients were members of a local self-help group participating in the reliability study and 45 of these patients (88.2%) returned at 2 weeks. The remaining 108 patients did not participate in the reliability study and were only asked to return at 6 months; 88 patients (81.5%) participated in the 6-month evaluation.

The majority of patients were male ($n = 133$; 83.6%) with a mean age of 49.0 yr (S.D. 12.1; range 20–74 yr). The mean symptom duration of participants was 20.9 yr (S.D. 11.9; range 1–58 yr). This suggests a broad spectrum of disease presentation.

The descriptive statistics for the ROM measures are shown in Table 3. Although a value of 0 cm describes maximal limitation of both Crot and lumbar flexion when measured by the MSI, the same value represents maximal possible movement when assessed by FFD; that is, the patient touches the floor. The distance following maximal lateral lumbar flexion was recorded for LLF.

Values for cervical rotation approximated the normal distribution. There was a wide range of values for FFD (0–60.0 cm); although many patients were capable of touching the floor ($n = 35$, 22.2%), 11.4% ($n = 16$) achieved scores greater than 40 cm. There was a slight skew towards more limited movement for LLF. There was a wide range of values (0–9.00 cm) and a bimodal distribution for the MSI; a cluster of patients had very limited lumbar flexion (0–2.9 cm) [10.7% ($n = 17$) with score 0–1.0 cm], with a second cluster having greater movement (3.0–9.0 cm) [3.3% ($n = 7$) with score 8.0–9.0 cm] (Table 3). There was a large range of values for TWD (9.1–44.5 cm), with results highly skewed towards scores which suggest increased spinal deformity; 7.6% of the population recorded values in excess of 30.0 cm.

Reliability

The ICCs are shown in Table 4. High levels of inter- and intra-observer reliability were found, with levels greater than 0.90 for most measurements. The starting position for the measurement of cervical rotation did not produce a level of inter-observer reliability necessary for measurement in individuals or groups of patients. The 95% limits of agreement are shown in Table 4.

TABLE 2. Hypothesized associations between spinal mobility measures and patient-assessed health instruments

	Range of movement				Patient-assessed health instruments		
Measure[a]	Crot	FFD	LLF	MSI (15 cm)	ASQoL	BASDAI	RLDQ
Crot					+/++	+/++	+/++
FFD	++				+/++	+/++	+/++
LLF	+	+			+	+	+
MSI (15 cm)	+	+	++/+++		+	+	+
TWD	++	+	++/+++	++/+++	+	+	+

Scale of association: +++, large (>0.70); ++, moderate (0.50); +, small (<0.30).
[a]A description of each measurement can be found in Table 1.

TABLE 3. Descriptive statistics of spinal mobility measures at baseline ($n = 159$)

Range of movement[a]	Mean (S.D.) (cm)	Median (cm)	Minimum score		Maximum score	
			cm	%	cm	%
Cervical rotation						
Left	7.36 (3.74)	7.45	0.0–1.0	4.3	14.0–15.0	1.2
Right	7.28 (3.57)	7.00	0.0–1.0	2.4	14.0–15.0	1.9
Fingertip-to-floor distance	19.71 (15.73)	17.70	0.0	22.2	60.8	0.6
Lateral lumbar flexion						
Left	53.10 (6.59)	54.10	25.8	0.6	68.4	0.6
Right	52.21 (6.30)	52.85	21.3	0.6	65.0	0.6
Modified Schober index (15 cm)	4.02 (2.30)	4.05	0.0–1.0	10.8	8.0–9.0	3.3
Tragus-to-wall distance	17.94 (7.11)	16.70	9.1–10.0	3.2	35.0–44.5	3.2

[a]A description of each measurement can be found in Table 1.

TABLE 4. Inter-observer ($n = 51$) and intra-observer ($n = 26$) reliability of measures of spinal mobility

