RMD Open
Rheumatic & Musculoskeletal Diseases

ORIGINAL ARTICLE

# Development of one general and six country-specific algorithms to assess societal health utilities based on ASAS HI

Ivette Essers,[1] Mickael Hiligsmann,[2] Uta Kiltz,[3] Nick Bansback,[4] Juergen Braun,[5] Desirée van der Heijde,[6] Annelies Boonen[7]

**To cite:** Essers I, Hiligsmann M, Kiltz U, *et al*. Development of one general and six country-specific algorithms to assess societal health utilities based on ASAS HI. *RMD Open* 2019;**5**:e000872. doi:10.1136/rmdopen-2018-000872

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136rmdopen-2018-000872).

Received 2 December 2018
Accepted 29 April 2019



For numbered affiliations see end of article.

**Correspondence to**
Dr Ivette Essers;
ivette.essers@mumc.nl

## ABSTRACT

**Objective** Health utilities represent preference values that persons attach to health states. This study aims to develop one general and six country-specific algorithms to calculate societal preference values for health of patients with spondyloarthritis (SpA), as assessed by the disease-specific Assessment of SpondyloArthritis international Society Health Index (ASAS HI).

**Methods** A survey was performed in random population samples from six European countries. In a best-worst choice experiment, subjects were asked to indicate repeatedly which of 4 random aspects of the 17-item ASAS HI was were most and least important. Bayesian analysis provided the relative importance of each of the 17 items. To rescale the relative importance scores on the absolute utility scale between 0 and 1, participants additionally completed two lead time trade-off experiments, one for 'severe SpA' and one for 'best health' without SpA. Six country-specific algorithms and one general algorithm were derived. The general algorithm was tested in 199 patients with axial SpA (axSpA).

**Results** 3039 subjects, mean age 47 years (SD 15) and 52% female completed the experiments. The population's health utility value for SpA varied between − 0.24 for 'worst' SpA (country range −0.35 to 0.03), and 0.88 for 'best' health (country range 0.81 to 0.90). Among 199 patients with axSpA, the mean utility was 0.36 (SD 0.30, range −0.24 to 0.88) and discriminated well between patients having high (Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) ≥ 4) or low (BASDAI < 4) disease activity (0.18 (SD 0.24) vs 0.51(SD 0.27), p<0.01).

**Conclusion** One general and six country-specific algorithms are available to convert scores from the ASAS HI into disease-specific societal utility values.

## Key messages

### Whatis already known about this subject?

► The ASAS-HI offers the opportunity to develop disease-specific health utility-valuations for patients with spondyloarthritis.

### What does this study add?

► Several algorithms are proposed to convert scores on the ASAS-HI into societal health utilities. These algorithms perform equal in non-radiographic and radiographic SpA and discriminated between levels of disease activity.

### How might this impact on clinical practice?

► These algorithms can be used to derive societal health utilities in patients with spondyloarthritis.

## INTRODUCTION

The aim of diagnosing and treating patients with spondyloarthritis (SpA) is to increase the patients' health for the remaining part of their life. Although no cure exists for patients with SpA, earlier diagnosis and innovations in pharmacological treatment substantially improved the outcome of SpA.[1] To assess the impact of the disease and its treatment on *overall functioning and health* of patients with SpA, disease-specific instruments to assess overall health such as the Ankylosing Spondylitis Quality of Life Questionnaire and the Assessment of SpondyloArthritis international Society Health Index (ASAS HI) have been developed.[2 3] However, these traditional outcome instruments do not reveal the *'preferences' or 'values'* for different health states, that can be represented by the combinations of items in these questionnaires. Different (combinations of) items might not be equally important for functioning and health. Preference-based valuations of health are based on choice experiments that force respondents to indicate their preference for different health states against each other. For application in healthcare, *health utilities* are preference-based valuation methods that anchor the value of preference of health states on a 0 to 1 scale, in which 0 corresponds a state equivalent to death and 1 to full health. This common scaling between 0 and 1 allows comparison

of health valuations between conditions. Health utilities are often integrated over time and quality-adjusted life years (QALYs) allow assessing the life impact of the disease. QALYs are useful when rational choices have to be made by decision makers when allocating resources to healthcare and to research,[4 5] but are increasingly also considered in clinical care in the context of value-based healthcare.[6]

When allocating *societal* resources, it is considered appropriate to consider the *societal point of view or perspective* on the 'value' for health.[7] In other words, persons that do not suffer from health problems are invited to value the health of persons with impairments. Arguments favouring a societal perspective indicate that those who do not have the disease did not adapt and can better judge the true value of health, are free of self-interest and have a democratic right to decide on distribution of (tax-based) societal resources.[7 8] Of interest, previous research revealed that cultural differences exist in the value or preference for health and functioning, and recommend country-specific societal valuations.[9]

To assess societal health utilities in SpA, generic approaches are available among which the EuroQoL five dimensions (EQ-5D) and the Short Form six dimensions (SF-6D) are the best known.[10 11] For these patient-reported measures of overall health, algorithms are available that can convert the patients' reported scores into a societal health utility value.[12] As these utility values are obtained by applying a societally derived algorithm to a patient-reported outcome (and not through direct descriptions of profiles of health impairments), these approaches to asses societal utility are called 'indirect' methods. However, using patient-reported measures as a basis for indirect health utility valuation may not accurately reflect the effect of interventions on specific impairments and limitations typically experienced by these patients. It is increasingly recognised that disease-specific measures of overall health better reflect health problems of patients with specific conditions.[13] This holds especially true when choices between interventions within the same disease have to be made. In SpA, the ASAS HI has recently been developed to assess the broad influence of SpA on functioning and health. The ASAS HI therefore offers an appropriate starting point to develop a disease-specific utility index, as it was developed with the specific aim to reflect aspects of health important and typical for patients with axial SpA (axSpA).[14] Moreover, the ASAS HI has been developed in parallel in several countries, and validated in additional countries.[15]

The objective of this study was to develop an algorithm to calculate the societal health utilities specific for SpA based on scores on the ASAS HI. As there can be cultural differences in preference for health, it was decided to develop a general as well as several country-specific algorithms. A secondary aim was to assess the new societal utility based on ASAS HI among patients with SpA.

