The gestalt of spondyloarthritis: From early recognition to long-term imaging outcomes

Sepriano, A.R.

Chapter 3

Performance of the ASAS classification criteria for axial and peripheral spondyloarthritis: a systematic literature review and meta-analysis

Alexandre Sepriano, Roxana Rubio, Sofia Ramiro, Robert Landewé, Désirée van der Heijde

Ann Rheum Dis. 2017 May;76(5):886-890

ABSTRACT

Objective: To summarize the evidence on the performance of the Assessment of SpondyloArthritis international Society (ASAS) classification criteria for axial spondyloarthritis (axSpA) (also imaging and clinical arm separately), peripheral (p)SpA and the entire set, when tested against the rheumatologist's diagnosis ('reference standard').

Methods: A systematic literature review was performed to identify eligible studies. Raw data on SpA diagnosis and classification were extracted or, if necessary, obtained from the authors of the selected publications. A meta-analysis was performed to obtain pooled estimates for sensitivity, specificity, positive and negative likelihood ratios, by fitting random effects models.

Results: Nine papers fulfilled the inclusion criteria (N=5,739 patients). The entire set of the ASAS SpA criteria yielded a high pooled sensitivity (73%) and specificity (88%). Similarly good results were found for the axSpA criteria (sensitivity: 82%; specificity: 88%). Splitting the axSpA criteria in 'imaging arm only' and 'clinical arm only' resulted in much lower sensitivity (30% and 23% respectively) but very high specificity was retained (97% and 94% respectively). The pSpA criteria were less often tested than the axSpA criteria and showed a similarly high pooled specificity (87%) but lower sensitivity (63%).

Conclusions: Accumulated evidence from studies with more than 5,500 patients confirms the good performance of the various ASAS SpA criteria as tested against the rheumatologist's diagnosis.

INTRODUCTION

The Assessment of SpondyloArthritis international Society (ASAS) has developed and validated criteria (ASAS-cohort) for spondyloarthritis (SpA), as well as for their subsets axial (axSpA) and peripheral SpA (pSpA).[1, 2] As in other rheumatic diseases,[3] in the absence of a 'true' gold-standard expert opinion has been used as an external 'anchor' to develop and test the SpA classification criteria. In the original validation studies, the ASAS criteria outperformed other classification criteria.

After their publication, the performance of the ASAS SpA criteria has been tested, all over the world, in different cohorts using the same approach. Some of these cohorts are expectedly similar to the ASAS cohort, while others differ (e.g. setting, inclusion criteria, disease duration). Appropriate data pooling and exploring relevant between-study differences yields unique insights into the criteria performance and applicability in a broad population of patients.

The aim of this systematic literature review is to summarise the published data pertaining to the performance of the ASAS classification criteria for axSpA (also 'imaging arm' and 'clinical arm' separately), pSpA and the entire SpA set when tested against the rheumatologist's diagnosis.

METHODS

Literature search

The scope of the literature search was defined according to the PICO format (patients, intervention, comparator, outcomes; online supplementary table S1).[4] MEDLINE and EMBASE databases were searched without language restriction. Eligible studies were observational cohorts assessing the performance of the ASAS SpA criteria against the rheumatologist's diagnosis, published from March 2009 (date of the axSpA ASAS criteria release) up to August 2016. Studies in which the primary aim was not assessing the performance of the ASAS criteria but still provided enough data to allow such an analysis were also included. In order to retrieve additional references, abstracts from the American College of Rheumatology and European League Against Rheumatism annual conferences (2014 and 2015) were searched. Only studies with full-text available were included, since abstracts neither provide appropriate detail for risk of bias (RoB) assessment nor appropriate data for analysis. Details on the search strategy are provided in online supplementary text 1.

