Instrument selection for the ASAS core outcome set for axial spondyloarthritis

Victoria Navarro-Compán[1*], Anne Boel[2*], Annelies Boonen[3], Philip Mease[4], Maxime Dougados[5], Uta Kiltz[6], Robert Landewé[7], Xenofon Baraliakos[6], Wilson Bautista-Molano[8], Praveena Chiowchanwisawakit[9], Hanne Dagfinrud[10], Lara Fallon[11], Marco Garrido-Cumbrera[12], Lianne S. Gensler[13], Bassel El-Zorkany[14], Nigil Haroon[15], Yu Heng Kwan[16], Pedro Machado[17], Walter P. Maksymowych[18], Anna Molto[5], Natasha de Peyrecave[19], Denis Poddubnyy[20], Mikhail Protopopov[20], Sofia Ramiro[21], In-Ho Song[22], Salima van Weely[23], Désirée van der Heijde[2]

*Shared first-authorship

[1] Rheumatology Service, Hospital Universitario la Paz-IdiPaz, Madrid, Spain, ORCID: 0000-0002-4527-852X

[2] Department of Rheumatology, Leiden University Medical Center, Leiden, the Netherlands, (Anne Boel Orchid ID: 0000-0003-2016-1744 and Désirée van der Heijde ORCID: 0000-0002-5781-158X)

[3] Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Center, the Netherlands and Care and Public Health Research Institute (CAPHRI), Maastricht University, the Netherlands. ORCID: 0000-0003-0682-9533

[4] Division of Rheumatology, Swedish Medical Center/Providence St. Joseph Health and University of Washington, Seattle, WA, USA. ORCID: 0000-0002-6620-0457

[5] Université Paris-Cité, Department of Rheumatology - Hôpital Cochin. Assistance Publique - Hôpitaux de Paris INSERM (U1153): Clinical epidemiology and biostatistics, PRES Sorbonne Paris-Cité. Paris, France.

[6] Rheumazentrum Ruhrgebiet Herne, Ruhr-Universität Bochum, Germany. Uta Kiltz (ORCID: 0000-0001-5668-4497); Xenofon Baraliakos (ORCID: 0000-0002-9475-9362)

[7] Department of rheumatology & clinical immunology, Amsterdam University Medical Center | loc. amC, Amsterdam & Zuyderland MC |loc. Heerlen, The Netherlands. ORCID: 0000-0002-0577-6620

[8] Rheumatology Department, University Hospital Fundación Santa Fe de Bogotá and School of Medicine Universidad El Bosque. Bogotá, Colombia ORCID: 0000-0003-0684-9542

[9] Faculty of Medicine Siriraj Hospital, Mahidol University, Thailand. ORCID: 0000-0002-4253-9229

[10] Dept of Rheumatology, Diakonhjemmet Hospital, Oslo, Norway.

[11] Inflammation and Immunology – Global Medical Affairs, Pfizer Inc, Kirkland, Quebec, Canada.

[12] Health & Territory Research (HTR), Universidad de Sevilla, Seville, Spain. Axial Spondyloarthritis International Federation (ASIF), London, UK. ORCID: 0000-0001-9727-1189

[13] Division of Rheumatology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. ORCID: 0000-0001-6314-5336

[14] Rheumatology department, Cairo University. ORCID: 0000-0003-2704-9712

[15] University of Toronto, Department of Medicine, University Health Network, Schroder Artritis Institute, Toronto. ORCID: 0000-0003-3210-4771

[16] Program in Health Systems and Services Research, Duke-NUS Medical School, Department of Pharmacy, National University of Singapore, Department of Rheumatology and Immunology, Singapore General Hospital, ORCID ID: 0000-0001-7802-9696

[17] Centre for Rheumatology & Department of Neuromuscular Diseases, University College London, London, United Kingdom; National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre, University College London Hospitals NHS Foundation Trust, London, UK; Department of Rheumatology, Northwick Park Hospital, London North West University Healthcare NHS Trust, London, UK. ORCID: 0000-0002-8411-7972

[18] Department of Medicine, University of Alberta, Edmonton, Canada. ORCID: 0000-0002-1291-1755

[19] Rheumatology Global Medical Affairs, UBC Pharma, Brussels, Belgium. OCRID: 0000-0001-5300-9226

[20] Department of Gastroenterology, Infectious Diseases and Rheumatology, Charité – Universitätsmedizin Berlin, Berlin, Germany, (Denis Poddubnyy ORCID: 0000-0002-4537-6015; Mikhail Protopopov ORCID ID: 0000-0003-4840-5069)

[21] Department of Rheumatology, Leiden University Medical Center, Leiden, the Netherlands; Department of Rheumatology, Zuyderland Medical Center, Heerlen, the Netherlands, ORCID ID: 0000-0002-8899-9087

[22] AbbVie, Immunology Clinical Development, 1 North Waukegan Road Building AP31-2, North Chicago, IL 60064, USA.

[23] Department of Orthopaedics, Rehabilitation and Physical Therapy, Leiden University Medical Center, Leiden, the Netherlands. ORCID: 0000-0001-8560-4687

Corresponding autor

Victoria Navarro-Compán, Rheumatology service, Hospital Universitario la Paz-IdiPaz, Madrid, Spain. Paseo de la Castellana, 261, Madrid, 28046, Spain. mvictoria.navarroc@gmail.com

Wordcount: 5020

References:

Tables and Figures: 2 figures and 4 tables, supplementary material (5 tables) and supplementary files (1-26)

Keywords: Instrument, outcome, core outcome set, axial spondyloarthritis

ABSTRACT

Objectives: To define the instruments for the ASAS-OMERACT core domain set for axial spondyloarthritis (axSpA).

Methods: An international working group representing key stakeholders selected the core outcome instruments following a predefined process: i) Identifying candidate instruments using a systematic literature review; ii) Reducing the list of candidate instruments by the working group, iii) Assessing the instruments' psychometric properties following OMERACT Filter 2.2, iv) Selection of the core instruments by the working group; v) Voting and endorsement by ASAS.

Results: The updated core set for axSpA includes seven instruments for the domains that are mandatory for all trials: ASDAS and NRS patient global assessment of disease activity; NRS total back pain; average NRS of duration and severity of morning stiffness; NRS fatigue; BASFI; and ASAS Health Index. There are 9 additional instruments considered mandatory for disease modifying drugs (DMARDs) trials: MRI activity SPARCC sacroiliac joints and SPARCC spine, uveitis, IBD and psoriasis assessed as recommended by ASAS, 44 swollen joint count, MASES, dactylitis count, and mSASSS. The imaging outcomes are considered mandatory to be included in at least one trial for a drug tested for DMARD-properties. Furthermore, 11 additional instruments were also endorsed by ASAS, which can be used in axSpA trials on top of the core instruments.

Conclusions: The selection of the instruments for the ASAS-OMERACT core domain set completes the update of the core outcome set for axSpA, which should be used in all trials.

Background

Efficacy and safety of any therapy should be demonstrated in randomised controlled trials. Therefore, it is important that all studies assess the same outcome domains and measurement instruments to facilitate comparison of results and to ensure that all relevant endpoints are reported. The use of core outcome sets (COS), which describe the minimum set of measures that should be used in all studies, is recommended to facilitate the comparability of results on efficacy and safety of therapies. For the development of any COS there is a specific procedure, that mainly consists of two consecutive phases: to determine the core domain set (*what to measure -selection of the domains-*) and the core measurement set (*how to measure -selection of the instruments-*). In addition, it is important to update the COS as the field develops.

The Assessment of SpondyloArthritis international Society-Outcomes Measures in Rheumatology (ASAS-OMERACT) COS for ankylosing spondylitis (AS) was developed more than two decades ago,[1-4]. Given the progress made since then, both in the knowledge of the disease and in the methodology for developing a COS, ASAS decided to update the original COS for AS into a COS for axial spondyloarthritis (axSpA). As a first phase of this process the ASAS-OMERACT core domain set has recently been updated and published,[5]. It includes 7 mandatory domains for all studies and 3 additional mandatory domains for studies evaluating disease-modifying antirheumatic drugs (DMARDs). The mandatory domains for all trials are: disease activity, pain, morning stiffness, fatigue, physical functioning, overall functioning and health, and adverse events including death. As additional mandatory domains for DMARDs, extra-musculoskeletal manifestations (EMMs), peripheral manifestations and structural damage have been included.

There are specific procedures available on how to define the core measurement set, mainly those by OMERACT and Core Outcome Measures in Effectiveness Trials (COMET),[6-8]. These enable standardised data collection and objective data-driven selection of instruments. The aim of this article is to report on the outcome of the instrument selection for the updated COS for axSpA.

