ORIGINAL ARTICLE

# Validity and reliability of the Turkish version of the Health Assessment Questionnaire for the Spondyloarthropathies

**Esra Ozcan · Ozlem Yilmaz · Ahmet Tutoglu · Hatice Bodur**

**Abstract** The Health Assessment Questionnaire for Spondyloarthropathies (HAQ-S) was built by adding five new questions related with cervical and lumbar spine functions to HAQ by Daltroy et al. (in J Rheumatol 17(7):946–945, 1990). The aim of this study was to adapt the added five items into Turkish and then to test its reliability and validity. New questions were adapted to Turkish according to 'translation-back translation' method. Seventy-nine patients with ankylosing spondylitis were asked with the Turkish version of HAQ-S (HAQ-S TV). To assess, construct validity patients were evaluated by HAQ, Bath AS Functional Index (BASFI), Bath AS Disease Activity Index (BASDAI), Bath AS Metrology Index (BASMI), Maastricht AS Enthesitis Score (MASES), Ankylosing Spondylitis Quality of Life (ASQoL), and laboratory variables (erythrocyte sedimentation rate, C-reactive protein). Construct validity was investigated with Spearmann's rank correlation coefficient. Reliability of HAQ-S TV was assessed by internal consistency. Inter- and intra-observer reliability were tested with Cronbach's alpha score. HAQ-S TV met set criteria of reliability and validity. Reliability of adapted version was found good with high internal consistency value. (Cronbach's alpha = 0.89). In addition to this intra-observer and inter-observer reliability were found adequate for total score of HAQ-S (Cronbach's alphas: 0.999 and 0.998 in order) and also for each added five questions (Cronbach's alphas >0.7). Positive correlations were found between HAQ-S TV and HAQ, BASFI, BASDAI, BASMI, MASES, ASQoL, ESR, and CRP ($P < 0.05$). The results of this study showed that HAQ-S TV was reliable and valid in patients with AS.

**Keywords** Health Assessment Questionnaire for the Spondyloarthropathies · HAQ-S, Turkish · Validity · Reliability

## Introduction

Health Assessment Questionnaires (HAQ) are generic instruments which are used for evaluation of functional status of patients with arthritis. The Health Assessment Questionnaire for Spondyloarthropathies (HAQ-S) was built by adding five new questions related with cervical and lumbar spine functions to HAQ by Daltroy et al. [1] in 1990. HAQ was translated into Turkish by Küçükdeveci et al. [2] in 2004. The aim of this study was to adapt the added five items into Turkish and then to test its reliability and validity.

## Materials and methods

We obtained the permission of Daltroy LH who developed HAQ-S before starting the study. Since Daltroy LH passed away, we communicated with Liang MH to obtain permission

E. Ozcan
Department of Physical Medicine and Rehabilitation, Midyat Government Hospital, Mardin, Turkey

O. Yilmaz
Camlica M. PTT Loj. A Blok No: 23 06200, Yenimahalle, Ankara, Turkey

O. Yilmaz (✉) · H. Bodur
Department of Physical Medicine and Rehabilitation, Ankara Numune Education and Research Hospital, Ankara, Turkey
e-mail: dr.ozlemyilmaz@gmail.com

A. Tutoglu
Department of Physical Medicine and Rehabilitation, Urfa Government Hospital, Urfa, Turkey

for translation of the scale into Turkish and to perform tests on its validity and reliability.

Later on, two subcategories with five questions that were added to HAQ were translated into Turkish independently by a metallurgy engineer with a good knowledge of English, by a physiatrist, and by a teacher living in Turkey whose native language was English. Then, these three translators prepared a collective translation. The translators stated that they had difficulty in the translation of the first question; they had a trouble to get an agreement in the translation of "grocery bag" into Turkish, and at last they decided to translate it as "alışveriş poşetleri" (shopping bags).

The prepared text was translated into English by a professional translator and two English language teachers, all unaware of the original questions.
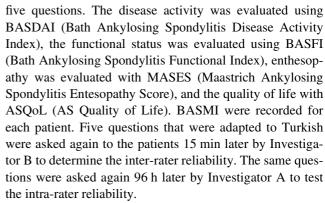
English translation was compared with the original text by a committee of physiatrists. The committee decided that the translation matched up with the original text. The questions were analyzed for their cultural content by the same committee. The word "sıra" (desk) in the third question (Düz bir masa ya da *sıra* başında çalışabiliyor musunuz?) was changed as "tezgah" (workbench) to make it more understandable. A pre-final version was established.

