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

U. Kiltz U[1], D. van der Heijde[2], A. Boonen[3], N. Akkoc[4], W. Bautista-Molano[5], R. Burgos Vargas[6], J. C. C Wei[7], P. Chiowchanwisawakit[8], M. Dougados[9], MT. Duruoz.[10] B. Elzorkany[11] I. Gaydukova[12], L. S. Gensler[13], M. Gilio[14], S. Grazio[15], J. Gu[16], R. Inman [17], T. Kim[18], V. Navarro-Compán[19], H. Marzo-Ortega[20], S. Ozgocmen[21], F. Pimentel dos Santos[22], M. Schirmer[23], S. Stebbings[24], F. van den Bosch[25, 26], A. van Tubergen[3], Braun J[1].

*[1]Rheumazentrum Ruhrgebiet, Herne, Germany; [2]Leiden University Medical Center, Department Of Rheumatology, Leiden, Netherlands; [3]Maastricht University Medical Center, Division Of Rheumatology, Maastricht, Netherlands; [4]Izmir, Turkey; [5] School of Medicine, Universidad Militar Nueva Granada and Rheumatology Department Hospital Militar, , Bogotá, Colombia; [6] Hospital General de Mexico Universidad Nacional Autónoma, Mexico;[7] Division of Allergy, Immunology and Rheumatology. Chung Shan Medical University Hospital, Institute of Medicine, , Graduate Institute of Integrated Medicine, China Medical University, Taichung, Taiwan; [8] Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand, [9] Paris-Descartes University, Department of Rheumatology, Hôpital Cochin. Assistance Publique - Hôpitaux de Paris INSERM (U1153): Clinical epidemiology and biostatistics, PRES Sorbonne Paris-Cité, - H, Paris, France; [10]Marmara University School of Medicine, PMR Department, Rheumatology Division, Istanbul, Turkey; [11]Rheumatology Department, Cairo University, Egypt, [20] North-Western State Medical University named after I I Mechnikov, St. Peterburg, , Russian Federation; [31]University Of California, San Francisco, Division Of Rheumatology, USA; [14]Rheumatology Institute of Lucania (IRel) And Rheumatology Department 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; [15]*Sestre milosrdnice University Hospital Center, Zagreb, Croatia*, [16]University Guangzhou, China; [17] Canada; [18]Chonnam Nat`L University Medical School&Hospital, South Korea; [19]University Hospital La Paz, Madrid, IdiPaz, Spain;[20]NIHR Leeds Biomedical Research Centre, Chapel Allerton Hospital, Leeds, and Leeds Institute Of Rheumatic And Musculoskeletal Medicine, University of Leeds, Leeds, UK; [21] Department of Rheumatology, Medicalpark Gaziosmanpasa Hospital, Istanbul, Turkey; [22]Universidade Nova De Lisboa, Lisboa, Portugal; [23]Innsbruck Medical University , Department of Internal Medicine, Clinic II, Austria; [24]University Of Otago, Dunedin School Of Medicine, Dunedin, New Zealand; [25]VIB Center for Inflammation Research, Ghent, Belgium, [26] Department of Rheumatology Ghent University Hospital, Ghent, Belgium.*

Corresponding author:

Dr. Uta Kiltz
Rheumazentrum Ruhrgebiet
Claudiusstr 45
44649 Herne
Germany
Tel 0049 2325 592 705
Fax 0049 2325 592 136
uta.kiltz@elisabethgruppe.de

Word count: 3293

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 ASAS Health Index (HI) in 23 countries worldwide in patients with Spondyloarthritis (SpA).

Methods: A representative sample of SpA patients fullfilling the ASAS classification criteria for axial (axSpA) or peripheral SpA (pSpA) 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 (ICC) 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 standardized response mean (SRM).

Results: Among the 1548 patients 64.9% were male, with a mean (SD) age 42.0 (13.4) years. Construct validity ranged from low (age: 0.10) to high (BASFI: 0.71). Internal consistency was high (Cronbachs-α of 0.93). The reliability among 578 patients was good (ICC=0.87 (95%CI 0.84-0.89)). Responsiveness among 246 patients was moderate-large (SRM= -0.44 for NSAIDs, -0.69 for csDMARD, and -0.85 for TNFi). The smallest detectable change was 3.0. Values ≤5.0 have balanced specificity to distinguish good health opposed to moderate health and values ≥12.0 are specific to represent poor health 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.

