

EXTENDED REPORT

Measurement properties of the ASAS Health Index: results of a global study in patients with axial and peripheral spondyloarthritis

Uta Kiltz,¹ Désirée van der Heijde,² Annelies Boonen,³ Nurullah Akkoc,⁴ Wilson Bautista-Molano,⁵ Ruben Burgos-Vargas,⁶ James Cheng-Chung Wei,⁷ Praveena Chiowchanwisawakit,⁸ Maxime Dougados,⁹ M Tuncay Duruoz,¹⁰ Bassel Kamal Elzorkany,¹¹ Inna Gaydukova,¹² Lianne S Gensler,¹³ Michele Gilio,¹⁴ Simeon Grazio,¹⁵ Jieruo Gu,¹⁶ Robert D Inman,¹⁷ Tae-Jong Kim,¹⁸ Victoria Navarro-Compan,¹⁹ Helena Marzo-Ortega,²⁰ Salih Ozgocmen,²¹ Fernando Pimentel dos Santos,²² Michael Schirmer,²³ Simon Stebbings,²⁴ Filip E Van den Bosch,^{25,26} Astrid van Tubergen,³ Juergen Braun¹

Handling editor Tore K Kvien

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/annrheumdis-2017-212076>).

For numbered affiliations see end of article.

Correspondence to

PD Dr Uta Kiltz, Rheumazentrum Ruhrgebiet, 44649 Herne, Germany; uta.kiltz@elisabethgruppe.de

Received 14 July 2017
Revised 12 April 2018
Accepted 20 April 2018

ABSTRACT

Objectives To evaluate construct validity, interpretability, reliability and responsiveness as well as determination of cut-off points for good and poor health within the original English version and the 18 translations of the disease-specific Assessment of Spondyloarthritis international Society Health Index (ASAS HI) in 23 countries worldwide in patients with spondyloarthritis (SpA).

Methods A representative sample of patients with SpA fulfilling the ASAS classification criteria for axial (axSpA) or peripheral SpA was used. The construct validity of the ASAS HI was tested using Spearman correlation with several standard health outcomes for axSpA. Test-retest reliability was assessed by intraclass correlation coefficients (ICCs) in patients with stable disease (interval 4–7 days). In patients who required an escalation of therapy because of high disease activity, responsiveness was tested after 2–24 weeks using standardised response mean (SRM).

Results Among the 1548 patients, 64.9% were men, with a mean (SD) age 42.0 (13.4) years. Construct validity ranged from low (age: 0.10) to high (Bath Ankylosing Spondylitis Functioning Index: 0.71). Internal consistency was high (Cronbach's α of 0.93). The reliability among 578 patients was good (ICC=0.87 (95% CI 0.84 to 0.89)). Responsiveness among 246 patients was moderate-large (SRM=−0.44 for non-steroidal anti-inflammatory drugs, −0.69 for conventional synthetic disease-modifying antirheumatic drug and −0.85 for tumour necrosis factor inhibitor). The smallest detectable change was 3.0. Values ≤ 5.0 have balanced specificity to distinguish good health as opposed to moderate health, and values ≥ 12.0 are specific to represent poor health as opposed to moderate health.

Conclusions The ASAS HI proved to be valid, reliable and responsive. It can be used to evaluate the impact of SpA and its treatment on functioning and health. Furthermore, comparison of disease impact between populations is possible.

INTRODUCTION

Spondyloarthritis (SpA) is characterised by inflammation and new bone formation in the axial skeleton and joints.¹ Patients with SpA suffer from axial and peripheral symptoms resulting in pain, spinal stiffness, sleep problems and fatigue.^{2–4} Peripheral manifestations (arthritis, dactylitis or enthesitis) and extra-articular manifestations such as uveitis, psoriasis and inflammatory bowel disease may add to the burden of disease in a substantial number of patients but are less well studied.^{5,6} The course of disease varies, but many patients experience functional disability and limitation in activities and social participation. The influence of the disease on health-related quality of life and functional status has been well characterised in patients with ankylosing spondylitis (AS), and to a lesser extent for patients with non-radiographic axial SpA (nr-axSpA), but there are little data relating to patients with peripheral SpA (pSpA).^{4,6}

The Assessment of Spondyloarthritis international Society Health Index (ASAS HI) has been developed to measure functioning and health in patients with SpA with the aim of defining and comparing the impact of the disease and health in this patient group.⁷ Initial phases in the development of the ASAS HI focused on investigating functional impairments from the patients' perspective using both qualitative and quantitative approaches.

The biopsychosocial model of disease proposed by the International Classification of Functioning, Disability and Health (ICF) was used as the basis for the development of the ASAS HI. The ICF is accompanied by a classification of categories, called factors, that allow description of functioning, disability and health in individuals in a systematic and inclusive way.⁸ The comprehensive ICF Core Set for AS is a disease-specific selection of the ICF factors that are typical and relevant for AS, and has served as the underlying construct of the ASAS HI since the whole range of functioning, disability and health of patients with AS was captured.⁹ Patients, rheumatologists and methodologists were involved



To cite: Kiltz U, van der Heijde D, Boonen A, et al. *Ann Rheum Dis* Epub ahead of print: [please include Day Month Year]. doi:10.1136/annrheumdis-2017-212076

in the further reduction of categories using qualitative and quantitative methods and resulting finally in the ASAS HI.⁷ The 17 dichotomous items of the ASAS HI address aspects of pain, emotional functions, sleep, sexual functions, mobility, self-care and community life representing a wide spectrum of different levels of functioning, disability and health in patients with SpA. The sum score of the ASAS HI ranges from 0 to 17, with a lower score indicating a better health status. Preliminary validity and feasibility (time of completion) have already been assessed in a field test during the final steps of the development phase.⁷ Cognitive debriefing was undertaken in patients with AS and nr-axSpA and patients with peripheral manifestations aiming at assessment of a broad impact of health on all patients with SpA.^{7 10} The ASAS HI was originally developed in parallel in English-speaking countries (Australia, Canada, Ireland, UK, USA), and it has later been translated and cross-culturally adapted into 18 languages worldwide.¹⁰

The objective of the current paper is to evaluate construct validity, interpretability, reliability and responsiveness as well as determination of cut-off points for good and poor health within the original English version and the 18 translations of the disease-specific ASAS HI in 23 countries worldwide.

