Criterion-concurrent Validity of Spinal Mobility Tests in Ankylosing Spondylitis: A Systematic Review of the Literature

Marcelo P. Castro, Simon M. Stebbings, Stephan Milosavljevic, and Melanie D. Bussey

ABSTRACT. ***Objective.*** To examine the level of evidence for criterion-concurrent validity of spinal mobility assessments in patients with ankylosing spondylitis (AS).

Methods. Guidelines proposed in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses were used to undertake a search strategy involving 3 sets of keywords: accura*, truth, valid*; ankylosing spondylitis, spondyloarthritis, spondyloarthropathy, spondylarthritis; mobility, spinal measure*, (a further 16 keywords with similar meaning were used). Seven databases were searched from their inception to February 2014: AMED, Embase, ProQuest, PubMed, Science Direct, Scopus, and Web of Science. The Quality Assessment of Diagnostic Accuracy Studies (with modifications) was used to assess the quality of articles reviewed. An article was considered high quality when it received "yes" in at least 9 of the 13 items.

Results. From the 741 records initially identified, 10 articles were retained for our systematic review. Only 1 article was classified as high quality, and this article suggests that 3 variants of the Schober test (original, modified, and modified-modified) poorly reflect lumbar range of motion where radiographs were used as the reference standard.

Conclusion. The level of evidence considering criterion-concurrent validity of clinical tests used to assess spinal mobility in patients with AS is low. Clinicians should be aware that current practice when measuring spinal mobility in AS may not accurately reflect true spinal mobility. (First Release Nov 15 2014; J Rheumatol 2015;42:243–51; doi:10.3899/jrheum.140901)

Key Index Terms:
ANKYLOSING SPONDYLITIS SPINAL MOBILITY SPONDYLOARTHRITIS
SYSTEMATIC REVIEW VALIDATION STUDY

The importance of assessing spinal mobility in patients with ankylosing spondylitis (AS) was emphasized after its recommendation as an inclusion criterion for diagnosing the disease in 1966 in the New York symposium[1]. Since then, measurements of spinal mobility have been widely used in the assessment of patients with AS, assisting with diagnosis, monitoring disease progression, and determining the efficacy of treatment interventions[2,3,4]. Limitation of spinal mobility may be a predictor of poor outcome in AS[5]. Structural damage, inflammation, and age have already been shown to affect spinal mobility[6,7]. However, before a clinical test is accepted as an assessment tool, it should demonstrate acceptable reliability, responsiveness, and validity. The latter is a particularly important feature for clinical tests[8]. It is further defined as face validity and content validity, subjective measures representing the concept of the test, often used to assess questionnaire-based assessments. Construct validity represents the ability of an instrument to measure an abstract construct, such as the level of health, capacity, or physical function. Criterion-related validity is divided into criterion-concurrent validity, when 2 tests or instruments — the criterion (a reference standard) and the target (index test) — are performed concurrently. Finally, criterion-predictive validity establishes how successful the outcome of the target test is as a predictor of a future status[8].

Following the Outcome Measures in Rheumatoid Arthritis Clinical Trials, "truth" has been identified as 1 of the 3 key criteria for any outcome measure (the other components are discrimination and feasibility)[9,10]. The truth of a measure represents the ability or accuracy of an instrument or clinical test to assess the intended variable[9,10]. Spinal mobility tests are used on the assumption that they reflect spinal range of motion. Although the term "truth" may cover both face and content validity, the most objective way of assessing the "truth" of a clinical test or instrument

From the School of Physical Education, Sport and Exercise Science, and the Dunedin School of Medicine, University of Otago, Dunedin, New Zealand; the School of Physical Therapy, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Supported by the University of Otago (PhD Scholarship).

M.P. Castro, PhD, School of Physical Education, Sport and Exercise Science, and Dunedin School of Medicine, University of Otago; S.M. Stebbings, MD, Dunedin School of Medicine, University of Otago; S. Milosavljevic, PhD, School of Physical Therapy, University of Saskatchewan; M.D. Bussey, PhD, School of Physical Education, Sport and Exercise Science, University of Otago.

Address correspondence to Dr. M.P. Castro, School of Physical Education, Sport and Exercise Science, University of Otago, 46 Union St. West, PO Box 56, University of Otago, Dunedin, New Zealand. E-mail: marcelo.peduzzi.castro@gmail.com

Accepted for publication October 3, 2014.

is through criterion-concurrent validity, which requires that a measure be compared to a reference standard[8]. The reference standard for range of motion is widely acknowledged to be radiographic measurements[11,12,13,14,15]. However, this approach is relatively time-consuming and expensive, and exposes the patients to radiation[11]. As a consequence, some noninvasive low-cost methods that are easy to apply and interpret, such as goniometry[16,17], tape measures[12,17,18,19,20,21], and inclinometry[18] have been used. These measurements are frequently used when assessing spinal mobility in patients with AS, together with cervical rotation, tragus-to-wall distance, lateral lumbar flexion, modified Schober test, chest expansion, and finger-tip-to-floor distance. Some indices combine several clinical measures to provide a composite clinical index and an assessment of spinal movement as a whole[2,16,22,23]. These include the Bath Ankylosing Spondylitis Metrology Index (BASMI)[23], the Edmonton Ankylosing Spondylitis Metrology Index[16], and the University of Cordoba Ankylosing Spondylitis Metrology Index[22].

