The Minimally Important Difference for Patient-reported Outcomes in Spondyloarthropathies including Pain, Fatigue, Sleep, and Health Assessment Questionnaire

LAURA WHEATON and JANET POPE

ABSTRACT. *Objective.* To study minimal important differences (MID) in spondyloarthropathies (SpA). MID are important in determining clinically relevant changes and for interpretation of trials and treating patients. MID have been widely studied in rheumatoid arthritis, but less so in SpA.
Methods. Patients with SpA had to be seen for 2 consecutive visits and have completed the Health Assessment Questionnaire (HAQ) and 100 mm visual analog scale on both visits for fatigue, pain, sleep, and global assessment. At the second visit they had to answer a question regarding any change in their overall health (from last visit), responding with much better, better, same, worse, or much worse. The MID were the mean changes for those who were either better or worse.
Results. Our study involved 140 eligible patients with a SpA: 69% were men, the mean age was 45 years, and the mean disease duration was 14.5 years. Almost half the patients rated themselves as unchanged from the previous visit but the remainder were either better or worse, with a minority rating themselves as much better or much worse. The MID for better and worse outcomes were HAQ (–0.136; 0.220), pain (–6.93; 18.97), fatigue (–1.43; 14.42), and sleep (–2.23; 10.76). No gender differences were observed.
Conclusion. Our results demonstrate that the MID vary depending on better versus worse (bidirectionally different). MID may be smaller in clinical practice than what is observed in trials. (First Release Feb 1 2010; J Rheumatol 2010;37:816–22; doi:10.3899/jrheum.090086)

Key Indexing Terms:
MINIMALLY IMPORTANT DIFFERENCE   ANKYLOSING SPONDYLITIS
SERONEGATIVE SPONDYLOARTHROPATHY   PATIENT OUTCOMES
HEALTH ASSESSMENT QUESTIONNAIRE   VISUAL ANALOG SCALE

Ankylosing spondylitis (AS) is a chronic and painful inflammatory arthritis of the spine and sacroiliac joints[1]. Spondyloarthropathies (SpA) consist of various seronegative inflammatory spondyloarthritis diseases including AS, Crohn's spondylitis, and reactive arthritis with spondyloarthropathy[2]. SpA may result in lifelong physical impairment and functional disability. Because of the chronic character of SpA and the extensive negative effect on patients' daily lives, patient-reported outcomes can be measured such as pain, function, fatigue, sleep, and global health[3-9]. Visual analog scales (VAS) can also evaluate patient-reported outcomes; often rated on a scale from 0 (no problem) to 100 (severe problem), they can be used for several outcomes[9,10]. The Health Assessment Questionnaire-Disability Index (HAQ-DI) assesses the self-reported functional status for performing activities of daily living and is scored from 0 (no disability) to 3 (severe disability)[11]. Higher scores (> 1) are predictive of worse outcomes in rheumatoid arthritis (RA)[12]. There is a modification of the HAQ-DI used specifically in spondylitis called the Health Assessment Questionnaire for the Spondyloarthropathies (HAQ-S)[13]. The HAQ-DI was initially developed in RA and is weighted toward evaluating peripheral joint problems. The HAQ-S incorporates 5 additional items into the HAQ-DI, evaluating spinal function (such as the ability to carry heavy packages, to sit for long periods of time, and to rotate the neck to look behind when driving a car in reverse)[13]. The Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and the Bath Ankylosing Spondylitis Functional Index (BASFI) are also commonly used to evaluate SpA activity, function, and progression[14,15].

In patient-reported measures, it is important to determine whether differences are meaningful to individuals. Minimally important difference (MID) is the smallest difference in a measure that patients perceive as change[16-18]. It is the smallest change in a score that would be considered relevant, and MID are useful for determining relevance of interventions or changes in status in randomized controlled trials and clinical practice. In order to determine the MID,

From the Schulich School of Medicine and Dentistry, St. Joseph's Health Care, University of Western Ontario, London, Ontario, Canada.

