

## EXTENDED REPORT

# Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis

L C Doward, A Spoorenberg, S A Cook, D Whalley, P S Helliwell, L J Kay, S P McKenna, A Tennant, D van der Heijde, M A Chamberlain

*Ann Rheum Dis* 2003;**62**:20–26

See end of article for authors' affiliations

Correspondence to:  
Dr L C Doward, Galen Research, Enterprise House, Manchester Science Park, Lloyd Street North, Manchester M15 6SE, UK; lcd@galen.eng.net

Accepted 8 May 2002

**Background:** Although disease-specific health status measures are available for ankylosing spondylitis (AS), no instrument exists for assessing quality of life (QoL) in the condition.

**Objective:** To produce an AS-specific QoL measure that would be relevant and acceptable to respondents, valid, and reliable.

**Methods:** The ASQoL employs the needs-based model of QoL and was developed in parallel in the UK and the Netherlands (NL). Content was derived from interviews with patients in each country. Face and content validity were assessed through patient field test interviews (UK and NL). A postal survey in the UK produced a more efficient version of the ASQoL, which was tested for scaling properties, reliability, internal consistency, and validity in a further postal survey in each country.

**Results:** A 41 item questionnaire was derived from interview transcripts. Field testing interviews confirmed acceptability. Rasch analysis of data from the first survey (n=121) produced a 26 item questionnaire. Rasch analysis of data from the second survey (UK: n=164; NL: n=154) showed some item misfit, but showed that items formed a hierarchical order and were stable over time. Problematic items were removed giving an 18 item scale. Both language versions had excellent internal consistency ( $\alpha=0.89-0.91$ ), test-retest reliability ( $r_s=0.92$  UK and  $r_s=0.91$  NL), and validity.

**Conclusions:** The ASQoL provides a valuable tool for assessing the impact of interventions for AS and for evaluating models of service delivery. It is well accepted by patients, taking about four minutes to complete, and has excellent scaling and psychometric properties.

Ankylosing spondylitis (AS) is a chronic inflammatory rheumatic condition affecting the sacroiliac joints, the spinal column to a varying degree, and to a smaller extent the peripheral joints. Patients have pain, morning stiffness, and disability which increases with duration of disease. A number of patients also experience extra-spinal and extra-articular manifestations as acute anterior uveitis and inflammatory bowel disease.<sup>1</sup> Population studies report a prevalence of AS of between 0.5% and 1.6% and it is more commonly found in men than women.<sup>2,3</sup> The pattern and rate of disease progression are variable but may be independent of disease duration.<sup>4</sup> Although major advances in the understanding of the disease pathogenesis have occurred in recent years, the optimal strategy for treatment is still unknown. Disease onset is generally in late adolescence or early adulthood and, consequently, the effects are present for most of the patient's life. Progression may continue through what should be economically active years.<sup>5</sup> Chamberlain reported that two thirds of male patients have difficulty at work, one third have social problems, and up to two thirds report having difficulty with sexual activity.<sup>6</sup> Reactive depression and frustration are noted, together with impaired self esteem and social skills.<sup>6</sup> Energy related problems are also widely reported.<sup>6</sup> All these features denote significant effects of the disease on lifestyle.

There is a growing interest in the assessment of quality of life (QoL), particularly in chronic disabling conditions. It is becoming relatively common to measure QoL in studies designed to assess the impact of new pharmaceutical products or to compare different treatment regimens. Although the concept has existed for many years, it is only within the past few decades that attempts have been made to operationalise QoL into a construct that can be measured in a meaningful way.<sup>7</sup>

Instruments currently available for use with patients with AS focus predominantly on symptoms (impairment) or functioning (disability), or both, and are used to assess the

presence or absence of disease and its consequences in these terms. Such instruments include the Bath Ankylosing Spondylitis Functional Index (BASFI)<sup>8</sup>; the Leeds Disability Questionnaire (LDQ)<sup>9</sup>; the Ankylosing Spondylitis Assessment Questionnaire (ASAQ)<sup>10</sup>; the Dougados Functional Index (DFI)<sup>11</sup>; a version of the Stanford Health Assessment Questionnaire modified for the spondyloarthropathies (HAQ-S)<sup>12</sup>; and, a modified version of the Arthritis Impact Measurement Scales 2 specific to AS (AS-AIMS2).<sup>13</sup> Although such measures provide important information about the degree of impairment and disability experienced by patients, they do not inform on the impact of the condition on QoL. The construct of QoL differs from impairment and disability insofar as it concerns the impact of disease from the patient's (rather than a clinical) perspective. By investigating how patient's lives are affected by impairment, disability, and other influences it provides an outcome that is complementary to the traditionally assessed impacts of disease.<sup>14,15</sup> Generic health status instruments such as the Nottingham Health Profile, Short Form-36 (SF-36), and EuroQoL also concentrate on impairment and disability rather than QoL. Furthermore, they have been shown to lack the responsiveness necessary to detect real changes in health status associated with effective treatment.<sup>8,16</sup>