BASMI is recommended by the Assessment of SpondyloArthritis International Society[2,24,25] and has been widely used[4,6,25,26,27,28,29,30,31,32,33]. The initial study that proposed the BASMI had 5 clinical tests that include the index, and considered these the most accurate to reflect spine mobility[23]. However, the term "accurate" was not defined, and although the BASMI study appears to deal with face validity, its aims were to calculate the reproducibility and responsiveness of these clinical tests, but no data considering any kind of validity was presented[23]. Nonetheless, based on this study[23], many authors have erroneously claimed the validity of the BASMI[4,16,25,28,29,34]. Moreover, previous studies observed associations between individual spinal mobility tests or compound indices and compound radiological indices using plain radiographic scoring systems such as the modified Stokes Ankylosing Spondylitis Spinal Score or the Bath Ankylosing Spondylitis Radiology Index[35,36,37,38,39]. However, although these findings are undoubtedly important, these studies deal with construct validity and were not designed to assess the extent to which compound indices or individual spinal mobility tests reflect true spinal movement. Thus, there is a common misconception regarding the criterion-concurrent validity of mobility tests in patients with AS. In addition, criterion-concurrent validity in the context of spinal measures can only be assessed for individual tests. Thus, the analysis of compound indices is not appropriate when criterion-concurrent validity is the aim. Our current systematic review aims to examine the level of evidence for criterion-concurrent validity of spinal mobility assessments in patients with AS.

MATERIALS AND METHODS

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines were used as the basis for our systematic review[40].

Literature search strategy. Based on the current literature, 3 sets of keywords were derived: (1) accura* OR truth OR valid*; (2) "ankylosing spondylitis" OR spondyloarthritis OR spondyloarthropathy OR spondylarthritis; (3) "mobility" OR BASMI OR "spinal measure*" OR "hip measure*" OR Goniomet* OR Inclinomet* OR "tape measure*" OR "cervical rotation" OR "tragus to wall" OR "lumbar flexion" OR Schober OR "intermalleolar distance" OR "chest expansion" OR "finger* to floor" OR "finger* to ground" OR "internal rotation" OR "range of motion" OR "range of movement". Initially, keywords were entered into the Cochrane database to identify any previous systematic reviews with a similar aim to our present study. Subsequently, 7 databases were searched from their inception to February 2014: AMED, Embase, ProQuest, PubMed, Science Direct, Scopus, and Web of Science. Some MEdical Subject Headings terms and filters were applied (Figure 1).

Eligibility criteria. Studies were considered if they met the following inclusion criteria: (1) assessing human adult subjects (> 18 yrs old); (2) assessing participants with a diagnosis of AS; (3) a design that assessed criterion-concurrent validity of spine mobility measures; (4) full-text availability; (5) assessing individual tests of spinal mobility; and (6) articles in peer-reviewed journals.

Articles were excluded when (1) the spinal mobility was related only to measures of structural damage or quality of life, and (2) results were presented only as total scores from compound indices.

To avoid missing relevant articles, there were no restrictions on language and publication dates.

Study selection. All articles retrieved from the database searches were imported into EndNote X4 (Thomson Reuters). One author (MPC) removed duplicated references, editorials, letters to the editor, short reports, abstracts, and reviews. Two independent reviewers (MPC and MDB) screened the titles and abstracts to identify the articles that would potentially meet the inclusion criteria. The full text of articles was then reviewed and those that met the criteria for inclusion were determined. Finally, the reference lists from included articles were screened to identify further relevant articles. Disagreement between the 2 reviewers regarding the relevance of a study for inclusion was settled with a consensus-agreement approach, and if consensus could not be reached, a third reviewer provided arbitration.

Methodological quality assessment and risk of bias. Each article included from the databases was rated for methodological quality and risk bias by 2 examiners (MPC and MDB). The Quality Assessment of Diagnostic Accuracy Studies (QUADAS)[41] was used. The scale was composed of 14 items, answering "yes", "no", or "unclear", and covered 3 topics: (1) reporting of selection criteria; (2) selection, execution, and interpretation of the index test and reference standard; and (3) data analysis. In accordance with the suggestions of the authors of QUADAS[41], the scale required adaptation depending on the nature of the study. Because our present study aimed to assess criterion-concurrent validity of spinal measurement rather than diagnostic accuracy, adaptations were made to QUADAS (Table 1): 3 items were removed (items 7, 12, and 13), 2 items were modified (items 3 and 5), and 2 items were added, which were derived from a scale of quality previously proposed to evaluate criterion-concurrent validity of cervical range of motion[42]. Therefore, the scale used was composed of 13 items.

Included articles were rated independently for quality by 2 reviewers (MPC and MDB). A maximum score of 13 points was assigned to each article. After initial scoring, the articles were classified as either low or high quality. An article was considered high quality when it met 2 criteria: (1) receiving "yes" for item 3 (Is the reference standard likely to correctly measure the target joint range of motion?). We considered radiographic analyses as reference standard for spinal range of motion; and (2) receiving "yes" for at least 9 out of the 13 items assessed. We set this threshold because previous studies have indicated that about 70% of positive scores (10 out of the 14 QUADAS items) discriminate between high-quality and low-quality studies regarding diagnostic accuracy[43,44].