The pre-final version was first presented to 25 clinicians who were interested in ankylosing spondylitis (AS). They were asked to mention their opinions on the questions and the points that could cause problems in the daily practice. The clinicians mentioned that the questions were suitable for investigating the functional disabilities of the patients, and they suggested some corrections related to sentence structures. The scale was examined and corrected by a Turkish Language and Literature teacher who was having an academic career. The corrected form was presented to 25 patients who had the diagnosis of AS according to New York modified criteria. The patients were graduated from the primary school as a minimum. They were asked to indicate the points that they did not understand as well as the activities they were not familiar with. There were no negative feedbacks. Subsequently, we advanced to the testing stage.

In the testing stage, 80 patients who had the diagnosis of AS according to New York modified criteria were enrolled in the study. One patient had the diagnosis of small-cell lung carcinoma during the study period and excluded from the study.

First, the patients were informed about the aim of the study and their informed consents were obtained. The age, gender, educational level, and disease duration of the patients were recorded. As laboratory tests, C-reactive protein (CRP) was measured using the nephelometric method ($n$: 0–5.0 mg/l) and erythrocyte sedimentation rate (ESR) was measured using Westergren method.

The patients were first evaluated by Investigator A. They were asked to reply 20 questions in HAQ as well as added five questions. The disease activity was evaluated using BASDAI (Bath Ankylosing Spondylitis Disease Activity Index), the functional status was evaluated using BASFI (Bath Ankylosing Spondylitis Functional Index), enthesopathy was evaluated with MASES (Maastrich Ankylosing Spondylitis Entesopathy Score), and the quality of life with ASQoL (AS Quality of Life). BASMI were recorded for each patient. Five questions that were adapted to Turkish were asked again to the patients 15 min later by Investigator B to determine the inter-rater reliability. The same questions were asked again 96 h later by Investigator A to test the intra-rater reliability.

Construct validity was tested using "criterian validity" method. For this purpose, HAQ-S was compared with HAQ and BASFI that also measure physical limitation. At the same time, the correlation with the clinical parameters and scales related to functionality was investigated (severity of morning stiffness, severity of pain, BASDAI, BASMI, ASQoL, and MASES). The correlation of the scale with the ESR and CRP, parameters with questionable reliability in the assessment and follow-up of AS, was also investigated.

Statistical analysis

The statistical analysis of the data was performed using SPSS (Statistical Package for the Social Sciences) for Windows 11.5 package program. Shapiro–Wilk test was used to investigate normality of the distribution of the continuous variables. The descriptive statistics were given as the mean ± standard deviation or median (minimum–maximum) for the continuous variables and as number of patients and per cent (%) for the categorical variables. The inter- and intra-rater reliability levels were determined by Cronbach's alpha (internal consistency) coefficient. The internal consistency of HAQ-S was also investigated using Cronbach's alpha coefficient. A Cronbach's alpha coefficient >0.7 was regarded as satisfactory [3]. The construct validity was investigated using Spearman's correlation test. A correlation coefficient ($r$) greater than 0.6 was regarded as a strong correlation, $r = 0.3 – 0.6$ was regarded as a moderate correlation, and $r < 0.3$ was regarded as a weak correlation [4]. The significance of the difference between independent groups with regard to continuous variables was investigated with Mann–Whitney $U$ test. $P < 0.05$ was regarded as statistically significant.

**Results**

A total of 79 patients were enrolled in the study. Other demographic characteristics of the patients are presented in Table 1.