**Abbreviations:** AS, ankylosing spondylitis; BASFI, Bath Ankylosing Spondylitis Functional Index; DFI, Dougados Functional Index; DIF, differential item functioning; INFIT, information-weighted fit statistic; IQR, interquartile range; LDQ, Leeds Disability Questionnaire; MNSQ, mean square; NHP, Nottingham Health Profile; NL, Netherlands; OUTFIT, outlier-sensitive fit statistic; QoL, quality of life

There is a clear need for a valid and reliable disease-specific instrument for assessing the impact of AS on QoL that is suitable for use in clinical practice. This paper describes the development of such a measure, the Ankylosing Spondylitis Quality of Life Questionnaire (ASQoL). The instrument was required to be suitable for monitoring patients, evaluating alternative treatment regimens, new pharmaceutical products and/or models of service delivery from the patient's perception. The development methodology employed is based on recent advances in the recognition and understanding of the conceptual and practical basis of measurement. The process combines the theoretical strengths of the needs-based QoL model<sup>17</sup> with the statistical and diagnostic power of the Rasch model.<sup>18</sup> The needs-based model of QoL postulates that life gains its quality from the ability of the individual to satisfy his or her needs. QoL is high when these needs are fulfilled and low when few needs are satisfied. The model is well established and has been applied successfully in the development of a large number of disease-specific QoL instruments, several of which have become established as the preferred outcome instrument for clinical trials and studies.<sup>17, 19–26</sup> Application of the Rasch model ensures that the fundamental scaling properties of the instrument (for example, unidimensionality and level of measurement) are assessed in addition to the traditional psychometric assessments of reliability and construct validity. Such basic measurement properties were considered at each stage of the development of the ASQoL.

## PATIENTS AND METHODS

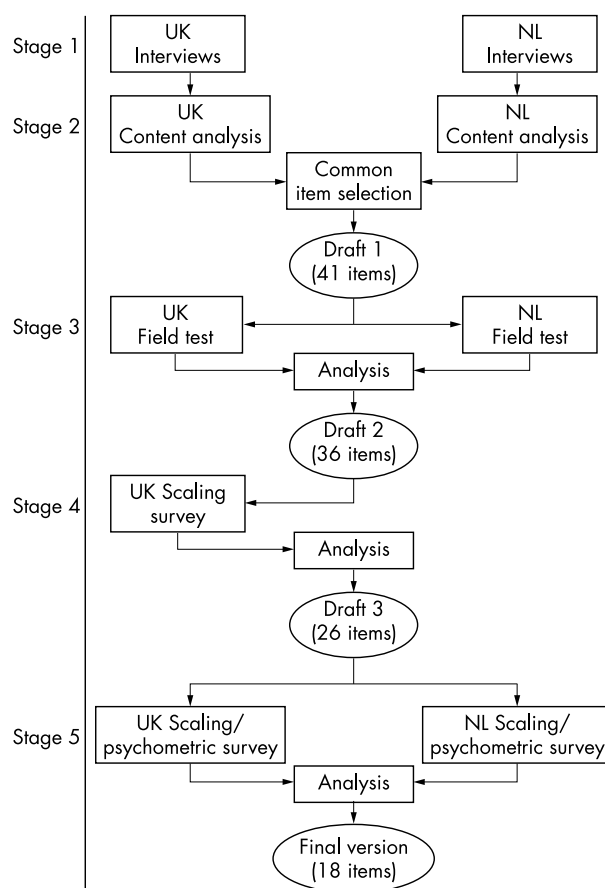
Figure 1 sets out the stages in the development of the ASQoL. The intention was to produce an instrument that would be equivalent for both the UK and the Netherlands (NL). Consequently, all stages were conducted simultaneously in both countries, with the exception of stage 4, which took place in the UK only. The purpose of stage 4 was to produce a more efficient instrument for final testing, by removing clearly problematic items.

### Patient samples

The study was approved by ethics committees in both countries and participants gave their written informed consent. All participating patients fulfilled the modified New York criteria for AS.<sup>27, 28</sup> Patients with significant comorbidity such as psychiatric disorders, cancer, or fibromyalgia were excluded. To ensure that a wide spectrum of clinical features was represented, each sample included patients with both axial and peripheral disease, a range of disease duration, and patients with uveitis or inflammatory bowel disease, or both. Patients were recruited from three hospitals in the north of England and from three in the south of the Netherlands. In both countries, different patients participated at each stage of the study.

### Stage 1: Interviews with patients

Deriving the content of a measure from subjects who are representative of the target population ensures that only relevant topics are included and that areas important to QoL are not omitted. For the ASQoL, the content of the questionnaire was derived from unstructured, qualitative interviews with relevant patients in both countries, conducted by experienced qualitative researchers. The interviews, took the form of informal, focused conversations. They were designed to explore the impact of AS on the patient, with emphasis on the person's ability to fulfil his or her needs. For example, where interviewees indicated functional limitations associated with AS they were prompted to consider how such restrictions impacted on their lives—particularly, how they prevented the fulfilment of their needs. The interviews were audio recorded with permission of the interviewee. Transcripts were produced from the tapes, which were then wiped clean. All traces of the interviewee's identity were omitted from the transcripts to maintain anonymity.