Range of movement[a]	Inter-observer reliability (baseline values)		Intra-observer reliability (2 weeks retest; AS same)	
	ICC (95% CI)	95% limits of agreement	ICC (95% CI)	95% limits of agreement[b]
Cervical rotation: start				
Left	0.65 (0.25–0.82)	–	0.79 (0.58–90)	–
Right	0.68 (0.50–0.80)	–	0.88 (0.75–0.94)	–
Cervical rotation: difference				
Left	0.94 (0.89–0.96)	−2.91 to 2.53	0.94 (0.88–0.97)	−2.78 to 2.32
Right	0.90 (0.84–0.94)	−3.78 to 3.11	0.95 (0.89 –0.98)	−2.10 to 2.32
Fingertip-to-floor distance	0.96 (0.94–0.98)	−6.06 to 6.05	0.98 (0.96–0.99)	−6.19 to 8.11
Lateral lumbar flexion				
Left	0.95 (0.94–0.98)	−0.11 to 0.32	0.95 (0.89–0.98)	−3.05 to 3.33
Right	0.98 (0.97–0.99)	−3.78 to 4.92	0.98 (0.95–0.99)	−4.83 to 4.33
Modified Schober index (15 cm)	0.90 (0.83–0.94)	−1.79 to 2.01	0.94 (0.88–0.97)	−1.03 to 1.71
Tragus-to-wall distance	0.98 (0.97–0.99)	−1.70 to 1.34	0.98 (0.96–0.99)	−2.73 to 3.03

[a]A description of each measurement can be found in Table 1.
[b]Mean difference ± 1.96 (S.D. of the mean difference).

TABLE 5. Correlation between spinal mobility and patient-assessed measures (Pearson's correlation) ($n = 128$)

Measure[a]	Crot-left	Crot-right	FFD	LLF-left	LLF-right	MSI (15 cm)	ASQoL	BASDAI	RLDQ
Crot-left							−0.47	−0.45	−0.66
Crot-right	0.89						−0.51	−0.46	−0.63
FFD	−0.43	−0.49					0.37	0.33	0.46
LLF-left	−0.28	−0.27	0.42				0.23	0.17	0.29
LLF-right	−0.25	−0.27	0.46	0.89			0.26	0.21	0.31
MSI (15 cm)	0.49	0.53	−0.52	−0.54	−0.51		−0.24	−0.20	−0.48
TWD	−0.49	−0.45	0.29	0.41	0.35	−0.67	0.18	0.12	0.45

[a]A description of each measurement can be found in Table 1.

Validity

As hypothesized, the correlations between the ROM measures were of a small to moderate size (Table 5). The MSI produced the largest correlations, which, with the exception of the fairly large correlation with TWD, was as hypothesized. The relationship between FFD and LLF was larger than hypothesized. Small levels of correlation were found between LLF and both cervical rotation and TWD, as hypothesized.

As hypothesized, the level of correlation between scores for ROM measures and patient-assessed health were small to moderate (Table 5). Crot had the largest level of correlation with the scores for the three instruments (range −0.45 to −0.66). The ROM measures consistently had the largest level of correlation with scores for the RLDQ, followed by those for the ASQoL and BASDAI. The change scores for the five ROM measures do not reflect the categories of the health transition question (Table 6).

Responsiveness

The results of responsiveness testing are shown in Table 6. The most consistent results for the mobility measures were found for the FFD, which produced small to moderate levels of responsiveness for groups of patients whose AS-specific health had improved or deteriorated (Table 6). The Crot (left) was the most responsive measure for the group whose health improved but only small levels of responsiveness were found for the group whose health

TABLE 6. Mean score changes (S.D.S) and modified standardized response mean (MSRM) at 6 months ($n = 66$ listwise)

	AS health transition					
	Better ($n = 9$)		Same ($n = 38$)	Worse ($n = 19$)		
Measure[a]	Mean change (S.D.)	MSRM[b]	Mean change (S.D.)	Mean change (S.D.)	MSRM[b]	P*
Spinal mobility						
Crot-left	1.18 (2.04)	0.75	0.31 (1.57)	0.26 (2.02)	0.17	0.39
Crot-right	0.32 (1.85)	0.21	0.65 (1.52)	0.36 (2.81)	0.24	0.85
FFD	−3.44 (11.06)	−0.54	2.78 (6.41)	2.06 (7.07)	0.32	0.79
LLF-left	−0.33 (1.82)	−0.10	1.81 (3.73)	1.73 (4.40)	0.46	0.29
LLF-right	−0.42 (2.80)	−0.13	0.84 (3.12)	1.30 (3.74)	0.42	0.43
MSI	0.37 (0.69)	0.53	−0.07 (0.70)	0.10 (0.77)	0.14	0.24
TWD	−0.60 (1.17)	−0.42	−0.38 (1.43)	−0.04 (1.52)	−0.03	0.57
Patient–assessed health						
ASQoL	−3.24 (3.42)	−1.73	−0.06 (1.87)	1.07 (3.23)	0.57	0.0001
BASDAI	−2.35 (2.05)	−1.68	−0.15 (1.40)	0.60 (1.55)	0.43	0.0001
RLDQ	−0.22 (5.21)	−0.05	−0.50 (4.50)	0.40 (5.94)	0.10	0.82