Data analysis. To assess the level of association between the reference

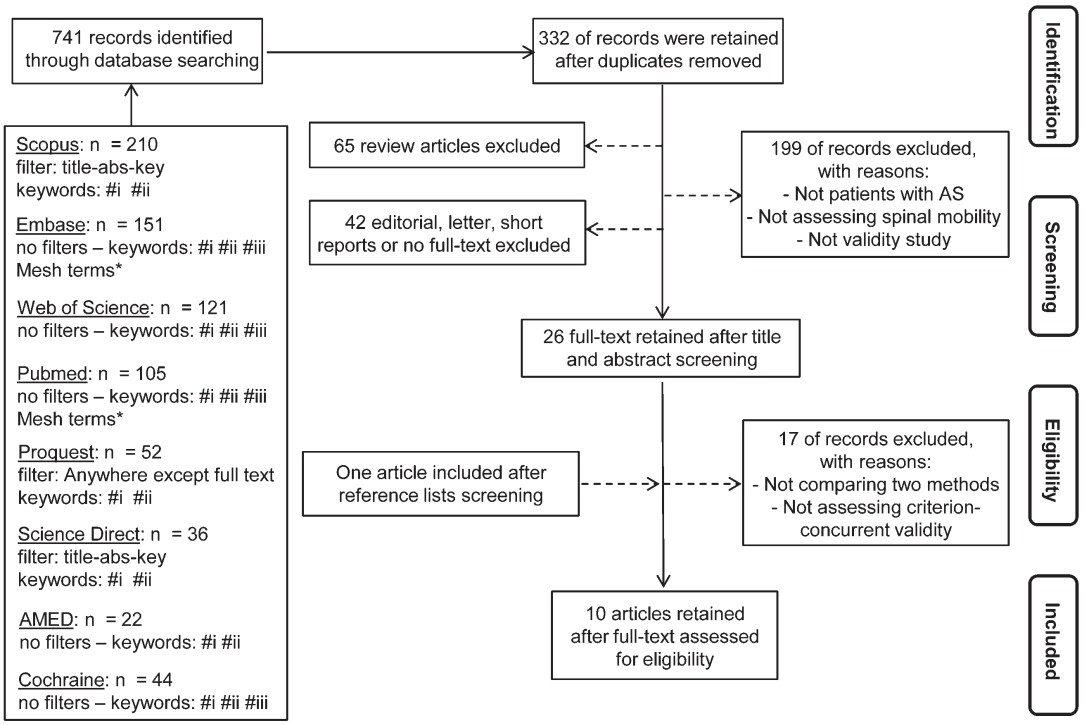

Figure 1. MEdical Subject Headings terms and filters. AS: ankylosing spondylitis.

Table 1. Implementation of the quality scale for the present study.

Items	Descriptions
Adapted items	
QUADAS 3	Original: Is the reference standard likely to correctly classify the target condition?
	New: Is the reference standard likely to correctly measure the target joint range of motion?
QUADAS 5	Original: Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?
	New: Did the whole sample or a random selection of the sample receive verification using a reference standard of measurement?
Removed items	
QUADAS 7	Was the reference standard independent of the index test (i.e., the index test did not form part of the reference standard)?
QUADAS 12	Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
QUADAS 13	Were uninterpretable/intermediate test results reported?
Added items based on Williams, *et al*[42]	
	Were appropriate descriptive statistics presented (means, SD, or SEM)?
	Were appropriate inferential statistics presented (correlation coefficient or agreement stats with CI)?

From Whiting, *et al*. BMC Med Res Methodol 2003;3:25; with Creative Commons Attribution License, CC-BY 4.0. From Williams, *et al*. J Manipulative Physiol Ther 2010;33:138-55; with permission. QUADAS: Quality Assessment of Diagnostic Accuracy Studies; SEM: standard error of the mean.

standard and the index test, either Pearson or Spearman correlation coefficients r values between 0 and 0.20 were designated as a low level of agreement, 0.21 and 0.40 as a fair agreement, 0.41 and 0.60 as a moderate agreement, 0.61 and 0.80 as a substantial agreement, and higher than 0.81 as excellent agreement[45].

RESULTS

Study selection. No systematic reviews addressing the question posed by our present paper were discovered. The primary search yielded a total of 741 articles. After title, abstract, and/or full-text screening, 715 were excluded as they were either duplicates, review articles, were not original full-text articles, or were not directly related to the topic of the present review. Twenty-six articles were retained for full-text analysis. Seventeen articles were excluded because they did not compare 2 methods or tests for assessing spinal mobility, or they did not assess criterion-concurrent validity. Thus, 9 articles[12,18,19,20,21,46,47,48,49] were retained for our current systematic review. After screening the reference list from these selected articles, another article[17] met the inclusion criteria and was also included (Figure 1).

Study characteristics. The earliest studies identified were

published in the 1970s[47,49], while the most recent study was published in 2012[12]. The number of participants assessed in the studies ranged from 3[17] to 263[21]. All studies used tape measures as one of the methods, in some cases as the index test[12,17,18,19,20,21,47,49], while in others as the reference standard[19,46,48]. One study used an electromagnetic 3-D tracking system in the index test[48]. Radiographic analysis was the most common reference standard used[12,20,47,49], while inclinometry[18] and goniometry[17,21] were also used (Table 2).

The range of motion of different segments of the spine was assessed. The thoraco-lumbar mobility was assessed by thoraco-lumbar lateral flexion[49], thoraco-lumbar extension[47], thoraco-lumbar forward flexion[12,17,20], Schober indices[12,17,19,20,48], fingertip-to-floor distance[17,46,48], and thoraco-lumbar rotation[19]. The upper thoracic and lower cervical mobility were assessed by the tragus (occiput)-to-wall distance[46], and the cervical spine mobility was assessed by cervical lateral flexion[18,48], cervical rotation[18,21,48], cervical flexion[18,48] (chin-to-chest distance[18]), and cervical extension[18,48] (Table 2).