Study selection, data extraction and assessment of risk of bias

Two reviewers (AS and RR) independently screened all titles and abstracts to identify eligible studies fulfilling the inclusion criteria followed by full-text review if appropriate (articles excluded and reason thereof in online supplementary table S2). Both reviewers independently extracted data on the studies' main characteristics, patient characteristics and disease characteristics and criteria performance (i.e. sensitivity, specificity, likelihood ratios of the ASAS criteria against the rheumatologist's diagnosis). Authors of the selected publications were contacted to obtain raw data (2X2 tables necessary for meta-analysis) on criteria performance, when this information was not available in the publication. The same two reviewers

independently assessed the RoB of each study using the Quality Assessment of Diagnostic Accuracy Studies 2 tool (QUADAS-2).[5] Disagreements were resolved by consensus and a third review-author was involved when necessary (DvdH).

Data analysis

Pooled sensitivity and specificity were estimated by random-effects bivariate generalised linear mixed models. Parameter estimates from each model were used to derive the positive likelihood ratio (LR+) and negative LR (LR-) and 95% CIs. In case of limited data, two univariate random-effects models were used by assuming no correlation between sensitivity and specificity.[6] Separate models were fit for the axSpA criteria, the pSpA criteria and the SpA criteria. The 'imaging arm' and the 'clinical arm' of the axSpA criteria were analysed separately using two approaches: (i) considering all patients that fulfil each arm irrespective of fulfilment of the other; and (ii) considering patients that fulfil one arm exclusively.

A series of sensitivity analyses was performed (whenever possible and appropriate) to assess the effect of the following on the criteria performance: (i) target population (original validation study inclusion criteria vs different inclusion criteria); (ii) risk of bias (low vs high RoB); (iii) study's main aim (criteria performance assessment vs other); (iv) setting (hospital vs community); and (v) symptom duration (< 2 years vs ≥ 2 years).

All analyses were performed in Stata V.12.1. The Cochrane Collaboration's Review Manager Software V.5.3 was used to build forest plots.

RESULTS

Of 1,486 screened articles (after deduplication) 9 fulfilled the inclusion criteria (table 1).[1, 2, 7-13] All but one study were considered to be at low RoB (see online supplementary table S3). In total 5,739 patients (range: 157-1,210) had been included, and 2,936 (51.2%; range: 25.2%-69.4%) had been diagnosed by the rheumatologist as SpA.

Study populations

This literature review included the original studies in which the axSpA criteria and the pSpA criteria (also the entire set) were validated.[1, 2] In addition, five studies assessed the ASAS axSpA criteria,[8-10, 12, 13] one study assessed the pSpA criteria,[7] and one study the SpA criteria (providing separate data also for the axSpA and pSpA criteria).[11] Raw data on the criteria performance were obtained from all, except two studies.[12, 13]

In table 1, main patient characteristics and disease characteristics per study are shown. The majority of the studies assessing the axSpA criteria had similar inclusion-criteria compared with the original validation study.[8-10, 12, 13] However, in one study inflammatory back pain was required, or otherwise patients had to have one additional SpA feature.[11]

Table 1. Main study characteristics

Study reference	Cohort	Sample size	Population (inclusion criteria)			SpA‡ prevalence N (%)	Males (%)	Disease duration	HLA-B27 (%)	mNY (%)	MRI-SI (%)	Risk of bias
			Symptoms	Age symptoms onset	Symptoms duration (years)							
Rudwaleit 2009 [1]	ASAS	649	Any CBP (> 3 months)	< 45	No limit	391 (60.2)	52.4	6.1 (7.6) years	65.9	29.7	64.7Ω	Low
Rudwaleit 2011 [2]	ASAS	266	Arthritis/ enthesitis/dactylitis	< 45	No limit	176 (66.2)	63.1	10.3 (18.6) months	47.2	19.5	44.0Ω	Low
van den Berg 2012[7]	EAC	302*	Peripheral arthritis	NR	< 2	76 (25.2)	48.7	22.8 (37.3) weeks	47.5	34.6	NR	Low
Moltó 2013 [8]	DECLIC	1,210	Any CBP (> 3 months)	< 45	No limit	425 (35.1)	56.0	1.08 years (0.16, 3.90)**	60.1	49.2	25.2Ω	Low
van den Berg 2013[9]	SPACE	157	Any CBP (> 3 months)	< 45	< 2	65 (41.4)	48.3	13.4 (7.7) months	79.7	18.3	41.7∑	Low
Strand 2013[10]	USA	816	Any CBP (> 3 months)	< 45	No limit	491 (60.2)	68.0	NR	NR	NR	NR	Low
Tomero 2014[11]	ESPERANZA	775	IBP /asymmetrical arthritis†	< 45	< 2	538 (69.4)	61.0	12.1 (6.8) months	56.0	19.0	24.0∑	Low
Lin 2014[12]	China	867	Any CBP (> 3 months)	< 45	No limit	455 (52.5)	68.1	2.6 (3.2) years	72.3	NA	70.5∑	High
Deodhar 2016[13]	PROSpA	697	Any CBP†† (> 3 months)	< 45	No limit	319 (45.8)	49.8	14.0 years	48.9	31.7	37.9∑	Low

