Article:
Capelusnik, D., López-Medina, C., Weber, C. et al. (2026) Measurement properties of instruments assessing peripheral arthritis disease activity in spondyloarthritis: a systematic literature review. RMD Open, 12 (2). e006900. ISSN: 2056-5933

https://doi.org/10.1136/rmdopen-2026-006900

RMD Open
Rheumatic & Musculoskeletal Diseases

ORIGINAL RESEARCH

Measurement properties of instruments assessing peripheral arthritis disease activity in spondyloarthritis: a systematic literature review

Dafne Capelusnik ,[1,2] Clementina López-Medina ,[3] Casper Weber,[1,4] Augusta Ortolan,[5] Louise Falzon ,[6] Désirée van der Heijde ,[7] Robert Landewé ,[8,9] Philip Mease ,[10,11] Anna Molto ,[12,13] Sofia Ramiro [7,14]

To cite: Capelusnik D, López-Medina C, Weber C, *et al*. Measurement properties of instruments assessing peripheral arthritis disease activity in spondyloarthritis: a systematic literature review. *RMD Open* 2026;**12**:e006900. doi:10.1136/rmdopen-2026-006900

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/rmdopen-2026-006900).

Received 3 March 2026
Accepted 1 June 2026

For numbered affiliations see end of article.

Correspondence to
Dr Clementina López-Medina; h72lomee@uco.es

ABSTRACT

Objectives The Assessment of SpondyloArthritis International Society-Spondyloarthritis and Peripheral Arthritis Disease Activity Instrument Selection and Evaluation (ASAS-SPARADISE) project aims to identify instruments to assess disease activity due to peripheral arthritis in axial (axSpA) and peripheral spondyloarthritis (pSpA). We aimed to summarise and compare the measurement properties of instruments used for this purpose as part of the ASAS-SPARADISE.

Methods A systematic literature review (Medline/EMBASE/Cochrane; until February 2025) was conducted following the Population-Instrument-Measurement framework. Eligible studies included adults with axSpA or pSpA and peripheral arthritis. Instruments retained from a previous review, after domain-match and feasibility assessment, included: Patient Global Assessment (PGA), swollen joint count (66/44), tender joint count (68/44), C reactive protein (CRP), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Axial Spondyloarthritis Disease Activity Score (ASDAS), Disease Activity Index for Psoriatic Arthritis (DAPSA) and Disease Activity Score using 44 joints. Measurement properties evaluated were construct validity (correlation hypotheses; known-groups discrimination) and discrimination (test-retest reliability, responsiveness; clinical trial discrimination; thresholds of meaning). Risk of bias was assessed using the Outcome Measures on Rheumatology (OMERACT) Good Methods Checklist and evidence was synthesised following OMERACT rules (GREEN—good; AMBER—adequate; RED—poor performance).

Results Of 4,364 records screened, eight studies were eligible. Evidence in axSpA was limited to PGA, CRP and BASDAI, rating mostly AMBER for construct validity, longitudinal construct validity and discrimination in clinical trials. In pSpA, composite instruments showed better results, with BASDAI, ASDAS and DAPSA rating mostly GREEN. No study provided test-retest reliability or thresholds of meaning.

Conclusions Evidence supporting instruments to assess peripheral arthritis in spondyloarthritis remains scarce, especially in axSpA. Composite instruments in pSpA performed better, but evidence gaps persist, highlighting the need for validated, domain-specific instruments.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ In axial spondyloarthritis (axSpA), swollen joint count using 44 joints is the recommended instrument, but it has poor trial discrimination, while some studies have shown promising performance of composite instruments for the assessment of peripheral arthritis.

⇒ In peripheral spondyloarthritis (pSpA), no consensus core outcome set for peripheral arthritis activity currently exists.

WHAT THIS STUDY ADDS

⇒ This systematic review provides a comprehensive evaluation of measurement properties of instruments used to assess peripheral arthritis in axSpA and pSpA.

⇒ In axSpA, available evidence is scarce and largely restricted to instruments not specifically developed to capture peripheral arthritis.

⇒ Composite instruments, such as Axial Spondyloarthritis Disease Activity Score, Bath Ankylosing Spondylitis Disease Activity Index, Disease Activity Index for Psoriatic Arthritis using 44 joints and Disease Activity Score using 44 joints, generally demonstrate better performance than single-item measures in capturing peripheral arthritis activity in pSpA.

PROSPERO registration number CRD420251177596.

INTRODUCTION

Patients with spondyloarthritis (SpA) can present with several phenotypes or clinical forms, including axial SpA (axSpA) and peripheral SpA (pSpA). Peripheral musculoskeletal manifestations such as arthritis, dactylitis and enthesitis may occur concomitantly with axial disease, independently (ie,

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Although some instruments demonstrate acceptable measurement properties, important evidence gaps remain, preventing full endorsement according to predefined criteria.

⇒ These findings highlight the need for further validation studies and guide future research within the Assessment of SpondyloArthritis International Society-Spondyloarthritis and Peripheral Arthritis Disease Activity Instrument Selection and Evaluation initiative to identify the most appropriate instrument for assessing disease activity related to peripheral arthritis in axSpA and pSpA.

without axial involvement) or even precede the onset of axial manifestations.[1 2] These heterogeneous presentations increase the complexity of disease assessment and the selection of outcome measures that adequately capture both axial and peripheral disease activity.

Recently, the Assessment of SpondyloArthritis International Society (ASAS) core outcome set (COS) for axSpA was updated. Within the 'peripheral manifestations' domain, only two instruments were evaluated for peripheral arthritis: the swollen joint count (SJC) using either 66 or 44 joints (SJC66 and SJC44). Measurement properties performance was comparable between the two, although both demonstrated inadequate clinical trial discrimination. Nevertheless, both were endorsed by ASAS members to promote standardised data collection and facilitate future validation efforts. The SJC44 was ultimately selected as the preferred instrument for inclusion in the COS.[3] Importantly, composite instruments were not assessed, leaving unclear whether they may better reflect overall disease activity in patients with peripheral articular involvement.

Emerging evidence, although limited, suggests that composite instruments may outperform single-domain measures in assessing peripheral disease activity: in pSpA, Turina *et al* reported that the Axial Spondyloarthritis Disease Activity Score (ASDAS),[4] Bath Ankylosing Spondylitis Disease Activity Index (BASDAI),[5] Patient Global Assessment (PGA) and Physician Global Assessment (PhGA) showed the highest sensitivity to change and discriminatory capacity, whereas the SJC, tender joint count (TJC) and C reactive protein (CRP) performed insufficiently.[6] Also in pSpA, Beckers *et al* found good construct validity for the Disease Activity Index for Psoriatic Arthritis (DAPSA),[7] the Psoriatic Arthritis Disease Activity Score (PASDAS)[8] and the ASDAS, although agreement among these instruments in classifying disease activity states was insufficient, highlighting the need for further evaluation of thresholds for disease activity in pSpA.[9]