*F-test for linearity.
[a]Measures defined in Table 1.
[b]Mean change in scores (6 months minus baseline) divided by S.D. of change scores in patients defined as stable.

deteriorated, or for those whose right Crot improved. The LLF produced small to moderate levels of responsiveness for the group that deteriorated but was unresponsive to improvements in health. The MSI and TWD produced small to moderate levels of responsiveness for those who improved, but were unresponsive to deterioration in health.

More consistent results were found for two AS-specific patient-assessed health instruments, the ASQoL and BASDAI, which produced high levels of responsiveness for the group whose health improved, and moderate levels for those whose health deteriorated (Table 6). The RLDQ was unresponsive.

Discussion

The assessment of spinal mobility is a common component of AS evaluation [7, 8, 11]. There is a wide range of measures of spinal mobility, and poor standardization and limited evidence of measurement properties can make it difficult to selection measurements within routine practice or clinical research. Measurement selection following clear evidence of measurement properties, patient acceptability and clinician feasibility will enhance evaluation [10, 63, 64]. This study represents a comparative evaluation of an evidence-based selection of measures of spinal mobility in a representative population of AS out-patients.

Most patients in this study had established and well-controlled AS. Although representing a wide range of disease presentations similar to that reported in other hospital-based studies [59, 65, 66], patients were predominantly male. Survey response rates compare favourably with other studies [58, 67]. However, responders were significantly older than non-responders. Similar differences between responders and non-responders have been reported elsewhere [68].

Both FFD and LLF were measured as a reflection of the distance between the tip of the middle finger and the floor following the described movement, and are therefore informative at the individual level only. The wide range of values recorded for the MSI in the study population ranged from extreme limitation to levels comparable with the normal population [69].

With the exception of the starting position for cervical rotation, all measures had estimates of intra- and inter-observer reliability, as assessed by correlation, that support their use in individual evaluation [53]. However, wide value ranges, when judged by 95% limits of agreement, were reported. For example, although a mean improvement in FFD of more than 3 cm has been described as clinically significant [70], this study suggests that a change of ±6 cm can be expected in patients who report no change in health over a 2-week period. Although the current study sample is small, similar results have been reported elsewhere following the completion of patient-assessed health instruments by AS patients [71]. The 95% limits of agreement describe a statistical range of values, not an accepted definition of agreement [55]. Interpretation of the range of values supports the evaluation of the smallest detectable difference or minimal clinically important difference in score [71], but lacks consensus and is influenced by the clinical circumstances of the evaluation [55, 71].

The moderate correlation between the MSI and the other ROM measures suggests that the MSI may be sensitive to aspects of mobility assessed by Crot, FFD, LLF and TWD. The relatively strong correlation between the MSI and TWD suggests that they measure related aspects of spinal mobility in patients with AS. The moderate correlation between the MSI and other measures supports the inclusion of this measure rather than TWD.

The ROM measures had small to moderate levels of correlation with the patient-assessed health instruments meeting *a priori* hypotheses. These results suggest that ROM measures assess different aspects of disease impact and should be included for purposes of evaluation. The moderate to large levels of correlation between the RLDQ and Crot, MSI and TWD were not predicted, and are indicative of the strong emphasis on spinal mobility, particularly cervical mobility, addressed by items within the RLDQ.

All ROM measures had a weak association with AS-specific health transition, which suggests that changes in spinal mobility have little impact on self-reported change in the disease. It follows that most ROM measures were not responsive to patient-reported change in health status over 6 months. The FFD showed moderate levels of responsiveness for self-perceived improvement or deterioration in health. Although the associated mean change may be of clinical significance [70], this value is smaller than the range of values calculated for the 95% limits of agreement in patients reporting no change in health at 2 weeks. Similarly, the MSI showed moderate levels of responsiveness to improvements in health. However, the associated mean change in movement was negligible, and is unlikely to be of clinical significance [43]. Although the Crot showed responsiveness to improvements in health, only small levels were recorded for deterioration in health. Inconsistencies between the results for right and left rotation reduce confidence in the results. This may be a reflection of the poor levels of reliability recorded for the measurement starting

position, supporting the need to modify the measurement method. Furthermore, the sample sizes for the evaluation of responsiveness were small and therefore caution should be exercised in the interpretation of results. Future studies should consider the appropriateness of health transition questions as external criteria in the assessment of measures of spinal mobility. More appropriate criteria might include radiographic change or change in alternative clinical criteria.