**Figure 1** Stages in production of the ASQoL

### Stage 2: Selection of items and response format for the draft questionnaire

In both countries, the interview transcripts were subjected to independent content analysis to identify statements relating to need satisfaction. As far as possible, the actual words used by interviewees were selected for the questionnaire. Duplicate and idiosyncratic items were removed and the list was subjected to further scrutiny, with items retained if they were applicable to all potential respondents, reflected a single idea, were unambiguous, and were short and simple. The item lists from each country were then compared at a meeting between the English and Dutch researchers. The purpose of this meeting was to decide on the content for the first draft of the questionnaire and to identify a response system that would be suitable for both languages.

A yes/no response system was selected for the draft measure as previous experience had indicated that this maximises language equivalence and ease of scoring and minimises respondent burden. In the development of the rheumatoid arthritis-specific instrument (the RAQoL) it was shown that a yes/no response format was more sensitive to change than a four-response Likert-type format.<sup>22</sup>

### Stage 3: Field testing for face and content validity

The purpose of this exercise was to test the applicability, comprehensibility, relevance, and comprehensiveness of the ASQoL with patients with AS. Participants completed the questionnaire in the presence of an interviewer. They were then asked to comment on its ease of completion and on the appropriateness of the instructions, items, and response format. Items found to be problematic in either country were removed. Items were considered problematic if respondents found them ambiguous or difficult to understand. Results

from this stage were used to compile a second draft version of the measure.

#### Stage 4: Postal survey 1 (UK)

The new draft ASQoL was sent by post to patients in the UK. Analyses were performed on the resulting data in order to identify items that failed to fit onto the underlying measurement construct and/or that worked differently by age (above or below the median), gender, AS diagnosis (axial only or axial with peripheral involvement), or disease duration (above or below the median). Such differential item functioning (DIF) would indicate that an item is valued differently by subgroups of patients. For example, in a disability measure it might be suggested that an item such as “I am unable to travel to my workplace” would be affirmed less often by respondents who had reached retirement age. Therefore, regardless of their level of disability, this item would appear to be less severe for younger respondents. DIF was identified through the application of the one parameter logistic item response theory model—the Rasch model.<sup>18</sup> In the context of a QoL scale, the Rasch model applies the premise that the likelihood of a person affirming a particular item depends on the level of QoL of the person and on the level of QoL represented by that item. The analysis provides estimates for the item and person parameters in log-odds units (logits). Such estimates are based on the assumption that the scale is indeed measuring a single underlying construct—that is, that the items form a unidimensional scale. The extent to which this assumption is justified is indicated by item fit statistics. For the present analysis, Rasch mean square (MNSQ) item fit statistics were identified through application of the computer program WINSTEPS.<sup>29</sup> Two MNSQ statistics are given; an information-weighted fit statistic (INFIT) and an outlier-sensitive fit statistic (OUTFIT). OUTFIT is more sensitive to inconsistencies in the extreme responses, that is those made to items far removed from the individual person’s level of QoL. The INFIT statistic is weighted so that these outliers have less impact and is, thus, more sensitive to non-extreme responses. Taken together, these two MNSQ item fit statistics provide information on the extent to which the individual items map onto the underlying measurement construct, in this case, the QoL. Given the present sample sizes, MNSQ values between 0.7 and 1.3 were taken to reflect adequate fit to the model.<sup>30</sup> As no Dutch data were included in stage 4, only those items that were clearly problematic were removed. The third draft of the questionnaire was produced on the basis of these analyses and used in the subsequent postal survey in both countries.

#### Stage 5: Postal survey 2 (UK and NL)

The purpose of the final postal survey was to assess the scaling properties, reliability, internal consistency, and construct validity of the ASQoL in each country. Patients in both countries were sent a package consisting of the ASQoL, a demographic questionnaire, additional comparator measures, and a reply paid envelope. Patients who completed and returned the first pack were sent a similar package timed to arrive two weeks later. The demographic questionnaire, which was consistent across countries, included questions on patient perceived disease activity and severity of illness. The Nottingham Health Profile (NHP)<sup>31</sup> and the BASFI were used as comparator measures in both countries. In addition, the LDQ was used in the UK and the DFI<sup>32</sup> was selected in the Netherlands. The NHP is a measure of perceived distress and provides a profile of scores in six sections: physical mobility, energy level, pain, emotional reactions, social isolation, and sleep. It is scored out of a maximum of 100 for each of the sections, with a higher score indicating greater distress. The BASFI, the LDQ, and the DFI each yield a single score. Scores on the BASFI can range from 0 to 100, on the LDQ, from 0 to 48, and on the DFI from 0 to 40. For each of these scales, a high score indicates

greater disability. Each item on the ASQoL is given a score of “1” or “0”. A score of “1” is given where the item is affirmed, indicating adverse QoL. All item scores are summed to give a total score or index, with a high score indicating a worse QoL. Questionnaires with missing data were omitted from the analysis. The following properties of the two versions of the ASQoL were assessed: scaling properties, reliability, internal consistency and construct validity.