More recently, data from the ASAS-perSpA study, which included patients with axSpA, pSpA and psoriatic arthritis (PsA), allowed for a comprehensive assessment of the construct validity of several instruments,[10] including single-domain measures (PGA, SJC, TJC and CRP) as well as composite instruments validated for SpA (ASDAS, BASDAI and DAPSA) and for other rheumatic diseases (Disease Activity Score (DAS)28 and DAS44).[11 12] This ancillary analysis demonstrated that composite instruments, including joint counts, particularly DAS28 and DAPSA, showed superior construct validity by better discriminating between active and inactive disease.[10] In parallel, data from the Clinical REmission in peripheral SPondyloArthritis (CRESPA) trial were used to evaluate the measurement properties of nine measurement instruments in patients with pSpA. The study concluded that both ASDAS and DAPSA showed good longitudinal construct validity and good discrimination in clinical trials, with DAPSA performing numerically better.[13]

Overall, these findings support the value of composite instruments for assessing peripheral arthritis in SpA and highlight the need for a systematic synthesis of available data on the measurement properties of all instruments measuring peripheral arthritis in axSpA and pSpA. In response to this gap, the ASAS-Spondyloarthritis and Peripheral Arthritis Disease Activity Instrument Selection and Evaluation (ASAS-SPARADISE) project was designed to identify and select the most appropriate instruments to assess peripheral arthritis activity in patients with axSpA and pSpA. Understanding the existing evidence is essential to compare measurement properties and identify remaining gaps that may guide future instrument selection and validation.

As part of the ASAS-SPARADISE project, the objective of this systematic literature review (SLR) was to summarise and compare the available evidence on the measurement properties of instruments used to assess disease activity related to peripheral arthritis in axSpA and pSpA.

METHODS

A dedicated Working Group was formed, comprising 29 international experts in SpA and two patient research partners (PRPs). Within this group, a Steering Committee (two principal investigators, four fellows and three additional members) provided methodological oversight and guided the project's development. The SLR was registered in PROSPERO (registration number: CRD420251177596). The methods outlined in the OMERACT Handbook on Instrument Selection for Core Outcome Measurement Sets were applied, which are based on the three pillars of truth, feasibility and discrimination.[14] This involved obtaining agreement from the working group that the instrument matched the target domain and was feasible, followed by a systematic review synthesising the measurement properties of the candidate instrument. Results were reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses-COnsensus-based Standards for the selection of health Measurement INstruments (PRISMA-COSMIN) guidance for reporting systematic reviews of outcome measurement instruments.[15]

Literature search

The search was performed in MEDLINE, EMBASE and Cochrane CENTRAL by an expert librarian (LF), covering all publications up to February 2025, without language restrictions. Reference lists of included studies and relevant review articles were also screened. The full search strategy is provided in online supplemental text S1. The scope and eligibility criteria of this SLR were defined using the Population-Instrument-Measurement properties (PIM) framework.[14]

The *Population* of interest comprised adults (aged ≥18 years) patients with axSpA or pSpA who had active peripheral arthritis. Peripheral arthritis was defined based on clinical assessment in the included studies. While this in principle corresponded to the presence of at least one peripheral joint with active arthritis at baseline, we accepted other specific study definitions for the presence of peripheral arthritis to be inclusive. The requirement for peripheral arthritis was essential to ensure that the evaluated instruments specifically assessed peripheral disease activity, rather than axial disease activity in axSpA. Studies focusing exclusively on PsA were excluded. Data for axSpA and pSpA were expected to be extractable and assessed separately.

The *Instruments* assessed were those identified in the previous ASAS-SPARADISE SLR, in which the measurement instruments assessing peripheral arthritis in axSpA and pSpA were reviewed.[16] Instruments used for peripheral arthritis assessment in other rheumatic diseases were also added. From this pool, the ASAS-SPARADISE Working Group selected instruments meeting adequate domain match and feasibility according to the OMERACT filter[17] for the assessment of the 'peripheral manifestations' domain, specifically peripheral arthritis. This selection was based on the discussion of the results of a survey completed by 30 Working Group members and 15 PRPs (online supplemental table S1). The final list of instruments passing the first part of the OMERACT filter included: PGA, SJC66/44, TJC68/44, CRP, BASDAI, ASDAS, DAPSA, DAS44. Three instruments (DAS28, Simple Disease Activity Index and Clinical Disease Activity Index)[11 18 19] did not pass the OMERACT filter because they include only 28 joints, which was considered inadequate for SpA given the frequent involvement of joints outside the 28-joint set, namely the feet.

The *Measurement properties* evaluated followed the OMERACT framework and included truth/construct validity and discrimination (test-retest reliability, responsiveness or longitudinal construct validity, discrimination in clinical trials and thresholds of meaning).[17] Randomised controlled trials (RCTs) and observational studies fulfilling the above-mentioned PIM framework were included.

Instruments description

PGA is a patient-reported outcome collected with the question "How active was your rheumatic disease on average during the last week?" on a Numeric Rating Scale (NRS) (range 0–10) (10=higher well-being).

The SJC44/66 and TJC44/68 include the assessment of the respective number of peripheral joints from the upper and lower limbs (as well as the temporomandibular, sternoclavicular and acromioclavicular joints when applicable). Each TJC is scored as 1 point; therefore, the total score ranges from 0 to 44 or 0 to 68. Similarly, each SJC is scored as 1 point, with total scores ranging from 0 to 44 or 0 to 66, respectively.

The acute phase reactant CRP (mg/L) was also assessed as a potential separate outcome measurement instrument for disease activity.

The BASDAI is a disease activity measure including patient-reported levels of fatigue, back pain, peripheral joint pain/swelling, localised tenderness and severity and duration of morning stiffness. Each question uses an NRS (range 0–10) (10=very severe), and the scores are averaged (questions 5 and 6 first and then with the remaining 4) giving a final score from 0 to 10 (10=maximal disease activity).[5]

The ASDAS is a disease activity measure including patient-reported overall back pain, peripheral pain/swelling, duration of morning stiffness, PGA, ranging from 0 to 10 on NRS and the CRP (mg/L) as a measure of inflammation. All these elements are combined in a weighted equation with higher values indicating higher disease activity.[4]

The DAPSA is a composite instrument of disease activity. It was originally developed for reactive arthritis as the Disease Activity Index for Reactive Arthritis and is currently widely used to assess disease activity in PsA. DAPSA includes PGA, pain assessment (0–10), SJC66, TJC68 and CRP (mg/dL). The score is calculated as the linear sum of all components, with higher scores indicating higher disease activity.[7 20]

The DAS44 is a composite instrument measuring disease activity in rheumatoid arthritis that combines information from swollen joints, tender joints, acute phase reactants and PGA. It is calculated with a weighted equation with higher values indicating higher disease activity.[12]

Study selection, data extraction and risk of bias assessment

Two reviewers (DC and CL-M) independently performed title/abstract screening, full-text review, data extraction and risk of bias (RoB) assessment. Prior to full screening, 20% of the records were evaluated by both reviewers to assess agreement, which was high (κ >0.80), allowing each reviewer to continue the screening and data extraction of part of the articles alone. Discrepancies were resolved through discussion with the convenors (SR and AM). Data extraction was carried out using predesigned and tested data extraction forms based on the templates recommended by OMERACT.