Key words: spondyloarthritis, ankylosing spondylitis, Outcome research, validation research

Introduction:

Spondyloarthritis (SpA) is characterized 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 (IBD) 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 characterized in patients with ankylosing spondylitis (AS), and to a lesser extent for patients with non-radiographic axial SpA (nr-axSpA) but there is little data relating to patients with peripheral SpA (pSpA).[4, 6]

The ASAS 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 bio-psychosocial 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 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 SpA patients fulfilling the ASAS classification criteria for either axial (axSpA) or pSpA was 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 co-existent 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 which 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 (CRP) levels, imaging results and current medications were also recorded. Physicians judgment 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 (HAD-S) [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 ASDAS sum score was calculated and patients were categorized 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 HI: 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 users manual for handling missing items (supplement 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 supplement 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 tumor necrosis factor inhibitor (TNFi) therapy over 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 analyzed

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 examined according to the OMERACT filter. [22] Descriptive statistics were used to characterize the sample. According to the COMSIN checklist, interpretability is being summarized 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 hypothesized 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 ≥ 0.8. [23] Internal consistency was evaluated using Cronbachs 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 to indicate excellent reliability. Agreement across the scale of the ASAS HI was visualized by Bland and Altman plot. Measurement error was assessed by analyzing the smallest detectable change (SDC) based on the 95% limits of agreement by using the formula: SDC=1.96*standard deviation of the mean difference in ASAS HI of the two assessments in the reliability sample/√2. [24] Responsiveness was tested with standardized 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/standard deviation 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 ANOVA. To distinguish between 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 (MCII). Final choice was based on a consensus during ASAS meeting in June 2017 (74 participants, 100% agreement). A p value ≤0.05 were considered significant. Statistical analyses were performed using SPSS (Chicago, Illinois, USA) version 23.

Results:

1.Sample characteristics

In total, 1593 patients participated in the international validation study (sample size per country varied between 15 and 130) (see supplement 3). Of these, 1548 had analyzable data (45 patients were excluded because of major incomplete data): 64.9% were male, 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 (29.0%) nr-axSpA and 917 (71.0%) AS patients) 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 supplement 4). As expected, the patients in the responsiveness sample have a higher level of disease activity at baseline.

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, 1st visit (n=578)**	**Reliability, 2nd visit (n=578)**	**Responsiveness 1st visit (n=246)§**	**Responsiveness 2nd 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)	
Extraarticular 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

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

§ Only data analyzed from those patients who stated that they improved during the time interval

&calculated for employed patients (n=961), see supplement 4

ASAS HI= ASAS Health Index; ASDAS=Ankylosing Spondylitis Disease Activity Index; 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 Dimension; HLA=Human leukocyte antigen; HADS=Hospital Anxiety Depression Scale; IBD=Inflammatory Bowel Disease; MSC=Mental component summary score; NSAIDs=Non-steroidal antirheumatic drug; NRS=numerical rating scale; PASS=Patient Acceptable Symptom State; PSC=Physical component summary score; NA= not applicable; SF-36=Short-form 36; TNFi=Tumuor necrosis factor inhibitor; VAS=Visual analogue scale; WPAI=Work Productivity and Impairment Scale.

2. Psychometric properties of the ASAS HI

2.1. 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 Users manual published in supplement 1). Numbers of missing values were limited and occurred between 0.1 and 0.3% (supplement 5). Floor (percentage of the respondents who had the lowest possible (total) score) 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).

2.2. 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 hypothesized, 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), HAD-S (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.

Table 2: Spearman correlation between ASAS HI 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

ASDAS=Ankylosing Spondylitis Disease Activity Index; BASDAI= Bath Ankylosing Spondylitis Disease Activity Index; BASFI=Bath Ankylosing Spondylitis Functional Index; EQ-5D=Euro Quality of Life 5 Dimension; HADS=Hospital Anxiety Depression Scale; MSC=Mental sum component; PSC=Physical sum component; SF-36=Short-form 36; WPAI=Work Productivity and Impairment Scale. *Column indicates whether hypothesis generated prior to analysis about magnitude and direction of correlation was confirmed in the specific variable.