## MATERIALS AND METHODS

### Subjects

A cross-sectional multinational online survey was conducted among general population subjects 18 years or older from the UK, France, Germany, the Netherlands, Spain and Italy. Recruitment of subjects and distribution of the questionnaire was performed by IPSOS, an international institution specialised in data collection for market and population research.[16] IPSOS approached subjects that were voluntarily subscribed to their panel. Participation for each subsequent survey is again on a completely voluntary basis. Surveys were conducted online through a password-protected website. The aim was to include 500 subjects per country to ensure a sufficient sample size. Within each country, inclusion was stratified by IPSOS for age categories, gender and region to guarantee a representative sample of each country's general population. The online survey was distributed between March and May 2015. Subjects were informed that they participated in a survey that aimed to assess the value of disease burden of patients with a chronic condition, and were not aware of the specific inflammatory rheumatic disease that was considered. IPSOS complies to the International Code on Market, Opinion and Social Research and Data Analytics (ICC/ESOMAR code)[17] on protection of personal data, but no ethics' approval was required for this general population survey.

### Sociodemographic background and health

Subjects provided sociodemographic characteristics including age, gender, region of residence, highest achieved educational level (university/non-university higher professional education, secondary school/low and middle professional education, or primary school) and work status (ie, having paid work or not). The presence of comorbidities was assessed with the Self-Administered Comorbidity Questionnaire modified for patients with SpA.[18] The EuroQoL Visual Analogue Scale (EQ-VAS) was used to measure overall health,[19 20] as we wanted to be informed about the self-perceived health (and not the societal utility valuation of reported health, as represented by the 5 domains of the EuroQoL (EQ-5D)). Finally, subjects completed the ASAS HI[3 14] to assess presence of impairments typical and relevant for patients with SpA (0–17, 17=most/worst influence on functioning and health).

### Choice experiments

The survey included two experiments. First, a best-worst scaling (BWS) experiment estimated the relative importance of each item of the ASAS HI on a latent scale of functioning and health (preference-based valuation). The BWS experiment consists of 17 choice tasks, each with 4 randomly selected aspects of the 17 items contained in the ASAS HI (see example in the online supplementary figure 1). One additional choice task was added to check for reliability. In each choice task, subjects had to indicate which of the four presented items they considered

Table 1 Demographics of the included subjects for the total group as well as for the individual countries

| | Overall (N=3039) | UK (N=504) | France (N=501) | Germany (N=506) | The Netherlands (N=507) | Spain (N=510) | Italy (N=511) |
|---|---|---|---|---|---|---|---|
| Women | 1556 (51.2%) | 259 (51.4%) | 261 (52.1%) | 257 (50.8%) | 260 (51.3%) | 253 (49.6%) | 266 (52.1%) |
| Age (years) | 46.5 (15.2) | 45.5 (15.8) | 47.1 (15.7) | 47.3 (15.0) | 46.7 (15.9) | 45.2 (14.6) | 47.3 (14.4) |
| Paid work in subjects ≤65 years | 1558 (57.6%) | 276 (62.3%) | 250 (57.9%) | 266 (58.0%) | 244 (55.3%) | 258 (54.9%) | 264 (57.5%) |
| Higher educational level * | 1241 (40.8%) | 249 (58.3%) | 250 (49.9%) | 119 (23.5%) | 178 (35.1%) | 209 (41.0%) | 191 (37.4%) |
| At least one morbidity † | 2104 (69.2%) | 354 (70.2%) | 336 (67.1%) | 373 (73.7%) | 345 (68.0%) | 348 (68.2%) | 348 (68.1%) |
| EQ-VAS (0–100) | 76.0 (20.0) | 74.5 (20.8) | 75.2 (19.5) | 74.5 (21.4) | 76.3 (20.1) | 76.1 (19.8) | 79.1 (18.1) |
| ASAS HI (0–17) | 4.4 (4.2) | 4.8 (4.6) | 4.6 (4.1) | 4.3 (4.1) | 4.3 (4.0) | 4.1 (4.0) | 4.5 (4.1) |

Mean (SD) or number of patients (%).
*University or non-university higher professional educational level.
†At least one condition measured with the self-administered comorbidity questionnaire modified for patients with spondyloarthritis (SpA).
ASAS HI, ASAS Health index; EQ-VAS, EuroQoL Visual Analogue Scale.

the most important and least important for overall functioning and health. Two different questionnaire sets were generated by the Sawtooth Software using an efficient approach taking into account balanced frequency (each item appeared an equal number of times), orthogonality (each item is paired an equal number of times with each other's attributes) and positional balance (each item appeared in each position across the choice sets, for example, first, second, third and fourth positions). Within each country, the two versions were randomly assigned to the subjects, but further equal distribution among age categories, gender and regional representation was not specifically accounted for.