Methods

Working group

The axSpA working group included 28 participants representing different stakeholders (rheumatologists and other health professional experts in axSpA, patient representatives, pharmaceutical industry representatives, drug regulation officer, and methodologists). The main task of this working group was to select at least one instrument for each of the mandatory core domains included in the updated core set for axSpA,[5]. A summary of the instrument selection process is depicted in figure 1.

[FIGURE 1]

Identify candidate instruments and reduce the list

A systematic literature review (SLR) was performed to identify all instruments that have been assessed in clinical trials in axSpA. For this, the SLR performed by Bautista-Molano et al,[9] formed the basis, which was used to update the literature search up to August 2018. The results from both SLRs were combined into a list of unique candidate instruments. Following a discussion in the working group, a reduced and more feasible list of candidate items was proposed. Instruments were excluded whenever experts agreed based on their experience and knowledge of the literature and of the instruments that lacked validity or had insufficient information on truth and discrimination.

Psychometric properties assessment

In order to collect information about all psychometric properties in a standardised manner, the OMERACT guidelines as described in the OMERACT Handbook were used,[10]. The assessment of psychometric properties consists of two consecutive steps: i) Assess domain match and feasibility; ii) Assess truth and discrimination. After completing the first step, it should be decided if the evaluation of the candidate instrument should continue (figure 2).

[FIGURE 2]

In order to move forward, the instrument should achieve at least 70% agreement (either 'good to go', or 'some cautions but okay to use'). If less than that, the instrument should be excluded from further properties assessment.

Step 1: Domain match and feasibility

Domain match (content and face) validity and feasibility were assessed by all members in the working group for each of the candidate instruments using standardised questionnaires provided in the OMERACT handbook,[10] . The last question in these questionnaires was a final conclusion with three answer options: 1. the instrument was considered 'good to go'; 2. there were some cautions, but it is 'okay to use the instrument'; or 3. the instrument was 'not right' for this application. Due to the high number of instruments to assess, it was decided that each instrument would be assessed by half of the working group members, with each subgroup representing all stakeholders and at least three different geographical regions. Additionally, 8-14 patients (from Colombia, the Netherlands, Singapore, Spain, and United States) were asked to rate all patient reported outcomes (PROs) for domain match and feasibility. Furthermore, a review of raw data was performed using data gathered in two observational studies,[11 12], which provided insight in the percentage of missing data, as well as possible floor and ceiling effects for each instrument. After completion of the questionnaires and data analyses, a virtual working group meeting was organised to discuss the results and decide which instruments would be further assessed.

Step 2: Truth and discrimination

To assess construct validity, the steering committee defined hypotheses regarding the expected strength of the correlation between the assessed instrument and other instruments. Here, due to lack of evidence, we deviated from the OMERACT-procedure which requires the expected correlations to be described within the manuscript that holds the data. Instead, Spearman or Pearson correlation coefficients were extracted to describe construct validity (see supplementary table S1 for interpretation of the level of the correlation coefficients).

Test-retest reliability was assessed by intraclass correlation coefficients (ICC) for all continuous scores and by (weighted) kappa statistics for binary and ordinal scores. Furthermore, the data extracted from the articles was used to calculate three measures of longitudinal construct validity [1. Guyatt's effect size (Guyatt's ES); 2. Standardized response mean (SRM); and 3. Effect size (ES) (Supplementary tables S1 and S2)] and two measures of discrimination evaluating the ability to differentiate change in the outcome between the arms in clinical trials: 1. Standardized mean difference (SMD); and 2. Standardized mean difference of improvement (SMD_{imp}) (Supplementary tables S1 and S2). The final psychometric property to be assessed was thresholds of meaning, which includes thresholds like a minimally clinical important difference (MCID) or improvement (MCII), or a patient acceptable symptom state (PASS). These are compared to an external anchor (e.g. patient defined improvement).

Data collection

In order to ensure a standardized manner of data collection for construct validity, test-retest reliability and thresholds of meaning, we used OMERACT search strings,[10] to collect the data adapted to fit our study population (i.e. "axial spondyloarthritis" OR "ankylosing spondylitis" OR "axial SpA" OR axSpA OR AS). All search results were assessed and articles that contained data on the psychometrics of each instrument were saved. All data was extracted by a fellow (AB) using a specific extraction file developed for this purpose and then also checked by a second reviewer (VNC).

For longitudinal construct validity and clinical trial discrimination, the steering committee pre-selected seven recent placebo-controlled trials in axSpA covering the entire spectrum of the disease and different drugs: INFAST,[13], RAPID-axSpA,[14], ASCEND,[15], COAST-V,[16], SELECT-AXIS,[17], ABILITY-1,[18], and COAST-X,[19]. All manuscripts that published data on these trials were collected and all data were extracted for the calculations.

Data overview and synthesis

The OMERACT summary of measurement property (SOMP) tables,[10] were used to summarise all psychometric properties assessment results for each instrument. These tables provide an overview of all the studies that reported data on one or more of the psychometric properties. A detailed explanation of the SOMPs is provided in Supplementary table S3.

Working group proposal

The working group discussed the instrument selection per domain in a two-day virtual meeting. Several principles were applied: first, at least one suitable instrument had to be selected for each mandatory domain in the COS. Second, it was important to be selective and to create a concise list of instruments to be assessed in every trial in axSpA. It was decided upfront that the decision to include an instrument would be based on the data collected, as well as the collective experience of the working group. Therefore, an instrument could still be included in the COS, even if it was not endorsed according to the OMERACT algorithm. Furthermore, if an instrument was included in the original COS for AS, there should be convincing new scientific evidence for it to be replaced by another instrument.

A two-step approach was taken in the selection of instruments for the COS. First, the working group decided for each instrument whether it was valid to assess the corresponding domain in clinical trials and should be endorsed by ASAS. Second, -for those instruments considered valid- the working group decided on inclusion in the COS, using a parsimonious approach ensuring the final product will be feasible and implementable. All decisions were voted on by all attendees. For the instruments assessing the three additional mandatory domains for DMARDs an additional vote was performed, regarding the frequency of assessment: the instrument should be assessed in all studies or at least in one study during the drug development programme.

ASAS voting

The proposal from the working group was taken to the entire ASAS community in the 2022 annual workshop, which was held in a virtual format. Here, a summary was provided describing all the steps leading to the proposal. Thereafter, the preliminary instruments for the COS were presented and discussed per domain by ASAS members. A formal voting was performed per domain applying the same cut-offs for agreement as described in the working group voting procedure applied for acceptance of the proposal by the ASAS community.

Results

A total of 24 participants took part in the working group meetings and 107 full members were present at the ASAS meeting.

Identify candidate instruments and reduce the list

The search to update the SLR up to August 2018 retrieved 320 records (supplementary figure S2). A total of 296 records were screened (AB), 81 articles were included for data-extraction, from which 67 unique candidate instruments were preselected and reviewed by the steering committee and proposed to the working group. Instruments were taken of the list if they were considered not feasible (n=15, e.g. too time-consuming, copyright costs), their performance was proven inferior compared to other candidate instruments (n=14), or had insufficient domain match (n=7). Finally, the list was reduced to a total of 31 instruments (table 1).

Table 1 Candidate instruments to be considered for the updated COS for axial spondyloarthritis.

Disease activity (n=10)
Patient global assessment for disease activity during last week (PtGA), on a NRS using the question “How active was your rheumatic disease on average during the last week?”
Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)
Ankylosing Spondylitis Disease Activity Score (ASDAS)
C-reactive protein (CRP)
Spondyloarthritis Research Consortium of Canada (SPARCC) MRI activity of the SIJ
Spondyloarthritis Research Consortium of Canada (SPARCC) MRI activity of the spine
Berlin MRI activity of the SIJ
Berlin MRI activity of the spine
Canada-Denmark (CAN-DEN) MRI activity of the spine
Ankylosing Spondylitis spine MRI activity (ASspiMRI-a)
Pain (n=2)
Total back pain in the past week, on a NRS using question two of the BASDAI “How would you describe the overall level of neck, back or hip pain you have had in the past week?”
Back pain at night in the past week, on a NRS using the question “How much pain of your spine due to axSpA do you have at night?”
Morning stiffness (n=3)
Severity of morning stiffness measured on an NRS (BASDAI Q5)
Duration of morning stiffness measured on an NRS (BASDAI Q6)
Combined average score including severity and duration of morning stiffness measured on an NRS [(BASDAI Q5 + BASDAI Q6)/2)
Fatigue (n=2)
Fatigue as assessed by BASDAI Q1 on a NRS
Functional Assessment of Chronic Illness Therapy (FACIT)-fatigue
Physical function (n=1)
Bath Ankylosing Spondylitis Function Index (BASFI)
Overall functioning & health (n=2)
ASAS Health Index (ASAS-HI)
36-Item Short Form Health Survey (SF-36)
Extra-musculoskeletal manifestations (n=3)
ASAS CRF uveitis
ASAS CRF psoriasis
ASAS CRF inflammatory bowel disease
Peripheral manifestations (n=5)
44 swollen joint counts
66 swollen joint counts
Maastricht Ankylosing Spondylitis Enthesitis Score (MASES) enthesitis score
Spondyloarthritis Research Consortium of Canada (SPARCC) enthesitis score
Dactylitis count as recommended by ASAS