RoB for each measurement property was assessed using the COSMIN-OMERACT 'Good Methods Checklist' (version 27 March 2025),[14] rating each property as:

'yes' (likely low risk); 'some cautions' and 'no' (high risk; should not be used as evidence). Components of evidence rated as GREEN or AMBER on the COSMIN-OMERACT Good Methods Checklists were included for data extraction and further review of performance adequacy.

Assessment of measurement properties

Truth or construct validity, defined as the extent to which a measurement instrument truly reflects the theoretical concept it intends to measure, was evaluated by extracting the Spearman's or Pearson's correlation coefficients between the assessed instrument and other measures of related disease domains (eg, Bath Ankylosing Spondylitis Functional Index, ASAS Health Index, Ankylosing Spondylitis Quality of Life, EuroQoL). Discrimination between known groups (eg, active vs inactive disease) was also assessed when available, using t-tests, analysis of variance or standardised mean differences (SMD). Data were extracted or calculated when possible.

The discrimination domain comprises: (i) test-retest reliability, defined as the stability of the instrument when no real change is expected, expressed as the intraclass correlation coefficient (ICC); (ii) responsiveness or longitudinal construct validity, defined as the ability of an instrument to detect meaningful change over time, assessed using the Guyatt's Responsiveness Index (GRI), standardised response mean (SRM) and effect size (ES), either extracted or calculated when data were available; (iii) clinical trial discrimination defined as the ability to detect between-arms differences in RCTs, expressed as the SMD, also either extracted or calculated when data were available; (iv) thresholds of meaning or thresholds defining a clinically meaningful change or disease activity state, including the patient acceptable symptom state, minimal important difference and minimal detectable change were also assessed.

Quantity, consistency and adequacy of results (performance)

Quantity, consistency and adequacy of the results (performance) for each measurement property were evaluated in accordance with OMERACT recommendations. For construct validity, predefined hypotheses regarding the expected strength of correlations were derived from the ASAS-COS project[3] and supplemented by those developed for the ASAS-perSpA ancillary analysis.[10] In line with the OMERACT Handbook, construct validity was determined by the proportion of hypotheses confirmed rather than the magnitude of the correlations, and was rated as good when ≥75% of hypotheses were met, adequate when 50%–75% were met and poor when <50% were met.[3 17] Correlation strengths were categorised as weak (<0.30), moderate (0.30–0.69) or strong (≥0.70).[21]

For the remaining measurement properties, an a priori methodological decision was taken to require that all instruments demonstrate at least adequate performance. This general requirement was applied to all instruments to prevent situations in which instruments with anticipated poor performance could formally 'meet' weaker or less demanding hypotheses. By applying the same minimum performance threshold across instruments, the comparison was made fair and directly aligned with the study objective of identifying the best-performing instruments for assessing peripheral arthritis activity.

The cut-offs used for the properties were defined as follows. Discrimination between known groups and clinical trial discrimination were rated as good when SMD ≥0.80, adequate when 0.50–0.79and poor when <0.50.[22 23] Test-retest reliability was considered good when ICC ≥0.75, adequate when 0.50–0.74 and poor when <0.50. Longitudinal construct validity was rated as good when GRI, SRM or ES ≥0.80, adequate when 0.50–0.79 and poor when <0.50.[3] For thresholds of meaning, no predefined cut-offs were required.[3 22 23]

All extracted data were summarised in tables organised by disease phenotype (axSpA and pSpA) and measurement property. Performance for each property was classified as '+' when above the predefined threshold (adequate or better), '±' when equivocal and '–' when below the threshold. As no explicit definition of consistency is provided in the OMERACT Handbook, consensus was reached to define consistency as at least 50% of studies demonstrating a '+' rating.[14 24]

Finally, following OMERACT synthesis rules, each instrument received an overall rating combining methodological quality with the quantity, consistency and adequacy of results: GREEN when good methods were used in at least two studies with consistent adequate or better performance; RED when good methods were used in one or two studies with consistent or questionable inadequate performance; WHITE when no evidence was available and AMBER in all other scenarios.[14]

RESULTS

From a total of 4364 records identified, 24 were selected for full-text review and eight studies met the eligibility criteria. Two publications reported results from the same study (the Tnf Inhibition in PEripheral SpondyloArthritis -TIPES- study), presenting largely overlapping analyses and findings. The study selection process is detailed in the flow chart provided in online supplemental figure S1.

Most publications showed a low RoB at the publication level (five of eight publications) or for at least one measurement property (13 of 17 assessed measurement properties in total). Among the four axSpA studies, two were cross-sectional analyses based on registry data,[25 26] one was a post hoc analysis of an RCT[27] and one combined registry data with a post hoc RCT analysis.[28] The four pSpA studies included three post hoc analyses of RCTs[6 13 29] and one cross-sectional analysis based on registry data.[9]

Across included studies, peripheral arthritis was defined based on clinical assessment. In most studies, this was operationalised as the presence of at least one peripheral joint with active arthritis at baseline,[6 9 13 25 26 28 29] whereas

one study reported peripheral arthritis as a binary variable (presence/absence) without further specification.[27]

Measurement property data were available for all selected instruments except SJC44, TJC44 and DAS44. Construct validity was assessed in seven studies (three in axSpA[25 26 28] and four in pSpA),[6 9 13 29] longitudinal construct validity in three studies (one in axSpA[28] and two in pSpA)[13 29] and clinical trial discrimination in three studies (one axSpA[27] and two pSpA[6 13 29]). No data were identified for test-retest reliability or thresholds of meaning (table 1).

Construct validity

In patients with axSpA and peripheral arthritis, correlation-based construct validity data were available only for PGA, CRP and BASDAI. Based on correlation analyses, all three instruments fulfilled criteria for a GREEN rating, with evidence derived from four studies for PGA, three for CRP and five for BASDAI (online supplemental table S2).

In pSpA, correlation data were available for a broader set of instruments. PGA, SJC66 and CRP achieved AMBER ratings while TJC68 was rated RED. Composite instruments showed better performance: ASDAS and DAPSA were rated GREEN, each supported by two studies, while the remaining instruments were supported by a single study (online supplemental table S2).

Additional construct validity evidence in pSpA was obtained by calculating discrimination between known-groups using SMD, based on active vs inactive disease status defined by PGA or Physician Global Assessment as anchor. These analyses showed good discrimination (GREEN) for PGA (n=3), SJC66 (n=2), TJC68 (n=2), BASDAI (n=3), ASDAS (n=3) and DAPSA (n=1), while CRP showed adequate discrimination (AMBER; n=3) (table 2).

Furthermore, for ASDAS and DAPSA, analyses of variance were available, demonstrating increasing mean values across disease activity categories when stratifying ASDAS by DAPSA disease activity states and vice versa. These analyses supported a positive ('+') performance for both instruments.

When combining all available construct validity approaches, the final synthesis showed that in axSpA with peripheral arthritis, PGA and BASDAI were rated AMBER and CRP reached GREEN, with no evidence for ASDAS and DAPSA (white), and in pSpA, PGA, SJC66, TJC68 and CRP were rated AMBER, while BASDAI, ASDAS and DAPSA were rated GREEN (online supplemental table S3).

Discrimination

Data on longitudinal construct validity (responsiveness) were available from three studies: one conducted in axSpA and two in pSpA.