Second, two lead time trade-off (LTTO) experiments were performed to anchor two health states of the ASAS HI on the 0 to 1 health utility scale. The first LTTO experiment (LTTO-1) aimed to reflect 'severe SpA' and the second (LTTO-2) a 'common health' of a person without SpA. In each LTTO experiment, subjects had to choose repeatedly whether they preferred life A, life B or were indifferent. Life A was characterised by 10 years in full health (lead time) followed by 10 years either in a severe ('bad') health state (LTTO-1), or 10 years in their own health (LTTO-2). Life B was also characterised by 10 years in full health (lead time), but followed by immediate death in the first choice task, but by an increasing number of life years in full health (max 10 years) in the following choice tasks. An example of the LTTO and the LTTO protocol is presented in online supplementary figure 2. The point of indifference or infliction between life A and life B is used to derive the person's preference valuation or utility for health state in life A. It reflects the number of healthy years (out of 10 years life expectancy) the respondent is willing to give up, to avoid living 10 years in the health state described in life A.[21] The 'severe health state' in LTTO-1 was defined as impairments in the eight aspects of health of the ASAS HI that mattered most to patients, as revealed in a preceding study: *pain, sleeping, being exhausted, standing, motivation to do anything that requires physical effort, getting frustrated, problems with running and restrictions in travelling*.[22] Data of subjects that indicated no problems at all on the eight aspects of ASAS HI were used to assess the utility value of the participants' 'best health state' in LTTO-2.

After finishing the experiments, subjects were also asked to rate each experiment's difficulty (Numeric Rating Scale (NRS) 0–10, 10=extremely difficult)

## Analyses

The BWS experiment was analysed using the Sawtooth Software. Based on the raw coefficients of the preference function, the relative importance scores of each aspect of health were estimated using the hierarchical Bayes method. They represent the probability that a subject would choose the specific aspect as most important. Overall, the relative importance scores sum to 100. If all aspects of health would be equally important, each aspect's relative importance score would be 5.88 (100/17). Fit statistics was used to identify unreliable responders, which should be higher than the chance that a patient gives random answers (>0.25).[23] Reliability was also assessed by the number of patients that indicated the same 'most/least' important aspect as in the identical choice task.

To obtain health utility values on the QALY Scale for 'severe SpA' and 'best health', results of LTTO-1 and LTTO-2 were used, respectively. A person's health utility was retrieved by identifying the point in which life B (giving up life years) was preferred, or in which the subject became indifferent for life A or life B. The health utility was than calculated as the number of years willing

**Table 2** Mean relative importance of the items of ASAS HI

| | UK | France | Germany | The Netherlands | Spain | Italy | Overall |
|---|---|---|---|---|---|---|---|
| **Item** | **Mean (95% CI)** | **Mean (95% CI)** | **Mean (95% CI)** | **Mean (95% CI)** | **Mean (95% CI)** | **Mean (95% CI)** | **Mean (95% CI)** |
| Pain (Q1) | 9.74 (9.35 to 10.13) | 9.9 (9.49 to 10.30) | 9.02 (8.58 to 9.45) | 11.10 (10.7 to 11.51) | 10.52 (10.09 to 10.96) | 9.33 (8.93 to 9.74) | 9.97 (9.79 to 10.14) |
| Sleeping (Q16) | 9.08 (8.63 to 9.54) | 10.49 (9.98 to 11.00) | 8.84 (8.38 to 9.3) | 10.48 (10.05 to 10.91) | 8.82 (8.35 to 9.3) | 9.39 (8.95 to 9.83) | 9.50 (9.31 to 9.69) |
| Being exhausted (Q5) | 8.02 (7.64 to 8.39) | 8.61 (8.21 to 9.00) | 7.10 (6.74 to 7.47) | 9.01 (8.67 to 9.35) | 7.40 (7.03 to 7.78) | 8.17 (7.80 to 8.53) | 8.03 (7.88 to 8.19) |
| Overcome difficulties (Q17) | 7.82 (7.43 to 8.21) | 6.86 (6.54 to 7.17) | 7.96 (7.55 to 8.37) | 7.02 (6.65 to 7.40) | 8.30 (7.97 to 8.64) | 7.11 (6.76 to 7.47) | 7.55 (7.39 to 7.70) |
| Concentration (Q11) | 6.11 (5.76 to 6.46) | 6.74 (6.35 to 7.13) | 7.47 (7.1 to 7.85) | 7.05 (6.72 to 7.38) | 6.02 (5.66 to 6.37) | 7.21 (6.85 to 7.57) | 6.78 (6.63 to 6.93) |
| Standing (Q2) | 5.43 (5.09 to 5.77) | 7.34 (6.94 to 7.73) | 6.17 (5.77 to 6.57) | 5.91 (5.51 to 6.32) | 6.64 (6.24 to 7.04) | 7.14 (6.76 to 7.52) | 6.45 (6.29 to 6.61) |
| Financial changes (Q15) | 7.02 (6.58 to 7.46) | 5.29 (4.92 to 5.65) | 7.60 (7.11 to 8.08) | 6.15 (5.76 to 6.55) | 6.61 (6.20 to 7.03) | 5.88 (5.50 to 6.25) | 6.42 (6.25 to 6.59) |
| Toileting (Q4) | 7.18 (6.7 to 7.66) | 4.8 (4.45 to 5.14) | 6.75 (6.32 to 7.17) | 4.99 (4.67 to 5.31) | 4.40 (4.09 to 4.70) | 5.59 (5.24 to 5.94) | 5.64 (5.48 to 5.80) |
| Frustrated (Q13) | 4.51 (4.15 to 4.87) | 4.16 (3.83 to 4.48) | 5.13 (4.76 to 5.49) | 5.72 (5.37 to 6.08) | 6.44 (6.05 to 6.83) | 7.02 (6.62 to 7.43) | 5.48 (5.33 to 5.64) |
| Walking outdoors (Q10) | 6.51 (6.12 to 6.89) | 5.82 (5.45 to 6.18) | 4.93 (4.61 to 5.25) | 4.39 (4.09 to 4.69) | 5.61 (5.25 to 5.96) | 5.53 (5.18 to 5.88) | 5.45 (5.30 to 5.59) |
| Motivation for physical effort (Q6) | 5.48 (5.11 to 5.85) | 5.98 (5.6 to 6.35) | 4.82 (4.46 to 5.17) | 5.74 (5.35 to 6.12) | 5.5 (5.13 to 5.88) | 4.53 (4.17 to 4.88) | 5.31 (5.16 to 5.47) |
| Sexual relationship (Q7) | 4.58 (4.12 to 5.04) | 5.07 (4.65 to 5.49) | 4.97 (4.52 to 5.42) | 4.47 (4.04 to 4.91) | 5.57 (5.11 to 6.04) | 5.13 (4.71 to 5.55) | 4.97 (4.79 to 5.15) |
| Travelling (Q12) | 5.88 (5.51 to 6.26) | 4.34 (4.03 to 4.66) | 3.88 (3.53 to 4.23) | 4.1 (3.77 to 4.43) | 4.29 (3.96 to 4.62) | 4.15 (3.88 to 4.42) | 4.42 (4.29 to 4.56) |
| Contact with people (Q9) | 4.38 (4.03 to 4.73) | 3.78 (3.48 to 4.09) | 4.71 (4.37 to 5.04) | 4.65 (4.35 to 4.95) | 4.47 (4.13 to 4.8) | 3.01 (2.75 to 3.28) | 4.16 (4.02 to 4.29) |
| Driving (Q8) | 3.72 (3.36 to 4.08) | 4.21 (3.9 to 4.52) | 4.10 (3.72 to 4.47) | 3.55 (3.24 to 3.86) | 3.71 (3.39 to 4.03) | 4.35 (4.03 to 4.67) | 3.94 (3.80 to 4.08) |
| Running (Q3) | 2.65 (2.29 to 3.00) | 4.19 (3.8 to 4.58) | 4.08 (3.71 to 4.46) | 3.41 (3.00 to 3.82) | 3.60 (3.26 to 3.95) | 3.93 (3.59 to 4.27) | 3.65 (3.50 to 3.80) |
| Washing hair (Q14) | 1.91 (1.65 to 2.16) | 2.45 (2.21 to 2.69) | 2.48 (2.22 to 2.75) | 2.25 (2.02 to 2.49) | 2.09 (1.86 to 2.32) | 2.53 (2.28 to 2.78) | 2.28 (2.18 to 2.39) |