L. Wheaton, BSc, Medical Student; J. Pope, MD, MPH, FRCPC, Schulich School of Medicine and Dentistry, University of Western Ontario.

Address correspondence to Dr. J. Pope, St. Joseph's Health Care, 268 Grosvenor St., London, ON N6A 4V2, Canada.
E-mail: janet.pope@sjhc.london.on.ca

Accepted for publication November 12, 2009.

an external anchor such as a physician or patient-reported global assessment is used. The MID may reflect either an improvement or a worsening of the disease state. The minimal clinically important improvement is defined as the smallest change in measurement that signifies an important improvement. A related concept is the patient acceptable symptom state (PASS), which has been defined as the highest level of symptoms where patients consider themselves well, and beyond which they would not be well[19,20]. Changes of 10% or about 10 on a 100 mm VAS correspond to the MID for patient-reported measures across many studies. Dougados, *et al* found that the mean change in BASDAI over 12 weeks for the patient to report feeling well was –3.5 (SD 2.3) on the BASDAI (0–10 cm VAS) and –2.4 (SD 2.0) on the BASFI (0–10 cm VAS)[21]. This result appears to be significantly greater than the results found for the MID of BASDAI and BASFI, which were 1.0 cm and 0.7 cm, respectively[22]. These scores were found to be independent of the patients' baseline scores. However, there are differences from a patient perspective in detecting minimal change in an outcome compared to feeling well.

We studied a large clinical practice of patients with seronegative SpA (SpA including AS, inflammatory bowel disease-related spondylitis, and reactive arthritis with spondylitis) at a university outpatient clinic (where SpA was diagnosed by the treating rheumatologist) to determine the MID estimates (for improvement and worsening) for HAQ-DI and pain, fatigue, sleep, and patient-reported global health measurements on a 100 mm VAS. We hypothesized that MID scores would be different bidirectionally (improvement and worsening), as shown in RA and other musculoskeletal diseases[23-25]. We did not include BASDAI or BASFI measurements in this study, since few of the patients included had consecutive BASDAI/BASFI data, and the MID were previously determined[22].

MATERIALS AND METHODS

Our study was approved by the University of Western Ontario (UWO) ethics board. Multiple data are collected routinely on patients seen at St. Joseph's Hospital Rheumatology Clinic, which is affiliated with the UWO, Schulich School of Medicine and Dentistry, and serves a region of > 1 million people. The data are from patients diagnosed with SpA who had been seen by 5 rheumatologists and for whom there were data on the outcomes of interest for at least 2 consecutive visits, which were not more than 16 months apart. Data were extracted from medical charts by a trained data extractor and entered into Excel and SPSS databases. Patients completed the HAQ-DI, pain, fatigue, sleep, and global health VAS, with scales from 0 (no problem) to 100 (very severe), and responded to a 5-point health status Likert scale question: "How would you describe your overall status since the last visit?". The possible responses were much better, better, same, worse, much worse. Patients who reported better or worse were defined as the minimally changed subgroups. The changes in the HAQ-DI and pain, fatigue, sleep, and global health VAS scores for the groups that indicated better or worse on the 5-point Likert scale at their next visit were determined in order to estimate the MID. This was compared to change scores for the groups that reported same, much better, or much worse. The scales are used for all patients at each visit and were not specifically vali-

dated for our study. However, we have used the same scales for MID calculations in RA[10,25].

The HAQ-DI change scores were calculated (HAQ from most recent minus previous visit), so a negative score implied an improved HAQ and a positive score implied worsening.

To assess the usefulness of an anchor, change in the anchor and change in the patient-reported outcome scores should have a correlation coefficient of at least 0.37[26]. Thus Spearman correlation coefficients (for change in anchor) of overall health status (Likert ordinal scale) and change in the HAQ-DI and pain, fatigue, sleep, and global health scores were calculated.