#### Scaling properties

Rasch analyses were conducted to confirm that items mapped onto the same underlying construct (unidimensionality), that they represented different amounts of the construct (hierarchical ordering), and that they worked in same way across different patient groups (DIF). The level of measurement (that is, ordinal or interval level) provided by the measure was also examined.

#### Reliability

The reliability of the ASQoL was assessed by the test-retest method. This is an estimate of the instrument’s reproducibility over time, assuming that no change in condition has taken place. For each country, ASQoL scores from each administration were correlated. Patients were excluded from these analyses if they reported significant changes to their perceived general health, severity of illness, or perceived disease activity (that is, whether or not the patients considered their disease to be active at the time of completing the questionnaire) between administrations. Where an instrument is required for use in a clinical trial or for monitoring individual patients, a correlation coefficient of at least 0.85 is required.<sup>33</sup> Owing to the ordinal nature of the data, Spearman rank correlation coefficients were produced (intraclass correlation coefficients are also reported for information only).

#### Internal consistency

Internal consistency was assessed by Cronbach’s  $\alpha$  coefficients. This statistic indicates the degree of relatedness between items. A value of 0.70 or above was taken as reflecting adequate internal consistency.<sup>34</sup>

#### Construct validity

ASQoL scores were related to the comparator instruments and to patient perceived general health and severity of illness and patient perceived disease activity (that is, whether or not the patients considered their disease to be active at the time of completing the questionnaire). Patients describe disease activity in terms of whether they are having a “good day” or a “bad day”. This terminology is used throughout the “Results” section. It was predicted that there would be a moderate association between the ASQoL and the comparator measures indicating that they assess different but related constructs. It was also suggested that QoL would be worse for respondents experiencing a bad day (active disease), those reporting poorer general health, or those describing their AS as severe.

## RESULTS

### Findings from the interviews (stage 1)

Thirty patients were interviewed in the UK and 25 in the Netherlands. Patient samples were comparable in each country. About two thirds of those interviewed were male and a third reported having peripheral arthritis. The age of those interviewed ranged from 18 to 78 years, with disease duration ranging from 1.5 to 44 years. Interviews lasted for between 30 minutes and two hours with a median length of one hour and 10 minutes. All respondents chose to be interviewed in their own homes and all gave consent for the interview to be audio recorded.

Similar findings emerged from the Dutch and UK interviews. Respondents commented on the impact of pain

and its effect on sleep, mood, motivation, and ability to cope with the day ahead. One of the greatest fears expressed was that of losing independence. Many reported that they required some degree of assistance with everyday tasks such as dressing, washing, and shopping (particularly for foodstuffs). In addition, many reported feeling that they were no longer in control of their own personal hygiene or grooming. A particular concern was about the future, especially in relation to uncertainties surrounding disease progression.

The AS had a major impact on interviewees' ability to meet their needs for stimulation and exploration, gender role fulfilment, and feelings of worth. Major impacts were also reported on self image and self esteem, resulting from concerns about appearing slouched or slovenly.

AS had a profound impact on relationships with family members and friends, and social life was severely limited. For example, several interviewees commented that they chose places they could visit on the basis of how tolerable they found the seating. The condition was often cited as a major source of family tension and some interviewees reported taking out their frustration and anger on those closest to them.

### Development of the draft questionnaire (stage 2)

Items for the questionnaire consisted of actual quotations from the transcripts in a majority of cases. However, it was necessary to change the actual words used by interviewees for some of the items. For example, some were shortened, had the word order altered, or were changed so that they were expressed in the first person and/or in the present tense. The item pool from each country was compared and items selected for the draft questionnaire that covered issues raised in both countries. Forty one items were selected that best expressed the issues raised by the interviewees.

### Field testing for face and content validity (stage 3)

In the UK 10 patients were interviewed in clinic and 5 in their home. In the Netherlands all 15 patients were interviewed in

clinic. The ASQoL took between two and 16 minutes to complete (median four minutes in both the UK and NL). The measure was well accepted by interviewees in both countries, who generally found the items to be easily understood and relevant. Field testing of the questionnaire resulted in minor changes to the wording of two items and the removal of five more from both language versions. Items were removed because they were found to be problematic or were considered inappropriate by a number of respondents. For example, the item "I find it difficult to get moving in the morning" was among those deleted, as it was interpreted in different ways by UK respondents. The item "I often have to rest when doing jobs around the house" was removed because of gender bias. Although the item was intended to cover a range of household tasks, such as cooking, cleaning, decorating, or home maintenance, it was generally construed by patients in the UK to be solely related to housework. Many male respondents in the UK commented that they never undertook such tasks and, consequently, could not answer the question. After these changes, a 36 item version of the ASQoL was produced for use in the first postal survey.

### Testing the psychometric and scaling properties of the ASQoL

For both versions of the measure, a high score indicates worse QoL. For all tables in the following sections, n values deviating from the overall number are owing to individual missing responses.