Quality assessment. In the present systematic review, 82.2% of the items (107) were scored in agreement by the 2 examiners, while for the other 23 items, consensus agreement was reached after discussion (Table 3). The lowest quality rating was 4 points[17] and the highest 10[12] (out of a maximum of 13 points). Four studies included a very small number of participants (less than 8)[17,46,47,49], 2 studies considered male patients alone[18,19], and 3 studies considered a small number of female participants[12,20,48]. Therefore, only 1 study was considered to assess a representative spectrum of patients (Item 1)[21]. Further, most articles (7 out of 10) did not clearly describe their selection criteria of participants (Item 2)[17,18,19,21,46,47,49].

Four studies used radiographic analysis as the reference standard for assessing the criterion-concurrent validity of tape measures (Item 3)[12,20,47,49]. The reference standard used in the other selected studies included a goniometer[17,21], inclinometer[18], or tape measures[19,46,48]. None of the studies using radiographic analysis clearly described the time interval between the radiograph and index test (Item 4)[12,20,47,49].

All participants were assessed using both the reference standard and index test (Item 5), and were assessed using the same reference standard regardless of the index test used (Item 6). The execution of the index test was described with enough detail to be reproduced in 9 out of the 10 studies (Item 7)[12,18,19,20,21,46,47,48,49], while the description of the execution of the standard reference was clear in 6 articles (Item 8)[12,17,19,20,21,46]. Only 1 study measured or interpreted the index test without knowledge of the results of the reference standard (Item 9)[12]. In 3 studies, the reference standard was interpreted without knowledge of the index test[12,47,49], in another 3 studies, this information was not presented[20,21,48], and in 4 articles, the examiners were aware of the results of the index test (Item 10)[17,18,19,46].

In 5 studies, there were either no withdrawals or reasons for withdrawals from the study that were clearly presented[18,19,47,48,49], while in the other 5, there was insufficient detail regarding withdrawals (Item 11)[12,17,20,21,46]. Most studies clearly showed their descriptive statistics by mean, median, and/or variance measures[12,18,19,21,47,49]; however, in 4 articles, the results were not clearly presented (Item 12)[17,20,46,48]. Only the 3 most recently conducted studies used inferential statistics, presenting coefficients of correlation between the reference standard and the index test[12,21,48]. In the other studies, the data from patients and control groups were pooled[17,47,49], no statistical analysis was performed[18,19,46], or the procedures described in methods that did not match those presented in the results (Item 13)[20].

Criterion-concurrent validity. From the 4 studies using radiographic analysis as reference standard[12,20,47,49], only 1 was rated as high quality. This study by Rezvani, *et al*[12] assessed the correlation between 3 variants of the Schober test (original, modified, and modified-modified) and 2 techniques for calculating lumbar range of motion by radiography (the angle between L1 and S1, and between L3 and S1) in patients with AS and control participants. For the Schober tests, 2 reference points were marked on the patients' low back region while they were in erect position: original Schober (marks at the lumbosacral junction and 10 cm above), modified Schober (marks 5 cm below and 10 cm above the lumbosacral junction), and modified-modified Schober (marks at the lumbosacral junction and 15 cm above). Then the patients performed forward flexion and the distance between these marks was measured. Higher distances between marks suggested higher lumbar flexion. Poor correlations were observed for all analyses (Table 4)[12]. Another study that also assessed lumbar spine forward flexion, found moderate and substantial correlations between Schober tests (modified and original) and radiographic lumbar range of motion[20]. Finally, Moll, *et al*[47,49] assessed lumbar spine inclination and extension, and observed excellent correlation coefficients between the range of motion measured by tape measures and radiographs. However, it was not clear whether these coefficients related to patients with AS because analyses were performed on pooled data from both patients and control groups (Table 4)[47,49].

Radiographic analysis was not used as the reference standard in 6 studies. Miller, *et al*[17] assessed the validity of a new tape measure (three 10-cm segment method) used for recording thoraco-lumbar spinal range of motion in the sagittal plane and found substantial to excellent correlations between the new method and goniometry (r = 0.82), modified Schober test (0.77), and fingertip-to-floor distance (0.86). Stokes, *et al*[46] evaluated a new instrument developed

Table 2. Characteristics of individual studies.