The data were analyzed using SPSS software, where $p < 0.05$ was considered statistically significant. Results were reported in mean (SD) unless otherwise stated. In exploratory analyses we compared (1) MID in men and women; (2) subdividing patients with high versus low HAQ; and (3) high versus low pain, using the earlier visit for the stratification of patients into the high or low group. The HAQ-DI was divided at 1.0 as it can separate moderate versus low self-reported functional disability[27], and pain VAS was divided at pain in the lower one-third of the 100 mm VAS or pain greater than 33.4 mm to determine whether the MID were similar in various subsets (i.e., we compared low pain with moderate or worse pain to see if the MID were different at different points of the pain VAS scale). We also determined the proportion that were concordant with their assessment of same, better, worse, much better, or much worse with respect to HAQ scores.

RESULTS

There were 211 patients considered eligible for the study, but 71 patients did not have complete data and were excluded from the reported analysis, leaving a final sample of 140 patients. The 140 patients with SpA had long disease duration (mean 14.5 yrs) and an average age of 44.7 (11.8) years, and 69.3% were men. Most used nonsteroidal antiinflammatory drugs (NSAID), about 12% were taking anti-tumor necrosis factor (TNF) therapies, and those with peripheral arthritis were taking disease-modifying antirheumatic drugs such as sulfasalazine and methotrexate. The mean baseline HAQ-DI was 0.72 (0.57) with little mean change at followup (Table 1). Forty-one percent reported no change in overall health status at the followup visit, 25% reported they were better, and 7% reported much better; 24% reported worse and 3% much worse. The data curves for changes in all outcomes were bell-shaped (i.e., a large minority were unchanged, with the next highest groups being minimally changed and the least being much better or much worse). The Spearman correlation coefficient between the patient assessment of global change and change in the HAQ-DI was good [0.485 ($p < 0.01$)]. The HAQ-DI changed score for the group that described being better and worse was mean –0.136 (SD 0.228) and 0.220 (SD 0.354), respectively (Table 2). This difference was statistically significant compared to the no-change group [HAQ-DI mean 0.005 (SD 0.263); Table 2]. A greater change was required for women to feel much better or much worse, similar changes for the men to feel better and the same, and less change to feel worse. The Spearman correlation coefficient between the patient assessment change in overall status and change in the pain VAS was 0.394 ($p < 0.01$). Patients who answered better and worse on the overall status anchor changed by

Table 1. Characteristics of the study population of 140 patients at baseline and followup.

Characteristics	Baseline	Followup	Mean Change Between Visits
Age, yrs, mean (SD)	44.7 (11.8)		
Men, %	69.3	—	—
Disease duration, yrs, mean (SD)	14.5 (10.3, range 1–45)		
Visit time interval, mo, mean (SD)	5.6 (3.8, range 1–16)		
HAQ-DI (0–3), mean (SD)	0.72 (0.57, range 0–2)	0.73 (0.61)	+0.017 (0.329)
Pain VAS score (0–100), mean (SD)	45.72 (26.33)	46.70 (27.55)	1.31 (26.92)
Fatigue VAS score (0–100), mean (SD)	49.11 (27.28)	50.30 (29.56)	1.54 (23.80)
Sleep VAS score (0–100), mean (SD)	46.59 (30.23)	45.00 (30.49)	–0.92 (26.58)
Global health VAS score (0–100), mean (SD)	43.20 (24.34)	42.66 (25.12)	+0.07 (19.72)

HAQ-DI: Health Assessment Questionnaire-Disability Index; VAS: visual analog scale.

Table 2. Minimal important difference (MID) estimates for Health Assessment Questionnaire-Disability Index (HAQ-DI) and pain, fatigue, sleep, and global health 100 mm visual analog scale (VAS). HAQ-DI scored from 0 to 3; improved is –, worsening is +. For the VAS scales of pain, fatigue, sleep, and global health, 0 is no problem, 100 is worst ever; improved is –, worsened is +. MID is determined as the mean of those who are better or worse.