Although bony change in AS may not be detected radiographically over periods of less than 2 yr [4, 61], the strong relationship between AS-specific radiographic change and the MSI, LLF and TWD [22] supports the ability of these measures to reflect both structural change and the irreversible nature of AS. Large changes in these measures over the short term can influence decision making, acting as a trigger for further investigations and management. The low levels of responsiveness observed in the present study may reflect measurement insensitivity or, more likely, the slow rate of change in spinal mobility in patients with stable AS receiving usual care.

Measures with a lower correlation with radiographic change may be more reflective of reversible change in AS [56]. Crot and FFD are responsive to change in ROM over 6 months and capture the short-term effects of management. For example, following a 3-week in-patient rehabilitation programme for AS patients, FFD (method not specified) was found to be the most responsive mobility measure [26]. However, the evaluation of ROM measurement responsiveness following anti-TNF therapies is required. Following a meta-analysis of studies reporting the effectiveness of non-steroidal therapies in patients with AS, spinal mobility was omitted from the proposed response criteria due to poor levels of responsiveness [72]. This study was completed before the introduction of biological therapies. However, an update of the structured review found limited evidence to suggest minimal change in chest expansion, Schober 10 cm index, OWD and the Bath AS Metrology Index scores (BASMI) [14] [73–75].

A 'compact set' of eight measures of spinal mobility has recently been recommended [13], and the developers of an earlier composite index of AS axial status (BASMI) [14] included five mobility measures as the minimum number. Although similarity exists between measures included in the current study and those in the BASMI (Table 1), the BASMI uses a wide range of equipment, and requires the patient to lie supine for the assessment of cervical rotation [14]. The range of instrumentation was reduced for the present study to enhance the feasibility and cost of multiple ROM assessments, whilst retaining high levels of reliability [11, 41]. The present study suggests that the choice of ROM measures could be further reduced to the measurement of Crot and/or FFD, as a short-term reflection of reversible change in spinal mobility, and the MSI (15 cm) as a long-term reflection of irreversible change in spinal mobility [3, 22]. All measurements can be completed in less than 5 min and the only instrumentation required is a tape measure. Routine use of these approaches in clinical practice would suggest that they already have good levels of clinical acceptance [7, 8, 65].

To reflect the multidimensional nature of disease impact, there is a need to include both patient-assessed and ROM measures in the evaluation of patients with AS [5, 6]. This study has illustrated the limited relationship between selected measures of spinal mobility and patient-assessed health measures and the patient perceptions of disease impact. Additionally, the ASQoL and BASDAI demonstrated greater levels of responsiveness than all mobility measures.

Following the consensus opinion of gathered experts, ASAS recommended the assessment of chest expansion (method not specified), lumbar flexion (Schober 10 cm index) and occiput-to-wall distance [6] for the assessment of AS spinal mobility. Structured reviews of evidence are a prerequisite for the evidence-based selection of instruments for clinical trials or routine practice [64]. However, ASAS recommended outcome measures which have not been fully evaluated, but acknowledged that recommendations may change in the light of new evidence for measurement properties [6]. From our findings there is a need to evaluate by direct comparison the properties of the MSI and Schober (10 cm) index, the TWD and OWD, and multiple methods for measuring cervical rotation and chest expansion in AS.

In conclusion, this study provides evidence for the measurement properties of five measures of spinal mobility in a UK population of patients with AS. The inclusion of the measures was supported by a structured review of the literature. We would recommend the use of MSI (15 cm), Crot and FFD as reflecting spinal mobility in AS. The measures are inexpensive and simple to administer in clinical practice and have evidence for reliability and validity.

Rheumatology	Key messages
	• Numerous mobility measures and the limited evidence for their measurement properties have contributed to the lack of standardization in clinical practice and research. • Evidence of reliability and validity supports the selection of three mobility measures for routine AS assessment: MSI (15 cm) Crot and FFD (following lumbar forward flexion). • Concurrent evaluation of mobility measures in different settings with representative samples of patients and long-term follow-up is required.