METHODS

Study design

A cross-sectional international observational study with a longitudinal component for reliability and responsiveness of the ASAS HI was performed in 23 countries during 2014 and 2015.

Patients

A representative sample of patients with SpA fulfilling the ASAS classification criteria for either axial (axSpA) or pSpA were recruited.^{11 12} Each centre was asked to recruit a sample of patients, 80% of whom were to have axSpA and 20% pSpA with no more than 10% of all recruits having coexistent psoriasis. Of the axSpA subset, 40% were to have nr-axSpA and 60% AS. There was a target of 50–100 recruits per country to reach an overall sample size of 1700 patients. The aim was to include patients with a broad range of disease severity and a variety of treatments. Patients with severe concomitant diseases that may influence their functional status were excluded from participation together with patients who were unable to understand the objectives of the study or the various questionnaires. Centres were asked to include at least 25% of their sample in the reliability arm and 25% in the responsiveness arm. All centres received approval from their local ethics committee. Written informed consent was obtained from all respondents prior to the start of their participation.

Data collection

Demographic and clinical information was collected including age, gender, predominant presentation, presence of extra-articular manifestations, years of education and employment. C reactive protein levels, imaging results and current medications were also recorded. Physician's judgement of patients' overall functioning and health was assessed by a single global question ("Please score the overall status of the subject's signs and symptoms and the functional capacity of the subject") on a zero to 10 numerical rating scale (NRS) (10 representing severe impairment) and a Likert scale ("How do you rate the health of your patient today?") on a 4-point scale ranging from very poor to very good. Physician's opinion on the level of disease activity

was recorded by answering the question "How active was the spondyloarthritis of your patient during the last week?".

Patients completed a series of self-reported questionnaires: ASAS HI,⁷ Bath Ankylosing Spondylitis Disease Index (BASDAI),¹³ Bath Ankylosing Spondylitis Functioning Index (BASFI),¹⁴ EuroQol five dimensions questionnaire (EQ-5D-5L index and thermometer),¹⁵ Short Form Survey Instrument 36-Item (SF-36),¹⁶ Hospital Anxiety and Depression Scale (HADS),¹⁷ work productivity and activity impairment questionnaire (WPAI)¹⁸ and pain and spinal pain NRS (0–10 NRS; 10 representing severe pain). Patient's opinion on the level of disease activity was recorded by a single patient global question ("How active was your rheumatic disease on average during the last week?") on a NRS 0–10 and on the health status ("How do you rate your health today?") on a 4-point Likert scale ranging from very poor to very good. Based on collected data, the Ankylosing Spondylitis Disease Activity Score (ASDAS) sum score was calculated and patients were categorised into ASDAS status groups.^{19 20} EQ-5D index was calculated using the value set for UK except for France, Germany, Netherlands, Spain, Thailand and USA for which country-specific value sets were used.

ASAS Health Index

The ASAS HI contains 17 items (dichotomous response option: 'I agree' and 'I do not agree') addressing different aspects of functioning. A sum score is being calculated by summing up all responses to 'I agree' given a total ASAS HI score ranging from 0 to 17—with a lower score indicating a better and a higher score indicating an inferior health status (see also user's manual for handling missing items; online supplementary file 1).⁷

Variables were collected at baseline and longitudinally in stable patients (reliability arm) or in patients who required a therapeutic change because of high disease activity (responsiveness arm) (flow chart and patients' disposition in online supplementary file 2). Longitudinal assessments were performed in patients who were in a stable disease state (reliability arm) or in patients who required a therapeutic change because of high disease activity (responsiveness arm). Patients in the reliability arm were eligible for the analyses when they considered themselves in a stable disease state while on stable treatment (no change in non-steroidal anti-inflammatory drugs (NSAIDs) over the preceding week, with no change in conventional synthetic disease-modifying antirheumatic drug (csDMARD) or tumour necrosis factor inhibitor (TNFi) therapy over the last 4 weeks). Patients were invited to complete the questionnaire at home after an interval of 4–7 days to evaluate reproducibility. Patients in the responsiveness arm required therapeutic change initiated due to high disease activity. The therapeutic change could include initiation of NSAIDs, a csDMARD or a TNFi. Patients were reassessed 12–24 weeks (for NSAIDs 2–24 weeks) after the treatment change had been implemented. The patients with longitudinal assessments (reliability and responsiveness) were asked to answer a global question at the second assessment and respond as to whether their condition was stable, improved or had worsened compared with baseline assessments. Only those patients reporting improvement in response to the global change question were analysed to assess responsiveness. Results of the validation process and the psychometric properties of the ASAS HI were presented at various ASAS meetings. Votes were taken from ASAS members to confirm the thresholds of ASAS HI.