Studies	Aims	Participants	Instruments and Procedures	Outcomes, Unit
Moll, *et al* 1972[49]	To present and validate a new method for assessing spinal lateral flexion.	7 patients with AS and 36 controls.	Radiograph and a tape measure. Two marks were inked on the lateral thoraco-lumbar region and the distance between them was measured in neutral position and after lateral flexion.	ROM of lateral flexion measured radiologically (degrees); corresponding distance by tape measure between the 2 marks (cm).
Moll, *et al* 1972[47]	To present and validate a new method for assessing spinal extension.	6 patients with AS and 18 controls.	Radiographs, plumb-line and tape measure. Two marks were inked on the lateral trunk. The plumb-line was used to verify the horizontal distance between the marks after spinal extension.	ROM of thoraco-lumbar extension measured radiologically (degrees); corresponding distance between the 2 marks (cm).
Miller, *et al* 1984[17]	To compare spinal mobility in the sagittal plane between the 10-cm segment method and 3 established methods.	3 male patients with AS and 1 control.	Tape measure and goniometer. Four tests were performed: 10-cm segment method, finger-to-floor distance, modified Schober test, and goniometry.	10-cm segment test, modified Schober test, and finger-floor test (cm), and ROM of spinal flexion by goniometry (degrees).
Stokes, *et al* 1988[46]	To compare the reliability and validity of a portable spinal mobility scale against a flexible measuring tape to measure upper and lower spinal mobility.	5 patients with AS.	Tape measure and portable spinal mobility scale. Finger-to-floor and occiput-to-wall test measures with both instruments were performed.	Finger-to-floor distance (cm); occiput-to-wall distance (cm).
Viitanen, *et al* 1998[18]	Compare cervical ROM between 2 new tape measures (lateral flexion and rotation) with measures from Myrin inclinometer technique.	52 male patients with AS.	Tape measure and inclinometer. Nine measures: occiput-to-wall and tragus-to-wall distance, lateral flexion (tape and inclinometer), rotation (tape and inclinometer), chin-to-chest distance, flexion, and extension (inclinometer).	Occiput-to-wall and tragus-to-wall distance (mm); cervical lateral flexion, and cervical rotation (both in mm and degrees); chin-to-chest distance (mm); cervical flexion and cervical extension (degrees).
Viitanen, *et al* 1999[19]	To verify the reliability and validity of a tape measure method for measuring thoraco-lumbar rotation.	52 male patients with AS.	Tape measure. Four tests: Pavelka thoraco-lumbar rotation, thoraco-lumbar rotation (needle method), modified Schober test, thoraco-lumbar flexion.	Pavelka thoraco-lumbar rotation (cm); thoraco-lumbar rotation (degrees); modified Schober test (cm); thoraco-lumbar flexion (cm).
Rahali-Khachlouf, *et al* 2001[20]	To assess validity and reliability of clinical measures of spinal mobility and sagittal curvatures.	21 male and 1 female patients with AS.	Radiograph and tape measure. Schober and modified Schober test with extra marks for thoraco-lumbar and dorsal mobility. Radiograph (lumbo-pelvic and thoraco-lumbar).	Original and modified Schober (cm); Schober + 10, + 20, and + 30 (cm); finger-to-floor distance (cm); maximum lumbar flexion (degrees) maximum thoraco-lumbar flexion, dorsal, and global flexion (degrees).
Jordan, *et al* 2004[48]	To assess validity of cervical range of motion by comparing the Fastrak measurements to tape measures.	45 male and 5 female patients with AS.	Kinematics and tape measure. Kinematic: flexion, extension, lateral flexion, and rotation; 3 tape measures: modified Schober, finger-to-floor distance, and cervical rotation.	Nine cervical ROM (degrees); modified Schober (cm); finger-to-floor distance (cm); cervical spine rotation (cm).
Maksymowych, *et al* 2006[21]	To compare a tape-based tool for measuring cervical mobility with goniometer-based approach.	191 males and 72 females with AS.	Tape measure and goniometer. Cervical rotation was assessed with both instruments.	Cervical rotation (cm and degrees).
Rezvani, *et al* 2012[12]	To determine the extent the metric measurement methods reflect the lumbar ROM during flexion in patients with AS.	41 male and 9 female patients with AS, and 17 controls.	Radiograph and tape measure. Three tests: original, modified, and modified-modified Schober. Radiograph during neutral and maximal flexion position.	Original, modified, and modified-modified Schober (cm); lumbar erect position (degrees); lumbar flexion position (degrees).

AS: ankylosing spondylitis; ROM: range of motion.

Table 3. Methodological quality of the included studies.

Studies	Items												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Moll, *et al*[49]	N	N	Y	U	Y	Y	Y	U	U	Y	Y	Y	N
Moll, *et al*[47]	N	N	Y	U	Y	Y	Y	U	U	Y	Y	Y	N
Miller, *et al*[17]	N	N	N	Y	Y	Y	U*	Y	N	N	N*	N*	N
Stokes, *et al*[46]	N	U	N	Y	Y	Y	Y	Y	N	N	U	U	N*
Viitanen, *et al*[18]	N*	U*	N	Y	Y	Y	Y	U	N	N	Y*	Y*	N
Viitanen, *et al*[19]	N*	U*	N	Y*	Y	Y	Y	Y	N	N	Y	Y	N
Rahali-Khachlouf, *et al*[20]	N*	Y	Y	U	Y	Y	Y	Y	U	U	N*	U*	U*
Jordan, *et al*[48]	N*	Y	N	U*	Y	Y	Y	U	U	U	Y	N*	Y
Maksymowych, *et al*[21]	Y	N	N	U*	Y	U*	Y	Y	U	U	N*	Y	Y*
Rezvani, *et al*[12]	N*	Y	Y	U	Y	Y	Y	Y	Y	Y	U*	Y	Y

Item 1: Was the spectrum of patients representative of the patients who will receive the test in practice?
Item 2: Were selection criteria clearly described?
Item 3: Is the reference standard likely to correctly measure the target joint range of motion?
Item 4: Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
Item 5: Did the whole sample or a random selection of the sample receive verification using a reference standard of measurement?
Item 6: Did patients receive the same reference standard regardless of the index test result?
Item 7: Was the execution of the index test described in sufficient detail to permit replication of the test?
Item 8: Was the execution of the reference standard described in sufficient detail to permit its replication?
Item 9: Were the index test results interpreted without knowledge of the results of the reference standard?
Item 10: Were the reference standard results interpreted without knowledge of the results of the index test?
Item 11: Were withdrawals from the study explained?
Item 12: Were appropriate descriptive statistics presented?
Item 13: Were appropriate inferential statistics presented (correlation coefficient or agreement stats with CI)?

* Agreement after discussion. N: no; Y: yes; U: unclear.