A best-worst scaling was conducted to retrieve the mean relative importance of every item of the ASAS HI, rescaled to 0–100. In this table the results of all the participating subjects are presented.
ASAS HI, Assessment of SpondyloArthritis international Society Health Index.

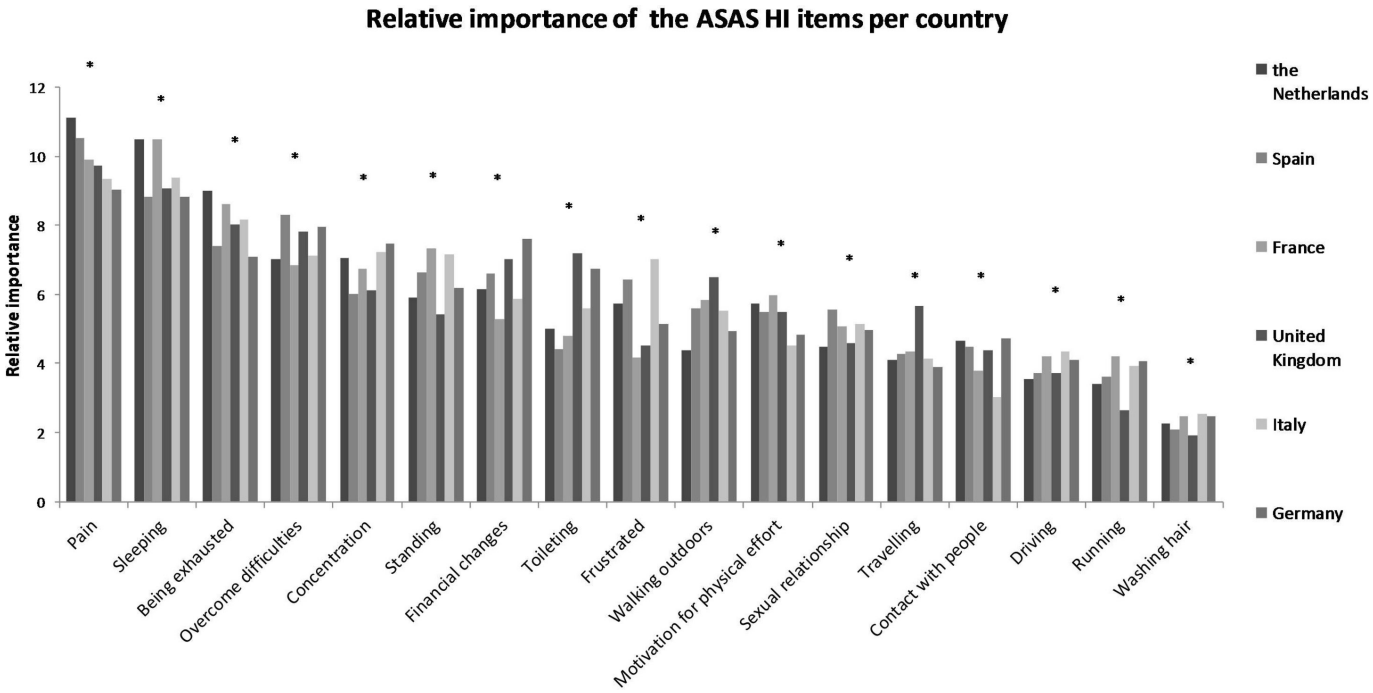


**Figure 1** Relative importance of the ASAS HI items according to the general population of six European countries *Statistically significant difference (p<0.003). A one-way analysis of variance or Kruskal-Wallis test was used, as appropriate, to compare the mean relative importance of the separate items of the ASAS HI according to country. ASAS HI, Assessment of SpondyloArthritis international Society Health Index.

to give up (life B) divided by 10 (online supplementary figure 3).

Finally, to develop the algorithm to convert the ASAS HI into a utility value, the relative importance scores of the 17 items of the ASAS HI as derived by BWS were anchored on the health utility scale based on the LTTO absolute values for 'severe SpA' and 'best health' on the absolute health utility scale.