Statistics

COSMIN recommendations were followed to test and report measurement properties.²¹ Psychometric properties were

Table 1 Patient characteristics, values of health status and composite indices at baseline and for reliability and responsiveness assessment

Patient characteristics	Baseline (n=1548)	Reliability, first visit (n=578)	Reliability, second visit (n=578)	Responsiveness first visit (n=246)*	Responsiveness second visit* (n=246)
Age (years)	42.0 (13.4)		45.3 (13.7)		37.2 (12.2)
Male, n (%)	1005 (64.9)		372 (64.4)		152 (61.8)
Symptom duration (years)	14.5 (11.4)		16.6 (11.9)		10.9 (9.5)
Extraspinal manifestation, current, n (%)					
Arthritis	301 (19.4)		99 (17.1)		43 (17.4)
Dactylitis	50 (3.2)		9 (1.6)		11 (4.5)
Enthesitis	261 (16.9)		80 (13.8)		44 (17.9)
Extra-articular manifestation, current, n (%)					
Uveitis	53 (3.4)		17 (2.9)		5 (2.0)
IBD	65 (4.2)		32 (5.5)		7 (2.8)
Skin psoriasis	110 (7.1)		51 (8.8)		3 (1.2)
HLA-B27 positive, n (%)	994 (77.0)		350 (73.9)		107 (71.9)
CRP (mg/L), sample n=1353	9.8 (16.01)	6.8 (10.2)	–	16.7 (22.6)	6.7 (11.4)
Elevated CRP (≥ 0.5 mg/L), n (%)	765 (49.4)	256 (44.3)	–	100 (64.5)	48 (31.0)
Current NSAID treatment	994 (64.2)		364 (63.0)		119 (76.8)
Current csDMARD treatment	402 (26.2)		166 (28.7)		35 (22.6)
Current TNFi treatment	591 (38.2)		240 (41.5)		23 (14.8)
ASAS HI (0–17)	6.7 (4.3)	6.2 (4.2)	6.0 (4.2)	8.2 (3.9)	5.7 (4.0)
ASDAS	2.5 (1.2)	2.6 (1.1)	–	3.3 (1.0)	1.9 (1.0)
BASDAI	4.1 (2.5)	3.7 (2.3)	3.4 (2.1)	5.4 (2.1)	3.0 (2.1)
BASFI	3.3 (2.8)	3.1 (2.7)	3.1 (2.7)	4.1 (2.7)	2.4 (2.3)
Pain, NRS 0–10	4.4 (2.9)	3.8 (2.6)	3.6 (2.4)	6.1 (2.4)	3.0 (2.1)
Physician global, NRS 0–10	3.7 (2.3)	3.1 (2.0)	–	5.6 (2.1)	2.4 (1.7)
Patient global, NRS 0–10	4.5 (2.8)	3.8 (2.5)	3.6 (2.3)	5.9 (2.4)	2.3 (1.7)
Well-being last week	4.4 (2.8)	3.8 (2.6)	3.5 (2.3)	5.8 (2.5)	2.4 (2.0)
PASS yes	801 (51.7)	368 (63.7)	401 (69.4)	45 (29.0)	130 (83.9)
HADS anxiety	17.6 (3.9)	18.0 (3.8)	18.3 (3.8)	16.6 (3.9)	18.5 (3.6)
HADS depression	15.3 (3.5)	15.9 (3.5)	15.7 (3.5)	14.4 (3.4)	16.0 (3.4)
EQ-5D VAS (0–100 mm)	61.6 (22.7)	63.9 (22.6)	65.3 (22.6)	57.5 (20.2)	67.3 (22.1)
EQ-5D (pooled)†	0.67 (0.2)	0.7 (0.2)	0.7 (0.2)	0.7 (0.2)	0.7 (0.2)
SF-36 PSC	38.9 (10.5)	40.4 (10.2)	40.6 (10.2)	34.4 (9.3)	41.3 (9.5)
SF-36 MSC	47.0 (11.5)	48.6 (11.2)	48.5 (11.2)	43.5 (10.9)	48.7 (11.0)
WPAI, presenteeism	29.2 (26.0)	23.4 (23.4)	21.9 (21.1)	39.3 (27.4)	23.6 (21.9)
WPAI, absenteeism‡	16.0 (32.3)	11.68±27.5	10.5 (27.3)	31.0±42.7	10.9±27.3

Values are presented as mean (SD) or absolute number (%). Percentages are % of available data. Fewer than 5% of the data were missing, except for HLA-B27 with 16.6%, CRP with 8.7% and EQ-5D with 9.7% at baseline visit and ASDAS (in responsiveness arm) with 7.3%. CRP and physician global were not measured in the reliability arm at the second visit.

*Only data analysed from those patients who stated that they improved during the time interval.

†The phrase *EQ pooled* means that EQ-5D-5L analysis was based on the five-level value set for UK except for France, Germany, Netherlands, Spain, Thailand and USA for which the country-specific value set was used.

‡Calculated for employed patients (n=961); see online supplementary file 4.