In patients with axSpA and peripheral arthritis, longitudinal construct validity was evaluated for PGA and BASDAI. Both instruments received an overall synthesis rating of AMBER, based on SRM and ES values exceeding the predefined thresholds; however, evidence for each instrument was limited to a single study.

In pSpA, longitudinal construct validity data were available for a broader set of instruments, with two studies contributing data for all instruments except DAPSA (n=1). PGA, SJC66, BASDAI and ASDAS demonstrated good responsiveness and achieved GREEN ratings, supported by GRI, SRM and/or ES values ≥0.80 in most analyses. TJC68 and CRP showed weaker performance in longitudinal construct validity and were rated AMBER. DAPSA was also rated AMBER, despite a GRI value ≥0.80, due to the availability of evidence from only one study (table 3).

Data on clinical trial discrimination were available from four studies, including one conducted in axSpA and three in pSpA.

In patients with axSpA and peripheral arthritis, clinical trial discrimination data were available for PGA, CRP, BASDAI and ASDAS. Although all instruments showed SMD values above the predefined threshold of 0.80, evidence was derived from a single study only; therefore, all were assigned an overall synthesis rating of AMBER.

In pSpA, a broader range of instruments was evaluated. PGA, SJC66, TJC68, CRP, BASDAI and ASDAS demonstrated good discrimination between treatment arms and achieved GREEN ratings. DAPSA was supported by a single study and, despite showing SMD values above the adequacy threshold and among the highest across all instruments, was rated AMBER due to the limited quantity of evidence (table 4).

No data were available for test-retest reliability or thresholds of meaning.

Table 5 summarises the final synthesis ratings of all measurement properties for each instrument. Overall, the available evidence was limited and, in axSpA, restricted to PGA, CRP and BASDAI, with insufficient data to support any instrument conclusively.

DISCUSSION

This SLR provides a comprehensive overview of the current evidence on the measurement properties of instruments used to assess disease activity due to peripheral arthritis in axSpA and pSpA. Overall, the findings demonstrate that the available evidence is limited, particularly in axSpA, and largely restricted to instruments not specifically developed to capture peripheral arthritis. Composite instruments tended to show more robust performance across measurement domains when data were available, whereas single-item measures showed more variable and often insufficient evidence.

A clear distinction emerged between axSpA and pSpA. In axSpA, peripheral arthritis is a relatively less frequent and often likely seen as a 'secondary' manifestation. Consequently, studies specifically evaluating the measurement properties of instruments in patients with axSpA

Table 1 Description of the studies included, instruments assessed, measurement properties and methods assessment

Study	Population	Study design and methods	Intervention	End point	Number of patients	Instrument(s) used	Measurement properties assessed	RoB
Hakkou *et al*[25]	r-axSpA with PA	Cross-sectional study from a prospective registry (Marrocco)	–	–	63	BASDAI	Construct validity	Some cautions
Heuft-Dorenbosch *et al*[26]	r-axSpA with PA	Cross-sectional study from a prospective registry (OASIS)	–	–	56	PGA BASDAI	Construct validity	Some cautions
Song *et al*[28]	r-axSpA with PA or enthesitis	Cross-sectional study from a prospective registry (GESPIC)	–	–	67	PGA CRP BASDAI	Construct validity	Some cautions
		RCT (IFX)	IFX	12 weeks	56	PGA CRP BASDAI	Construct validity	Some cautions
							Longitudinal construct validity	Low
van der Heijde *et al*[27]	axSpA with PA	RCT (TNFi)	TNFi	12 weeks	72	PGA CRP BASDAI ASDAS	Clinical trial discrimination	Low
Beckers *et al*[9]	pSpA	Cross-sectional study from a prospective registry (SpA-Net)	–	–	82	ASDAS DAPSA	Construct validity	Low
López-Medina *et al*[13]	pSpA	RCT (CRESPA)	GOL	12 weeks	60	PGA SJC66 TJC68 CRP BASDAI ASDAS DAPSA	Construct validity	Low
							Longitudinal construct validity	Low
							Clinical trial discrimination	Low
Paramarta *et al*[29]	pSpA	RCT (TIPES)	ADA	12 weeks	40	PGA SJC66 TJC68 CRP BASDAI ASDAS	Construct validity	Low
							Longitudinal construct validity	Low
							Clinical trial discrimination	Low

Continued

Table 1 Continued

Study	Population	Study design and methods	Intervention	End point	Number of patients	Instrument(s) used	Measurement properties assessed	RoB
Turina *et al*[6]	pSpA	RCT (Ability 2)	ADA	12 weeks	183	BASDAI ASDAS PGA SJC66 TJC68 CRP	Construct validity	Low
							Clinical trial discrimination	Low
		RCT (TIPES)	ADA	12 weeks	40	PGA SJC66 TJC68 CRP BASDAI ASDAS	Construct validity	Low
							Clinical trial discrimination	Low

All RCTs used placebo as the comparator.
In observational studies, patients were treated based on the treating physician's criteria.
ADA, adalimumab; ASDAS, Axial Spondyloarthritis Disease Activity Score; axSpA, axial spondyloarthritis; BASDAI, Bath Ankylosing Spondylitis Disease Activity Score; CRESPA, Clinical REmission in peripheral SPondyloArthritis; CRP, C reactive protein; DAPSA, Disease Activity in Psoriatic Arthritis; GESPIC, GErman SPondyloarthritis Inception Cohort; GOL, golimumab; IFX, infliximab; OASIS, Outcome in AS International Study; PA, peripheral arthritis; PGA, Patient Global Assessment; pSpA, peripheral spondyloarthritis; r-axSpA, radiographic axial spondyloarthritis; RCT, randomised controlled trial; RoB, risk of bias; SJC66, swollen joint count with 66 joint count; SpA-Net, eHealth system for spondyloarthritis in the Netherlands; TIPES, Tnf Inhibition in PEripheral SpondyloArthritis; TJC68, tender joint count with 68 joint count; TNFi, tumour necrosis factor inhibitor.

Table 2 Construct validity: discrimination between known groups (SMD) in pSpA population