Patient-rated Overall Status	HAQ-DI Change, mean (SD)	95% CI	Pain VAS Change, mean (SD)	95% CI	Fatigue VAS Change, mean (SD)	95% CI	Sleep VAS Change, mean (SD)	95% CI	Global Health Change, mean (SD)	95% CI
Much better (n = 10)	–0.250 (0.276)	–0.448, –0.052	–14.60 (26.31)	–33.42, 4.22	–14.30 (20.58)	–29.02, 0.42	–19.30 (30.16)	–40.87, 2.27	–8.60 (15.74)	–19.86, 2.66
Better (n = 35)	–0.136 (0.228)	–0.214, –0.058	–6.83 (27.14)	–16.15, 2.50	–1.43 (29.18)	–11.45, 8.60	–2.23 (29.35)	–12.31, 7.85	–4.03 (15.13)	–9.23, 1.17
Same (n = 58)	0.005 (0.263)	–0.064, 0.074	–2.03 (22.77)	–8.02, 3.95	–1.90 (19.92)	–7.13, 3.34	–3.29 (21.16)	–8.86, 2.27	–6.09 (18.29)	–10.89, –1.28
Worse (n = 33)	0.220 (0.354)	0.094, 0.345	18.97 (27.00)	9.40, 28.54	14.42 (20.45)	7.17, 21.67	10.76 (28.94)	0.50, 21.02	16.85 (18.90)	10.15, 23.55
Much worse (n = 4)	0.531 (0.483)	–0.237, 1.299	15.00 (4.76)	7.42, 22.58	10.75 (14.91)	–12.97, 34.47	–5.50 (5.26)	–13.87, 2.87	8.50 (11.82)	–10.31, 27.31

–6.83 (27.14) and 18.97 (27.00) mm on the VAS (Table 2). These estimates were larger than in the no-change group [–2.03 (22.77)]. The Spearman correlation coefficient for fatigue VAS was weaker (0.288; p < 0.01) than for HAQ-DI and pain, and for fatigue MID only the worse change was larger than the no-change group. The Spearman correlation coefficient between sleep and overall health was 0.223 (p < 0.01) and the MID estimates were different for better and worse (bidirectionally different). Baseline global health VAS was 43.20 (24.34); the Spearman correlation coefficient between the patient assessment of change in overall status and change in global health was 0.383 (p < 0.01). The

Table 3. Exploratory analyses for MID divided by baseline low and high HAQ scores and baseline low and high pain scores. Improved is –, worsened is +.

	HAQ-DI Baseline Score		Pain VAS Baseline Score	
	HAQ-DI < 1.0	HAQ-DI ≥ 1.0	Pain VAS < 33.3	Pain VAS ≥ 33.4
MID (SD) in the slightly improved group	–0.092 (0.207) (n = 27)	–0.281 (0.248) (n = 8)	–12.588 (22.020) (n = 17)	–25.167 (16.853) (n = 18)
MID (SD) in same group	0.053 (0.222) (n = 38)	–0.085 (0.314) (n = 20)	9.095 (22.176) (n = 21)	–8.351 (20.841) (n = 37)
MID (SD) in slightly worsened group	0.221 (0.267) (n = 22)	0.216 (0.500) (n = 11)	34.167 (21.049) (n = 12)	10.286 (26.550) (n = 21)

MID: minimally important difference; HAQ: Health Assessment Questionnaire; DI: Disability Index; VAS: visual analog scale.

worse change was larger than the no-change group, but the better group had a smaller change than the no-change group [–6.09 (18.29)]. Most patients (83%) had some morning stiffness at both baseline and followup visits. MID for morning stiffness was 17 minutes less for better and 56 minutes longer for worse. It was noted that many patients reported stiffness in 15-minute segments (i.e., 0, 15, 30, 45, 60 min). The MID results for HAQ-DI and pain did not differ when stratifying by gender in an exploratory analysis.