2.3. Internal consistency: ASAS HI scores showed a high Cronbachs-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).

2.4. 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 timeframe (n=64). Finally, 578 (75.1%) patients who considered themselves to be in a stable state were analyzed (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 sumscore at first and second assessment. No systematic differences in sumscore for the two measurement timepoints 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.

2.5. Responsiveness: 353 patients were allocated to the sensitivity to change arm because of initiation of a new treatment. 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 timeframe (n=1). Finally, 246 (69.7%) estimated themselves to have improved between visits and were analyzed. 78 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).

2.6. 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 3: Discriminant ability of the ASAS HI with respect to disease activity and physical functioning

Disease activity					**F test**	**p-value**
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.	p<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	p<0.001
Functioning					**F test**	**p-value**
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)

2.7. 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 analyzed scenarios with application of different external anchors and different methodological approaches are presented in Supplement 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.

Table 4: Analysis of cut-off values for 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
ASAS HI cut offs and external criterion	**n (P+N)**	**90%SP (SE/SP)**	**(0,1) (SE/SP)**	**AUC**
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

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

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 % 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 upon after discussion and voting by 74 ASAS members during their EULAR meeting 2017 (74 approval, 0 decline, 0 abstention).

Table 5: Discriminant ability of the health status groups

Health state (number, %patients)	Good ≤5.0 (n=553, 36 %)	Moderate <5.0 - <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)
SF36 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

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 to 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 pSpA patients 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 bDMARDs 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.

Acknowledgement:
The authors thank all patients who participated in the study. We would like to thank all national collaborators for their tremendous endeavor 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 Sofía, Córdoba), M. Aparicio and X. Juanola (University Hospital Bellvitge, Barcelona), R. Almodóvar and P. Zarco (University Hospital Foundation Alcorcón, Madrid), and C. Rodríguez (University Hospital Dr. Negrín, Gran Canaria) in Spain, E. Kilic, G. KIlic, D. Yıldız, and G. Kenar in Turkey, and 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); 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 you Joachim Listing from 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 following conferences and published as a conference abstract:

1. U Kiltz, D van der Heijde, A Boonen, J Braun, on behalf of the ASAS HI International Validation Study. VALIDATION OF THE ASAS Health index: RESULTS OF A MULTICENTER INTERNATIONAL STUDY IN 23 COUNTRIES. Annals of the Rheumatic Diseases, volume 75, supplement 2, year 2016, page 86.
2. Kiltz U, van der Heijde D, Boonen A, Braun J. Validation of the ASAS Health Index: Results of a Multicenter International Study in 23 Countries [abstract]. Arthritis Rheumatol. 2016; 68 (suppl 10): 2003.
3. Bautista-Molano W, Landewé R, Kiltz U, Valle-Oñate R, van der Heijde D. ASAS Health Index for Patients with Spondyloarthritis: Translation into Spanish, Validation, Reliability and Construct Validity [abstract]. *Arthritis Rheumatol.* 2016; 68 (suppl 10): 691.

Funding: This study was funded by Assessment of Spondyloarthritis international Society (ASAS).

Competing interests: None declared.

Ethics approval: National IRB in each country.

Data sharing statement: No additional data are available.

Contributorship Statement:
Study concept and design: U Kiltz, D. van der Heijde, A Boonen, J Braun
Acquisition of data: all authors
Analysis and interpretation of data: All authors
Writing of the manuscript: U. Kiltz
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.