Study	Population	External construct	Mean (SD) (group 1)	Mean (SD) (group 2)	SMD	Performance assessment	Synthesis rating
PGA (0–10)							
Ability 2, Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=15 vs 84)	55.3 (17.0)	71.0 (15.7)	**0.99**	+	GREEN
CRESPA, López-Medina *et al*[13]	pSpA	PhGA ≥4 vs <4 (n=18 vs 42)	7.0 (1.7)	2.5 (2.5)	**1.98**	+	
TIPES, Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=6 vs 12)	45.8 (13.4)	75.2 (15.1)	**−2.02**	+	
SJC66 (0–66)							
CRESPA, López-Medina *et al*[13]	pSpA	PhGA ≥4 vs <4 (n=18 vs 42)	6.7 (5.8)	0.5 (1.7)	**1.81**	+	GREEN
		PGA ≥4 vs <4 (n=27 vs 33)	4.5 (5.7)	0.6 (2.0)	**0.95**		
TIPES, Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=6 vs 12)	2.5 (2.7)	4.2 (3.4)	−0.53	+	
TJC68 (0–68)							
CRESPA, López-Medina *et al*[13]	pSpA	PhGA ≥4 vs <4 (n=18 vs 42)	13.7 (15.2)	1.0 (2.0)	**1.51**	+	GREEN
		PGA ≥4 vs <4 (n=27 vs 33)	9.3 (13.9)	1.2 (2.3)	**0.85**		
TIPES, Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=6 vs 12)	5.7 (3.8)	9.0 (6.5)	−0.57	+	
CRP (mg/L)							
Ability 2, Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=15 vs 84)	7.3 (8.9)	13.0 (20.7)	−0.29	–	AMBER
CRESPA, López-Medina *et al*[13]	pSpA	PhGA ≥4 vs <4 (n=18 vs 42)	26.8 (51.9)	3.2 (5.5)	**0.83**	+	
		PGA ≥4 vs <4 (n=27 vs 33)	19.1 (43.5)	3.1 (5.8)	0.54		
TIPES, Paramarta *et al*[29], Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=56 vs 12)	1.6 (1.1)	16.7 (31.0)	−0.59	+	
BASDAI (0–10)							
Ability 2, Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=25 vs 84)	4.2 (1.9)	6.0 (1.5)	**1.15**	+	GREEN
CRESPA, López-Medina *et al*[13]	pSpA	PhGA ≥4 vs <4 (n=18 vs 42)	6.2 (1.8)	2.3 (1.6)	**2.36**	+	
		PGA ≥4 vs <4 (n=27 vs 33)	5.5 (2.0)	1.7 (1.0)	**2.45**		
TIPES, Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=6 vs 12)	3.4 (1.5)	6.2 (1.7)	**1.71**	+	
ASDAS							
Ability-2, Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=25 vs 84)	2.3 (0.8)	3.2 (0.8)	**−1.16**	+	GREEN
CRESPA, López-Medina *et al*[13]	pSpA	PhGA ≥4 vs <4 (n=18 vs 42)	3.6 (1.2)	1.6 (0.7)	**2.32**	+	
		PGA ≥4 vs <4 (n=27 vs 33)	5.5 (2.0)	3.3 (1.1)	**2.30**		
TIPES, Turina *et al*[6]	pSpA	PhGA <40 vs ≥60 (n=6 vs 12)	1.9 (0.7)	3.4 (0.9)	**−1.84**	+	
DAPSA							
CRESPA, López-Medina *et al*[13]	pSpA	PhGA ≥4 vs <4 (n=18 vs 42)	36.5 (21.8)	6.3 (0.7)	**2.38**	+	GREEN
		PGA ≥4 vs <4 (n=27 vs 33)	28.2 (21.3)	4.9 (5.4)	**1.57**		

Continued

Table 2 Continued

Study	Population	External construct	Mean (SD) (group 1)	Mean (SD) (group 2)	SMD	Performance assessment	Synthesis rating

No data available for axSpA with peripheral arthritis.
Performance assessment: '+' when SMD ≥0.50 (good or adequate performance); '–' when SMD <0.50 (poor performance).
Synthesis rating: GREEN: good methods used, in at least two pieces of evidence, with consistent findings of adequate or better performance; RED, good methods used, in at least one or two pieces of evidence, with consistent or questionable findings of inadequate performance; WHITE, no evidence; AMBER, all other situations (not GREEN, RED or WHITE).
*Construct validity (discrimination between known groups): good (in bold), SMD ≥0.80; adequate, SMD between 0.50 and 0.79; low, SMD <0.50.
ASDAS, Axial Spondyloarthritis Disease Activity Score; axSpA, axial spondyloarthritis; BASDAI, Bath Ankylosing Spondylitis Disease Activity Score; CRESPA, Clinical REmission in peripheral SPondyloArthritis; CRP, C reactive protein; DAPSA, Disease Activity in Psoriatic Arthritis; PGA, Patient Global Assessment; PhGA, Physician Global Assessment; pSpA, peripheral spondyloarthritis; SDM, standardised mean difference; SJC66, swollen joint count (66 joints); TIPES, Tnf Inhibition in PEripheral SpondyloArthritis; TJC68, tender joint count (68 joints).

and peripheral arthritis, the population of interest when assessing the domain of peripheral manifestations, are scarce.[30] As a result, the available evidence in this population was limited and largely confined to traditional or generic disease activity measures such as PGA, CRP and BASDAI. Although some of these instruments demonstrated acceptable construct validity and discrimination, none could be supported conclusively due to limited data, underscoring the challenges of assessing peripheral arthritis in axSpA using existing instruments.

In contrast, pSpA offered a broader and more consistent evidence base. Peripheral arthritis represents a core manifestation in this phenotype, leading to a more frequent inclusion in clinical trials and registries and to the systematic assessment of peripheral arthritis outcomes in these settings. As a result, several instruments, particularly composite instruments such as ASDAS, DAPSA and BASDAI, demonstrated good construct validity, responsiveness and clinical trial discrimination in pSpA. These findings suggest, as previously reported,[10 13] that instruments integrating multiple items may better capture the complexity of peripheral arthritis activity than single-item measures.

It is also important to contextualise these findings in relation to previously published recommendations, such as those from the ASAS-COS,[3] which endorses the use of SJC44 and SJC66 for the assessment of peripheral arthritis in axSpA, despite previously reported limitations in their ability to discriminate treatment effects in clinical trials. In the present review, however, it was not possible to evaluate any measurement properties for these instruments. This was primarily due to one of the eligibility criteria applied to our population of interest, whereby analyses in axSpA were restricted to patients with documented peripheral arthritis rather than the broader axSpA population. As a result, the available studies did not provide sufficient or appropriate data to assess the performance of SJC44 or SJC66 within this specific subgroup. These findings do not challenge the conceptual relevance of joint counts as endorsed by ASAS-COS, but rather highlight an important evidence gap when such instruments are applied specifically to patients with axSpA and peripheral arthritis. In contrast, no COS currently exists for pSpA, further underscoring the need for systematic evaluation and selection of appropriate instruments in this population.

A major strength of this review is its strict domain-specific focus on peripheral arthritis. By restricting the axSpA analyses to patients with documented peripheral arthritis, the evaluation was conducted in the population truly relevant to the construct of interest, avoiding extrapolation from broader axSpA populations without peripheral arthritis. In addition, this study included a comprehensive set of outcome measurement instruments potentially suitable to assess peripheral arthritis. And lastly, the application of a rigorous and transparent OMERACT-based methodology, including predefined hypotheses, standardised performance thresholds and structured synthesis rules, ensured a reproducible and purpose-driven assessment aligned with the objectives of the ASAS-SPARADISE initiative.

An important methodological consideration of this study was the a priori decision to require that all instruments demonstrate at least adequate performance for measurement properties other than construct validity. This approach aligns with the objective of the ASAS-SPARADISE project to identify the best-performing instruments and avoids inadvertently favouring instruments for which weaker or 'easier' hypotheses could be formulated and subsequently met despite suboptimal performance. By applying uniform performance expectations across all instruments, the synthesis enabled a fair and meaningful comparison, going beyond the original intent of the OMERACT recommendations, which were not specifically designed for direct comparison of instrument performance.