Acknowledgements

We are grateful to the patients who willingly gave their time to participate in the study. We also thank all consultant rheumatologists from the Staffordshire Rheumatology Centre (SRC) for allowing access to AS patient databases, and all nursing and clinic staff from the SRC, with particular thanks to Ms Jackie Waterfield for assistance with data collection. This study was supported by a grant from the Arthritis Research Campaign and funding from the Staffordshire Rheumatology Centre.

The authors have declared no conflicts of interest.

References

1. van der Linden SJ, 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.
2. Lubrano E, Helliwell P. Deterioration in anthropometric measures over six years in patients with ankylosing spondylitis. Physiotherapy 1999;85:138–43.
3. Dawes P. Stoke Ankylosing Spondylitis Spine Score. J Rheumatol 1999;26:993–6.
4. van der Heijde D, Spoorenberg A. Plain radiographs as an outcome measure in ankylosing spondylitis. J Rheumatol 1999;26:985–7.
5. van der Heijde D, Bellamy N, Calin A, Dougadas M, Khan MA, van der Linden S. Preliminary core sets for endpoints in ankylosing spondylitis. J Rheumatol 1997;24:2225–9.
6. van der Heijde D, Calin A, Dougadas M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing

spondylitis. Progress report of the ASAS Working Group. Assessments in Ankylosing Spondylitis. J Rheumatol 1999;26:951–4.
7. Bellamy N, Kaloni S, Pope J, Coulter K, Campbell J. Quantitative rheumatology: a survey of outcome measurement procedures in routine rheumatology outpatient practice in Canada. J Rheumatol 1998;25:852–8.
8. Bellamy N, Muirden KD, Brooks PM, Barraclough D, Tellus MM, Campbell J. A survey of outcome measurement procedures in routine rheumatology outpatient practice in Australia. J Rheumatol 1999;26:1593–9.
9. Whitham J. Health outcomes in ankylosing spondylitis: A survey of the use of patient-assessed and anthropometric measures by physiotherapists in the UK. BSc dissertation, Coventry University, Coventry, UK, 2002.
10. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technology Assessment 1998;2.
11. Haywood KL. Health outcomes in ankylosing spondylitis: an evaluation of patient-based and anthropometric measures. DPhil thesis, University of York, York, UK, 2000.
12. Viitanen JV, Kokko ML, Heikkila S, Kautiainen H. Neck mobility assessment in ankylosing spondylitis: a clinical study of nine measurements including new tape methods for cervical rotation and lateral flexion. Br J Rheumatol 1998;37:377–81.
13. Viitanen JV, Heikkila S, Kokko ML, Kautiainen H. Clinical assessment of spinal mobility measurements in ankylosing spondylitis: a compact set for follow-up and trials? Clin Rheumatol 2000; 19:131–7.
14. Jenkinson TR, Mallorie PA, Whitelock HC, Kennedy GL, Garrett SL, Calin A. Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. J Rheumatol 1994;21:1694–8.
15. Viitenan JV, Kautiainen H, Suni J, Kokko ML, Lehtinen K. The relative value of spinal and thoracic mobility measurements in ankylosing spondylitis. Scand J Rheumatol 1995;24:94–7.
16. Viitanen JV, Lehtinen K, Suni J, Kautiainen H. Fifteen months' follow-up of intensive inpatient physiotherapy and exercise in ankylosing spondylitis. Clin Rheumatol 1995;14:413–9.
17. Helliwell PS, Abbott CA, Chamberlain MA. A randomised trial of three different physiotherapy regimes in ankylosing spondylitis. Physiotherapy 1996;82:85–90.
18. Bellamy N, Buchanan WW, Esdaile JM *et al*. Ankylosing spondylitis antirheumatic drug trials I: Effects of standardisation procedures on observer dependent outcome measures. J Rheumatol 1991;18:1701–8.
19. Tomlinson MJ, Barefoot J, Dixon AS. Intensive in-patient physiotherapy courses improve movement and posture in ankylosing spondylitis. Physiotherapy 1986;72:238–40.
20. 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;30:330–5.
21. Abbott CA, Helliwell, PS, Chamberlain MA. Functional assessment in ankylosing spondylitis–Evaluation of a new self-administered questionnaire and correlation with anthropometric variables. Br J Rheumatol 1994;33:1060–6.
22. 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.
23. 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.
24. Ward MM, Kuzis S. Validity and sensitivity to change of spondylitis-specific measures of functional disability. J Rheumatol 1999;26:121–7.