### Results of the first postal survey (UK) (stage 4)

Questionnaire packs were distributed to 180 people and returned by 121, a response rate of 67%. Table 1 shows the demographic details of the sample. Rasch analyses were performed on the data to identify items that were problematic because of misfit or DIF. Although a number of items were found to misfit, DIF was minimal. As a result of these analyses, 10 items were removed from the measure, leaving a 26 item version of the ASQoL. This version was taken forward for further testing in each country.

**Table 1** Demographic and disease information (postal surveys). Results are shown as No (%)

	First postal survey (stage 4) UK (n=121)	Second postal survey (stage 5)	
		UK (n=210)	The Netherlands (n=154)
<b>Demographic details</b>			
Men (%)	92 (76)	150 (72)	110 (71)
Women (%)	29 (24)	59 (28)	44 (29)
Age range (years)	21–77	19–82	20–79
Mean age (SD) (years)	47.6 (12.4)	46.1 (12.4)	47.6 (11.8)
Married or living as married (%)	85 (72)	144 (69)	126 (82)
<b>Disease information</b>			
Range of duration of illness (years)	1.5–50	1–62	3–51
Median (mean) duration of illness in years	15 (16.3)	18 (19.6)	19 (20.8)
No reporting peripheral involvement (%)	85 (70)	174 (83)	112 (71)
No reporting uveitis (%)	30 (25)	48 (23)	36 (23)
No reporting IBD (%)	14 (12)	31 (15)	22 (14)
<b>Perceived AS severity</b>			
Mild (%)	18 (15)	22 (11)	36 (25)
Moderate (%)	50 (42)	76 (37)	69 (47)
Quite severe (%)	45 (38)	93 (45)	35 (24)
Very severe (%)	5 (4)	16 (8)	6 (4)
<b>Perceived general health status</b>			
Excellent/very good (%)	12 (10)	10 (5)	7 (5)
Good (%)	48 (39)	58 (28)	48 (32)
Fair (%)	43 (36)	100 (49)	83 (55)
Poor (%)	18 (15)	38 (18)	14 (9)
<b>Perception of today</b>			
Very good (%)	11 (9)	11 (5)	13 (8)
Good (%)	71 (61)	101 (50)	95 (62)
Bad (%)	34 (29)	83 (41)	42 (27)
Very bad (%)	1 (1)	8 (4)	4 (3)

**Table 2** Rasch item statistics for the 18 item ASQoL in the UK and Netherlands

Item number	Mean square fit statistic (MNSQ)			
	UK		Netherlands	
	INFIT	OUTFIT	INFIT	OUTFIT
1	0.90	0.74	0.91	0.78
2	1.21	1.22	0.93	0.84
3	0.95	0.87	0.97	0.88
4	0.79	0.77	0.84	0.63
5	1.02	0.70	1.25	1.23
6	0.88	0.75	0.87	0.71
7	<b>1.36</b>	<b>1.87</b>	1.00	1.06
8	0.83	0.65	0.93	0.68
9	0.99	0.96	0.99	0.91
10	1.13	0.84	1.08	1.25
11	0.86	0.71	1.20	1.15
12	1.20	<b>1.40</b>	1.06	1.12
13	0.99	0.90	0.89	0.76
14	0.88	0.92	1.37	1.16
15	0.98	0.90	0.90	0.74
16	1.16	1.06	0.98	1.08
17	1.11	1.02	0.76	0.76
18	0.87	0.73	1.19	1.16

**Bold italics**, MNSQ values above 1.3; *italics*, MNSQ values below 0.7.

**Table 3** Correlations\* between scores on the 18 item ASQoL and those on the comparator measures

Comparator measure	UK Time 1	Netherlands Time 1
NHP sections		
Physical mobility	0.78	0.79
Energy	0.74	0.73
Pain	0.81	0.79
Emotional reactions	0.72	0.73
Sleep	0.54	0.59
Social isolation	0.53	0.50
BASFI	0.72	0.75
LDQ	0.70	–
DFI	–	0.80

\*Spearman rank correlation coefficients.

### Results of the second postal survey (stage 5)

In the UK, 288 questionnaires were distributed at time 1 and 210 were returned, a response rate of 73%. Of these, 157 (75%) were returned at time 2. In the NL, 180 questionnaires were distributed at time 1 and 158 were returned, giving a response rate of 88%. Of these, 139 (88%) were returned at time 2. Four

questionnaire sets from the Dutch sample were returned too late to be included in the analyses. Table 1 shows demographic details of the samples at time 1 in the UK and the NL. It can be seen from the table that the samples included in the postal surveys were similar demographically. Demographic characteristics of respondents at time 2 were also comparable. Table 1 also provides information on the respondents' perceived health status. The table shows that the UK respondents rated their health status worse than the Dutch participants. Respondents' scores on the comparator instruments showed that, with the exception of social isolation, perceived distress (as shown by NHP section scores) is high for this patient sample and higher in the UK than in the NL (extra web table W1).