Despite its strengths, this review has several limitations. First, the number of eligible studies was small, and the number of studies contributing evidence for each instrument and each measurement property was highly heterogeneous, which may have resulted in uneven levels of support across instruments and limited the comparability of their performance. Furthermore, no data were identified for key measurement properties such as test-retest reliability or thresholds of meaning, which are essential for the final evaluation of instruments according to OMERACT recommendations. In addition, PsA-specific studies were not included, as this SLR is part of the ASAS-SPARADISE project focusing on peripheral arthritis in

Table 3 Longitudinal construct validity (responsiveness)

Study	Population	Treatment	Placebo	Interval	No of treatment	No of placebo	SRM, GRI, ES	Performance assessment	Synthesis rating
axSpA with peripheral arthritis									
PGA (0–10)									
Song *et al*[28]	axSpA with peripheral arthritis	TNFi	Placebo	12 weeks	91	NA	GRI: - **SRM: –8.2** **ES: –2.7**	+	AMBER
BASDAI (0–10)									
Song *et al*[28]	axSpA with peripheral arthritis	TNFi	Placebo	12 weeks	91	NA	GRI: - **SRM: –4.9** **ES: –3.8**	+	AMBER
pSpA									
PGA (0–10)									
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**GRI: –1.2** **SRM: –1.1** **ES: –1.8**	+	GREEN
TIPES, Paramarta *et al*[29]	pSpA	ADA	Placebo	12 weeks	20	20	**GRI: –1.5** **SRM: –1.4** **ES: –1.8**	+	
SJC66 (0–66)									
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**GRI: –1.4** **SRM: –1.1** **ES: –0.8**	+	GREEN
TIPES, Paramarta *et al*[29]	pSpA	ADA	Placebo	12 weeks	20	20	**GRI: –1.4** SRM: –0.6 ES: –0.6	+	
TJC68 (0–68)									
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**GRI: –1.1** **SRM: –1.0** ES: –0.7	+	AMBER
TIPES, Paramarta *et al*[29]	pSpA	ADA	Placebo	12 weeks	20	20	GRI: –0.3 SRM: –0.2 ES: –0.2	–	

Continued

Table 3 Continued

Study	Population	Treatment	Placebo	Interval	No of treatment	No of placebo	SRM, GRI, ES	Performance assessment	Synthesis rating
CRP (mg/L)									
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	GRI: −0.6	+/−	AMBER
							SRM: −0.4		
							ES: −0.5		
TIPES, Paramarta *et al*[29]	pSpA	ADA	Placebo	12 weeks	20	20	GRI: −0.2	−	
							SRM: −0.5		
							ES: −0.4		
BASDAI (0–10)									
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**GRI: −1.6**	+	GREEN
							SRM: −1.0		
							ES: −1.5		
TIPES, Paramarta *et al*[29]	pSpA	ADA	Placebo	12 weeks	20	20	**GRI: −1.2**	+	
							SRM: −0.7		
							ES: −0.8		
ASDAS									
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**GRI: −1.6**	+	GREEN
							SRM: −1.1		
							ES: −1.6		
TIPES, Paramarta *et al*[29]	pSpA	ADA	Placebo	12 weeks	20	20	**GRI: −1.2**	+	
							SRM: −0.9		
							ES: −1.1		
DAPSA									
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**GRI: −1.7**	+	AMBER
							SRM: −1.3		
							ES: −1.3		

Longitudinal construct validity: good (in bold), SRM, GRI, ES ≥0.80; adequate, SRM, GRI, ES between 0.50 and 0.79; low, SRM, GRI, ES <0.50.
Performance assessment: '+' when SRM, GRI and/or ES ≥0.50 (good or adequate performance); '−' when SRM, GRI and/or ES <0.50 (poor performance).
Synthesis rating: GREEN: good methods used, in at least two pieces of evidence, with consistent findings of adequate or better performance; RED, good methods used, in at least one or two pieces of evidence, with consistent or questionable findings of inadequate performance; WHITE, no evidence; AMBER, all other situations (not GREEN, RED or WHITE).
ADA, adalimumab; ASDAS, Axial Spondyloarthritis Disease Activity Score; axSpA, axial spondyloarthritis; BASDAI, Bath Ankylosing Spondylitis Disease Activity Score; CRESPA, Clinical REmission in peripheral SPondyloArthritis; CRP, C reactive protein; DAPSA, Disease Activity in Psoriatic Arthritis; ES, effect size; GOL, golimumab; GRI, Guyatt's Responsiveness Index; NA, not available; PGA, Patient Global Assessment; pSpA, peripheral spondyloarthritis; SJC66, swollen joint count (66 joints); SRM, standardised response mean; TIPES, Tnf Inhibition in PEripheral SpondyloArthritis; TJC68, tender joint count (68 joints); TNFi, tumour necrosis factor inhibitor.

Table 4 Clinical trial discrimination

Study	Population	Treatment	Placebo	Interval	No of treatment	No of placebo	SMD	Discrimination in clinical trial	Synthesis rating
axSpA with peripheral arthritis									
PGA (0–10)									
van der Heijde *et al*[27]	axSpA with peripheral arthritis	TNFi	Placebo	12 weeks	37	35	**1.24**	+	AMBER
CRP (mg/L)									
van der Heijde *et al*[27]	axSpA with peripheral arthritis	TNFi	Placebo	12 weeks	37	35	**1.06**	+	AMBER
BASDAI (0–10)									
van der Heijde *et al*[27]	axSpA with peripheral arthritis	TNFi	Placebo	12 weeks	37	35	**1.12**	+	AMBER
ASDAS									
van der Heijde *et al*[27]	axSpA with PA peripheral arthritis	TNFi	Placebo	12 weeks	37	35	**1.49**	+	AMBER
pSpA									
PGA (0–10)									
Ability 2, Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	82	81	−0.47	–	GREEN
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**−0.83**	+	
TIPES, Paramarta *et al*[29], Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	19	19	**−1.16**	+	
SJC66 (0–66)									
Ability 2, Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	82	81	−0.10	–	GREEN
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**−1.26**	+	
TIPES, Paramarta *et al*[29], Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	20	20	−0.68	+	
TJC68 (0–68)									
Ability 2, Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	82	81	−0.50	+	GREEN
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**−1.25**	+	
TIPES, Paramarta *et al*[29], Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	20	20	−0.44	–	
CRP (mg/L)									

Continued

Table 4 Continued

Study	Population	Treatment	Placebo	Interval	No of treatment	No of placebo	SMD	Discrimination in clinical trial	Synthesis rating
Ability 2, Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	82	81	−0.18	–	GREEN
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**−0.86**	+	
TIPES, Paramarta *et al*[29], Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	20	20	−0.53	+	
BASDAI (0–10)									
Ability 2, Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	82	81	−0.50	+	GREEN
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**−0.92**	+	
TIPES, Paramarta *et al*[29], Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	20	20	−0.71	+	
ASDAS									
Ability 2, Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	82	81	−0.63	+	GREEN
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**−0.83**	+	
TIPES, Paramarta *et al*[29], Turina *et al*[6]	pSpA	ADA	Placebo	12 weeks	20/19	20/19	**−1.04**	+	
DAPSA									
CRESPA, López-Medina *et al*[13]	pSpA	GOL	Placebo	12 weeks	40	20	**−1.21**	+	AMBER