**Table 1** Demographic and clinical characteristics of the patients

|  | $n = 79$ |
|---|---|
| Age | $40.2 \pm 10.6$ (17–66) |
| Gender |  |
| Female | 16 (20.3%) |
| Male | 63 (79.7%) |
| Marital status |  |
| Married | 65 (82.3%) |
| Single | 14 (17.7%) |
| Education level |  |
| Primary school | 29 (36.7%) |
| Middle school | 18 (22.8%) |
| High school | 25 (31.6%) |
| Collage | 7 (8.9%) |
| Using motor-vehicle |  |
| Actively drivers in daily living | 50 (63.3%) |
| Duration of disease (year) | $10.6 \pm 7.5$ (1–32) |
| Severity of morning stiffness (VAS) | 13 (0–100) |
| Severity of pain (VAS) | 46 (0–100) |

The mean HAQ-S score ($0.77 \pm 0.63$) was higher than the mean HAQ score ($0.64 \pm 0.59$). This suggested that HAQ-S determined the difficult situations of the AS patients better then HAQ.

The mean scores of added two subcategories were $1.33 \pm 1.03$ and $1.24 \pm 1.20$, respectively. The highest score was obtained in the category related to lumbar spine functions (additional activities). The mean scores of the subcategories added to HAQ were higher than the other eight categories of HAQ (Table 2).

When the mean scores of the added five questions were considered, the highest score ($1.28 \pm 1.28$) was found to be related to the fifth question (Are you able to turn your head to drive in reverse?) The mean score of question 4 (Are you able to look in the rear-view mirror?) was $0.48 \pm 0.84$ that lowest among the scores of the five questions (Table 3).

**Table 2** Mean scores of HAQ and HAQ-S subunits

| Categories | Mean score |
|---|---|
| 1. Dressing and grooming | $0.97 \pm 0.98$ |
| 2. Eating | $0.19 \pm 0.42$ |
| 3. Arising | $0.89 \pm 0.90$ |
| 4. Walking | $0.46 \pm 0.73$ |
| 5. Hygiene | $0.51 \pm 0.73$ |
| 6. Grip | $0.15 \pm 0.48$ |
| 7. Reach | $1.11 \pm 1.05$ |
| 8. Activities | $0.92 \pm 0.91$ |
| 9. Added activities | $1.33 \pm 1.03$ |
| 10. Driving | $1.24 \pm 1.20$ |

**Table 3** Inter-rater and intra-rater reliability of HAQ-S

|  | Mean score | Inter-rater Cronbach's alpha | Intra-rater Cronbach's alpha |
|---|---|---|---|
| Total HAQ-S | $0.77 \pm 0.64$ | 0.998 | 0.999 |
| Question 1 | $0.86 \pm 0.83$ | 0.984 | 0.957 |
| Question 2 | $1.15 \pm 0.97$ | 0.703 | 0.943 |
| Question 3 | $0.95 \pm 0.99$ | 0.951 | 0.977 |
| Question 4 | $0.48 \pm 0.84$ | 0.932 | 0.934 |
| Question 5 | $1.28 \pm 1.28$ | 0.976 | 0.981 |

Intra-rater and inter-rater reliability was tested using Cronbach's alpha coefficient and found for the total HAQ-S score as 0.999 and 0.998, respectively. On the basis of the individual questions, the Cronbach's alpha coefficients ranged between 0.943 and 0.981 (>0.7) for intra-rater and 0.703–0.984 (>0.7) for inter-rater reliability. Those values showed that both intra- and inter-rater reliabilities of the scale were high (Table 3).

The Cronbach's alpha coefficient calculated to estimate the internal consistency of HAQ-S was 0.890 (>0.7). The Cronbach's alpha coefficients of the added two subcategories were 0.870 and 0.876, respectively (>0.7). The Cronbach's alpha coefficients of the other categories were also greater than 0.7 (Table 4).

When the internal consistency was tested using item statistics, the mean HAQ-S score was found to be significantly positively correlated with all of the mean subcategory scores ($P < 0.01$; $P < 0.001$). The correlation coefficients were above 0.2, which is the value used in practice. The strongest correlation was with the third category that tested the arising function ($r = 0.86$; $P < 0.001$); and the weakest correlation was with the sixth category testing the gripping function ($r = 0.42$; $P < 0.01$) (Table 5).