Rasch analyses were conducted on the data from each country. Eight items were removed as they were shown to misfit in one or both countries. The fit of the final 18 item ASQoL was good in both countries, with most MNSQ values within the required 0.7–1.3 range (table 2). Item stability over time was excellent in both countries, with Rasch item parameter estimates similar at times 1 and 2 (within 95% confidence intervals). Items were not equally spaced along the measurement continuum, indicating that the 18 item ASQoL produces raw scores at the ordinal level of measurement.

Scores on the 18 item ASQoL can range from 0 to 18. Median scores for the UK were 10.0 (interquartile range (IQR) 5.0–14.0; mean 9.5, standard deviation (SD) 5.3) at time 1 and 9.0 (IQR 4.0–14.0; mean 8.8, SD 5.7) at time 2. For the NL, median scores were 6.0 (IQR 2.0–10.0; mean 6.7, SD 4.8) at time 1 and 6.0 (IQR 1.5–9.0; mean 6.2, SD 4.8) at time 2. Relatively few respondents scored at the extremes, although the basement effect was greater in the NL.

### Association with additional factors

ASQoL scores were not related to duration of illness or to the presence of uveitis. Patients with inflammatory bowel disease scored higher on the measure (indicating worse QoL) than those without (UK  $p < 0.01$ , NL  $p < 0.005$ ; Mann-Whitney U test).

### Reliability and internal consistency of the ASQoL

The Spearman rank correlation coefficient for the test-retest reliability of the 18 item ASQoL was 0.92 in the UK ( $n = 129$ ) and 0.91 ( $n = 119$ ) in the NL, indicating that the measure has excellent reliability, producing low levels of random measurement error. Identical intraclass correlation coefficients were obtained (0.92 in the UK and 0.91 in the NL). Very few patients (two in the UK and one in the NL) reported any significant change in perceived general health, severity of illness, or perceived disease activity. Therefore, removing such patients made little difference to the results obtained. The

**Table 4** ASQoL scores by specified groups

Grouping factor	UK			Netherlands		
	n	Median	IQR	n	Median	IQR
Disease activity (good day/bad day)			$p < 0.001^*$			$p < 0.001^*$
Very good/good	102	7.0	3.0–11.3	101	5.0	1.5–9.0
Very bad/bad	81	13.0	10.0–15.5	40	10.0	8.0–12.7
Perceived general health			$p < 0.001^{**}$			$p < 0.001^{**}$
Excellent/good	61	3.0	1.0–7.0	52	2.0	0–4.0
Fair	91	11.0	8.0–13.0	76	9.0	6.0–10.7
Poor	34	15.0	14.0–17.0	12	13.5	12.0–14.7
Perceived AS severity			$p < 0.001^{**}$			$p < 0.001^{**}$
Mild	19	3.0	1.0–5.0	34	1.5	0–4.2
Moderate	73	7.0	3.0–10.0	61	6.0	4.0–10.0
Quite/very severe	95	13.0	11.0–16.0	39	9.0	8.0–12.0

\*Mann-Whitney U test; \*\*Kruskal-Wallis one way analysis of variance.

ASQoL also has good internal consistency in both countries (0.91 at time 1 and 0.92 at time 2 in the UK and 0.89 at time 1 and 0.90 at time 2 in the NL).

### Validity of the ASQoL

Evidence of construct validity was provided by examining the levels of association between the ASQoL and the comparator instruments. Moderate to high correlations were found between the ASQoL and all the comparator instruments (table 3). The pattern of association between the NHP section scores and the ASQoL was as expected, with the highest correlations being with the physical mobility, pain, and energy level sections. The correlations with the emotional reactions section were also high. Further evidence of the validity of the ASQoL was gained by investigating the measure's ability to distinguish between specified groups of patients (known groups validity). Table 4 shows that ASQoL scores differed significantly by whether the respondent was having a good or a bad day (disease activity), self perceived general health status, and self perceived AS severity.

### DISCUSSION

The efficient and cost effective management of any disease requires competing treatment regimens to be evaluated for their ability both to control the disease and improve the QoL of patients. Existing instruments for use with subjects with AS focus on symptoms and functioning. Although these provide important information they do not provide information about the overall impact of the condition and its treatment on the patient's QoL. The ASQoL is based on a clear, conceptual model of the QoL that has been successfully employed in the development of several other disease-specific QoL instruments.<sup>17–19–26</sup> The development process was conducted in parallel in the UK and the NL. Consequently, it was possible to remove items that were problematic in one or other language version of the instrument at each stage of the testing procedure. This method of development is preferable to the standard one, in which an instrument is produced in one country and then adapted for use in other languages. Such sequential development cannot overcome cultural and linguistic differences between countries.

The content of the measure was derived from interviews with subjects diagnosed with AS in the UK and the NL. For each language version, the items are expressed (as far as possible) in the original words of the patients. Consequently, respondents find the instrument acceptable, comprehensive, and relevant to their condition. The ASQoL is quick and easy to complete (taking less than five minutes), making it suitable for use in clinical settings.