Performance assessment: '+' when SMD ≥0.50 (good or adequate performance); '−' when SMD <0.50 (poor performance).
Synthesis rating: GREEN: good methods used, in at least two pieces of evidence, with consistent findings of adequate or better performance; RED, good methods used, in at least one or two pieces of evidence, with consistent or questionable findings of inadequate performance; WHITE, no evidence; AMBER, all other situations (not GREEN, RED or WHITE).
*Discrimination in clinical trial: good (in bold), SMD ≥0.80; adequate, SMD between 0.50 and 0.79; low, SMD <0.50.
ADA, adalimumab; ASDAS, Axial Spondyloarthritis Disease Activity Score; axSpA, axial spondyloarthritis; BASDAI, Bath Ankylosing Spondylitis Disease Activity Score; CRESPA, Clinical REmission in peripheral SPondyloArthritis; CRP, C reactive protein; DAPSA, Disease Activity in Psoriatic Arthritis; GOL, golimumab; PA, peripheral arthritis; PGA, Patient Global Assessment; pSpA, peripheral spondyloarthritis; SJC66, swollen joint count (66 joints); SMD, standardised mean difference; TIPES, Tnf Inhibition in PEripheral SpondyloArthritis; TJC68, tender joint count (68 joints); TNFi, tumour necrosis factor inhibitor.

Table 5 Synthesis rating of the measurement properties for each instrument

	Construct validity		Longitudinal construct validity		Discrimination in clinical trials	
	axSpA with peripheral arthritis	pSpA	axSpA with peripheral arthritis	pSpA	axSpA with peripheral arthritis	pSpA
PGA	AMBER	AMBER	AMBER	GREEN	AMBER	GREEN
SJC66	WHITE	AMBER	WHITE	GREEN	WHITE	GREEN
TJC68	WHITE	AMBER	WHITE	AMBER	WHITE	GREEN
CRP	GREEN	AMBER	WHITE	AMBER	AMBER	GREEN
BASDAI	AMBER	GREEN	AMBER	GREEN	AMBER	GREEN
ASDAS	WHITE	GREEN	WHITE	GREEN	AMBER	GREEN
DAPSA	WHITE	GREEN	WHITE	AMBER	WHITE	AMBER
DAS44	WHITE	WHITE	WHITE	WHITE	WHITE	WHITE

Synthesis rating: GREEN: good methods used, in at least two pieces of evidence, with consistent findings of adequate or better performance; RED, good methods used, in at least one or two pieces of evidence, with consistent or questionable findings of inadequate performance; WHITE, no evidence; AMBER, all other situations (not GREEN, RED or WHITE).
Not available data for the SJC44/TJC44 or DAS44.
Not available data on the assessment of test-retest reliability or thresholds of meaning.
ASDAS, Axial Spondyloarthritis Disease Activity Score; axSpA, axial spondyloarthritis; BASDAI, Bath Ankylosing Spondylitis Disease Activity Score; CRP, C reactive protein; DAPSA, Disease Activity in Psoriatic Arthritis; PGA, Patient Global Assessment; pSpA, peripheral spondyloarthritis; SJC66, swollen joint count (44/66 joints); TJC68, tender joint count (44/68 joints).

axSpA and pSpA. This decision was made to preserve comparability across instruments within the target populations, as inclusion of PsA cohorts would have introduced a disproportionate amount of evidence for certain instruments.

Within the ASAS-SPARADISE project, this SLR represents a subsequent phase following instrument identification and selection based on domain match and feasibility. The next steps will focus on additional analyses of randomised controlled trial datasets and a structured consensus process involving clinical experts and PRPs, alongside further refinement and validation of candidate composite indices.

In conclusion, this SLR highlights substantial gaps in the evidence supporting instruments used to assess disease activity due to peripheral arthritis in SpA, particularly in axSpA. While composite instruments show promise, especially in pSpA, no instrument can currently be recommended unequivocally based on the available evidence. From a practical perspective, although composite instruments such as ASDAS and DAPSA showed more consistent performance in pSpA, the available evidence remains insufficient to support formal recommendation of specific instruments, and in axSpA with peripheral arthritis, the evidence base is too limited to support even pragmatic recommendations. These findings directly inform the next phases of the ASAS-SPARADISE initiative, emphasising the need to strengthen the evidence base on the measurement properties of instruments assessing peripheral disease activity. Moreover, there is a need for prospective studies explicitly designed to evaluate peripheral arthritis and to comprehensively assess measurement properties in accordance with OMERACT standards.

Author affiliations
[1]Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands
[2]Department of Rheumatology, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel
[3]Rheumatology, University of Cordoba, Reina Sofia University Hospital, Maimonides Biomedical Research Institute of Cordoba (IMIBIC), Cordoba, Spain
[4]Department of Medicine, Division of Rheumatology, Maastricht University Medical Centre, Maastricht, The Netherlands
[5]Department of Rheumatology, Fondazione Policlinico Universitario Agostino Gemelli IRCSS, Catholic University of the Sacred Heart, Rome, Italy
[6]Health Economics and Decision Science, Sheffield Centre for Health and Related Research, University of Sheffield, Sheffield, UK
[7]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands
[8]Department of Rheumatology & Clinical Immunology, Amsterdam University Medical Centres, Amsterdam, The Netherlands
[9]Department of Rheumatology, Zuyderland Medical Center, Heerlen, The Netherlands
[10]Department of Rheumatology, Providence Swedish Medical Center, Seattle, Washington, USA
[11]University of Washington, Seattle, Washington, USA
[12]INSERM, INRAE, Centre for Research in Epidemiology and Statistics, Université Paris Cité and Université Sorbonne Paris Nord, Paris, France
[13]Rheumatology Department, Hospital Bichat - Claude-Bernard, Paris, France
[14]Department of Rheumatology, Zuyderland Medical Centre Heerlen, Heerlen, The Netherlands

Contributors SR and AM conceived the study, participated in the design and project coordination and critically revised the manuscript. DC and CL-M performed the screening and data collection. DC performed the statistical analysis and interpretation and drafted the manuscript. All authors critically interpreted the results and revised the manuscript. DC is the guarantor. AM and SR share last authorship.

Funding This SLR was funded by the Assessment of SpondyloArthritis International Society (ASAS) as part of the ASAS-SPARADISE project.