Table 4. Correlation coefficients between radiography and tape measures.

Variables	#Moll, *et al*[49]	#Moll, *et al*[47]	Rahali-Khachlouf, *et al*[20]	*Rezvani, *et al*[12]
Lumbar spine lateral flexion	0.79			
Lumbar spine extension		0.75		
Original Schober			0.68	0.33
Modified Schober			0.56	0.36
Fingertip-to-floor distance			0.72	

Data from patients and control group pooled. * High-quality study.

to measure the fingertip-to-floor distance and occiput-to-wall distance (an "L" scale) by comparing it with tape measures. No correlation coefficients were presented, but the authors observed differences between these 2 instruments and stated they were not interchangeable[46]. Viitanen, *et al*[19] described a new tape measure method based on the measurement of the distance between the tip of the xiphoid process and the first sacral spinous process before and after rotation (Pavelka rotation method) and compared it with the needle rotation method, modified Schober test, and whole thoracolumbar spine[19]. While the authors claimed that the Pavelka rotation method was valid, no statistical analysis verifying the relationships among the 4 assessed methods was provided[19].

The remaining 3 studies explored the motion of the cervical spine. Viitanen, *et al*[18] evaluated 9 tests (6 by tape measures and 3 by goniometry), but no statistical data regarding the relationship between those tests were presented[18]. Jordan, *et al*[48] observed moderate to substantial correlations between 3-D kinematics of measuring cervical spine movement and other 2 tape measures (modified Schober test and fingertip-to-floor distance). Finally, Maksymowych, *et al*[21] found moderate correlation between cervical spinal rotation recorded by goniometry (reference standard) and tape measure (index test).

DISCUSSION

Although a wide range of clinical tests using simple measurement procedures such as goniometry[17,21,23], inclinometry[18,23], and tape measures[12,18,19,20,21] have been used

to assess spinal mobility in people with AS, literature-based evidence regarding criterion-concurrent validity reflecting true spinal mobility is unclear. Therefore, the objective for our systematic review was to investigate the level of evidence for criterion-concurrent validity for spinal mobility tests in patients with AS.

Only 1 study[12] met the 2 criteria required for high quality. This study used radiograph assessment of range of motion as the reference standard, and met 10 of the 13 items suggested by the modified QUADAS quality assessment tool. Although classified as high quality, it is important to consider the 3 items where the study scored poorly. Rezvani, *et al*[12] assessed a representative population of males, with patients covering a wide range of disease severity and an adequate number of participants (n = 41); however, only a small number of female patients were included (n = 9). Therefore, the generalizability of their findings for a population of patients with AS is uncertain. Two other issues regarding this article were a lack of clarity regarding the time interval between the radiographs and the tape measures, and the number of withdrawals from the study. Although radiological and physical measures for AS are not likely to demonstrate any substantial day-to-day or short-term differences, a potential source of error relates to the known diurnal variability of both symptoms and physical measures in AS if performed at different periods of the day. Variation in time of physical measurement may have affected the results of the study, even though the risk of such bias may be acceptable. Clear documentation for consistency and clarity of time of day for recording of symptoms and range of motion would have been desirable. Given these limitations, the authors concluded that tape measures poorly reflect lumbar spine mobility[12].

Contrasting results were observed by Rahali-Khachlouf, *et al*[20], who suggested tape measures have acceptable psychometric properties to assess patients with AS. However, it is important to note that this study failed in most items (2 "no" and 5 "unclear"). The study was rated as having a high risk of bias and their conclusions appear to be fragile[20]. The radiographic analysis was also used as a reference standard to criterion-concurrent validation of tape measures in 2 further studies; Moll, *et al* assessed the lumbar extension[47] and thoraco-lumbar lateral flexion[49] in 2 unrepresentative samples (6 and 7 patients, respectively) without mentioning the sex of the participants. Moreover, the data recorded from the patients with AS were pooled with those recorded in a larger sample of control participants (18[47] and 36[49] control participants). Therefore, the substantial association between the clinical tests and radiograph presented by the authors does not appear to represent criterion-concurrent validity of lumbar extension or lumbar lateral flexion specifically for patients with AS.

The remaining 6 studies included in our systematic review did not meet the specified standard in both adopted criteria required to be considered high quality[17,18,19,21,46,48]; their reference standards were not radiograph analysis, and they received "yes" in fewer than 10 items. The former criterion was established because radiograph analyses are widely accepted as the ideal reference standard for measuring range of motion[11,12,13,14,15]. Although inclinometry[18], goniometry[17,21], and tape measures[19,46,48] were used as reference standards, these measures have never been subject to criterion-concurrent validation as assessments of spinal mobility in AS. Overall, these studies observed substantial to excellent relationships between goniometry and tape measures to assess spinal range of motion in the sagittal plane[17], moderate association to record cervical mobility[21], and excellent associations between an electromagnetic 3-D tracking system and tape measures to record cervical range of motion[48]. However, they did not reflect acceptable data regarding criterion-concurrent validity.

There appears to be no robust evidence regarding criterion-concurrent validity for clinical tests used to measure spinal mobility in patients with AS. Three out of the 4 studies using a proper reference standard included in our systematic review were classified as low quality[20,47,49], and the only high-quality study[12] suggested clinical tests for assessing mobility of lumbar spine (original, modified, and modified-modified Schober test) poorly reflect the range of motion of this spinal segment in patients with AS. This is of concern because these mobility tests are widely used in routine clinical practice and research[2] where they are considered as validated tools[4,16,25,28,29,34].