All steps described above were performed in the total sample and repeated in the country-specific subsamples. Country-specific values (of relative importance, anchors or utility) were compared using the one-way analysis of variance (ANOVA) test or Kruskall-Wallis test, using SPSS V.20.0. If adjustment for multiple testing was necessary, a Bonferroni correction was used.

### Application

To apply the algorithm in patients with axSpA, data from an international worldwide survey that comprised the ASAS HI were used.[13] First, the distribution of utility values derived for the ASAS HI was explored (mean, SD, median, minimum, maximum, quartiles). Discriminative ability between patients with high or low disease activity (Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)≤4 or BASDAI>4) or SpA phenotype (axSpA vs non-radiographic axSpA (nr-axSpA)) was explored using one-way ANOVA, with a significance level of 0.05.

All statistical analyses were performed using SPSS V.20.0.

## RESULTS

In total, 3039 subjects were included. The mean age was 46.5 years (SD 15.2), and 1556 were women (51.2%) (table 1). Depression (18.6%), hypertension (24.0%), and pain in joints other than the back (25.1%) were the three most frequently reported health problems. EQ-VAS was 76.0 (SD 20.0) and the mean ASAS HI was 4.4 (SD 4.2). These results were comparable across countries.

### Development

#### Relative importance of the different aspects derived from ASAS HI

Of the 17 aspects of health typical and important to patients with SpA, the general population considered that pain and sleep were the most important impairments, followed by being exhausted, overcoming difficulties, concentrating, standing and experiencing financial changes (table 2). There was no aspect that was not considered important. The relative importance differed significantly for several aspects of health across countries (p<0.003) (table 2 and figure 1).

Subjects rated the difficulty of the BWS experiment as 3.7 (SD 2.8). The fit statistic was good (0.46). In the reliability test, 2102 (69.2%) and 1858 (61.1%) of the subjects indicated the same most important and least important items, respectively.

#### Anchor values for 'best health' and 'severe SpA'

Of the 3039 participants, 715 (23%) indicated to prefer 'immediate death' when valuing their own health state. However, their ASAS HI, EQ-VAS and the EQ-5D utility were actually moderate to good and fully comparable to

the EQ-VAS and EQ-5D of persons with similar good score on the ASAS HI but who did actually trade-off between immediate death or living longer. As this seemed implausible, these persons must have misunderstood the experiment and were excluded. Of note, excluded persons had attained less frequently a university or non-university higher professional degree (35.2% vs 42.6%), and had less frequently paid work (48.5% vs 53.0%).

After excluding the 715 (23%) persons with implausible data, the value for 'best health' in 462 subjects without a problem in any of the most relevant aspects of health of the ASAS HI was 0.88 (SD 0.18). The value for 'severe SpA' in 2.324 persons of was 0.29 (SD 0.34). Differences among the countries tested were statistically significant (table 3). Difficulty of the LTTO experiment was rated as 3.4 (SD 2.9).

### Scaling the relative importance of the attributes using anchors

By scaling the relative importance of the 17 items of ASAS HI between the absolute 'anchors' of the health utility scale, the algorithms of the general and six country-specific utility valuations can be easily computed (table 4). The general population assigns a utility value of −0.24 (country range −0.35 to 0.03) to a hypothetical patient with problems on all aspects of ASAS HI. As an example, a patient with 'mild' SpA who would have problems with pain, sleeping and standing, but no limitations in any other item of ASAS HI, would be assigned a utility value of 0.59 (country range 0.54 to 0.64).

## Application of the algorithm

The mean age of 199 patients with axSpA was 42.3 years (SD 13.6), mean disease duration 11.1 years (SD 11.2) and 117 (58.8%) of the patients were male. AS was present in 130 (65.3%) patients and 67 (33.7%) had nr-axSpA. The mean BASDAI was 3.8 (SD 2.3) and the mean ASAS HI 7.2 (SD 4.5). Figure 2 indicates a normal distribution of the utility values with a mean of 0.36 (SD 0.30, range −0.24 to 0.88). The average utility value was significantly higher in patients with BASDAI <4 (0.67 (SD 0.15) compared with patients with BASDAI ≥4 (0.49, (SD 0.14), p<0.01). Average utility values were not significantly different between patients with AS (0.37 (SD 0.31)) and nr-axSpA (0.34 (SD 0.29), p=0.56).

## DISCUSSION

This study provides one general and six country-specific algorithms to convert patients' scores on ASAS HI into societal health utility values. The theoretical range of the obtained health utilities varied from −0.24, for a patient with worst SpA, to 0.88, representing best health. When applying the algorithm in patients with axSpA, mean utility was 0.36 and the full range of the theoretical scale was represented. Comparable utility values were obtained for patients with AS and nr-axSpA indicating the algorithm can be applied in both phenotypes. Importantly, a good discriminative ability between patients with low and high disease activity was observed.

**Table 3** Highest and lowest anchors of the utility scores based on lead time trade-off (LTTO)

| Anchor | UK | France | Germany | The Netherlands | Spain | Italy | Overall | P values |
|---|---|---|---|---|---|---|---|---|
| Value for 'health without SpA' (ASAS HI=0) | 0.88 (SD 0.16) | 0.87 (SD 0.17) | 0.81 (SD 0.21) | 0.90 (SD 0.15) | 0.91 (SD 0.13) | 0.90 (SD 0.14) | 0.88 (SD 0.16) | <0.01 |
| Value for 'Severe SpA' | 0.30 (SD 0.34) | 0.41 (SD 0.36) | 0.26 (SD 0.33) | 0.24 (SD 0.32) | 0.27 (SD 0.34) | 0.23 (SD 0.32) | 0.29 (SD 0.34) | <0.01 |

For health without SpA, the following number of subjects contributed to the lowest anchor: all subjects N=462, UK N=84; France N=70; Germany N=75; the Netherlands N=90; Spain N=79; and Italy N=70. For 'severe SpA', the following number of subjects contributed to the lowest anchor: all subjects N=1289; UK N=241; France N=287; Germany N=182; the Netherlands N=199; Spain N=197; and Italy N=183. A one-way analysis of variance was performed to compare the highest and lower anchors across countries.
ASAS HI, Assessment of SpondyloArthritis international Society Health Index; SpA, spondyloarthritis.