Legends:
Figure 1: Score distribution (0-17) of the ASAS HI at baseline

Figure 2: Bland and Altman Plot. The differences between total sum score of ASAS HI at two time points were plotted against the mean of the two values together with the SDC

Literature:

1. Dougados M, Baeten D. Spondyloarthritis. Lancet. 2011 Jun 18; 377(9783):2127-2137.
2. Ward MM. Health-related quality of life in ankylosing spondylitis: a survey of 175 patients. Arthritis care and research : the official journal of the Arthritis Health Professions Association. 1999 Aug; 12(4):247-255.
3. van Tubergen A, Coenen J, Landewe R, Spoorenberg A, Chorus A, Boonen A, et al. Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis. Arthritis Rheum. 2002 Feb; 47(1):8-16.
4. Kiltz U, Baraliakos X, Karakostas P, Igelmann M, Kalthoff L, Klink C, et al. Do patients with non-radiographic axial spondylarthritis differ from patients with ankylosing spondylitis? Arthritis Care Res (Hoboken). 2012 Sep; 64(9):1415-1422.
5. Stolwijk C, Essers I, van Tubergen A, Boonen A, Bazelier MT, De Bruin ML, et al. The epidemiology of extra-articular manifestations in ankylosing spondylitis: a population-based matched cohort study. Ann Rheum Dis. 2015 Jul; 74(7):1373-1378.
6. Kristensen LE, Karlsson JA, Englund M, Petersson IF, Saxne T, Geborek P. 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 (Hoboken). 2010 Oct; 62(10):1362-1369.
7. Kiltz U, van der Heijde D, Boonen A, Cieza A, Stucki G, Khan MA, 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 May; 74(5):830-835.
8. (WHO) WHO. The International Classification of Functioning Disability and Health. Geneva, 2001.
9. Boonen A, Braun J, van der Horst Bruinsma IE, Huang F, Maksymowych W, Kostanjsek N, 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 Jan; 69(1):102-107.
10. Kiltz U, van der Heijde D, Boonen A, Bautista-Molano W, Burgos-Vargas R, Chiowchanwisawakit P, 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(2):e000311.
11. Rudwaleit M, van der Heijde D, Landewe R, Listing J, Akkoc N, Brandt J, 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 Jun; 68(6):777-783.
12. Rudwaleit M, van der Heijde D, Landewe R, Akkoc N, Brandt J, Chou CT, et al. The Assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. Ann Rheum Dis. 2011 Jan; 70(1):25-31.
13. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol. 1994 Dec; 21(12):2286-2291.
14. Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol. 1994 Dec; 21(12):2281-2285.

15. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. Health policy. 1990 Dec; 16(3):199-208.
16. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical care. 1992 Jun; 30(6):473-483.
17. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta psychiatrica Scandinavica. 1983 Jun; 67(6):361-370.
18. Reilly MC, Gooch KL, Wong RL, Kupper H, van der Heijde D. Validity, reliability and responsiveness of the Work Productivity and Activity Impairment Questionnaire in ankylosing spondylitis. Rheumatology (Oxford). 2010 Apr; 49(4):812-819.
19. van der Heijde D, Lie E, Kvien TK, Sieper J, Van den Bosch F, Listing J, et al. ASDAS, a highly discriminatory ASAS-endorsed disease activity score in patients with ankylosing spondylitis. Ann Rheum Dis. 2009 Dec; 68(12):1811-1818.
20. Machado P, Landewe R, Lie E, Kvien TK, Braun J, Baker D, et al. Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. Ann Rheum Dis. 2011 Jan; 70(1):47-53.
21. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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 Jul; 63(7):737-745.
22. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol. 1998 Feb; 25(2):198-199.
23. Fleiss JL. The Design and Analysis of Clinical Experiments. New York: John Wiley & Sons, 1999.
24. Bruynesteyn K, Boers M, Kostense P, van der Linden S, van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis. 2005 Feb; 64(2):179-182.
25. Coffin M, Sukhatme S. Receiver operating characteristic studies and measurement errors. Biometrics. 1997 Sep; 53(3):823-837.
26. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. American journal of epidemiology. 2006 Apr 1; 163(7):670-675.
27. Barton JL, Imboden J, Graf J, Glidden D, Yelin EH, Schillinger D. Patient-physician discordance in assessments of global disease severity in rheumatoid arthritis. Arthritis Care Res (Hoboken). 2010 Jun; 62(6):857-864.
28. Desthieux C, Molto A, Granger B, Saraux A, Fautrel B, Gossec L. Patient-physician discordance in global assessment in early spondyloarthritis and its change over time: the DESIR cohort. Ann Rheum Dis. 2016 Sep; 75(9):1661-1666.
29. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. BMC medical research methodology. 2011 Nov 18; 11:152; author reply 152.