*Number of patients used in the analysis from a total 2011 patients included in the cohort; ‡ According to the rheumatologist's diagnosis (for van der Berg 2012, prevalence of pSpA was calculated considering the 302 patients included in the analysis (prevalence in entire cohort: 76/2011= 3.8%); † in absence of IBP or arthralgia only (without arthritis), one additional SpA feature required: psoriasis, inflammatory bowel disease, uveitis, radiographic sacroiliitis, positivity for HLA-B27 or a family history of SpA; †† and ≥1 of the following: HLA–B27 positivity, current IBP, and prior imaging (MRI or radiographic) evidence of sacroiliitis.**median (interquartile range); Ω typical signs of active inflammation (no formal definition); ∑ ASAS/OMERACT definition. For longitudinal studies the baseline characteristics are shown. Characteristics are referring to SpA patients according to the rheumatologist except for: van den Berg 2013 (according to ASAS axSpA criteria) and Strand 2013 (SpA and no-SpA); ASAS, Assessment of SpondyloArthritis international Society; SPACE, SpondyloArthritis Caught Early; EAC, Early Arthritis Clinic; PROSpA, Prevalence of Axial SpA; USA, United States of America; SpA, spondyloarthritis; SI, sacroiliitis; mNY, modified New York criteria; MRI, magnetic resonance imaging; CBP, chronic back pain, IBP, inflammatory back pain; NA, not applicable; NR, not reported.