Structural damage (n=3)
Modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS)
Modified New York (mNY) sacroiliitis score
SPARCC MRI Sacroiliac joint Structural Score (SPARCC SSS) for erosion

NRS: Numerical Rate Scale; **MRI:** magnetic resonance imaging; **SIJ:** sacroiliac joints; **ASAS:** Assessment of SpondyloArthritis international Society; **Q:** question

Measurement properties assessment

Based on domain match and feasibility results the working group decided to exclude three instruments: Canada-Denmark (CAN-DEN) MRI activity of the spine,[20], Ankylosing Spondylitis spine MRI activity (ASspiMRI-a),[21] and Functional Assessment of Chronic Illness Therapy (FACIT)-fatigue,[22]) and to further assess the additional measurement properties from step 2 (truth and discrimination) in the 28 remaining candidate instruments. The results for each of the measurement property assessment are presented in detailed for all these instruments in Supplementary files 1-26). In addition, a summary overview table following the SOMPs format is included at the end of the files for all instruments. For the purpose of providing an example table 2 shows the summary table for one instrument: the Ankylosing Spondylitis Disease Activity Score (ASDAS),[23].

Voting results of the working group members for the proposal in the selection of instruments in the COS are presented in Supplementary table S4. Furthermore, the final voting results at ASAS 2022 annual workshop are displayed in Supplementary table S5.

Table 2: Example of an OMERACT summary of measurement properties (SOMPs) table for one of the instruments: the Ankylosing Spondylitis Disease Activity Score. The table provides an overview of the studies that reported data on one or more of the psychometric properties.

Instrument: ASDAS **Domain**: Disease activity								**Date completed**: 13 Dec 2021
Population: Axial spondyloarthritis	**Intervention**: drugs		**Control**: placebo/drug		**Type of studies**: clinical trials			
		Truth		**Truth**	**Discrimination**			
Author/year	**Study population**	**Domain match**	**Feasibility**	**Construct validity**	**Test-retest reliability**	**Longitudinal construct validity**	**Clinical trial discrimination**	**Threshold of meaning**
Working Group Appraisal (n=29 including 2 PRPs)		🟩	🟩	▒	▒	▒	▒	▒
Lukas et al. (2009)[1]	r-axSpA			🟩				
van der Heijde et al. (2009)[2]	r-axSpA			🟩				
El Miedany et al. (2010)[3]	r-axSpA			🟧	▤			
Pedersen et al. (2010)[4]	all axSpA			🟧				
van Tubergen et al. (2015)[5]	all axSpA			🟩				
Kiltz et al. (2018)[6]	all axSpA			🟩				
Lopez-Medina et al. (2018)[7]	all axSpA			🟧				
Kwan et al. (2019)[8]	all axSpA			🟩				
Hoepken et al. (2021)[9]	all axSpA			🟩				
Boel et al. (2021)[10] *COAST*	all axSpA				🟩			
Boel et al. (2021)[10] *RAPID-axSpA*	all axSpA				🟩			
van der Heijde et al. (2012)[12]	r-axSpA					🟩	🟩	
Sieper et al. (2013)[15]	nr-axSpA					▤		
Landewé et al. (2014)[11]	all axSpA					🟩	🟩	
van der Heijde et al. (2018)[13]	r-axSpA					🟩	🟩	
Van der Heijde et al. (2019)[14]	r-axSpA					🟩	🟩	
Deodhar et al. (2020)[16]	nr-axSpA					🟩	🟧	
Machado et al. (2011)[17]	r-axSpA					🟧		🟩
Machado et al. (2018)[18]	axSpA							🟩
Molto et al. (2018)[19]	all axSpA							🟩
Total available studies for each property				**9**	**3**	**7**	**6**	**3**

Total studies available for synthesis				**9**	**3**	**7**	**6**	**3**
Synthesis Rating		**GREEN from WG**	**GREEN from WG**	**GREEN**	**GREEN**	**GREEN**	**GREEN**	**GREEN**
OMERACT Endorsement	Based on the OMERACT algorithm this instrument is: Endorsed *More research could be performed to strengthen test-retest reliability of the ASDAS, especially in the nr-axSpA subgroup*							

r-axSpA: radiographic axial spondyloarthritis; **axSpA:** axial spondyloarthritis; **nr-axSpA: non radiographic axial spondyloarthritis; WG:** working group.

Note: SOMP table also includes a synthesis rating per psychometric property. Again, a colour system is used to visualise the conclusion for each measurement property based on the collected data:

- Green: adequate or good performance of the psychometric property, at least two good quality papers showing consistent results.
- Red: inadequate performance of the psychometric property, only studies of poor quality.
- White: no evidence found for this psychometric property.
- Amber: all other instances (e.g. inconsistent results in good quality papers, only moderate quality papers with consistent results, only one paper was available, which was a good quality paper).

In order to get a full OMERACT endorsement (green), all psychometric properties had to have a green synthesis rating. If there is a mix of green and amber in the synthesis rating (e.g. all green, except for one), this results in provisional OMERACT endorsement (amber). Finally, if any of the psychometric properties had a red or white (i.e. no information available) synthesis rating, the final conclusion according to the OMERACT algorithm would be that the instrument was not endorsed (red).

Mandatory domains for all trials

Disease activity

Ten candidate instruments were investigated for domain match and feasibility, including two patient reported outcomes (PROs), one composite measure (of PROs and inflammation biomarker) and seven objective measures of disease activity (table 1). As the CAN-DEN MRI activity of the spine,[20] and the ASspiMRI-a,[21] did not pass the domain match and feasibility requirements, the psychometric properties of the remaining eight instruments were assessed. The SOMPs are presented per instrument in the supplement (Supplementary files 1-8). For ASDAS,[23] and patient global assessment for disease activity during last week (PtGA) there was sufficient evidence for all psychometric properties to support the use of the instrument in clinical trials. Bath Ankylosing Spondylitis Disease Activity Index (BASDAI),[24] performed well for discrimination, but there was inconsistent data regarding the truth aspect. C-reactive protein (CRP) performed well with regards to the truth aspect, contrary to the assessment of discrimination, which showed poor performance, even though we know from experience that CRP is highly discriminative in clinical trials. This can be explained by the fact the calculations proposed by OMERACT assume normal data distribution, making them unsuitable to assess discrimination in non-normal distributed data such as CRP. Psychometric properties of the instruments used for the assessment of activity on MRI of the SIJ and spine (i.e. Spondyloarthritis Research Consortium of Canada -SPARCC,[25 26]- and Berlin MRI scores,[27] [28]) were comparable between the two scoring methods. There was more information available for the SPARCC- compared to the Berlin assessments, and the SPARCC has a defined cut-off indicating minimally important change while this was not available for the Berlin scores.

All eight investigated instruments for the domain disease activity were finally endorsed by the ASAS community (tables 3 and 4).

Table 3 Instruments for updated COS for axial spondyloarthritis

<table>
<tr><th colspan="2">Mandatory instruments for all trials</th></tr>
<tr><td><b>Domain</b></td><td><b>Instrument</b></td></tr>
<tr><td>Disease activity</td><td>ASDAS</td></tr>
<tr><td></td><td>Patient global assessment of disease activity (NRS)</td></tr>
<tr><td>Pain</td><td>NRS total back pain (BASDAI Q2)</td></tr>
<tr><td>Morning stiffness</td><td>Severity and duration of stiffness (BASDAI (Q5+Q6)/2))</td></tr>
<tr><td>Fatigue</td><td>NRS fatigue (BASDAI Q1)</td></tr>
<tr><td>Physical function</td><td>BASFI</td></tr>
<tr><td>Overall functioning & health</td><td>ASAS-HI</td></tr>
<tr><th colspan="2">Additional mandatory instruments for disease modifying drugs trials</th></tr>
<tr><td><b>Domain</b></td><td><b>Instrument</b></td></tr>
<tr><td>Disease activity</td><td>SPARCC MRI-SIJ*</td></tr>
<tr><td></td><td>SPARCC MRI-spine*</td></tr>
<tr><td>Extra-musculoskeletal manifestations</td><td>Acute anterior uveitis[†‡]</td></tr>
<tr><td></td><td>Psoriasis[†§]</td></tr>
<tr><td></td><td>Inflammatory bowel disease[†‖]</td></tr>
<tr><td>Peripheral manifestations</td><td>44 swollen joint count</td></tr>
<tr><td></td><td>MASES</td></tr>
<tr><td></td><td>Dactylitis count (including active fingers and/or toes)</td></tr>
<tr><td>Structural damage</td><td>mSASSS*</td></tr>
</table>

*Needs to be assessed at least once in a disease modifying drug programme; † According to ASAS recommendations: diagnosis has never been made, was known at the preceding visit or has been made since the last visit; ‡ In case of diagnosis: the number of episodes since the last visit and corresponding treatment; § In case of diagnosis: percentage of skin area with psoriasis and treatment yes/no; ‖ In case of diagnosis: subtype and treatment yes/no; **ASDAS**: Ankylosing Spondylitis Disease Activity Score; **NRS**: Numerical Rate Scale; **BASDAI**: Bath Ankylosing Spondylitis Disease Activity Index; **Q**: question; **BASFI**: Bath Ankylosing Spondylitis Functional Index; **SPARCC**: SpondyloArthritis Research Consortium of Canada Scoring System; **MRI**: Magnetic Resonance Imaging; **SIJ**: Sacroiliac Joint; **MASES**: Maastricht Ankylosing Spondylitis Enthesitis Score; **mSASSS**: modified Stoke Ankylosing Spondylitis Spinal Score.