ASAS HI, ASAS Health Index; ASDAS, Ankylosing Spondylitis Disease Activity Score; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; CRP, C reactive protein; csDMARD, conventional synthetic disease-modifying antirheumatic drug; EQ-5D, Euro Quality of Life 5 Dimensions; HADS, Hospital Anxiety Depression Scale; HLA, human leucocyte antigen; IBD, inflammatory bowel disease; MSC, mental component summary score; NRS, numerical rating scale; NSAID, non-steroidal anti-inflammatory drug; PASS, Patient Acceptable Symptom State; PSC, physical component summary score; SF-36, Short Form 36; TNFi, tumour necrosis factor inhibitor; VAS, Visual Analogue Scale; WPAI, Work Productivity and Impairment Scale.

examined according to the OMERACT filter.²² Descriptive statistics were used to characterise the sample. According to the COMSIN checklist, interpretability is being summarised as information about percentage of missing items and description of how missing items were handled as well as distribution of the (total) ASAS HI score including floor and ceiling effects. Distributions of scores were examined for identification of floor and ceiling effects. Construct validity was evaluated against other health outcomes (including patient and physician global assessment, ASDAS, BASDAI, BASFI, HADS, WPAI, SF-36

summary values (physical component summary score (PCS) and mental component summary score (MCS), EQ-5D) in a cross-sectional analysis using Spearman correlation. Prior to the analysis, we hypothesised magnitude and direction of correlations, and correlation were considered low if ≤ 0.30 , moderate if > 0.30 and ≤ 0.50 , high if > 0.50 and < 0.80 , and very high if ≥ 0.80 .²³ Internal consistency was evaluated using Cronbach's α coefficient (adequate: ≥ 0.70). Test-retest reliability was assessed by intraclass correlation coefficient (ICC) (two-way model, single measure) with a 95% CI. An ICC of ≥ 0.8 was considered

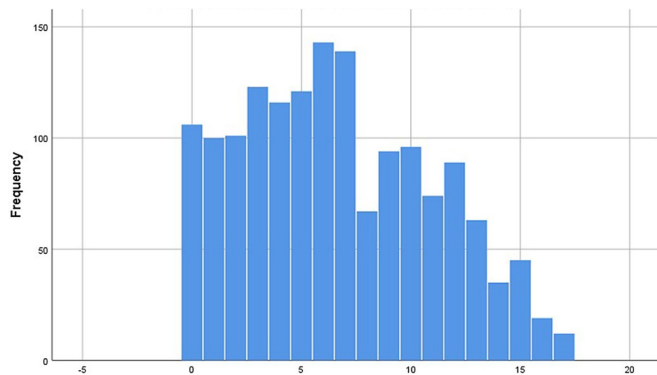


Figure 1 Score distribution (0–17) of the ASAS Health Index (ASAS HI) at baseline.

to indicate excellent reliability. Agreement across the scale of the ASAS HI was visualised by Bland and Altman plot. Measurement error was assessed by analysing the smallest detectable change (SDC) based on the 95% limits of agreement by using the formula: $SDC = 1.96 \times SD$ of the mean difference in ASAS HI of the two assessments in the reliability sample $\sqrt{2}$.²⁴ Responsiveness was tested with standardised response mean (SRM) after 2–24 weeks depending on the type of medication. SRM was assessed by using the following formula: $SRM = ASAS\ HI\ mean\ difference / SD\ of\ ASAS\ HI\ mean\ difference$. A $SRM < 0.4$ was considered to represent a low effect, 0.4–0.79 a moderate effect and ≥ 0.8 a large effect. The discriminant ability of the ASAS HI was assessed by calculating ASAS HI mean scores for predefined status groups (ASDAS status groups (inactive, moderate, high and very high), BASDAI and BASFI thresholds (< 2.0 , 2.0–3.99, 4.0–5.99, ≥ 6.0)) by analysis of variance. To distinguish between

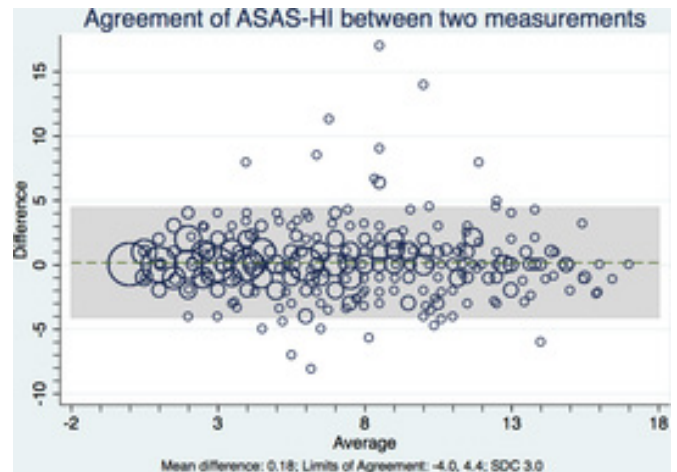


Figure 2 Bland and Altman plot. The differences between total sum score of ASAS Health Index (ASAS HI) at two time points were plotted against the mean of the two values together with the smallest detectable change (SDC).

relevant health states (an additional relevant aspect of interpretability), two different methods were applied: fixed 90% specificity and the closest point to (0,1).^{25 26} We used the patient global assessment at predefined levels (< 3 and > 6 on NRS and cut-off between good and poor on Likert scale) as external constructs for ‘poor’, ‘moderate’ and ‘good’ health status. We used a global rating of change question (Likert scale) as external construct to assess change perceived by the patient. A cut-off between ‘improved’ versus ‘no change’ or ‘worse’ was used to determine minimal clinically important improvement. Final choice was based on a consensus during the ASAS meeting in June 2017 (74 participants, 100% agreement). A $p\ value \leq 0.05$ was considered significant. Statistical analyses were performed using SPSS V.23.