**Table 4** Country-specific and overall utility algorithm for patients with SpA based on ASAS HI

| | Contribution in the utility algorithm | | | | | | |
|---|---|---|---|---|---|---|---|
| **Item** | **UK** | **France** | **Germany** | **The Netherlands** | **Spain** | **Italy** | **Overall** |
| Best health (ASAS HI=0) | 0.88 | 0.87 | 0.81 | 0.90 | 0.91 | 0.90 | 0.88 |
| *Pain (Q1)* | −0.11 | −0.08 | −0.10 | −0.13 | −0.13 | −0.12 | −0.11 |
| *Sleeping (Q16)* | −0.10 | −0.09 | −0.10 | −0.12 | −0.11 | −0.12 | −0.11 |
| *Being exhausted (Q5)* | −0.09 | −0.07 | −0.08 | −0.11 | −0.09 | −0.07 | −0.09 |
| *Overcome difficulties (Q17)* | −0.09 | −0.06 | −0.09 | −0.08 | −0.10 | −0.09 | −0.08 |
| *Concentration (Q11)* | −0.07 | −0.06 | −0.08 | −0.08 | −0.07 | −0.09 | −0.08 |
| *Standing (Q2)* | −0.06 | −0.06 | −0.07 | −0.07 | −0.08 | −0.09 | −0.07 |
| *Financial changes (Q15)* | −0.08 | −0.04 | −0.09 | −0.07 | −0.08 | −0.07 | −0.07 |
| *Toileting (Q4)* | −0.08 | −0.04 | −0.08 | −0.06 | −0.05 | −0.07 | −0.06 |
| *Frustrated (Q13)* | −0.05 | −0.03 | −0.06 | −0.07 | −0.08 | −0.09 | −0.06 |
| *Walking outdoors (Q10)* | −0.07 | −0.05 | −0.06 | −0.05 | −0.07 | −0.07 | −0.06 |
| *Motivation for physical effort (Q6)* | −0.06 | −0.05 | −0.05 | −0.07 | −0.07 | −0.06 | −0.06 |
| *Sexual relationship (Q7)* | −0.05 | −0.04 | −0.06 | −0.05 | −0.07 | −0.06 | −0.06 |
| *Travelling (Q12)* | −0.07 | −0.04 | −0.04 | −0.05 | −0.05 | −0.05 | −0.05 |
| *Contact with people (Q9)* | −0.05 | −0.03 | −0.05 | −0.07 | −0.05 | −0.04 | −0.05 |
| *Driving (Q8)* | −0.04 | −0.04 | −0.05 | −0.04 | −0.04 | −0.05 | −0.04 |
| *Running (Q3)* | −0.03 | −0.04 | −0.05 | −0.04 | −0.04 | −0.05 | −0.04 |
| *Washing hair (Q14)* | −0.02 | −0.02 | −0.03 | −0.03 | −0.08 | −0.03 | −0.03 |
| Worst SpA | −0.26 | 0.03 | −0.31 | −0.29 | −0.29 | −0.35 | −0.24 |
| Example for mild SpA* | 0.61 | 0.64 | 0.54 | 0.58 | 0.59 | 0.57 | 0.59 |

*Italic*: Aspects (n=8) included in the health state 'severe SpA', which were defined in a pilot study in patients with axial SpA.
To calculate the utilities for a specific health state, the intercept, for a person with no complaints (ASAS HI=0), should be subtracted with the value or the aspects of health with problems. The example (*) illustrates the utility values for a patient with problems of pain, sleeping and standing.
ASAS HI, Assessment of SpondyloArthritis international Society Health Index; SpA, spondyloarthritis.

The availability of ASAS HI was essential for the development of the disease-specific utility algorithm, as description of health states 'typical and relevant for SpA' is a core aspect of a preference-based valuation. Of note, ASAS HI is grounded in the universally endorsed International Classification of Functioning and Health, and accounted in each step of the development from the perspective of the patients, ensuring all aspects of health important for patients with SpA are represented.[3 14]

BWS and LTTO were combined to compute the conversion algorithm. BWS is an elegant approach to understand the relative importance of a large number of aspects of health in valuations of health state.[24] On that line, all 17 items of the ASAS HI could be considered. All 17 items were also shown to be relevant when valuing health and remained included in the utility algorithm. Further advantages of BWS are the ease of completion for respondents, absence of scale bias and effective discrimination between items. Recently, BWS was also used to value the child health utility, 9D.[25] However, BWS provides the relative importance of the items on a latent scale, while the utility requires values that are expressed on an absolute scale in which 'zero' presents death and 'one' full health. Although a subject of discussion, conventional standard gamble and TTO are still the preferred options to define absolute anchors of the utility scale, as they remain consistent with the utility theory.[26–29] In LTTO, trading between quality and quantity of life is postponed to a later period in life (after a lead time of 10 years in the current study), as this seems more realistic for persons with chronic diseases that are less likely to take immediate life risks. Notwithstanding, in the current online survey, a considerable number of subjects seemed not to have understood the experiment, as they chose to die immediately after the lead time, despite the absence of severe health impairments. Although, a previous study revealed that interview-based TTO provided overall similar results as an online unassisted experiment,[30] we cannot exclude that LTTO is more challenging in an unassisted online environment. Of interest, persons with implausible choices had lower level of education in our study. Notwithstanding, by excluding subjects that chose immediate death despite 'good health', we cannot exclude the fact that some persons made their choice consciously and therefore our anchors are overestimated. However, during cognitive debriefing interviews among