The investigation of the concordance among the added five questions revealed moderately or strongly positive

**Table 4** Internal consistency tested by Cronbach's alpha coefficient of HAQ-S

| Categories | Cronbach's alpha |
|---|---|
| Total HAQ-S | 0.890 |
| Dressing and grooming | 0.864 |
| Eating | 0.891 |
| Arising | 0.867 |
| Walking | 0.891 |
| Hygiene | 0.887 |
| Grip | 0.898 |
| Reach | 0.870 |
| Activities | 0.872 |
| Added activities | 0.870 |
| Driving | 0.876 |

**Table 5** Internal consistency tested by item statistics of HAQ-S

| | Dressing | Eating | Arising | Walking | Hygiene | Grip | Reach | Activities | Added activities | Driving | HAQ-S |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dressing | 1.00 | 0.43*** | 0.74*** | 0.43*** | 0.40*** | 0.23* | 0.76*** | 0.63*** | 0.61*** | 0.64*** | 0.83*** |
| Eating | | 1.00 | 0.42*** | 0.40*** | 0.38*** | 0.57*** | 0.31** | 0.45*** | 0.43*** | 0.41** | 0.55*** |
| Arising | | | 1.00 | 0.50*** | 0.48*** | 0.32** | 0.72*** | 0.74*** | 0.65*** | 0.58*** | 0.86*** |
| Walking | | | | 1.00 | 0.27* | 0.23* | 0.45*** | 0.41*** | 0.45*** | 0.37** | 0.59*** |
| Hygiene | | | | | 1.00 | 0.53*** | 0.40*** | 0.48*** | 0.40*** | 0.32* | 0.59*** |
| Grip | | | | | | 1.00 | 0.28** | 0.31** | 0.30** | 0.19 | 0.42** |
| Reach | | | | | | | 1.00 | 0.60*** | 0.57*** | 0.52*** | 0.79*** |
| Activities | | | | | | | | 1.00 | 0.72*** | 0.58*** | 0.82*** |
| Added activities | | | | | | | | | 1.00 | 0.64*** | 0.81*** |
| Driving | | | | | | | | | | 1.00 | 0.78*** |
| Total HAQ-S | | | | | | | | | | | 1.00 |

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

**Table 6** Internal consistency of added questions

| | Question 1 | Question 2 | Question 3 | Question 4 |
|---|---|---|---|---|
| Correlation coefficient ($r$) | | | | |
| Question 2 ($n = 79$) | 0.56*** | | | |
| Question 3 ($n = 79$) | 0.61*** | 0.77*** | | |
| Question 4 ($n = 50$) | 0.54*** | 0.55*** | 0.58*** | |
| Question 5 ($n = 50$) | 0.48*** | 0.60*** | 0.60*** | 0.71*** |

*** $P < 0.001$

**Table 7** Construct validity of HAQ-S

| | HAQ-S | | BASFI | | HAQ | |
|---|---|---|---|---|---|---|
| | $r$ | $P$ | $r$ | $P$ | $r$ | $P$ |
| HAQ-S | | | 0.88 | <0.001 | 0.97 | <0.001 |
| BASFI | 0.88 | <0.001 | | | 0.89 | <0.001 |
| BASMI | 0.42 | <0.001 | 0.46 | <0.001 | 0.37 | <0.001 |
| BASDAI | 0.69 | <0.001 | 0.72 | <0.001 | 0.73 | <0.001 |
| Pain | 0.71 | <0.001 | 0.67 | <0.001 | 0.73 | <0.001 |
| Morning stiffness | 0.44 | <0.001 | 0.44 | <0.001 | 0.42 | <0.001 |
| ASQoL | 0.80 | <0.001 | 0.76 | <0.001 | 0.80 | <0.001 |
| MASES | 0.43 | <0.001 | 0.41 | <0.001 | 0.42 | <0.001 |
| ESH | 0.29 | <0.05 | 0.23 | 0.043 | 0.25 | <0.05 |
| CRP | 0.26 | <0.05 | 0.22 | 0.052* | 0.22 | <0.05 |

* $P > 0.05$

correlations among the questions ($P < 0.001$). These values showed the internal consistency of the scale was satisfactory (Table 6).

The relation of HAQ-S with other clinical parameters (HAQ, BASFI, BASDAI, ASQoL, MASES, BASMI, morning stiffness, and severity of pain) was analyzed to test its construct validity. There were strongly positive correlations between HAQ-S and HAQ, BASFI, BASDAI, ASQoL, and severity of pain. There was a moderately positive correlation between HAQ-S and the severity of the morning stiffness, MASES, and BASMI. HAQ-S was found to be weakly positively correlated with the ESR and the CRP level (Table 7).