Application of item response theory in the form of the one parameter Rasch model showed that the ASQoL was unidimensional, had good item stability over time, and had minimal DIF. The reliability of each language version of the measure has been shown to be excellent—the test-retest reliability coefficients obtained indicate that the ASQoL is suitable for use in routine clinical practice or for monitoring the progress of individual patients. Internal consistency was also adequate. It is essential to establish that a new instrument has construct validity—that is, that it is measuring the intended construct. Two prerequisites for this are that the instrument is based on a model of the construct assessed and that it has good reliability.<sup>35</sup> These requirements were met in both countries and hence, it is possible to infer that the ASQoL provides a valid assessment of the construct defined in the model. However, it is also necessary to determine construct validity formally through association with instruments measuring related constructs (convergent validity) and by comparing scores of patients at different stages of disease activity or with different disease severity (known groups validity). For the ASQoL, formal assessment was undertaken

by correlating scores on the ASQoL with those on the NHP and the BASFI. ASQoL scores in the UK were also correlated with the LDQ and in the NL with the DFI. These comparator instruments measure a range of constructs; the NHP assesses perceived distress, whereas the BASFI, LDQ, and DFI measure AS-specific disability. The relatively high levels of association between the ASQoL and these different constructs reflect the multifaceted nature of the impact of the disease on the patient. For example, pain, being a prominent feature of AS, would be expected to have a major influence on the QoL of the patient and, indeed, the correlation between these two measurements indicates approximately 66% shared variance. Similarly, QoL was moderately highly correlated with physical disability, energy, and emotional reactions sections of the NHP. The results obtained show that the ASQoL and comparator instruments measure different though related constructs. Taken together, they provide a more complete picture of the impact of AS than any single measure can give alone.

The psychometric and scaling properties of the ASQoL suggest that researchers and clinicians can have confidence in the scores obtained by respondents on the measure. Further assessments of the instrument's validity will be possible as it is used in clinical studies. In addition, it is recommended that future studies are carried out to assess responsiveness, the instruments ability to detect meaningful changes in QoL.

The decision to adopt a dichotomous response system for the instrument was driven by practical issues related to language equivalence and ease of completion and scoring. There is often an assumption that such simplification is at the cost of some loss of sensitivity because it is presumed that multiple response items are able to provide more detailed information about the variable of interest. However, this assumption is not necessarily correct.<sup>22</sup> The ASQoL comprises 18 dichotomous items that have been shown through Rasch analysis to form a single scale. Furthermore, the results from the assessment of known groups validity suggest that this scale can measure the QoL associated with a wide range of perceived disease severity and activity.

The ASQoL will serve as a valuable tool for assessing the impact of AS and its treatment on QoL in clinical settings and research studies. Such an instrument will allow accurate assessment of the effectiveness of interventions from the patient's perspective.

### ACKNOWLEDGEMENTS

We thank Vicky Wilkinson at the University of Leeds and Gisela Mulder at the University Hospital Maastricht for their assistance in administering the postal surveys.

This work in the UK was funded by the NHS Research and Development Programme.

.....

#### Authors' affiliations

**L C Doward, S A Cook, D Whalley, S P McKenna**, Galen Research, Enterprise House, Manchester Science Park, Lloyd Street North, Manchester M15 6SE, UK

**A Spoorenberg, D van der Heijde**, University Hospital, Maastricht, The Netherlands

**P S Helliwell, A Tennant, M A Chamberlain**, Rheumatology and Rehabilitation Research Unit, University of Leeds, Clarendon Road, Leeds, LS2 9NZ, UK

**L J Kay**, Department of Rheumatology, Freeman Hospital, Newcastle-upon-Tyne, Newcastle NE7 7DN, UK



Two appendices—The ASQoL questionnaire and a note on scoring the ASQoL—and an additional table can be found on the web page at [www.annrheumdis.com](http://www.annrheumdis.com)

### REFERENCES

- 1 Ryall NH, Helliwell PS. A critical review of ankylosing spondylitis. *Critical Reviews in Physical and Rehabilitation* 1998;10:265–301.