Table 2 Spearman correlation between ASAS Health Index scores and other PRO

	Hypothesis	Spearman correlation	Confirmation*
Pain	High	0.60	Yes
Spinal pain	High	0.54	Yes
Patient global	High	0.57	Yes
Physician global	Moderate	0.49	Yes
ASDAS	High	0.61	Yes
BASDAI	High	0.70	Yes
BASFI	High	0.71	Yes
HADS anxiety	Moderate	-0.55	No
HADS depression	Moderate	-0.57	No
EQ-5D VAS (0–100 mm)	High	0.45	No
EQ-5D	High	-0.72	Yes
SF-36 PSC	High	-0.73	Yes
SF-36 MSC	Moderate	-0.59	No
WPAI presenteeism	Moderate	0.60	No
WPAI absenteeism	Moderate	0.38	Yes
Well-being last week	High	0.61	Yes

*Column indicates whether hypothesis generated prior to analysis about magnitude and direction of correlation was confirmed in the specific variable.

ASDAS, Ankylosing Spondylitis Disease Activity Score; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; EQ-5D, Euro Quality of Life 5 Dimensions; HADS, Hospital Anxiety Depression Scale; MSC, mental sum component; PSC, physical sum component; SF-36, Short Form 36; VAS, Visual Analogue Scale; WPAI, Work Productivity and Impairment Scale.

RESULTS

Sample characteristics

In total, 1593 patients participated in the international validation study (sample size per country varied between 15 and 130) (see online supplementary file 3). Of these, 1548 had analysable data (45 patients were excluded because of major incomplete data): 64.9% were men, mean age 42.0 (SD 13.4) years, mean symptom duration 14.5 (11.4) (table 1). There were 1292 (83.5%) patients with axSpA (375 patients (29.0%) with nr-axSpA and 917 (71.0%) with AS) and 256 (16.5%) patients with pSpA. Patients had, on average, moderate disease activity as measured by ASDAS and BASDAI, with 64.2% treated with NSAIDs and 38.2% were treated with TNFi (table 1; additional detailed patients’ characteristics of the whole cohort are presented in online supplementary file 4). As expected, the patients in the responsiveness sample have a higher level of disease activity at baseline.

Psychometric properties of the ASAS HI

Interpretability

The mean total score in the population sampled for the ASAS HI was 6.7 (SD 4.3). A total score was calculated for respondents in which not more than 20% of the data were missing (see also user’s manual published in online supplementary file 1). Numbers of missing values were limited and occurred between 0.1% and 0.3% (online supplementary file 5). Floor (percentage of the respondents who had the lowest possible (total) score)

Table 3 Discriminant ability of the ASAS Health Index (ASAS HI) with respect to disease activity and physical functioning

Disease activity					F test	P values
ASDAS thresholds	Inactive (n=245)	Moderate (n=283)	High (n=500)	Very high (n=289)		
ASAS HI	2.9 (3.1)	5.1 (3.5)	7.3 (3.6)	10.4 (3.5)	230.	<0.001
BASDAI thresholds	<2.0 (n=372)	2.0–3.9 (n=405)	4.0–5.9 (n=347)	≥6.0 (n=414)		
ASAS HI	2.8 (2.9)	5.2 (3.1)	7.8 (3.3)	10.5 (3.4)	421.4	<0.001
Functioning					F test	P values
BASFI thresholds	<2.0 (n=633)	2.0–3.9 (n=322)	4.0–5.9 (n=258)	≥6.0 (n=323)		
ASAS HI	3.7 (3.1)	6.5 (3.1)	8.6 (3.4)	11.2 (3.1.6)	438.0	P<0.001

*All values given as mean (SD).

ASDAS, Ankylosing Spondylitis Disease Activity Score; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index.

or ceiling effects (percentage of the respondents who had the highest possible (total) score) of the ASAS HI in this analysis were acceptable (6.9% and 0.8%, respectively) (figure 1).

Construct validity

Construct validity showed Spearman correlation coefficient ranging from moderate (WPAI absenteeism: 0.38) to high (BASFI: 0.71 or SF-36 PSC 0.73).

As hypothesised, the ASAS HI had high correlation with patient global ($r=0.57$), pain ($r=0.60$), spinal pain ($r=0.54$), SF-36 MCS ($r=0.59$), HADS ($r=-0.55$ and -0.57), BASFI, ASDAS ($r=0.61$), presenteeism ($r=0.60$), BASDAI, BASFI, EQ-5D and SF-36 PCS ($r>0.70$). The correlation of ASAS HI with physician global ($r=0.49$) and absenteeism ($r=0.38$) was moderate (table 2). Of note, correlation of ASAS HI with age was weak ($r=0.10$). Hypothesis about magnitude and direction of correlation was confirmed in 68.7% of variables.

Internal consistency

ASAS HI scores showed a high Cronbach's α of 0.93. Internal consistency of ASAS HI did not vary much across different disease groups (0.93 for AS, 0.94 for nr-axSpA and 0.91 for pSpA).

Reliability and measurement error

A total of 770 patients had a second assessment for reliability. Of these, 192 patients had to be excluded because of missing data ($n=54$), patients not being stable ($n=74$) or second assessment performed outside the time frame ($n=64$). Finally, 578 (75.1%) patients who considered themselves to be in a stable state were analysed (table 1). The mean (SD) baseline ASAS HI was 6.2 (4.2) and the second ASAS HI was 6.0 (4.2). Reliability was excellent with an ICC of 0.87 (95% CI 0.84 to 0.89) and ICCs were comparably high in all disease subtypes (AS 0.87 (95% CI

0.84 to 0.89); nr-axSpA 0.89 (95% CI 0.85 to 0.93); pSpA 0.83 (95% CI 0.75 to 0.88)). Bland-Altman plot shows a good agreement between ASAS HI sum score at first and second assessment. No systematic differences in sum score for the two measurement time points were found. Calculation of the limits of agreement (and the SDC) was based on the assumption that reliability was homoscedastic over the entire range of ASAS HI although this was not completely the case as the variation was somewhat more pronounced in the middle of the range (figure 2). The SDC was calculated as 3.0, which corresponds to the minimum change beyond measurement error that can be detected in an individual patient over time.