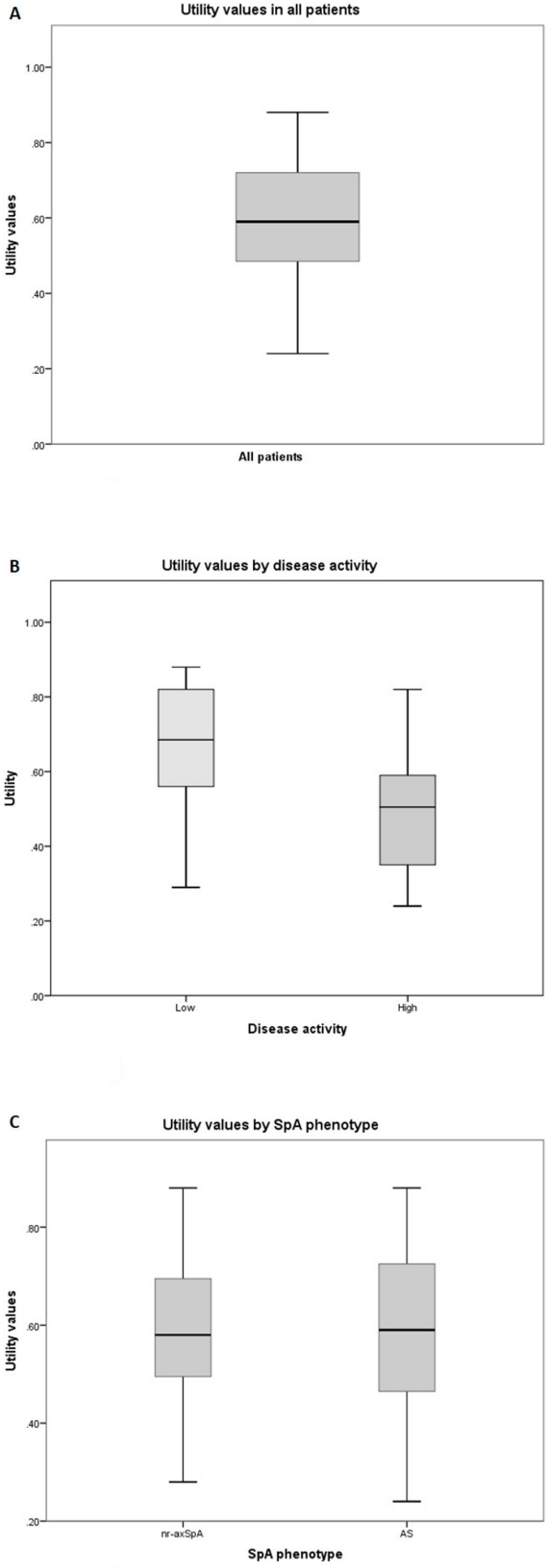


**Figure 2** Distribution of the utility values in patients with axial spondyloarthritis (axSpA) using the overall utility algorithm. (A) Distribution of utility values in patients with axSpA. (B) Distribution of utility values by low (BASDAI <4) or high (BASDAI ≥4) disease activity. (C) Distribution of utility values by SpA phenotype. AS, akylosing spondylitis; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index.

12 persons without chronic disease and no impairments on the most important items of ASAS HI (50% female, average age 47 years (range 21years to 74 years)), the interviewer indicated that about 40% of subjects needed some support (sometimes a simple confirmation they were thinking correctly) and none of them chose to die immediately when valuing own health, providing support to the fact that our exclusion criteria was valid. Notwithstanding, (L)-TTO remains a challenging experiment, in which training of investigators and subjects is necessary to perform the experiment accurately. Cognitive debriefing seems insufficient as a precaution and a pilot study before the actual experiment might have pinpointed the misunderstanding.

The main interest of this study can be found in the availability of an algorithm that enables estimation of societal health utilities and QALYs specific for patients with SpA. Generic utilities, such as EQ-5D and SF-6D, might not capture all the aspects of health important for patients with SpA.[31] Generic instruments might therefore discriminate less between relevant subgroups and be less sensitive to changes over time.[31] Examples of other disease-specific utility measures are the preference-based index for visual function and the utility valuation based on the European Organisation for Research and Treatment of Cancer Core Quality of Life Questionnaire.[32 33] Although, there might be doubts about the comparability of disease-specific utility values across disease, it can also be argued that the key towards comparability is found in a common methodological approach to (A) Select aspects of health that are most typical and relevant. (B) Value their relative importance. (C) Scale these aspects between absolute anchors on the health utility scale. Notwithstanding, the main application of disease-specific utilities would serve decisions within one disease. Further research is now required to compare the ASAS HI utility with generic utility measures regarding construct validity, sensitivity to change and influence on cost-effectiveness.

Our study is one of the first to study cultural differences in preferences for health using the same design and sampling across countries. We now provide algorithms on the societal perspective of six western European countries, and hopefully other countries are encouraged to use our methodology for new country-specific algorithms. We confirm differences in preference for health are related to the relative importance of the aspects of health, as well as to the absolute values for severe and best health on the 'zero to one' utility scale. It seems that these differences did not have a major influence on the value of mild health states, but impact on sensitivity to change and cost-effectiveness needs to be still explored. For the time being, we feel the country-specific utility algorithm should be used when available; otherwise the general algorithm seems appropriate.

In conclusion, one generic and six country-specific algorithms are now available to convert scores on ASAS HI into a utility of the societal perspective. This makes it possible to use disease-specific utilities and QALYs in

decision-making processes when comparing treatment strategies among patients with SpA.