- 2 **Gran JT**, Husby G. Ankylosing spondylitis: prevalence and demography. In: Klippel JH, Dieppe PA, eds. *Rheumatology*. St Louis: Mosby, 1998;6:15.1-6.
- 3 **Khan, MA**. Ankylosing spondylitis: clinical features. In: Klippel JH, Dieppe PA, eds. *Rheumatology*. St Louis: Mosby, 1994;3:25.1-10.
- 4 **Chamberlain MA**. Socio-domestic and psychological factors in management. In: Moll JMH, ed. *Ankylosing spondylitis*. London:Churchill Livingstone, 1980:227-35.
- 5 **Lubrano E**, Helliwell PS. Deterioration in anthropometric measures over six years in patients with ankylosing spondylitis: an initial comparison with disease duration and reported exercise frequency. *Physiotherapy* 1999;85:138-43.
- 6 **Chamberlain MA**. Socio-economic effects of ankylosing spondylitis in females: a comparison of 25 female with 25 male subjects. *Int Rehabil Med* 1983;5:149-53.
- 7 **Doward LC**, McKenna SP. Evolution of quality of life assessment. In: Rajagopalan R, Sheretz EF, Anderson RT, eds. *Care management of skin diseases: life quality and economic impact*. New York: Marcel Dekker, 1997:9-33.
- 8 **Calin A**, Garrett S, Whitelock L, Kennedy G, 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.
- 9 **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.
- 10 **Nemeth R**, Smith F, Elswood, J, Calin A. Ankylosing spondylitis - an approach to measurement of severity and outcome: ankylosing spondylitis assessment questionnaire (ASAQ) - a controlled study [abstract]. *Br J Rheumatol* 1987;26(suppl 1):69-70.
- 11 **Dougados M**, Gueugen A, Nakache JP, Nguyen M, Mery C, Amor B. Evaluation of the functional index and an articular index in ankylosing spondylitis. *J Rheumatol* 1988;15:302-7.
- 12 **Daltroy LH**, Larson MG, Liang MH. A modification of the Health Assessment Questionnaire for the Spondyloarthropathies. *J Rheumatol* 1990;17:946-50.
- 13 **Guillemin F**, Challier B, Urlacher F, Vançon G, Pourel J. Quality of life in ankylosing spondylitis: validation of the ankylosing spondylitis arthritis impact measurement scales 2, a modified arthritis impact measurement scales questionnaire. *Arthritis Care Res* 1999;12:157-62.
- 14 **Heinemann AW**, Whiteneck GG. Relationships among impairment, disability, handicap and life satisfaction in persons with traumatic brain injury. *J. Head Trauma Rehabil* 1995;10:54-63.
- 15 **McKenna SP**, Whalley D, Doward LC. Which outcomes are important in schizophrenia. *International Journal of Methods in Psychiatric Research* 2000;9(suppl):S58-67.
- 16 **McKenna SP**. Quality of life assessment in the conduct of economic evaluations of medicines. *British Journal of Medical Economics* 1995;8:33-8.
- 17 **Hunt SM**, McKenna SP. The QLDS: a scale for the measurement of quality of life in depression. *Health Policy* 1992;22:307-19.
- 18 **Rasch G**. *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press, 1980.
- 19 **McKenna SP**, Hunt SM. A new measure of quality of life in depression: testing the reliability and construct validity of the QLDS. *Health Policy* 1992;22:321-30.
- 20 **Holmes SJ**, McKenna SP, Doward LC, Shalet SM. Development of a questionnaire to assess the quality of life of adults with growth hormone deficiency. *Endocrinol Metab* 1995;2:63-9.
- 21 **Wagner TH**, Patrick DL, McKenna SP, Froese PS. Cross-cultural development of a quality of life measure for men with erection difficulties. *Qual Life Res* 1996;5:443-9.
- 22 **Whalley D**, McKenna SP, de Jong Z, van der Heijde D. Quality of life in rheumatoid arthritis. *Br J Rheumatol* 1997;36:884-8.
- 23 **de Jong Z**, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. *Br J Rheumatol* 1997;36:878-83.
- 24 **Doward LC**, McKenna SP, Kohlmann T, Niero M, Patrick D, Spencer B, et al. The international development of the RGHQoL: a quality of life measure for recurrent genital herpes. *Qual Life Res* 1998;7:143-53.
- 25 **McKenna SP**, Doward LC, Mackenzie Davey K. The development and psychometric properties of the MSQOL: a migraine-specific quality-of-life instrument. *Clinical Drug Investigation* 1998;15:413-23.
- 26 **McKenna SP**, Doward LC, Alonso J, Kohlmann T, Niero M, Prieto L, et al. The QoL-AGHDA: an instrument for the assessment of quality of life in adults with growth hormone deficiency. *Qual Life Res* 1999;8:373-83.
- 27 **Bennett PH**, Burch TA. New York symposium on population studies in rheumatic diseases: new diagnostic criteria. *Bull Rheum Dis* 1967;17:453-8.
- 28 **van der Linden S**, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. *Arthritis Rheum* 1984;27:361-8.
- 29 **Linacre JM**, Wright BD. *A user's guide to WINSTEPS*. Chicago: Messa Press, 1997.
- 30 **Smith RM**, Schumacker RE, Bush MJ. Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement* 1998;2:66-78.
- 31 **Hunt SM**, McKenna SP, McEwen J, Williams J, Papp E. The Nottingham Health Profile: subjective health status and medical consultations. *Soc Sci Med* 1981;15:221-9.
- 32 **Creemers MC**, Van 't Hof MA, Franssen MJ, van de Putte LB, Gribnau FW, Van Riel PL. A Dutch version of the functional index for ankylosing spondylitis: development and validation in a long-term study. *Br J Rheumatol* 1994;33:842-6.
- 33 **Weiner E**, Stewart B. *Assessing individuals*. Boston: Little Brown, 1984.
- 34 **Nunnally JC Jr**. *Psychometric theory*. 2nd ed. New York: McGraw-Hill, 1978.
- 35 **Streiner D**, Norman G. *Health measurement scales*. Oxford: Oxford University Press, 1989.