Responsiveness

A total of 353 patients were allocated to the sensitivity to change arm because of initiation of a new treatment. Also, 107 patients had to be excluded from the 353 initial patients because of missing data ($n=47$), patients deteriorating during time interval ($n=12$), patients not reporting a change in their disease state ($n=47$) and second assessment performed outside of the time frame ($n=1$). Finally, 246 (69.7%) estimated themselves to have improved between visits and were analysed. Seventy-eight patients started NSAIDs, 41 patients a csDMARD and 127 patients TNFi. The SRM was -0.44 for NSAIDs (moderate), -0.69 for csDMARDs (moderate) and -0.85 for TNFi (large).

Discriminant ability

The ASAS HI discriminated well between patients with different disease activity states (measured by ASDAS and BASDAI) and function (measured by BASFI) (table 3). The groups with greater disease activity and more impaired functioning had higher mean ASAS HI scores (indicating impaired health) than those with lower disease activity.

Table 4 Analysis of cut-off values for ASAS Health Index (ASAS HI) scores to define health status

ASAS HI cut-offs and external criterion	n (P+N)	90% SP (SE/SP)	(0,1) (SE/SP)	AUC
Cut-off between 'good' and 'moderate' functioning				
Patient global, Likert very good/good versus all others	1531 (624+907)	3.0 (49.5/90.0)	5.7 (73.5/75.9)	0.81
Patient global, NRS <3	1533 (435+1098)	2.3 (46.2/90.5)	5.0 (76.5/71.9)	0.80
Cut-off between 'moderate' and 'poor' functioning				
Patient global, Likert very poor/poor versus all others	1531 (304+1227)	12.0 (39.5/90.5)	7.4 (79.3/67.6)	0.80
Patient global, NRS >6	1533 (425+1108)	11.7 (34.3/89.6)	7.4 (71.3/69.7)	0.76

AUC, area under the curve; NRS, numerical rating scale; P+N, number of positive+negative results according to the external criterion; SE, sensitivity; SP, specificity; 90% SP, cut-off according to the 90% specificity criterion; (0,1), cut-off according to the closest point to (0,1) criterion.

Table 5 Discriminant ability of the health status groups

Health state (number, % patients)	Good ≤5.0 (n=553, 36%)	Moderate <5.0 to <12.0 (n=755, 49%)	Poor ≥12.0 (n=235, 15%)
ASAS HI	2.1 (1.5)	7.8 (2.0)	13.7 (1.5)
BASFI	1.2 (1.5)	3.8 (2.5)	6.3 (2.3)
BASDAI	2.1 (1.6)	4.8 (2.1)	6.6 (1.9)
ASDAS	1.7 (0.9)	2.6 (2.1)	3.7 (1.1)
SF-36 PSC	47.6 (7.1)	35.7 (8.8)	28.7 (6.6)
EQ-5D	0.8 (0.1)	0.6 (0.2)	0.4 (0.2)

*Values given as mean (SD) otherwise indicated.

ASAS HI, ASAS Health Index; ASDAS, Ankylosing Spondylitis Disease Activity Score; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; EQ-5D, Euro Quality of Life 5 Dimensions; PSC, physical sum component; SF-36, Short Form 36.

Cut-off values for interpreting health status based on ASAS HI scores

Final cut-offs for ASAS HI scores to distinguish poor versus moderate, and moderate versus good health are presented in table 4. All analysed scenarios with application of different external anchors and different methodological approaches are presented in online supplementary file 6. In order to balance sensitivity and specificity, a threshold of ASAS HI, which differentiated patients with 'good/very good' health from those with 'moderate' health state, was identified as being 5.0. In contrast, the 90% specificity criterion was considered to be the most clinically relevant threshold of ASAS HI for 'moderate' versus 'poor/very poor' health identified as a score of 12.0 or above.

Attempts to define a clinically important improvement proved an elusive target since scores were too heterogeneous. We therefore recommend using the SDC value of 3.0 to determine change in ASAS HI in individual patients and present the percentage of patients with a change of ≥3.0.

Applying these thresholds within the validation cohort, we were able to show that the three defined health status groups within ASAS HI could discriminate with respect to both disease activity, functioning and health measures (table 5). The two cut-off values delineating the three health statuses were agreed on after discussion and voting by 74 ASAS members during their European League Against Rheumatism meeting 2017 (74 approval, 0 decline, 0 abstention).

DISCUSSION

The manuscript presents the psychometric properties of the original English ASAS HI and its different translations, as obtained in a large international cohort. We show that the ASAS HI is a valid, reliable and responsive measure of functioning and health in patients with SpA on a global level. Interpretability was good as has been shown for different aspects. In this paper, we report the values for the entire cohort and country-specific results will be published separately in the language of the specific country. Generally, the results were similar in the various countries (data not shown).