Table 3 provides the mean (SD) MID scores for improvement in the HAQ-DI scores stratified by severity of the mild (HAQ-DI < 1.0) versus moderate/severe (HAQ-DI ≥ 1.0) groups at baseline. Although exploratory, the data suggest that patients with high baseline HAQ-DI scores required a larger improvement to be minimally improved, as exemplified by the HAQ-DI (MID for the minimally improved group = –0.09; high score group = –0.28). In contrast, the change required for the worse group with moderate to severe HAQ-DI (0.21) was very similar to that for the mild HAQ-DI group (0.22). A similar pattern was seen with the pain VAS for the better group, but in this case the change required for the worse group with moderate to severe pain VAS (10.3) was smaller than that for the mild pain VAS group (34.2). This pattern was also observed in the rest of the VAS scores.

We examined the proportion of changes in patient-reported global states that were concordant with change in HAQ (Table 4). As might be expected, for larger changes such as much better (60% concordance) or much worse (75% concordance), there was significant agreement. However, in better and worse the HAQ was concordant just as often (60%) and in the same group, the HAQ had no change (0 difference) in one-third. The rest had some change of the HAQ scores between visits.

DISCUSSION
MID estimates provide a benchmark by helping researchers and clinicians understand whether patient-reported outcome differences between 2 treatment groups are meaningful, or whether changes within 1 group over time are meaningful[3,4,16]. On average, our MID score for HAQ-DI improvement was –0.136 (SD 0.228). Worsening was a larger

change of 0.220 (0.354) and did not differ between men and women. These MID values are small and similar to reports in RA, indicating that a change of 1 to 2 items on the HAQ can be perceived by a patient as a relevant difference[25].

Pain (using 100 mm VAS) yielded MID scores of –6.83 (27.14) and 18.97 (27.00) for improvement and worsening, a finding that may imply bidirectional differences. The correlations, not surprisingly, were less for fatigue and sleep than for pain, as the former 2 may be more multidimensional and not exclusively related to SpA or disease activity[28]. It is interesting to note that in each patient-reported outcome, the MID for worsening is about 3 times as much as the MID for improvement. When the data were divided by sex, no clear trends were found. Of note, however, for pain it took more for women to feel better and less to feel worse than for men, but these analyses are only exploratory. Others have observed that women are more sensitive to pain than men[29].

In our study, few of the patients had BASDAI/BASFI data. We primarily use BASDAI to monitor response with anti-TNF biologics. However, in the study by Pavy, *et al* the MID for both of these indices were discovered to be 7 mm for the BASFI and 10 mm for the BASDAI[22]. These values are similar to our observations of pain, fatigue, sleep, and global health. We cannot compare the MID to BASDAI and BASFI results within our study as they were not routinely collected. Many patients did not report change between visits but still may have had slightly different scores on various scales, allowing one to interpret that any mean MID of slightly better or worse has a range around it. There could be floor effects, as the patients had a low baseline HAQ-DI, but it is unlikely as the MID was similar to what we have found in RA[10,25]. The MID must be interpreted in the context of what population is studied (baseline levels for the scales studied or disease severity and activity); and MID is dependent on the anchor used[30]. In this study we used the patient change in overall status, where the MID was used for those who were worse or better. Others have found that fatigue is somewhat related to pain, global assessment, and function in AS, but NSAID treatment in AS had less effect on this domain than function and pain[9].

As noted, the MID estimates may depend on the baseline scores. We saw this trend when people with higher baseline

Table 4. Proportion of change scores concordant with directional trend of category.