Some limitations should be considered. Although we adapted a well-accepted index of quality assessment (QUADAS), these adaptations have never been previously used or validated. Moreover, although some studies have supported a cutoff score of about 70% to consider a study as high quality[44,50], others have questioned this kind of quality score for classifying studies[41]. Finally, while there were no restrictions of language for our systematic review, only keywords in English were used.

The level of evidence considering criterion-concurrent validity of clinical tests commonly used to assess spinal mobility in patients with AS is low. There is only an acceptable level of evidence for criterion-concurrent validity for tests used to assess the lumbar spine, which suggests these tests poorly reflect the mobility of this segment. Based on current literature, there are no high-quality studies supporting the criterion-concurrent validity of clinical tests for spinal mobility in patients with AS. However, these clinical measures are in widespread use because they are highly feasible in that they are easy to perform, low in cost, and rapidly implemented. Therefore, we suggest further research addressing criterion-concurrent validity of mobility tests to establish which of these clinical tests most accurately reflects spinal mobility in patients with AS.

REFERENCES

1. van der Linden S, Valkenburg H, 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. Zochling J, Braun J. Assessments in ankylosing spondylitis. Best Pract Res Clin Rheumatol 2007;21:699-712.
3. Haywood KL, Garratt AM, Jordan K, Dziedzic K, Dawes PT. Spinal mobility in ankylosing spondylitis: reliability, validity and responsiveness. Rheumatology 2004;43:750-7.
4. Jauregui E, Conner-Spady B, Russell AS, Maksymowych WP. Clinimetric evaluation of the Bath Ankylosing Spondylitis Metrology Index in a controlled trial of pamidronate therapy. J Rheumatol 2004;31:2422-8.
5. Amor B, Santos RS, Nahal R, Listrat V, Dougados M. Predictive factors for the longterm outcome of spondyloarthropathies. J Rheumatol 1994;21:1883-7.
6. Machado P, Landewé R, Braun J, Hermann K-GA, Baker D, van der Heijde D. Both structural damage and inflammation of the spine contribute to impairment of spinal mobility in patients with ankylosing spondylitis. Ann Rheum Dis 2010;69:1465-70.
7. Calvo-Gutierrez J, Garrido-Castro JL, Gil-Cabezas J, Gonzalez-Navas C, Font Ugalde P, Carmona L, et al. Spinal mobility in patients with spondylitis is determined by age, structural damage, and inflammation. Arthritis Care Res 2014 Jul 21 (E-pub ahead of print).
8. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Upper Saddle River, NJ: Pearson & Prentice Hall; 2009.
9. Davis JC, Gladman DD. Spinal mobility measures in spondyloarthritis: application of the OMERACT filter. J Rheumatol 2007;34:666-70.
10. Boers M, Brooks P, Strand C, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol 1998;25:198-9.
11. Littlewood C, May S. Measurement of range of movement in the lumbar spine—what methods are valid? A systematic review. Physiotherapy 2007;93:201-11.
12. Rezvani A, Ergin O, Karacan I, Oncu M. Validity and reliability of the metric measurements in the assessment of lumbar spine motion in patients with ankylosing spondylitis. Spine 2012;37:1189-96.
13. Portek I, Pearcy MJ, Reader GP, Mowat AG. Correlation between radiographic and clinical measurement of lumbar spine movement. Rheumatology 1983;22:197-205.
14. Tousignant M, Poulin L, Marchand S, Viau A, Place C. The Modified-Modified Schober Test for range of motion assessment of lumbar flexion in patients with low back pain: A study of criterion validity, intra- and inter-rater reliability and minimum metrically detectable change. Disabil Rehabil 2005;27:553-9.
15. Norkin C, White D. Measurement of joint motion: a guide to goniometry. 4th ed. Philadelphia: F.A. Davis Company; 2009.
16. Maksymowych WP, Mallon C, Richardson R, Conner-Spady B, Jauregui E, Chung C, et al. Development and validation of the Edmonton Ankylosing Spondylitis Metrology Index. Arthritis Rheum 2006;55:575-82.
17. Miller MH, Lee P, Smythe HA, Goldsmith CH. Measurements of spinal mobility in the sagittal plane: new skin contraction technique compared with established methods. J Rheumatol 1984;11:507-11.
18. Viitanen JV, Kokko ML, Heikkilä 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. Rheumatology 1998;37:377-81.
19. Viitanen JV, Kokko ML, Heikkilä S, Kautiainen H. Assessment of thoracolumbar rotation in ankylosing spondylitis: a simple tape method. Clin Rheumatol 1999;18:152-7.
20. Rahali-Khachlouf H, Poiraudeau S, Fermanian J, Ben Salah FZ, Dziri C, Revel M. [Validity and reliability of spinal clinical measures in ankylosing spondylitis]. [Article in French] Ann Readapt Med Phys 2001;44:205-12.
21. Maksymowych WP, Mallon C, Richardson R, Conner-Spady B, Jauregui E, Chung C, et al. Development and validation of a simple tape-based measurement tool for recording cervical rotation in patients with ankylosing spondylitis: comparison with a goniometer-based approach. J Rheumatol 2006;33:2242-9.
22. Garrido-Castro J, Escudero A, Medina-Carnicer R, Galisteo A, Gonzalez-Navas C, Carmona L, et al. Validation of a new objective index to measure spinal mobility: the University of Cordoba Ankylosing Spondylitis Metrology Index (UCOASMI). Rheumatol Int 2013;34:401-6.
23. Jenkinson T, Mallorie P, Whitelock H, Kennedy L, Garrett S, Calin A. Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. J Rheumatol 1994;21:1694-8.
24. Jones S, Porter J, Garrett S, Kennedy L, Whitelock H, Calin A. A new scoring system for the Bath Ankylosing Spondylitis Metrology Index (BASMI). J Rheumatol 1995;22:1609.
25. van der Heijde D, Landewé R, Feldtkeller E. Proposal of a linear definition of the Bath Ankylosing Spondylitis Metrology Index (BASMI) and comparison with the 2-step and 10-step definitions. Ann Rheum Dis 2008;67:489-93.
26. Aggarwal R, Malaviya A. Clinical characteristics of patients with ankylosing spondylitis in India. Clin Rheumatol 2009;28:1199-205.
27. van der Heijde D, Deodhar A, Inman RD, Braun J, Hsu B, Mack M. Comparison of three methods for calculating the Bath Ankylosing Spondylitis Metrology Index in a randomized placebo-controlled study. Arthritis Care Res 2012;64:1919-22.
28. Karapolat H, Akkoc Y, Sarı İ, Eyigor S, Akar S, Kirazlı Y, et al. Comparison of group-based exercise versus home-based exercise in patients with ankylosing spondylitis: effects on Bath Ankylosing Spondylitis Indices, quality of life and depression. Clin Rheumatol 2008;27:695-700.
29. Fernandez-de-Las-Penas C, Alonso-Blanco C, Morales-Cabezas M, Miangolarra-Page JC. Two exercise interventions for the management of patients with ankylosing spondylitis: a randomized controlled trial. Am J Phys Med Rehabil 2005;84:407-19.
30. Martindale J, Sutton C, Goodacre L. An exploration of the inter- and intra-rater reliability of the Bath Ankylosing Spondylitis Metrology Index. Clin Rheumatol 2012;31:1627-31.
31. Maksymowych WP, Mallon C, Richardson R, Conner-Spady B, Chung C, Russell AS. Does height influence the assessment of spinal and hip mobility measures used in ankylosing spondylitis? J Rheumatol 2006;33:2035-40.
32. Garrido-Castro JL, Medina-Carnicer R, Schiottis R, Galisteo AM, Collantes-Estevez E, Gonzalez-Navas C. Assessment of spinal mobility in ankylosing spondylitis using a video-based motion capture system. Man Ther 2012;17:422-6.
33. Chilton-Mitchell L, Martindale J, Hart A, Goodacre L. Normative values for the Bath Ankylosing Spondylitis Metrology Index in a UK population. Rheumatology 2013;52:2086-90.
34. Kennedy LG, Jenkinson TR, Mallorie PA, Whitelock HC, Garrett SL, Calin A. Ankylosing spondylitis: the correlation between a new metrology score and radiology. Rheumatology 1995;34:767-70.
35. Wanders A, Landewé R, Dougados M, Mielants H, van der Linden S, 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.
36. Almodóvar R, Zarco P, Collantes E, González C, Mulero J, Fernández-Sueiro JL, et al. Relationship between spinal mobility and disease activity, function, quality of life and radiology. A cross-sectional Spanish registry of spondyloarthropathies (REGISPONSER). Clin Exp Rheumatol 2009;27:439-45.