Table 4 ASAS-endorsed instruments, which can be used in addition to the COS

Additional ASAS-endorsed instruments	
Domain	**Instrument**
Disease activity	BASDAI
	CRP
	Berlin MRI-SIJ
	Berlin MRI-spine
Pain	NRS back pain at night
Morning stiffness	Severity of morning stiffness (BASDAI Q5)
	Duration of morning stiffness (BASDAI Q6)
Overall functioning & health	SF-36
Peripheral manifestations	66 swollen joint count
	SPARCC Enthesitis
Structural damage	SPARCC MRI SIJ Structural Score (SPARCC SSS) for erosion

BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; **CRP**: C-Reactive Protein; **MRI**: Magnetic Resonance Imaging; **SIJ**: Sacroiliac Joint; **NRS**: Numerical Rate Scale; **Q**: question; **SF-36**: 36-Item Short Form Health Survey; **SPARCC**: SpondyloArthritis Research Consortium of Canada Scoring System; **SSS**: Sacroiliac joint Structural Score.

Out of these, a total of four instruments were selected in the COS (table 3): ASDAS and PtGA are mandatory to be assessed in all clinical trials while SPARCC MRI activity SIJ and SPARCC MRI activity spine are mandatory (at least in one trial in the development programme of a specific drug) for DMARD-trials.

Pain

In the domain pain two instruments were identified: total back pain and back pain at night in the past week,[24],[29]. As both instruments passed the domain match and feasibility requirements, all psychometric properties were assessed. For back pain at night all psychometrics achieved a good synthesis rating, indicating good performance and consistent results. Results for total back pain were similar, with the exception of construct validity which showed inconsistent results (Supplementary files 9-10). Subsequently, both instruments were endorsed by the ASAS members and total back pain was chosen to be included in the COS (tables 3 and 4). Total back pain was preferred as this is present in most patients, while night pain is not and the implementation of total back pain was in 96-100% of all studies while night pain was included in only 20-42%,[9].

Morning stiffness

Three instruments were identified for the domain morning stiffness: Severity of morning stiffness,[24], Duration of morning stiffness,[24]; and combined average score including both severity and duration of morning stiffness,[24]. All three instruments passed the domain match and feasibility requirements, and subsequently data was collected on all psychometric properties (Supplementary files 11-13). Psychometric properties were comparable across all three instruments. There was more information available on construct validity of the individual questions compared to the composite score, contrary, there was much more information on longitudinal construct validity and discrimination for the composite score. ASAS members endorsed all three instruments to assess morning stiffness (tables 3 and 4). Out of the three, the combined score was selected as the preferred instrument to be included in the COS.

Fatigue

The FACIT-fatigue,[22] measure was discussed within the working group, but it was decided this instrument did not have sufficient utilization at this time to assess feasibility requirements and was therefore set aside for future research agenda. Therefore, one instrument was assessed for the domain fatigue: question one of the BASDAI reflecting fatigue,[24]. This instrument was also included in the previous core set and was well implemented (84-100%,[9]). Good performance and consistent results were found for all psychometric properties except clinical trial discrimination (Supplementary file 14). ASAS members endorsed this instrument and voted for inclusion in the COS to assess the domain fatigue (Table 3).

Physical function

One instrument was investigated for the assessment of physical function: Bath Ankylosing Spondylitis Function Index (BASFI),[30]. There was inconsistent information regarding construct validity and clinical trial discrimination (Supplementary file 15); for the other psychometric properties BASFI showed good performance and has been well implemented (88-100%,[9]). BASFI was endorsed and voted to remain in the COS (Table 3).

Overall functioning & health

Two instruments were identified to assess overall functioning & health, one disease specific instrument: ASAS Health Index (ASAS-HI)[31], and one generic instrument: 36-Item Short Form Health Survey (SF-36),[32]. The ASAS-HI is a relatively new instrument developed by ASAS according to the latest insights in methodology, based on the International Classification of Functioning, Disability and Health. It is free for use and available in many languages. Both SF-36 and ASAS-HI showed comparable construct validity, but ASAS-HI performed better on test-retest reliability (Supplementary files 16-17). Contrary to the ASAS-HI, there is no sufficient disease specific information regarding the thresholds of meaning for the SF-36. The ASAS members endorsed both instruments, but preferred the ASAS-HI over the SF-36 for inclusion in the COS (tables 2 and 3).

Mandatory domains for DMARD trials

Extra-musculoskeletal manifestations

For the assessment of extra-musculoskeletal manifestations (EMMs) three instruments were identified to collect information on acute anterior uveitis (AAU), psoriasis and inflammatory bowel disease (IBD), based on previous ASAS recommendations,[33]. For all three EMMs, it is required to collect information on the diagnosis (has never been made, was known at the preceding visit or has been made since the last visit) and additional information such as extent and treatment on the EMM. For all EMMs only limited information was available regarding construct validity and discrimination (Supplementary file 18). Nonetheless, given the relevance of standardised information collection on EMMs, ASAS agreed to collect EMMs as an outcome measure, rather than as adverse events, which is currently common practice. Therefore, the instruments to assess AAU, psoriasis and IBD were endorsed and selected for the COS (table 3).

Peripheral manifestations

A total of five instruments were identified for the assessment of peripheral manifestations (Supplementary files 19-23), which included two instruments for the assessment of arthritis, two instruments for the assessment of enthesitis and one instrument for the assessment of dactylitis,[33] (table 1).

Psychometric properties were comparable for the 44 and 66 swollen joint counts, both showing inadequate performance for clinical trial discrimination. However, the inclusion criteria of current trials do not request a minimum number of involved joints, which hampers the discriminatory ability. In addition, the data are highly skewed, which makes the assessment of trial discrimination challenging. Moreover, there was no information available on thresholds of meaning. Nonetheless, both were endorsed by the ASAS members (tables 3 and 4), thereby ensuring standardised data collection that allows for future assessment of their performance. As the 44 swollen joint count performed slightly better and is included in the original COS for AS,[3], this was chosen as the preferred instrument for inclusion in the COS.

There was more information available regarding the psychometric properties of Maastricht Ankylosing Spondylitis Enthesitis Score (MASES)[34] than the SPARCC enthesitis score,[35], but overall, the performance of both was comparable. Similar to the swollen joint counts, the assessment of discriminatory ability is hampered by the fact that current trials do not request the presence of enthesitis and data are skewed. Here too, the ASAS members endorsed both instruments, but chose the MASES to be included in the COS, as this instrument is considered more specific for axSpA and was included in the previous core set (tables 3 and 4).

For dactylitis, there was little information available on any of the psychometric properties. However, as for the EMMs, the working group decided it would be of great value to start collecting information in a standardised manner. Therefore, the dactylitis count (per ASAS recommendations),[33] was endorsed and included in the COS (table 3).

Structural damage

Three instruments in the domain structural damage were investigated: modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS),[36] to assess the spine, the modified New York (mNY) score for the SIJ,[37], and the SPARCC MRI SIJ structural Score (SSS) for erosion,[38]. For this domain it was difficult to assess discrimination as it takes at least two years for radiographic changes to occur in axSpA (especially in early disease),[39],[40] trials.

Inter- and intra-rater reliability has been shown to be poor for the mNY score, which also has an impact on its potential to show change over time (Supplementary file 24). Therefore, the ASAS members did not endorse this instrument.

Test-retest reliability for both the mSASSS and SPARCC MRI SSS erosion was good, and there was information in support of construct validity (Supplementary files 25-26). Therefore, both mSASSS and SPARCC MRI SSS erosion were endorsed by the ASAS members. Yet, only the mSASSS was selected for inclusion in the COS (tables 3 and 4).