Since the ASAS HI contains only 17 items with a dichotomous response option addressing all important aspects of patient complaints, administration of the questionnaire is feasible as it has been shown in a previous field test.¹⁰ The calculation to obtain a single sum score is simple and quick to undertake. Floor effect was acceptable with almost no ceiling effect observed in our study. The scores have good face validity and the ASAS HI exhibited excellent correlation with other measures covering a range of health outcomes. Analysis of construct validity demonstrated a strong association between ASAS HI sum score and both disease activity

and functional disability, indicating that the ASAS HI is measuring a broader concept than just disease activity or physical functioning. In addition, the high correlation between ASAS HI and patient global assessment as well as generic health measures (such as SF-36) suggest that patients do not make substantial distinctions between disease-specific and more generally worded questionnaires. We noted in our cohort a weaker correlation between ASAS HI and physician global as well as discordance between physician and patient global scores at baseline. However, the discordance between patient responses and physician response is very small and not comparable with those reported in literature.^{27 28}

We were able to show that the ASAS HI is applicable in all patients with SpA irrespective of the disease subgroup. Similar results in internal consistency between AS, nr-axSpA and pSpA provide support for the use of these questionnaires in the whole group of SpA. This is an important finding as the ASAS HI was originally developed in patients with AS. However, use of the ASAS HI in patients with pSpA should be carefully checked and its applicability should be further investigated to gain more insights into this subgroup of patients.

There is a debate about which measure is suitable for assessing responsiveness. Our choice is SRM, which is not recommended according to the COSMIN guidelines.²¹ However, SRM is one of the widely used responsiveness measures and there is also critique published in the literature about this part of the COSMIN guidelines.²⁹ One of the arguments is that the SRM is more reflecting the magnitude of the event than providing information on the measure. Indeed, we do show that the SRM is better for start of biological DMARDs than for NSAIDs. However, providing the SRM is a useful information for researchers who want to use the ASAS HI as an outcome measure in a trial.

This study has clear strengths and weaknesses. Strengths include the involvement of 23 countries with 18 country-specific translations with different cultures and socioeconomic backgrounds within the validation process.¹⁰ Thus, the domains of functioning and health assessed in the questionnaire are likely to be relevant across countries and cultures. However, qualitative research about this issue is lacking. One relative weakness of this international validation study may be considered the small sample size in some countries, especially in the longitudinal arm. However, the results of the study do show that the psychometric properties are robust and meaningful. The ASAS HI can be used in clinical trials to evaluate the impact of SpA and its treatment on overall functioning and health in patients with SpA and also to compare disease impact in cohorts and populations. Further research is needed to address the question whether and how the ASAS HI is applicable in daily routine care to guide treatment decisions.

In conclusion, the ASAS HI proved to be a valid, interpretable, reliable and responsive questionnaire to assess overall functioning and health in this global international validation study including 19 languages.

Author affiliations

¹Rheumazentrum Ruhrgebiet, Herne, Germany

²Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

³Division of Rheumatology, Maastricht University Medical Center, Maastricht, The Netherlands

⁴Izmir, Turkey

⁵School of Medicine, Universidad Militar Nueva Granada and Rheumatology Department, Hospital Militar, Bogotá, Colombia

⁶Hospital General de Mexico Universidad Nacional Autónoma, Mexico

⁷Division of Allergy, Immunology and Rheumatology, Chung Shan Medical University Hospital, Institute of Medicine, Graduate Institute of Integrated Medicine, China Medical University, Taichung, Taiwan

- ⁸Faculty of Medicine, Siriraj Hospital, Mahidol University, Thailand
⁹Department of Rheumatology, Servicio de Reumatología, Hôpital Cochin, Assistance Publique—Hôpitaux de Paris INSERM (U1153): Clinical Epidemiology and Biostatistics, PRES Sorbonne Paris-Cité, H, Paris, France
¹⁰PMR Department, Rheumatology Division, Marmara University School of Medicine, Istanbul, Turkey
¹¹Rheumatology Department, Cairo University, Cairo, Egypt
¹²North-Western State Medical University named after I I Mechnikov, St. Petersburg, Russian Federation
¹³Division of Rheumatology, University of California, San Francisco, San Francisco, California, USA
¹⁴Department of Lucania, San Carlo Hospital of Potenza and Madonna delle Grazie Hospital of Matera, Potenza and Matera and PhD Scholarship in Life Sciences, Department of Health Sciences, University of Catanzaro 'Magna Graecia', Catanzaro, Italy
¹⁵Sestre Milosrdnice University Hospital Center, Zagreb, Croatia
¹⁶University of Guangzhou, Guangzhou, China
¹⁷Spondylitis, University of Toronto, Toronto, Ontario, Canada
¹⁸Chonnam National University Medical School and Hospital, South Korea
¹⁹University Hospital La Paz, IdiPaz, Madrid, Spain
²⁰NIHR Leeds Biomedical Research Centre, Chapel Allerton Hospital, Leeds, and Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK
²¹Department of Rheumatology, Medicalpark Gaziosmanpasa Hospital, Istanbul, Turkey
²²Universidade Nova De Lisboa, Lisbon, Portugal
²³Department of Internal Medicine, Innsbruck Medical University, Clinic II, Innsbruck, Austria
²⁴Dunedin School of Medicine, University of Otago, Dunedin, New Zealand
²⁵VIB Center for Inflammation Research, Unit for Molecular Immunology and Inflammation, Department of Internal Medicine, Ghent University, Ghent, Belgium
²⁶Department of Rheumatology, Ghent University Hospital, Ghent, Belgium