Overall Status (Direction)	N	Negative HAQ Change (Improved), %	Neutral HAQ Change (No Change), %	Positive HAQ Change (Worsened), %
Much better (–)	10	60	30	10
Better (–)	35	60	23	17
Same (0)	58	31	34	34
Worse (+)	33	9	27	64
Much worse (+)	4	25	0	75

HAQ: Health Assessment Questionnaire.

scores (defined as baseline score of HAQ-DI ≥ 1.0 and VAS scores ≥ 33.4) required a larger improvement in their HAQ-DI and VAS scores to be considered as minimally improved. Conversely, people with lower baseline scores (defined as baseline scores of HAQ-DI < 1.0 and VAS < 33.4) required a larger worsening to be considered as minimally worsened by the patients. We are not certain why this is the case. One would predict larger changes in both improved and worsened HAQ-DI and VAS scores at higher baseline values (and thus percentage changes would be similar), but this was not observed. In general, patients with SpA who scored better on their patient-reported outcomes may be more optimistic and require larger changes for worsening than improving. In treatment trials there is an expectation that improvement may occur and patients often are enrolled with a high level of disease activity. Thus, in general, trials do not study MID that are worsening (as even placebo-treated patients can improve somewhat or remain stable over the duration of the study), while active treatment on average improves outcomes[31]. So, the worsening of a MID may not be as relevant in a trial, but can be important in clinical practice to try to determine if a patient is a bit worse and then adjust treatment instead of waiting until a patient is much worse. The associations of change in disease status with fatigue and sleep were less than HAQ-DI and pain, as would be expected, so the MID may be less robust for those outcomes[32].

Not many patients were much better or much worse, so the mean changes in these groups for all outcomes are not necessarily larger than those who were minimally changed and that could be due to small sample size. However, the MID was calculated only on those who were minimally changed and the data were distributed as a bell curve, which allows us to have confidence in the results of the 2 minimally changed groups. The concordance is not 100% for the change in HAQ-DI being the same direction as the change in overall health status. These are real-world data and a large enough correlation to calculate the MID was obtained[26,33-35]. It is important to remember that a change in overall status is affected by many patient factors, and function as reported by the HAQ-DI is linked to disease activity (reversible component), damage (irreversible component), and other factors such as patient perception, age, and even other problems such as mechanical back problems that can be superimposed on the spondyloarthropathy[36].

We did not collect the characteristics of the patients who did not complete their forms, and they may have different results, thus there is a potential bias. The patients were primarily Caucasians and had to be able to read English in order to complete their forms. Most patients had long disease duration, so the results may not be generalizable to very early SpA. The time reference for the HAQ-DI is over the previous week and we asked for the Likert scale "since the last visit," which was usually 6 to 8 months before. The

duration between visits may have been a limitation. Retrospective ratings are susceptible to recall biases. In addition to forgetting, recall has been shown to be influenced by more salient events and by one's current mood. Although patients often ignore time anchors on questions, a longer visit interval could have increased recall bias, which could have resulted in increased random error and therefore, increased or decreased MID. The choice of anchors can influence the MID scores[34,35]. The sample sizes for a number of the subgroups were quite small, and this could unduly affect the results if these groups included outlying data. Also, treatment was not necessarily similar between the rheumatologists.

However, despite these possible limitations, the MID on several VAS scales is similar to observations using BAS-DAI. One may also question whether the changes are relevant because the perspective was anchored to the patient's change in health status, and we have not reported physician measures or changes in treatment because the study was done through a chart review, so the physician-collected outcomes had not been standardized. However, the data are similar to those found in other reports on VAS changes in SpA[3,18,22] that used different methods. Thus, the data presented may be interpreted with caution and the context needs to be remembered (patients seen over 2 sequential visits in a clinic setting, using only patient-reported outcomes). We did not standardize the outcomes in the physical examination measurements as the clinic has several assessors (including trainees in rheumatology for a month), so the physician outcomes were not included or used to calculate the MID. We should emphasize that the study had several limitations including variable followup between assessments, small sample sizes, and failure to compare those included in the study to those who were not included, as nearly everyone completed the forms and those who could not read did not. For these reasons the results cannot necessarily be generalized to all patients with AS who attend rheumatology clinics for regularly scheduled assessments.