37. Kaya T, Gelal F, Gunaydin R. The relationship between severity and extent of spinal involvement and spinal mobility and physical functioning in patients with ankylosing spondylitis. Clin Rheumatol 2006;25:835-9.
38. Viitanen JV, Kokko M-L, Lehtinen K, Suni J, Kautiainen H. Correlation between mobility restrictions and radiologic changes in ankylosing spondylitis. Spine 1995;20:492-6.
39. Viitanen JV, Heikkilä 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.
40. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Ann Intern Med 2009;151:264-9.
41. Whiting P, Rutjes A, Reitsma J, Bossuyt P, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
42. Williams MA, McCarthy CJ, Chorti A, Cooke MW, Gates S. A systematic review of reliability and validity studies of methods for measuring active and passive cervical range of motion. J Manipulative Physiol Ther 2010;33:138-55.
43. Cook C, Hegedus E. Diagnostic utility of clinical tests for spinal dysfunction. Man Ther 2011;16:21-5.
44. Cook C, Mabry L, Reiman MP, Hegedus EJ. Best tests/clinical findings for screening and diagnosis of patellofemoral pain syndrome: a systematic review. Physiotherapy 2012;98:93-100.
45. Williams R, Goldsmith C, Minuk T. Validity of the double inclinometer method for measuring lumbar flexion. Physiother Can 1998;50:147-52.
46. Stokes BA, Helewa A, Goldsmith CH, Groh JD, Kraag GR. Reliability of spinal mobility measurements in ankylosing spondylitis patients. Physiotherapy Canada 1988;40:338-44.
47. Moll JM, Liyanage SP, Wright V. An objective clinical method to measure spinal extension. Rheumatol Phys Med 1972;11:293-312.
48. Jordan K, Haywood KL, Dziedzic K, Garratt AM, Jones PW, Ong BN, et al. Assessment of the 3-dimensional Fastrak measurement system in measuring range of motion in ankylosing spondylitis. J Rheumatol 2004;31:2207-15.
49. Moll JM, Liyanage SP, Wright AV. An objective clinical method to measure lateral spinal flexion. Rheumatology 1972;11:225-39.
50. May S, Littlewood C, Bishop A. Reliability of procedures used in the physical examination of non-specific low back pain: a systematic review. Aust J Physiother 2006;52:91-102.