These values show that the construct validity level of HAQ-S tested with criterian validity method is sufficient.

## Discussion and conclusions

The scales used for functional evaluation may be based on the declaration or the performance of the patient. It is claimed that performance-based scales could not evaluate the functional capacity sufficiently. Declaration-based scales can be used easily, they are cheap, and a number of variables can be tested with only one scale [5]. AIMS

(Arthritis Impact Measurement Scales) and HAQ are declaration-based generic scales that are used for functional evaluation. However, these scales mainly evaluate the functional impairment related to the peripheral joint involvement. However, axial involvement is more pronounced in AS [6]. This is why Daltroy et al. [1] developed HAQ-S by adding two subcategories to HAQ related to lumbar and cervical spinal functions.

Adaptation of a scale from one language to another starts with its translation into the target language. There are a number of guidelines in the literature defining this translation stage. The most widely accepted one is AAOS (American Association of Orthopedic Surgeons), published in 2000 by Guillamin and Beaton and updated in the same year by the same authors [7, 8]. We based on the AAOS guideline during the adaptation stage of HAQ-S into Turkish.

Our translators stated that they had most difficulty in the translation of the question "Are you able to carry heavy packages such as grocery bags?" (question 1) into Turkish. The term "grocery bags" was translated as "alışveriş poşetleri" (shopping bags) by our translators. "Grocer" means "bakkal" in Turkish. However, there are general stores and supermarkets instead of groceries in most of the cities. In addition to that, there are many people who shop in the local markets. This is why we decided that the term "alışveriş poşetleri" was more convenient. A similar problem in the first question was reported in the Brazil–Portuguese adaptation of HAQ-S. Shinjo et al. [9] stated that the term "grocery bags" was not a term used in Brazil or Portugal, and they translated it into a term meaning "supermarket bags".

The term "a flat topped desk or table" was translated as "düz bir masa ya da sıra" by our translators. However, in the revision stage it was considered that the term "sırada çalışmak" was not widely used in our country and it may not be understood by the patients. Instead, more widely used "tezgah başında çalışmak" (work on a workbench) was preferred.

The total HAQ-S score was found 25% higher than the total HAQ score in the study of Daltroy et al. [1]. Similarly, Shinjo et al. [9] found the total HAQ-S score higher than the total HAQ score. Both Daltroy et al. [1] and Shinjo et al. [9] found that the mean scores of the added two subcategories were higher than the mean scores of the other categories. Similarly in our study, the mean scores of the added two subcategories were found to be higher than the mean scores of the other categories. As in other studies, our results suggest that functional impairment can be tested better by adding two new categories to HAQ evaluating lumbar and cervical spinal functions.

The intra-rater reliability was tested for total HAQ-S score and for the mean scores of each question, and the Cronbach's alphas were found much higher than the 0.7 limit [3]. These values indicate a high intra-rater reliability of our scale.

The Cronbach's alpha coefficients were obtained to test inter-rater reliability, and they were higher than 0.7 for each question and total HAQ-S scores. These results suggest that the inter-rater reliability of our scale is at a sufficient level.

The internal consistency of a scale means the concordance among its subunits. The most commonly used statistical method to determine the internal consistency is calculating the Cronbach's alpha coefficient [3]. In our study, the Cronbach's alpha coefficient for the internal consistency was found as 0.89. It was found as 0.9 in the German version of HAQ-S [10]. In our study, the Cronbach's alpha coefficients calculated for each category were found higher than 0.7. These values show the concordance of the subunits of the Turkish version of HAQ-S.

In addition to Cronbach's alpha coefficient, item statistics were used to analyze the internal consistency. The correlations between the mean total HAQ score and the mean scores of the subcategories were investigated, and all correlation coefficients were found to be higher than 0.2, a value used in practice. This result shows that the subunits of HAQ-S work in harmony and they are pieces of a whole that aim to test functionality.

"Criterian validity method" was used to test the construct validity. For this purpose, the correlation between HAQ-S and other valid and reliable scales used in AS was investigated.