25. Ozgocmen S, Ardicoglu O, Kaya A. The relationship of clinical and laboratory measurement to two different radiological scoring methods in ankylosing spondylitis. J Back Musculoskel Rehabil 2000;15;37–40.
26. Heikkila S, Viitanen JV, Kautiainen H, Kauppi M. Sensitivity to change of mobility tests: effect of short term intensive physiotherapy and exercise on spinal, hip, and shoulder measurements in spondyloarthropathy. J Rheumatol 2000;27:1251–6.
27. Moll JMH, Wright V. The pattern of chest and spinal mobility in ankylosing spondylitis. Rheumatol Rehabil 1973;12:115–34.
28. Franssen MJ, van Herwaarden CL, van de Putte PB. Lung function in patients with ankylosing spondylitis. A study of the influence of disease activity and treatment with antiinflammatory drugs. J Rheumatol 1986;13:936–40.
29. Fisher LR, Cawley MID, Holgate ST. Relation between chest expansion, pulmonary function, and exercise tolerance in patients with ankylosing spondylitis. Ann Rheum Dis 1990;49:921–5.
30. Seckin U, Bolukbasin N, Gursel G *et al*. Relationship between pulmonary function and exercise tolerance in patients with ankylosing spondylitis. Clin Exp Rheumatol 2000;18:503–6.
31. Macrae IF, Wright V. Measurement of back movement. Ann Rheum Dis 1969;28:584–9.
32. Reynolds PMG. Measurement of spinal mobility: a comparison of three methods. Rheumatol Rehabil 1975;14:180–5.
33. Miller MH, Lee P, Smythe HA, Goldsmith CH. Measurement of spinal mobility in the sagittal plane: new skin contraction technique compared with established methods. J Rheumatol 1984;11:507–11.
34. Dougadas M, Gueguen A, Nakache JP, Nguyen M, Mery C, Amor B. Evaluation of a functional index and an articular index in ankylosing spondylitis. J Rheumatol 1988;15:302–7.
35. Moll JMH, Liyange SP, Wright V. An objective clinical method to measure lateral spinal flexion. Rheumatol Phys Med 1972;11:225–39.
36. Battle-Gualda E, Figuero M, Ivorra J, Raber A. The efficacy and tolerability of aceclofenac in the in the treatment of patients with ankylosing spondylitis: a multicenter controlled clinical trial. Aceclofenac Indomethacin Study Group. J Rheumatol 1996;23: 1200–6.
37. Clegg DO, Reda DJ, Weisman MH *et al*. Comparison of sulphasalazine and placebo in the treatment of ankylosing spondylitis. A Department of Veterans Affairs cooperative study. Arthritis Rheum 1996;39:2004–12.
38. Tishler M, Brostovski Y, Yaron M. Effect of spa therapy in Tiberias on patients with ankylosing spondylitis. Clin Rheumatol 1995; 14:21–5.
39. Viitanen JV, Suni J, Kautianen H. Effect of physiotherapy on spinal mobility in ankylosing spondylitis. Scand J Rheumatol 1992;21:38–41.
40. Heikkila S, Viitanen JV, Kautiainen H, Kauppi M. Does improved spinal mobility correlate with functional changes in spondyloarthropathy after short-term physical therapy? J Rheumatol 2000;27:2942–4.
41. Haywood KL, Waterfield J, Jordan KP, Garratt AM, Dziedzic K, Dawes PT. Reliability of anthropomorphic measures in AS—a pilot study. Physiotherapy 1999;85:372–3.
42. O'Driscoll SL, Jatson MIV, Baddeley H. Neck movements in ankylosing spondylitis and its response to physiotherapy. Ann Rheum Dis 1978;37:64–6.
43. 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.
44. Stokes BA, Helewa A, Goldsmith CH, Groh JD, Kragg GR. Reliability of spinal mobility measurements in ankylosing spondylitis patients. Physiother Canada 1988;40:338–44.
45. Haywood KL, Garratt AM, Jordan K, Dziedzic K, Dawes PT. Disease-specific patient-assessed measures of health outcome in ankylosing spondylitis: reliability, validity and responsiveness. Rheumatology 2002;41:1295–302.
46. Haywood KL, Garratt AM, Dziedzic K, Dawes PT. Generic measures of health-related quality of life in ankylosing spondylitis: reliability, validity and responsiveness. Rheumatology 2002;41:1380–7.
47. Dziedzic K. Ankylosing spondylitis. In: David C, Lloyd J, eds. Rheumatological physiotherapy. London, UK: Mosby, 1998:97–114.
48. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use, 2nd edn. Oxford, UK: Oxford Medical Publications, 1995.
49. Fitzpatrick R, Zieblans S, Jenkinson C, Mowat A, Mowat A. Transition questions to assess outcomes in rheumatoid arthritis. Br J Rheumatol 1993;32:807–11.