Acknowledgements The authors thank all patients who participated in the study. We would like to thank all national collaborators for their tremendous endeavour to recruit patients and to document the results. We thank S Kemmerling in Austria; K Elgarf, M Shaaban and A Ahmed in Egypt; M Flörecke, J Meier and J Winter in Germany; G Marsico and I Olivieri in Italy; L Ward in New Zealand; A Akulova and A Rebrov in Russia; R Ortega (University Hospital Reina Sofia, Córdoba), M Aparicio and X Juanola (University Hospital Bellvitge, Barcelona), R Almodóvar and P Zarco (University Hospital Fundación Alcorcón, Madrid), and C. Rodríguez (University Hospital Dr. Negrín, Gran Canaria) in Spain; E Kilic, G Klic, D Yildiz and G Kenar in Turkey; S Bhalara (West Hertfordshire Hospitals NHS Trust), K Gaffney (Norfolk and Norwich University Hospital), P Helliwell (Bradford Institute for Health Research and University of Leeds), J Packham (Keele University) and CS Yee (Doncaster Royal Infirmary) in UK; and G Yoon and The Russell Engleman Rheumatology Research Center (UCSF) in USA for local data collection. We thank Joachim Listing from the German Rheumatism Research Centre Berlin (DRFZ) Germany, who did the threshold analysis and who was of enormous help in discussing the detailed analysis. This work was previously presented at the following conferences and published as a conference abstract:

Contributors Study concept and design: UKU, DvdH, AB, JB. Acquisition of data: all authors. Analysis and interpretation of data: all authors. Writing of the manuscript: UKU. Critical revision of the manuscript for important intellectual content: all authors. All authors had access to the data, commented on the report drafts and approved the final submitted version.

Funding This study was funded by the Assessment of Spondyloarthritis International Society (ASAS).

Competing interests None declared.

Patient consent Obtained.

Ethics approval All centres received approval from their local ethics committee.

Provenance and peer review Not commissioned; externally peer reviewed.

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Dougados M, Baeten D. Spondyloarthritis. *Lancet* 2011;377:2127–37.
- Ward MM. Health-related quality of life in ankylosing spondylitis: a survey of 175 patients. *Arthritis Care Res* 1999;12:247–55.
- van Tubergen A, Coenen J, Landewé R, *et al.* Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis. *Arthritis Rheum* 2002;47:8–16.
- Kiltz U, Baraliakos X, Karakostas P, *et al.* Do patients with non-radiographic axial spondylarthritis differ from patients with ankylosing spondylitis? *Arthritis Care Res* 2012;64:1415–22.
- Stolwijk C, Essers I, van Tubergen A, *et al.* The epidemiology of extra-articular manifestations in ankylosing spondylitis: a population-based matched cohort study. *Ann Rheum Dis* 2015;74:1373–8.
- Kristensen LE, Karlsson JA, Englund M, *et al.* Presence of peripheral arthritis and male sex predicting continuation of anti-tumor necrosis factor therapy in ankylosing spondylitis: an observational prospective cohort study from the South Swedish Arthritis Treatment Group Register. *Arthritis Care Res* 2010;62:1362–9.
- Kiltz U, van der Heijde D, Boonen A, *et al.* Development of a health index in patients with ankylosing spondylitis (ASAS HI): final result of a global initiative based on the ICF guided by ASAS. *Ann Rheum Dis* 2015;74:830–5.
- WHO. *The International Classification of Functioning Disability and Health*. Geneva: WHO, 2001.
- Boonen A, Braun J, van der Horst Bruinsma IE, *et al.* ASAS/WHO ICF Core Sets for ankylosing spondylitis (AS): how to classify the impact of AS on functioning and health. *Ann Rheum Dis* 2010;69:102–7.
- Kiltz U, van der Heijde D, Boonen A, *et al.* Measuring impairments of functioning and health in patients with axial spondyloarthritis by using the ASAS Health Index and the Environmental Item Set: translation and cross-cultural adaptation into 15 languages. *RMD Open* 2016;2:e000311.
- Rudwaleit M, van der Heijde D, Landewé R, *et al.* The development of assessment of spondyloarthritis international society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68:777–83.
- Rudwaleit M, van der Heijde D, Landewé R, *et al.* The assessment of Spondyloarthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. *Ann Rheum Dis* 2011;70:25–31.
- Garrett S, Jenkinson T, Kennedy LG, *et al.* A new approach to defining disease status in ankylosing spondylitis: the Bath ankylosing spondylitis disease activity index. *J Rheumatol* 1994;21:2286–91.
- Calin A, Garrett S, Whitelock H, *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.
- EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
- Reilly MC, Gooch KL, Wong RL, *et al.* Validity, reliability and responsiveness of the Work Productivity and Activity Impairment Questionnaire in ankylosing spondylitis. *Rheumatology* 2010;49:812–9.
- van der Heijde D, Lie E, Kvien TK, *et al.* ASDAS, a highly discriminatory ASAS-endorsed disease activity score in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68:1811–8.
- Machado P, Landewé R, Lie E, *et al.* Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. *Ann Rheum Dis* 2011;70:47–53.
- Mokkink LB, Terwee CB, Patrick DL, *et al.* The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- Boers M, Brooks P, Strand CV, *et al.* The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198–9.
- Fleiss JL. *The design and analysis of clinical experiments*. New York: John Wiley & Sons, 1999.
- Brunnesteyn K, Boers M, Kostense P, *et al.* Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. *Ann Rheum Dis* 2005;64:179–82.
- Coffin M, Sukhatme S. Receiver operating characteristic studies and measurement errors. *Biometrics* 1997;53:823–37.
- Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006;163:670–5.
- Barton JL, Imboden J, Graf J, *et al.* Patient-physician discordance in assessments of global disease severity in rheumatoid arthritis. *Arthritis Care Res* 2010;62:857–64.
- Desthieux C, Molto A, Granger B, *et al.* Patient-physician discordance in global assessment in early spondyloarthritis and its change over time: the DESIR cohort. *Ann Rheum Dis* 2016;75:1661–6.
- Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. *BMC Med Res Methodol* 2011;11:152.