The ASAS (Assessment of SpondyloArthritis International Society) suggests using BASFI and DFI (Dougados Functional Index) for functional evaluation in AS [10]. Although investigations point out that these two scales have equal capacities for showing functional impairment, BASFI is found superior for evaluating moderate impairment and it is more sensitive to the changes and is used more easily since it is short [11]. We used BASFI to test the construct validity, and a strong correlation was found between HAQ-S and BASFI. Similarly, these two scales were found to be strongly correlated in the Brazil–Portuguese version of HAQ-S [9]. However, the correlation coefficients in our study is higher ($r = 0.88$ and $r = 0.57$). In addition, a strong correlation was found between HAQ-S and HAQ. The same result was obtained in the Brazil–Portuguese version of HAQ-S [10]. Ward et al. [6] also reported a strong correlation between HAQ-S and HAQ. The strong correlation of Turkish version of HAQ-S with BASFI and HAQ, scales that evaluate functional impairment, shows its construct validity. The positive correlations of varying degrees between HAQ-S and BASDAI, ASQoL, BASMI, and MASES also support the construct validity of HAQ-S Turkish version.

It is known that acute phase reactants do not accurately reflect the clinical course of AS [6, 9, 12, 13]. In our study, we found that ESR and CRP levels were significantly but weakly correlated with HAQ-S. This weak correlation supports the divergent validity and signifies the convergent validity of HAQ-S Turkish version and indicates the strength of the construct validity.

The sensitivity of a disease-specific scale on the changes must be shown to make it a widely used scale. The most important limitation in our study is the lack of sensitivity analysis of HAQ-S Turkish version to changes. The sensitivity of HAQ-S to changes has not been reported in the literature up to date. Viitanen et al. [14] reported that HAQ-S scores showed improvement after a 3-week physiotherapy program; however, this change was not statistically significant.

In conclusion, Turkish version of HAQ-S is a valid and reliable scale that can be used by investigators. New studies are needed to investigate its sensitivity to changes.

# References

1. Daltroy LH, Larson MG, Roberts NW et al (1990) A modification of the Health Assessment Questionnaire for the Spondyloarthropathies. J Rheumatol 17(7):946–950
2. Küçükdeveci AA, Sahin H, Ataman S et al (2004) Issues in cross-cultural validity: example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. Arthritis Rheum 51(1):14–19
3. DeVon HA, Block ME, Moyle-Wright P (2007) A psychometric toolbox for testing validity and reliability. J Nurs Scholarsh 39(2):155–164
4. Hinkle DE, Wiersma W, Jurs SG (1998) Applied statistics for the behavioral sciences, 5th edn. Houghton Mifflin, Boston
5. Van Weely SF, van Denderen CJ, van der Horst-Bruinsma IE et al (2009) Reproducibility of performance measures of physical function based on the BASFI, in ankylosing spondylitis. Rheumatology (Oxford) 48(10):1254–1260
6. Ward MM, Kuzis S (1999) Validity and sensitivity to change of spondylitis-specific measures of functional disability. J Rheumatol 26(1):121–127
7. Guillemin F, Bombardier C, Beaton D (1993) Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 46(12):1417–1432
8. Acquadro C, Conway K, Hareendran A et al (2008) European Regulatory Issues and Quality of Life Assessment (ERIQA) Group. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. Value Health 11(3):509–521
9. Shinjo SK, Gonçalves R (2007) 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 26(8):1254–1258
10. Sieper J, Rudwaleit M, Baraliakos X et al (2009) The Assessment of Spondyloarthritis International Society (ASAS) handbook: a guide to assess spondyloarthritis. Ann Rheum Dis 68(sub II):1–44
11. Spoorenberg A, van der Heijde D (1999) A comparative study of the usefulness of the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index in the assessment of ankylosing spondylitis. J Rheumatol 26(4):961–965
12. Özgöçmen S, Godekmerdan A (2007) Acute-phase response, clinical measures and disease activity in ankylosing spondylitis. Joint Bone Spine 74:249–253
13. Zhao LK, Liao ZT, Li CH et al (2007) Evaluation of quality of life using ASQoL questionnaire in patients with ankylosing spondylitis in a Chinese population. Rheumatol Int 27(7):605–611
14. Viitanen JV, Heikkilä S (2001) Functional changes in patients with spondylarthropathy. A controlled trial of the effects of short-term rehabilitation and 3-year follow-up. Rheumatol Int 20(5):211–214