Voting and endorsement

At the 2022 ASAS annual meeting, the instruments selected by the working group were presented and discussed per domain, followed by a vote on the proposal. For each domain, there was only one round of voting required to obtain the 75 % cut-off (as specified in the methods section). The agreement varied between 80 and 97%. Detailed voting results can be found in Supplementary table S5.

Discussion

This manuscript presents the instruments selected to assess the ASAS-OMERACT core domains for axSpA. This is the final step of an extensive process to update the previous COS dating from 1999,[1-4]. In total, the COS includes seven instruments for the domains that are mandatory for all trials and nine additional instruments mandatory for studies evaluating DMARDs.

It is important to keep in mind that the objective of the COS is not to include everything that may be useful for assessing the efficacy and safety of a treatment within a study, but rather to define a minimum but mandatory set, considering that the final product must be feasible and implementable. Adhering to the principle of parsimony, only one instrument was selected for each domain, except for the disease activity domain, where two instruments were selected for all trials, and two more instruments were included for studies assessing DMARDs. This highlights the relevance of the disease activity domain when assessing the efficacy of therapies in patients with axSpA.

The previous core set was endorsed by OMERACT. We tried to follow the OMERACT filter 2.2,[41] as much as possible to select the instruments. However, strict application of this filter would have resulted in endorsement of only three out of the 28 instruments (Supplementary files 1-26). Instruments that are currently used (e.g. CRP) could not be fully endorsed by OMERACT, even though these instruments were used in the past to obtain drug regulatory agencies approval for currently used therapies. The consequence would be that we would not be able to recommend any instrument in the near future, and perhaps never, even though patients and physicians consider these domains important. Moreover, as the axSpA field is moving quickly, there is a high need for a speedy update of the core set. After discussion, the overall conclusion was that having a core domain set without instruments would be meaningless and potentially harmful for its final goal to standardise outcomes. Therefore, it was preferred to include less optimal instruments or instruments that are likely optimal but for which some information is missing, but which may also be cumbersome to obtain. This will at least enhance standardisation and will subsequently provide more information on these instruments. The decision is important as with some instruments, full or even provisional endorsement is very hard to obtain, since not all instruments are suitable for the process- and summary tables as requested by OMERACT,[10]. PROs are most suitable to follow the recommended process, but the process is less applicable to instruments whose data are highly skewed -such as structural damage- or instruments that pertain to a subgroup of patients, which the RCT is not powered on, such as swollen joints. However, the results of the OMERACT summary tables can be used to direct further research.

Compared to the original COS, the instruments set of the updated COS is more specific and precise, which will favour its implementation and help standardise the evaluation of outcomes in studies,[42]. After the publication of the original COS, some smaller adaptations had taken place. For example, in the original set visual analogue scales (VAS) were included. This was changed to numerical rating scales, which is now officially confirmed and was based on the scientific evidence that has emerged over the years demonstrating a preference for NRS,[43 44].

The following instruments were part of the original core set and remain: PtGA to assess disease activity (NRS), fatigue (NRS, Q1 BASDAI), total back pain (NRS, BASDAI Q2), BASFI and 44 swollen joint count (the latter only for DMARD trials). Five instruments that were part of the original core set have not been reselected: ESR, night pain, chest expansion, modified Schober, and occiput to wall distance. The latter three were not selected, because the domain spinal mobility was no longer included in the COS. The CRP needs to be assessed as this is part of the ASDAS, but ESR/CRP were not considered essential as separate outcome measures. With regards to instruments assessing pain, the fact that pain at night was not well implemented in the original core set (20-42%,[9]) in addition to the fact that this may be absent in patients with axSpA made the stakeholders regard total back pain as sufficient to assess the domain pain.

Three new instruments have been added for all trials: Severity and duration combined score of morning stiffness (BASDAI (Q5+Q6)/2) replacing the duration of morning stiffness, the ASDAS as part of the domain 'disease activity' and the ASAS-HI for the new domain 'overall functioning and health'. An important aspect for the implementation of a core set is feasibility. Although there are seven instruments listed to be included for all trials, actually only five instruments need to be collected: PtGA, CRP, BASDAI, BASFI and ASAS-HI. Two questions of the BASDAI together with CRP and PtGA are used to calculate the ASDAS; other separate questions from BASDAI are used as instruments for fatigue, total back pain and morning stiffness. The BASFI and ASAS-HI are two specific instruments developed to assess the respective domains. Although the information for the entire BASDAI and also CRP is available, these are not required to be present individually. The ASDAS has been shown to have better psychometric properties than the BASDAI and is therefore preferred and makes the BASDAI redundant,[45 46]. The CRP is less useful as a marker of inflammation as it is not elevated in most patients and for some interventions (e.g. physiotherapy) it is not expected that CRP will improve.

The most prominent changes are in the instruments selected for trials assessing DMARDs. By the selection of the domains, it was already made clear that all aspects of axSpA need to be assessed. Therefore, instruments had to be selected for three peripheral manifestations, three EMMs and structural damage. The ASAS community decided that it was also important to add two objective instruments to the domain disease activity: the SPARCC MRI SIJ and SPARCC MRI spine to assess inflammatory lesions on MRI. This underlines the importance of objectively assessing inflammation in this specific setting at least in one trial in the development program of a specific DMARD. The SPARCC instruments were selected over the Berlin instruments as there were more data available on the SPARCC instruments, including a defined cut-off indicating minimally important change.

To assess arthritis the 44 swollen joint count was maintained. Moreover, the choice of the MASES was also in agreement with the previous COS. For dactylitis, the dactylitis count -assessed according to ASAS recommendations,[33] was chosen. It was decided to count only digits with active dactylitis as this improves the performance of the instrument. ASAS has previously developed CRFs to assess uveitis, IBD and psoriasis,[33]. These are recommended as the optimal way to obtain information about EMMs. It is clear that such CRFs are not instruments as such, but they collect all information to present incidence rates in both patients known to have the respective EMM or as new onset. Although there is little information on the use of these CRFs, it was felt very important to implement them to improve collection of these (efficacy) outcomes, which are currently often only assessed as adverse events with insufficient information. Finally, the domain structural damage was already in the previous core set, but without a selected instrument, although in practice, the mSASSS was used for this. This is now officially endorsed. While the mSASSS assesses structural damage in the spine only, the SIJs are also important, but there was no instrument chosen for the assessment of structural damage in the SIJs. The mNY score on radiographs was not endorsed. The SPARCCC MRI SSS for erosion was endorsed for the assessment of erosions on MRI of the SIJs, but it was judged that it was too early to include this in the core set as mandatory instrument. The low-dose CT scan assessing the SIJs or the entire spine are promising tools under development, but there was insufficient information available to formally assess it.

Furthermore, 11 other instruments were also endorsed by ASAS. Both the working group and the entire ASAS community considered all these instruments valid for assessing the corresponding domain. They can be used in clinical trials, but always in addition to (and not as a substitute for) those already included in the COS.

In conclusion, the definition of the instruments for the ASAS-OMERACT core domain set is a milestone in the area of axSpA as it completes the update of the COS for axSpA. From now on, it should be used in all trials evaluating the efficacy and safety of any type of therapy in patients with axSpA. However, in order to make the COS update meaningful, it is necessary to work on further steps. First, it is essential to put efforts into dissemination and implementation of the COS. For this, ASAS intends to work following the same strategy as for other ASAS products, such as the classification criteria for axSpA, by maximising all its dissemination platforms (website, social media, courses, congresses, publications). Secondly, after defining the domains and instruments to be used in all studies, it is important to establish how the results of these individual measurements in the studies are to be reported. In this sense, the aim of ASAS is to establish a consensus that defines exactly which results are to be published and how this is to be done. Finally, as progress is made in the axSpA field, it will be necessary to consider the next update of the COS. However, in order for the COS to meet its final goal, it needs to remain unchanged for a certain period of time to allow time for implementation in studies.

Acknowledgements

We would like to thank all ASAS members for their participation in the project.

Figure legend

Figure 1 Development process to determine the instruments of the core outcome set

Figure 2: Psychometric property assessment two-steps process

Conflicts of interest:

V. Navarro-Compán: Grants/honoraria from AbbVie, Janssen, Lilly, Novartis, Pfizer, and UCB.

A. Boel: No competing interests to declare.

A. Boonen: Grants from AbbVie ad Novartis and honoraria for boards or lectures form Abbvie, Galapagos and Lilly.

P. Mease: Research grants, consultation fees, and/or speaker honoraria: Abbvie, Aclaris, Amgen, Bristol Myers, Boehringer-Ingelheim, Galapagos, Gilead, GlaxoSmithKline, Inmagene, Janssen, Lilly, Novartis, Pfizer, SUN Pharma, UCB.

M. Dougados: Consulting fees Pfizer, AbbVie, Amgen, BMS, Boehringer Ingelheim, Celgene, Galapagos, Gilead, Glaxo-Smith-Kline, Janssen, Lilly, Merck, Novartis, Pfizer, Regeneron, Roche, Sanofi, UCB Pharma.