50. Keller SD, Maijkut TC, Kosinski M, Ware JE Jr. Monitoring health outcomes among patients with arthritis using the SF-36 health survey. Med Care 1999;37(5 Suppl.):MS1–MS9.
51. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull 1979;86:420–8.
52. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials 1991;12:142S–158S.
53. Nunnally JC, Bernstein IH. Psychometric theory, 3rd edn. McGraw-Hill, 1994.
54. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;i:307–10.
55. Altman DG. Practical statistics for medical research. London: Chapman and Hall, 1996.
56. Roberts WN, Liang MH, Pallozzi LM, Daltroy LH. Effects of warming up on reliability of anthropometric techniques in ankylosing spondylitis. Arthritis Rheum 1988;31:549–52.
57. Jenkinson C, Fitzpatrick R, Peto V. The Parkinson's Disability Questionnaire. User manual for the PDQ-39, PDQ-8 and PDQ Summary Index. Oxford: Health Services Research Unit, Department of Public Health, University of Oxford, 1998.
58. Doward LC, Spoorenberg A, Cook SA *et al*. Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis. Ann Rheum Dis 2003;62:20–6.
59. Garrett S, Jenkinson T, Kennedy G, 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;1:2286–91.
60. van der Linden S, van der Heijde DM. Ankylosing spondylitis and other B27 related spondyloarthropathies. Baillieres Clin Rheumatol 1995;9:355–73.
61. 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.
62. Garratt AM, Ruta DA, Abdalla MI, Russell IT. Responsiveness of the SF-36 and a condition-specific measure of health for patients with varicose veins. Qual Life Res 1996;5:223–34.
63. McDowell I, Newell C. Measuring health, 2nd edn. A guide to rating scales and questionnaires. Oxford University Press, 1996.
64. Garratt AM, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ 2002;24:1417–9.
65. Lubrano E, Butterworth M, Heddelden A, Well S, Helliwell P. An audit of anthropometric measurements by medical and physiotherapy staff in patients with ankylosing spondylitis. Clin Rehabil 1998;12:216–20.
66. Uhrin Z, Kuzis S, Ward MM. Exercise and changes in health status in patients with ankylosing spondylitis. Arch Intern Med 2000;160: 2969–75.
67. McBeth J, Silman AJ, Macfarlane GJ. Association of widespread body pain with an increased risk of cancer and reduced cancer survival. Arthritis Rheum 2003;48:1686–92.
68. McBeth J, Macfarlane GJ, Silman AJ. Does chronic pain predict future psychological distress? Pain 2002;96:239–45.
69. Moll JMH, Liyange SP, Wright V. Normal range of spinal mobility–An objective clinical study. Ann Rheum Dis 1971;30: 381–6.
70. Kragg G, Stokes B, Groh J, Helewa A, Goldsmith C. The effects of comprehensive home physiotherapy and supervision on patients with ankylosing spondylitis–A randomised controlled trial. J Rheumatol 1990;17:228–33.
71. Auleley GR, Benbouazza K, Spoorenberg A *et al*. Evaluation of the smallest detectable difference in outcome or process variables in ankylosing spondylitis. Arthritis Rheum 2002;47:582–7.
72. Anderson JJ, Baron G, van der Heijde D. Ankylosing Spondylitis Assessment Group preliminary definition of short-term improvement in ankylosing spondylitis. Arthritis Rheum 2001;44:1876–86.
73. Braun J, Brandt J, Listing J *et al*. Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial. Lancet 2002;359:1187–93.
74. Gorman JD, Sack KE, Davis JC. Treatment of ankylosing spondylitis by inhibition of tumor necrosis factor alpha. N Engl J Med 2002;346:1349–56.
75. Maksymowych WP, Jhangri GS, Lambert RG *et al*. Infliximab in ankylosing spondylitis: a prospective observational inception cohort analysis of efficacy and safety. J Rheumatol 2002; 29:959–65.