Patients whose overall health status changed were not fully concordant with what they reported in their HAQ. The neutral HAQ was calculated in the patients who reported that they were overall "the same" as last visit, subtracting HAQ-DI at baseline visit from followup visit. So if a person had 1 different change of HAQ (0.125) or more, that person was considered changed. This may reflect heterogeneity in the responses (i.e., a moderate correlation between the 2 scales was considered enough to determine a MID); but it may also be because the HAQ reflects both damage and disease activity, so some of the HAQ is not reversible, despite a patient reporting improvement when damage is present[36]. Because of long disease duration, it is likely that some of the HAQ would not be reversible in our population. Unfortunately we did not have data on the HAQ-S, but can use the estimates obtained from the HAQ-DI to compare the

MID in SpA to other musculoskeletal diseases such as RA. Further studies could also include the PASS to determine how much minimal change is needed to have a patient exit from the PASS. The HAQ-S is more functionally related to SpA compared to the HAQ-DI[13,37], but the HAQ-DI and health transition had a good correlation in this study. The standard deviations of the calculated MID and the no-change group were large; and not all patients were concordant on the outcomes of interest and the global health changes, so our observations are "averages" where patients who rated themselves unchanged may have nevertheless achieved a MID. To further validate the results of this study, it would be beneficial to include objective or clinical measures of the patient's status, since this study relied solely upon patient-reported outcomes.

MID scores may be bidirectionally different in SpA and may vary depending on high versus low baseline scores. The MID scores may also be influenced by disease duration, and our study does not comment on very early SpA. MID for worsening usually need a larger value than improving. The MID results seem to be comparable to other studies that calculated MID using different anchors in SpA. The HAQ-DI MID in SpA is relatively small and approximates what has been observed in RA.