U. Kiltz: Grant and research support and consultancy fees from AbbVie, Amgen, Biogen, Chugai, Eli Lilly, Fresenius, Gilead, GSK, Janssen, MSD, Novartis, Pfizer, Roche and UCB.

R. Landewé: Honoraria from AbbVie, AstraZeneca, BMS, Boehringer Ingelheim, Celgene, Galapagos, Gilead, Glaxo-Smith-Kline, Janssen, Eli-Lilly, Novartis, Pfizer, UCB Pharma. Director of Rheumatology Consultancy BV.

X. Baraliakos: Consulting fees AbbVie, Amgen, BMS, Celgene, Galapagos, Gilead, Janssen, Lilly, MSD, Novartis, Pfizer, Roche, UCB.

W. Bautista-Molano: Research speaker´s fees from Janssen, Lilly, Novartis, Pfizer, and Biopas.

P. Chiowchanwisawakit: Research grants/honoraria from Novartis, Pfizer, Zuellig Pharma, Janssen.

Hanne Dagfinrud: No competing interests to declare.

L. Fallon: Employee and shareholder of Pfizer Inc.

M. Garrido-Cumbrera: Research collaboration with and provides services to Novartis Pharma AG.

LS. Gensler: Research grants/honoraria from AbbVie, Eli Lilly, Gilead, GSK, Janssen, Novartis, Pfizer, and UCB.

B. Elzorkany: Consultancy, research grants, and speaker's fees from: AbbVie, Amgen, BMS, Eva, Hekma, Janssen, Lilly, MSD, New Bridge, Novartis, Pfizer, Roche, Sanofi-Aventis, and Servier.

N. Haroon: Consultant for Amgen, Abbvie, Eli Lilly, Janssen, Merck, Novartis, Pfizer and UCB.

YH. Kwan: No competing interests to declare.

PM. Machado: Consulting/speaker's fees from Abbvie, BMS, Celgene, Eli Lilly, Janssen, MSD, Novartis, Orphazyme, Pfizer, Roche and UCB, all unrelated to this manuscript.

WP. Maksymowych: Consulting fees Abbvie, BMS, Boehringer, Celgene, Lilly, Janssen, Novartis, Pfizer, UCB; Research and/or Educational Grants Abbvie, Novartis, Pfizer; Chief Medical Officer CARE Arthritis Limited.

AM: Speaker honoraria for lectures, presentations from Abbvie, UCB, Novartis, Pfizer, Gilead, Janssen, MSD, BMS, Biogen, Lilly,

N. de Peyrecave: Employee of UCB Pharma.

D. Poddubnyy: Research support from: AbbVie, Lilly, MSD, Novartis, and Pfizer. Consulting fees from: AbbVie, Biocad, Gilead, GlaxoSmithKline, Eli Lilly, MSD, Novartis, Pfizer, Samsung Bioepis, and UCB. Speaker fees from: AbbVie, Bristol-Myers Squibb, Lilly, MSD, Novartis, Pfizer, and UCB.

M. Protopopov: Speaker honoraria for lectures, presentations from: Novartis and UCB,

S. Ramiro: Research grants from Galapagos, MSD, Novartis, Pfizer and consulting fees from AbbVie, Eli Lilly, MSD, Novartis, UCB and Sanofi.

I-H. Song, employee of AbbVie, Immunology Clinical Development, USA.

S. van Weely: No competing interests to declare.

D. van der Heijde: Consulting fees AbbVie, Amgen, Astellas, AstraZeneca, Bayer, BMS, Boehringer Ingelheim, Celgene, Cyxone, Daiichi, Eisai, Galapagos, Gilead, Glaxo-Smith-Kline, Janssen, Lilly, Merck, Novartis, Pfizer, Regeneron, Roche, Sanofi, Takeda, UCB Pharma. Director of Imaging Rheumatology bv.

References

1. van der Heijde D, Bellamy N, Calin A, et al. Preliminary core sets for endpoints in ankylosing spondylitis. Assessments in Ankylosing Spondylitis Working Group. J Rheumatol 1997;24(11):2225-9.
2. van der Heijde D, van der Linden S, Bellamy N, et al. Which domains should be included in a core set for endpoints in ankylosing spondylitis? Introduction to the ankylosing spondylitis module of OMERACT IV. J Rheumatol 1999;26(4):945-7.
3. van der Heijde D, van der Linden S, Dougados M, et al. Ankylosing spondylitis: plenary discussion and results of voting on selection of domains and some specific instruments. J Rheumatol 1999;26(4):1003-5.
4. van der Heijde D, Calin A, Dougados M, et al. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. Assessments in Ankylosing Spondylitis. J Rheumatol 1999;26(4):951-4.
5. Navarro-Compán V, Boel A, Boonen A, et al. The ASAS-OMERACT core domain set for axial spondyloarthritis. Semin Arthritis Rheum 2021;51(6):1342-49.
6. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. Trials 2017;18(Suppl 3):280.
7. Boers M, Idzerda L, Kirwan JR, et al. Toward a generalized framework of core measurement areas in clinical trials: a position paper for OMERACT 11. J Rheumatol 2014;41(5):978-85.
8. Boers M, Kirwan JR, Wells G, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol 2014;67(7):745-53.
9. Bautista-Molano W, Navarro-Compán V, Landewé RB, et al. How well are the ASAS/OMERACT Core Outcome Sets for Ankylosing Spondylitis implemented in randomized clinical trials? A systematic literature review. Clin Rheumatol 2014;33(9):1313-22.
10. Boers M, Kirwan JR, Tugwell P. OMERACT Handbook, 2018.
11. van den Berg R, de Hooge M, van Gaalen F, et al. Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. Rheumatology (Oxford) 2013;52(8):1492-9.
12. López-Medina C, Molto A, Sieper J, et al. Prevalence and distribution of peripheral musculoskeletal manifestations in spondyloarthritis including psoriatic arthritis: results of the worldwide, cross-sectional ASAS-PerSpA study. RMD Open 2021;7(1):e001450.
13. Sieper J, Lenaerts J, Wollenhaupt J, et al. Efficacy and safety of infliximab plus naproxen versus naproxen alone in patients with early, active axial spondyloarthritis: results from the double-blind, placebo-controlled INFAST study, Part 1. Ann Rheum Dis 2014;73(1):101-7.
14. Landewé R, Braun J, Deodhar A, et al. Efficacy of certolizumab pegol on signs and symptoms of axial spondyloarthritis including ankylosing spondylitis: 24-week results of a double-blind randomised placebo-controlled Phase 3 study. Ann Rheum Dis 2014;73(1):39-47.
15. van der Heijde D, Braun J, Dougados M, et al. Sensitivity and discriminatory ability of the Ankylosing Spondylitis Disease Activity Score in patients treated with etanercept or sulphasalazine in the ASCEND trial. Rheumatology (Oxford) 2012;51(10):1894-905.
16. van der Heijde D, Cheng-Chung Wei J, Dougados M, et al. Ixekizumab, an interleukin-17A antagonist in the treatment of ankylosing spondylitis or radiographic axial spondyloarthritis in patients previously untreated with biological disease-modifying anti-rheumatic drugs (COAST-V): 16 week results of a phase 3 randomised, double-blind, active-controlled and placebo-controlled trial. Lancet 2018;392(10163):2441-51.