**Author affiliations**
[1]Rheumatology, Maastricht University Medical Center, Maastricht, , The Netherlands
[2]Department of Health Services Research, Maastricht University, Maastricht, , The Netherlands
[3]Rheumazentrum Ruhrgebiet, Ruhr-University Bochum, Herne, Germany
[4]School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada
[5]Rheumazentrum Ruhrgebiet, Herne, Germany
[6]Rheumatology, LUMC, Leiden, The Netherlands
[7]Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center, and the Caphri Research Institute Maastricht University, Maastricht, The Netherlands

**Acknowledgements** The authors thank Sandrine Bours (French), Victoria Navarro (Spanish) and Michelle Gillio (Italian) for their help with the translation of the instructions for the online survey.

**Contributors** IE and AB contributed to the development of the study, statiscal analyses and writing proces. MH contributed to the development of the study and statistical analyses. All other coauthors contributed to the development of the study and reviewed the manuscript.

**Funding** An unconditional ASAS grant was received to conduct this study.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request.



## REFERENCES

1. van der Heijde, Ramiro L*et al*. Update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis* 2016;2017:978–91.
2. Doward LC, Spoorenberg A, Cook SA, *et al*. Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis. *Ann Rheum Dis* 2003;62:20–6.
3. Kiltz U, van der Heijde D, Boonen A. The ASAS Health Index (ASAS HI) - a new tool to assess the health status of patients with spondyloarthritis. *Clin Exp Rheumatol* 2014;32:S- 105–8.
4. Porter ME. Value-based health care delivery. *Ann Surg* 2008;248:144–50.
5. Neumann V. *Theory of games and economic behavior*. New York: Wiley, 1953.
6. Porter ME. A strategy for health care reform-toward a value-based system. *N Engl J Med* 2009;361:109–12.
7. Jönsson B. Ten arguments for a societal perspective in the economic evaluation of medical innovations. *Eur J Health Econ* 2009;10:357–9.
8. Ubel PA, Richardson J, Menzel P. Societal value, the person trade-off, and the dilemma of whose values to measure for cost-effectiveness analysis. *Health Econ* 2000;9:127–36.
9. Drummond M, Barbieri M, Cook J, *et al*. Transferability of economic evaluations across jurisdictions: ISPOR good research practices Task Force report. *Value Health* 2009;12:409–18.
10. Dolan P. Modeling Valuations for EuroQol health states. *Medical Care* 1997;35:1095–108.
11. Kharroubi SA, Brazier JE, Roberts J, *et al*. Modelling SF-6D health state preference data using a nonparametric Bayesian method. *J Health Econ* 2007;26:597–612.
12. Bansback N, Harrison M, Brazier J, *et al*. Health state utility values: a description of their development and application for rheumatic diseases. *Arthritis Rheum* 2008;59:1018–26.
13. Petrillo J, Cairns J. Converting condition-specific measures into preference-based outcomes for use in economic evaluation. *Expert Rev Pharmacoecon Outcomes Res* 2008;8:453–61.
14. 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.
15. 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.
16. IPSOS. IPSOS-nederland. Available: http://www.ipsos-nederland.nl/
17. ICC/ESOMAR. ICC/ESOMAR contents International Code on market, opinion and social research and data analytics, 2016. Available: https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ICCESOMAR_Code_English_.pdf [Accessed 12 May 2018].
18. Stolwijk C, van Tubergen A, Ramiro S, *et al*. Aspects of validity of the self-administered comorbidity questionnaire in patients with ankylosing spondylitis. *Rheumatology* 2014;53:1054–64.
19. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
20. Herdman M, Gudex C, Lloyd A, *et al*. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–36.
21. Augustovski F, Rey-Ares L, Irazola V, *et al*. Lead versus lag-time trade-off variants: does it make any difference? *Eur J Health Econ* 2013;14:25–31.
22. Kiltz U, Essers I, Hiligsmann M, *et al*. Which aspects of health are most important for patients with spondyloarthritis? a best worst scaling based on the ASAS health index. *Rheumatology* 2016;55:1771–6.
23. Sawtooth Software. Identifying 'bad' respondents: Fit Statistic and Identifying Random Responders, 2015. Available: https://www.sawtoothsoftware.com/help/issues/ssiweb/online_help/hid_web_maxdiff_badrespondents.htm [Accessed 25 Aug 2018].
24. Cheung KL, Wijnen BFM, Hollin IL, *et al*. Using Best-Worst scaling to investigate preferences in health care. *Pharmacoeconomics* 2016;34:1195–209.
25. Ratcliffe J, Huynh E, Chen G, *et al*. Valuing the child health utility 9D: using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. *Soc Sci Med* 2016;157:48–59.
26. Attema AE, Edelaar-Peeters Y, Versteegh MM, *et al*. Time trade-off: one methodology, different methods. *Eur J Health Econ* 2013;14:53–64.
27. Devlin NJ, Tsuchiya A, Buckingham K, *et al*. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. *Health Econ* 2011;20:348–61.
28. Gafni. The standard gamble method: what is being measured and how it is interpreted. *Health Serv Res* 1994;29:207–24.
29. van Osch SMC, Wakker PP, van den Hout WB, *et al*. Correcting biases in standard gamble and time tradeoff utilities. *Med Decis Making* 2004;24:511–7.
30. Buitinga L, Braakman-Jansen LMA, Taal E, *et al*. A computer time trade-off: a feasible and reliable alternative for the interview time trade-off in rheumatoid arthritis. *Clin Exp Rheumatol* 2011;29:783–9.
31. Lin, Longworth P. Evaluation of content on EQ-5D as compared to disease-specific utility measures. *Qual Life Res* 2013;22:853–74.
32. Rentz AM, Kowalski JW, Walt JG, *et al*. Development of a preference-based index from the National Eye Institute visual function Questionnaire-25. *JAMA Ophthalmol* 2014;132:310–8.
33. Rowen D, Brazier J, Young T, *et al*. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value Health* 2011;14:721–31.