ACKNOWLEDGMENT

REFERENCES

1. Martindale J, Smith J, Sutton CJ, Grennan D, Goodacre L, Goodacre JA. Disease and psychological status in ankylosing spondylitis. Rheumatology 2006;45:1288-93.
2. Forouzesh S, Bluestone R. The clinical spectrum of ankylosing spondylitis. Clin Orthop Relat Res 1979;143:53-8.
3. 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.
4. Haywood KL, Garratt AM, Dziedzic K, Dawes PT. Generic measures of health-related quality of life in ankylosing spondylitis: reliability, validity and responsiveness. Rheumatol 2002;41:1380-7.
5. Ward MM, Weisman MH, Davis JC, Reveille JD. Risk factors for functional limitations in patients with long-standing ankylosing spondylitis. Arthritis Rheum 2005;53:710-7.
6. Turan Y, Duruoz MT, Bal S, Guvenc A, Cerrahoglu L, Gurgan A. Assessment of fatigue in patients with ankylosing spondylitis. Rheumatol Int 2007;27:847-52.
7. Borman P. Functional disability and quality of life in patients with ankylosing spondylitis. Rheumatol Int 2004;24:59-60.
8. Haywood KL, Garratt AM, Dziedzic K, Dawes PT. Patient centered assessment of ankylosing spondylitis-specific health related quality of life: evaluation of the patient generated index. J Rheumatol 2003;30:764-73.
9. Dernis-Labous E, Messow M, Dougados M. Assessment of fatigue in the management of patients with ankylosing spondylitis. Rheumatology 2003;42:1523-8.
10. Khanna D, Pope J, Khanna PP, Maloney M, Samedi N, Norrie D, et al. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. J Rheumatol 2008;35:2339-43.
11. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
12. Wolfe F, Michaud K, Gefellar O, Choi HK. Predicting mortality in patients with rheumatoid arthritis. Arthritis Rheum 2003;48:1530-42.
13. Daltroy LH, Larson MG, Roberts NW, Liang MH. A modification of the Health Assessment Questionnaire for the spondyloarthropathies. J Rheumatol 1990;17:946-50.
14. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease activity in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.
15. Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.
16. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407-15.
17. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 2003;56:395-407.
18. Falgarone G, Zerkak D, Bauer C, Messow M, Dougados M. How to define a minimal clinically individual state (MCIS) with pain VAS in daily practice for patients suffering from musculoskeletal disorders. Clin Exp Rheumatol 2005;23:235-8.
19. Maksymowych WP, Richardson R, Mallon C, van der Heijde D, Boonen A. Evaluation and validation of the patient acceptable symptom state (PASS) in patients with ankylosing spondylitis. Arthritis Rheum 2007;57:133-9.
20. Tubach F, Pham T, Skomsvoll JF, Mikkelsen K, Bjorneboe O, Ravaud P, et al. Stability of the patient acceptable symptomatic state over time in outcome criteria in ankylosing spondylitis. Arthritis Rheum 2006;55:960-3.
21. Dougados M, Luo MP, Maksymowych WP, Chmiel JJ, Chen N, Wong RL, et al. Evaluation of the patient acceptable symptom state as an outcome measure in patients with ankylosing spondylitis: data from a randomized controlled trial. Arthritis Rheum 2008;59:553-60.
22. Pavy S, Brophy S, Calin A. Establishment of the minimum clinically important difference for the Bath ankylosing spondylitis indices: a prospective study. J Rheumatol 2005;32:80-5.
23. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478-87.
24. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004;8:283-91.
25. Pope JE, Khanna D, Norrie D, Ouimet J. The minimally important difference for the Health Assessment Questionnaire in rheumatoid arthritis clinical practice is smaller than in randomized controlled trials. J Rheumatol 2009;36:254-9.
26. Cohen J. A power primer. Psychol Bull 1992;112:155-9.
27. Davis JC, Revicki D, van der Heijde DMF, Rentz AM, Wong RL, Kupper H, et al. Health-related quality of life outcomes in patients with active ankylosing spondylitis treated with adalimumab: results from a randomized controlled study. Arthritis Rheum 2007;57:1050-7.
28. Ward MM. Outcome measurement: health status and quality of life. Curr Opin Rheumatol 2004;16:96-101.

29. Rustoen T, Wahl A, Hanestad B, Lerdal A, Paul S, Miaskowski C. Gender differences in chronic pain — findings from a population-based study of Norwegian adults. Pain Manag Nurs 2004;5:105-17.

30. Hays RD, Brodsky M, Johnston MF, Spritzer KL, Hui KK. Evaluating the statistical significance of health-related quality of life change in individual patients. Eval Health Prof 2005;28:160-71.

31. Gonzalez-Lopez L, Garcia-Gonzalez A, Vazquez-Del-Mercado M, Munoz-Valle JF, Gamez-Nava JI. Efficacy of methotrexate in ankylosing spondylitis: a randomized, double blind, placebo controlled trial. J Rheumatol 2004;31:1568-74.

32. Ortiz Z, Shea B, Garcia Dieguez M, Boers M, Tugwell P, Boonen A, et al. The responsiveness of generic quality of life instruments in rheumatic diseases. A systematic review of randomized controlled trials. J Rheumatol 1999;26:210-6.

33. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102-9.

34. Hays RD, Farivar S, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. COPD 2005;2:63-7.

35. Khanna D, Tsevat J. Health-related quality of life — an introduction. Am J Manag Care 2007;13:S218-23.

36. Aletaha D, Smolen J, Ward MM. Measuring function in rheumatoid arthritis: identifying reversible and irreversible components. Arthritis Rheum 2006;54:2784-92.

37. Shinjo SK, Goncalves R, Kowalski S, Goncalves CR. Brazilian-Portuguese version of the Health Assessment Questionnaire for Spondyloarthropathies (HAQ-S) in patients with ankylosing spondylitis: a translation, cross-cultural adaptation, and validation. Clin Rheumatol 2007;26:1254-8.