Competing interests DC and LF declare no conflicts of interest. CL-M reports consulting fees from Johnson & Johnson, UCB, Novartis, Eli Lilly and AbbVie; grants from UCB, Eli Lilly and AbbVie and speaking fees from Johnson & Johnson, UCB, Novartis, Eli Lilly and AbbVie. CW reports honoraria from Novartis. AO reports speaker or travel fees from AbbVie, AstraZeneca, Eli Lilly, Johnson & Johnson, GSK, Novartis and UCB. DvdH has received consultancy fees from Alfasigma, BMS, Grey Wolf Therapeutics, Janssen, Takeda and UCB Pharma and is Associate Editor of Annals of the Rheumatic Diseases and Director of Imaging Rheumatology BV. RL reports consultancy fees from AbbVie, UCB, Novartis, Galapagos and Johnson & Johnson. PM has received research grants from AbbVie, Amgen, AstraZeneca, Bristol Myers Squibb, Eli Lilly, Johnson & Johnson, MoonLake, Novartis, Sana, Takeda and UCB; consultancy fees from AbbVie, Amgen, Bristol Myers Squibb, Century, Cullinan, Eli Lilly, Inmagene, Johnson & Johnson, Merck, MoonLake, Novartis, Pfizer, Spyre, SUN Pharma, Takeda and UCB; speaker fees from AbbVie, Amgen, Eli Lilly, Johnson & Johnson, Novartis and UCB. AM reports consultancy fees from AbbVie, Alfasigma, Biogen, Eli Lilly, Johnson & Johnson, MSD, Novartis, Pfizer and UCB; speaker fees from Eli Lilly, Novartis and UCB and research grants from AbbVie, Biogen and UCB. SR reports consultancy fees from AbbVie, Alfasigma, Eli Lilly, Johnson & Johnson, MSD, Novartis, Pfizer, Takeda and UCB; speaker fees from Eli Lilly, Novartis and UCB and research grants from AbbVie, Alfasigma, Eli Lilly, MSD, Novartis, Pfizer and UCB.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Dafne Capelusnik https://orcid.org/0000-0001-9336-0416
Clementina López-Medina https://orcid.org/0000-0002-2309-5837
Louise Falzon https://orcid.org/0000-0002-9449-6056
Désirée van der Heijde https://orcid.org/0000-0002-5781-158X
Robert Landewé https://orcid.org/0000-0002-0577-6620
Philip Mease https://orcid.org/0000-0002-6620-0457
Anna Molto https://orcid.org/0000-0003-2246-1986
Sofia Ramiro https://orcid.org/0000-0002-8899-9087

REFERENCES

1 Rudwaleit M, van der Heijde D, Landewé R, *et al*. The Assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. *Ann Rheum Dis* 2011;70:25–31.
2 Molto A, Sieper J. Peripheral spondyloarthritis: concept, diagnosis and treatment. *Best Pract Res Clin Rheumatol* 2018;32:357–68.
3 Navarro-Compán V, Boel A, Boonen A, *et al*. Instrument selection for the ASAS core outcome set for axial spondyloarthritis. *Ann Rheum Dis* 2023;82:763–72.
4 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;68:18–24.
5 Garrett S, Jenkinson T, Kennedy LG, *et al*. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. *J Rheumatol* 1994;21:2286–91.
6 Turina MC, Ramiro S, Baeten DL, *et al*. A psychometric analysis of outcome measures in peripheral spondyloarthritis. *Ann Rheum Dis* 2016;75:1302–7.
7 Schoels M, Aletaha D, Funovits J, *et al*. Application of the DAREA/DAPSA score for assessment of disease activity in psoriatic arthritis. *Ann Rheum Dis* 2010;69:1441–7.
8 Tillett W, FitzGerald O, Coates LC, *et al*. Composite measures for clinical trials in psoriatic arthritis: testing pain and fatigue modifications in a UK multicenter study. *J Rheumatol* 2021;97:jrheum.201674.
9 Beckers E, Been M, Webers C, *et al*. Performance of 3 composite measures for disease activity in peripheral spondyloarthritis. *J Rheumatol* 2022;49:256–64.
10 Capelusnik D, Lopez-Medina C, van der Heijde D, *et al*. Evaluation of instruments assessing peripheral arthritis in spondyloarthritis: an analysis of the ASAS-PerSpA study. *Ann Rheum Dis* 2025;84:1324–34.
11 Fuchs HA, Brooks RH, Callahan LF, *et al*. A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis. *Arthritis Rheum* 1989;32:531–7.
12 van der Heijde DM, van 't Hof MA, van Riel PL, *et al*. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 1990;49:916–20.
13 López-Medina C, Capelusnik D, Webers C, *et al*. Measurement properties of disease activity instruments in peripheral spondyloarthritis: a post-hoc analysis of the CRESPA trial. *RMD Open* 2025;11:e005525.
14 Beaton D ML GS, Shea B, Tugwell P, eds. *The OMERACT handbook Version 2.1*.
15 Elsman EBM, Mokkink LB, Terwee CB, *et al*. Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *J Clin Epidemiol* 2024;173:111422.
16 Webers C, Ortolan A, Nikiphorou E, *et al*. Peripheral manifestations in spondyloarthritis: a systematic literature review on their assessment and the effect of biological/targeted synthetic DMARDs. *Rheumatology (Oxford)* 2026;65:keag042.
17 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;51:1320–30.
18 Smolen JS, Breedveld FC, Schiff MH, *et al*. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. *Rheumatology (Oxford)* 2003;42:244–57.
19 Aletaha D, Nell VPK, Stamm T, *et al*. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. *Arthritis Res Ther* 2005;7:R796–806.
20 Eberl G, Studnicka-Benke A, Hitzelhammer H, *et al*. Development of a disease activity index for the assessment of reactive arthritis (DAREA). *Rheumatology (Oxford)* 2000;39:148–55.
21 Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 2012;24:69–71.
22 Cohen J. *Statistical power analysis for the behavioral sciences*.2nd edn. Routledge, 1988.
23 Murad MH, Wang Z, Chu H, *et al*. When continuous outcomes are measured using different scales: guide for meta-analysis and interpretation. *BMJ* 2019;364:k4817.
24 Beaton D, Boers M, Bingham CO, *et al*. Summary of findings tables for measurement property reviews: The evolution and application of OMERACT's summary of measurement properties (SOMP) Table. *Semin Arthritis Rheum* 2025;72:152664.
25 Hakkou J, Rostom S, Aissaoui N, *et al*. Comparison of the BASDAI and the miniBASDAI in assessing disease activity in patients with ankylosing spondylitis. *Clin Rheumatol* 2012;31:441–5.
26 Heuft-Dorenbosch L, van Tubergen A, Spoorenberg A, *et al*. The influence of peripheral arthritis on disease activity in ankylosing spondylitis patients as measured with the Bath Ankylosing Spondylitis Disease Activity Index. *Arthritis Rheum* 2004;51:154–9.
27 van der Heijde D, Lie E, Kvien TK, *et al*. ASDAS, a highly discriminatory ASAS-endorsed disease activity score in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68:1811–8.
28 Song IH, Rudwaleit M, Listing J, *et al*. Comparison of the Bath Ankylosing Spondylitis Disease Activity Index and a modified version of the index in assessing disease activity in patients with ankylosing spondylitis without peripheral manifestations. *Ann Rheum Dis* 2009;68:1701–7.
29 Paramarta JE, De Rycke L, Heijda TF, *et al*. Efficacy and safety of adalimumab for the treatment of peripheral arthritis in spondyloarthritis patients without ankylosing spondylitis or psoriatic arthritis. *Ann Rheum Dis* 2013;72:1793–9.
30 van Gaalen FA, Navarro-Compán V, Baraliakos X, *et al*. ASAS recommendations on reporting axial spondyloarthritis clinical trials. *Ann Rheum Dis* 2025;84:1770–8.