17. van der Heijde D, Song IH, Pangan AL, et al. Efficacy and safety of upadacitinib in patients with active ankylosing spondylitis (SELECT-AXIS 1): a multicentre, randomised, double-blind, placebo-controlled, phase 2/3 trial. Lancet 2019;394(10214):2108-17.
18. Sieper J, van der Heijde D, Dougados M, et al. Efficacy and safety of adalimumab in patients with non-radiographic axial spondyloarthritis: results of a randomised placebo-controlled trial (ABILITY-1). Ann Rheum Dis 2013;72(6):815-22.
19. Deodhar A, van der Heijde D, Gensler LS, et al. Ixekizumab for patients with non-radiographic axial spondyloarthritis (COAST-X): a randomised, placebo-controlled trial. Lancet 2020;395(10217):53-64.
20. Lambert RGW, Pedersen SJ, Maksymowych WP, et al. Active Inflammatory Lesions Detected by Magnetic Resonance Imaging in the Spine of Patients with Spondyloarthritis – Definitions, Assessment System, and Reference Image Set. The Journal of Rheumatology 2009;843-17.
21. Braun J, Baraliakos X, Golder W, et al. Magnetic resonance imaging examinations of the spine in patients with ankylosing spondylitis, before and after successful therapy with infliximab: Evaluation of a new scoring system. Arthritis & Rheumatism 2003;48(4):1126-36.
22. Webster K, Cella D, Yost K. The F unctional A ssessment of C hronic I llness T herapy (FACIT) Measurement System: properties, applications, and interpretation. Health and Quality of Life Outcomes 2003;1(1):79.
23. Lukas C, Landewé R, Sieper J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. Ann Rheum Dis 2009;68(1):18-24.
24. 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(12):2286-91.
25. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Care & Research 2005;53(5):703-09.
26. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis research consortium of canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Care & Research 2005;53(4):502-09.
27. Landewé RB, Hermann KG, van der Heijde DM, et al. Scoring sacroiliac joints by magnetic resonance imaging. A multiple-reader reliability experiment. J Rheumatol 2005;32(10):2050-5.
28. Lukas C, Braun J, van der Heijde D, et al. Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: a multireader experiment. J Rheumatol 2007;34(4):862-70.
29. Sieper J, Rudwaleit M, Baraliakos X, et al. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. Ann Rheum Dis 2009;68 Suppl 2ii1-44.
30. 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(12):2281-5.
31. Kiltz U, van der Heijde D, Boonen A, et al. Measurement properties of the ASAS Health Index: results of a global study in patients with axial and peripheral spondyloarthritis. Ann Rheum Dis 2018;77(9):1311-17.
32. Ware JE, Jr., Kosinski M, Bayliss MS, et al. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. Med Care 1995;33(4 Suppl):As264-79.
33. Dougados M, Braun J, Vargas RB, et al. ASAS recommendations for variables to be collected in clinical trials/epidemiological studies of spondyloarthritis. Annals of the Rheumatic Diseases 2012;71(6):1103-04.

34. Heuft-Dorenbosch L, Spoorenberg A, van Tubergen A, et al. Assessment of enthesitis in ankylosing spondylitis. Ann Rheum Dis 2003;62(2):127-32.
35. Maksymowych WP, Mallon C, Morrow S, et al. Development and validation of the Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index. Ann Rheum Dis 2009;68(6):948-53.
36. Creemers MC, Franssen MJ, van't Hof MA, et al. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. Ann Rheum Dis 2005;64(1):127-9.
37. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984;27(4):361-8.
38. Maksymowych WP, Wichuk S, Chiowchanwisawakit P, et al. Development and preliminary validation of the spondyloarthritis research consortium of Canada magnetic resonance imaging sacroiliac joint structural score. J Rheumatol 2015;42(1):79-86.
39. Van der Heijde D, Landewe R, Spoorenberg A, et al. Can a historical cohort of patients with ankylosing spiondylitis containing data of 2-year radiographic progression serve as a control group to assess inhibition of structural damage by TNF-blockers? Arthritis and Rheumatism 2003;48(9):S441-S41.
40. Sepriano A, Ramiro S, van der Heijde D, et al. Biological DMARDs and disease modification in axial spondyloarthritis: a review through the lens of causal inference. RMD Open 2021;7(2):
41. Maxwell LJ, Beaton DE, Boers M, et al. The evolution of instrument selection for inclusion in core outcome sets at OMERACT: Filter 2.2. Semin Arthritis Rheum 2021;51(6):1320-30.
42. Andreasen RA, Kristensen LE, Baraliakos X, et al. Assessing the effect of interventions for axial spondyloarthritis according to the endorsed ASAS/OMERACT core outcome set: a meta-research study of trials included in Cochrane reviews. Arthritis Res Ther 2020;22(1):177.
43. Akad K, Solmaz D, Sari I, et al. Performance of response scales of activity and functional measures of ankylosing spondylitis: numerical rating scale versus visual analog scale. Rheumatol Int 2013;33(10):2617-23.
44. Van Tubergen A, Debats I, Ryser L, et al. Use of a numerical rating scale as an answer modality in ankylosing spondylitis-specific questionnaires. Arthritis Rheum 2002;47(3):242-8.
45. Kirkham J, Christensen R, Boers M. Use of composite outcomes facilitate core outcome set uptake in rheumatoid arthritis trials. Ann Rheum Dis 2020;79(2):301-02.
46. Landewé RBM, van der Heijde D. Use of multidimensional composite scores in rheumatology: parsimony versus subtlety. Ann Rheum Dis 2020;10.1136/annrheumdis-2020-216999

Funding

The Assessment of SpondyloArthritis international Society (ASAS) funded Anne Boel and Victoria Navarro-Compán to work on the project to update the core outcome set.

Author contribution

VN-C, AB and DvdH wrote the first draft of the manuscript. All authors participated actively in the project. All authors critically reviewed the manuscript for important intellectual contribution and approved the final version.

Supplementary Table S1 Cut-offs used to interpret the data on all psychometric properties.

Performance	*Construct validity*	*Test-retest reliability*	*Longitudinal construct validity*	*Clinical trial discrimination*	*Threshold of meaning*
Good	≥75% of hypotheses confirmed in the article	ICC ≥0.75	Guyatt's ES, SRM, ES ≥0.80	SMD or SMD_{imp} ≥0.80	External anchor is solid
Adequate	50%-75% of hypotheses confirmed in the article	ICC ≥0.50 & <0.75	Guyatt's ES, SRM, ES ≥0.50 & <0.80	SMD or SMD_{imp} ≥0.50 & <0.80	External anchor is not described in detail
Poor	<50% of hypotheses confirmed in the article	ICC <0.50	Guyatt's ES, SRM, ES <0.50	SMD or SMD_{imp} <0.50	External anchor does not make sense

Note: OMERACT uses '+', '+/-' and '-' to indicate if performance of the instrument was good, adequate or inadequate, based on predefined threshold for each property. Here we have deviated from the OMERACT data visualisation to improve understanding and instead of symbols we used a colour system to visualise performance of the instrument in a given psychometric property (good (green), adequate (amber) or inadequate (red)).

Supplementary Table S2 Calculations longitudinal construct validity and clinical trial discrimination

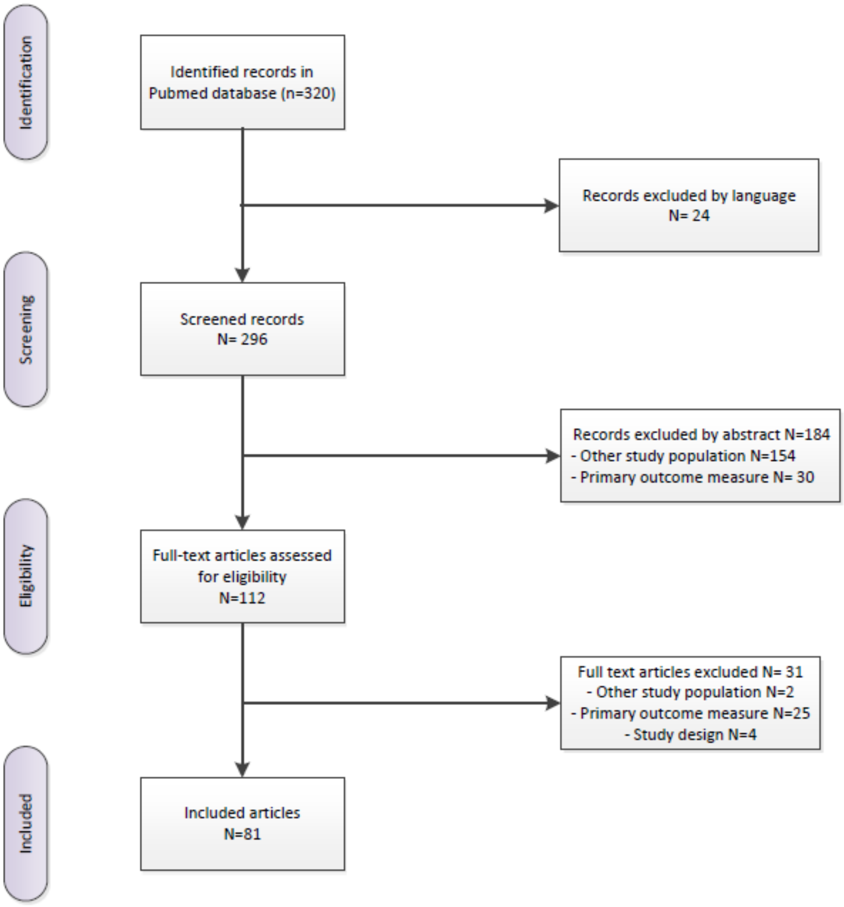
Longitudinal construct validity	$\text{Guyatt's ES} = \frac{\text{average change in treatment group}}{\text{SD change in comparator group}}$
	$\text{SRM} = \frac{\text{average change in treatment group}}{\text{SD change in treatment group}}$
	$\text{ES} = \frac{\text{average change in treatment group}}{\text{SD baseline in treatment group}}$
Discrimination in clinical trials	$SMD = \frac{\text{mean treatment} - \text{mean comparator}}{\frac{(\text{N treatment} * \text{SD after treatment}) + (\text{N comparator} * \text{SD after treatment})}{\text{N treatment} + \text{N comparator}}}$
	$\text{SMDimp} = \frac{\text{average change difference between treatment groups}}{\text{SD average change difference}}$

ES, Effect Size; **Guyatt's ES**, Guyatt's effect size; **SMD**, Standardized mean difference; **SMD_{imp},** Standardized mean difference of improvement; **SRM**, Standardized response mean

Supplementary Table S3 Methodological quality of included papers

	Good methods used
▤	Some cautions, but this will be used as evidence
▥	No, don't use this as evidence

Note: For methodological quality assessment OMERACT developed the checklist OMERACT-COSMIN Good Methods Checklist, which uses a colour code to define low, intermediate or high risk of bias, available from: https://omeract.org/instrument-selection/downloadable-forms/. As we preferred to use the colours to indicate the performance of the instrument, we used shading to indicate the methodological quality of the included papers.


Identification
Screening
Eligibility
Included
Identified records in Pubmed database (n=320)
Records excluded by language N= 24
Screened records N= 296
Records excluded by abstract N=184
- Other study population N=154
- Primary outcome measure N= 30
Full-text articles assessed for eligibility N=112
Full text articles excluded N= 31
- Other study population N=2
- Primary outcome measure N=25
- Study design N=4
Included articles N=81

Supplementary Figure S2 Flow diagram updated search strategy systematic literature review

Supplementary Table S4 Voting results two-day online working group meeting

Mandatory instruments for all trials

	Endorsed					Include in core measurement set				
	N	Yes	No	Abstained	%	N	Yes	No	Abstained	%
Domain: Disease activity										
Patient global assessment disease activity last week (NRS)	19	19	0	0	**100**	19	18	1	0	**95**
CRP	19	19	0	0	**100**	17	0	17	0	**0***
BASDAI	20	19	1	0	**95**	17	0	17	0	**0†**
ASDAS	20	20	0	0	**100**	20	20	0	0	**100**
SPARCC MRI activity SIJ‡§	20	19	0	1	**100**	18	12	0	6	**100**
Berlin MRI activity SIJ‡	20	16	1	3	**94**	18	0	12	6	**0**
SPARCC MRI activity spine‡§	22	21	0	1	**100**	22	15	1	6	**94**
Berlin MRI activity spine‡	22	19	2	1	**90**	22	1	15	6	**6**
Domain: Pain										
NRS total back pain past week (BASDAI Q2)	21	21	0	0	**100**	21	19	2	0	**90**
NRS back pain at night past week	21	21	0	0	**100**	21	1	20	0	**5**
Domain: Morning stiffness										
NRS Duration of morning stiffness (BASDAI Q6)	21	19	1	1	**95**	21	5	16	0	**24**
NRS Severity of morning stiffness (BASDAI Q5)	21	17	3	1	**85**	21	0	21	0	**0**
NRS Severity and duration (BASDAI (Q5+Q6)/2))	21	19	2	0	**90**	21	16	5	0	**76**
Domain: Fatigue										
NRS overall level of fatigue/tiredness (BASDAI Q1)	21	21	0	0	**100**	21	21	0	0	**100**
Domain: Physical function										
BASFI	21	21	0	0	**100**	21	21	0	0	**100**
Domain: Overall functioning & health										
ASAS Health Index	21	21	0	0	**100**	21	21	0	0	**100**
36-item Short-Form health survey (SF-36)	21	18	3	0	**86**	21	0	21	0	**0**

Additional mandatory instruments for disease modifying drugs trials

	Endorsed					Include in core measurement set				
	N	Yes	No	Abstained	%	N	Yes	No	Abstained	%
Domain: Extra-musculoskeletal manifestations										
Uveitis, psoriasis and IBD (ASAS recommendations)	20	20	0	0	**100**	20	18	2	0	**90**
Domain: Peripheral manifestations										
Arthritis: 44 swollen joint count	19	18	0	1	**100**	19	15	3	1	**83**
Arthritis: 66 swollen joint count	19	15	3	1	**83**	19	3	15	1	**17**
Enthesitis: MASES	19	17	1	1	**94**	20	15	3	2	**83**
Enthesitis: SPARCC	19	18	0	1	**100**	20	3	15	2	**17**
Dactylitis: number of affected digits (ASAS recommendations)	20	15	3	2	**83**	20	14	4	2	**78**
Domain: Structural damage										
mNY sacroiliitis	18	0	18	0	**0†**					
modified Stoke AS Spine Score (mSASSS) §	17	16	1	0	**94**	17	14	3	0	**82**
SPARCC MRI Sacroiliac Joint Structural Score for Erosion	17	13	4	0	**76**	17	3	14	0	**18**

Presented percentages are from the first voting round unless otherwise indicated. Percentages are calculated based on the yes and no votes only, abstained votes do not count towards the total. * Percentages from the third round of voting; † Percentages from the second round of voting; ‡ MRI SIJ and spine instruments will only be investigated as mandatory in disease modifying therapy trials; § Structural damage instruments will have to be investigated at least once in a disease modifying therapy programme

NRS: Numerical Rate Scale; **CRP**: C-Reactive Protein; **ASDAS**: Ankylosing Spondylitis Disease Activity Score; **BASDAI**: Bath Ankylosing Spondylitis Disease Activity Index; **SPARCC**: SpondyloArthritis Research Consortium of Canada Scoring System; **MRI**: Magnetic Resonance Imaging; **SIJ**: Sacroiliac Joint; **Q**: question; **BASFI**: Bath Ankylosing Spondylitis Functional Index; **IBD**: Inflammatory Bowel Disease; **MASES**: Maastricht Ankylosing Spondylitis Enthesitis Score; **mNY**: modified New York; **mSASSS**: modified Stoke Ankylosing Spondylitis Spinal Score; **SSS**: Sacroiliac joint Structural Score.

Note: For each vote, the outcome was accepted if ≥75% of attendees agreed, taking only the yes/no votes into account (i.e. excluding abstentions). If this percentage was not reached, the instrument was further discussed, followed by a second round of voting. In the second voting round, the outcome was accepted if ≥67% of attendees agreed. Again, further discussion and another round of voting followed if the percentage was not reached. In the third -and last- voting round, the outcome was accepted if ≥50% of attendees agreed. If this percentage was not reached, the instrument would not be included within the working group proposal to be endorsed/included in the COS. Data on the three candidate instruments for extra-musculoskeletal manifestations were jointly reported in the included manuscripts, hence they were voted on jointly.

Supplementary Table S5 Voting results ASAS 2022 annual workshop#

Mandatory instruments for all trials					
	N	Yes	No	Abstained	%
Disease activity	77	67	5	5	**93**
Pain	88	78	8	2	**93**
Morning stiffness	101	91	9	1	**97**
Fatigue	96	87	7	2	**93**
Physical function	99	93	4	2	**96**
Overall functioning & health	103	90	6	7	**94**
Additional mandatory instruments for disease modifying drugs trials					
Extra-musculoskeletal manifestations	98	85	9	4	**90**
Peripheral manifestations	98	75	13	10	**85**
Structural damage	93	69	17	7	**80**

Presented percentages are from the first voting round unless otherwise indicated. Percentages are calculated based on the yes and no votes only, abstained votes do not count towards the total.

Note: For each vote, the outcome was accepted if ≥75% of attendees agreed, taking only the yes/no votes into account (i.e. excluding abstentions). If this percentage was not reached, the instrument was further discussed, followed by a second round of voting. In the second voting round, the outcome was accepted if ≥67% of attendees agreed. Again, further discussion and another round of voting followed if the percentage was not reached. In the third -and last- voting round, the outcome was accepted if ≥50% of attendees agreed. If this percentage was not reached, the instrument would not be included in the updated core set. Data on the three candidate instruments for extra-musculoskeletal manifestations were jointly reported in the included manuscripts, hence they were voted on jointly.

Supplementary Files 1-26

Note: Each SOMP table also includes a synthesis rating per psychometric property. Again, a colour system is used to visualise the conclusion for each measurement property based on the collected data:

- Green: adequate or good performance of the psychometric property, at least two good quality papers showing consistent results.
- Red: inadequate performance of the psychometric property, only studies of poor quality.
- White: no evidence found for this psychometric property.
- Amber: all other instances (e.g. inconsistent results in good quality papers, only moderate quality papers with consistent results, only one paper was available, which was a good quality paper).

In order to get a full OMERACT endorsement (green), all psychometric properties had to have a green synthesis rating. If there is a mix of green and amber in the synthesis rating (e.g. all green, except for one), this results in provisional OMERACT endorsement (amber). Finally, if any of the psychometric properties had a red or white (i.e. no information available) synthesis rating, the final conclusion according to the OMERACT algorithm would be that the instrument was not endorsed (red).