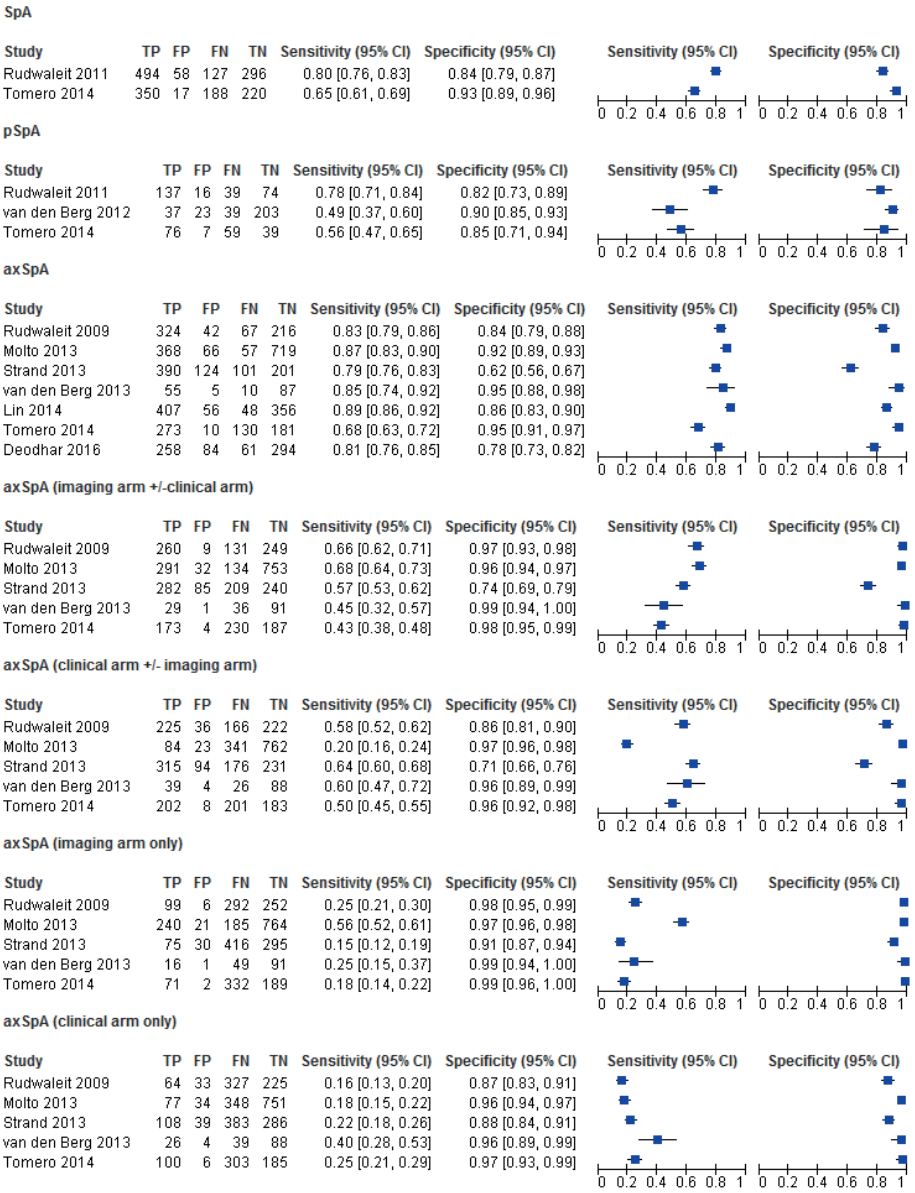

Figure 1. Performance of the ASAS SpA classification criteria across studies. ASAS, Assessment of SpondyloArthritis international Society; axSpA, axial spondyloarthritis; pSpA, peripheral spondyloarthritis; CI, confidence interval; TP, true positives, FP, false positives; FN, false negatives; TN, true negatives.

Two studies assessing the pSpA criteria used different inclusion criteria as compared with the ASAS cohort. In one study, only patients with peripheral arthritis were included (excluding those with only enthesitis or dactylitis),[7] while in another study patients had to have typical SpA arthritis (asymmetrical and predominantly in lower limbs) or arthralgia associated with one additional SpA feature (not including enthesitis and dactylitis).[11]

Performance of the ASAS SpA classification criteria

The sensitivity and specificity of the various criteria for each individual study is shown in figure 1 and the results of the meta-analysis in table 2. The ASAS SpA criteria were assessed in two studies (N=1,750) yielding a high pooled sensitivity and specificity (73%; 88%).[2, 11]

Three studies (N=749) assessed the ASAS pSpA criteria.[2, 7, 11] Although specificity was consistently high (82%-90%; pooled: 87%), sensitivity was much lower in the two studies with inclusion criteria differing from the original validation study (49%-56% *vs* 78%; pooled: 62%).

Seven studies, with 4,990 patients in total, together generated a very high pooled sensitivity and specificity (82% and 87% respectively) for the axSpA criteria with little variation across studies.[1, 8-13] The pooled sensitivity of the 'imaging arm' +/- 'clinical arm' and 'clinical arm' +/- 'imaging arm' was 57% and 49% respectively (26% and 23% when considering patients fulfilling each arm exclusively). High estimates of pooled specificity were found for both 'arms' irrespective of the definition (range: 92%-97%). However, the LR+ of the 'imaging arm' only was higher as compared with the 'clinical arm' only (9.6 *vs* 3.6).

Sensitivity analyses

The ASAS axSpA criteria performed similarly well irrespective of the population in which they were applied, the setting, symptom duration, RoB and study's main aim (sensitivity (range): 78%-85%, specificity (range): 80%-93%; online supplementary table S4). Due to a scarcity of data, sensitivity analyses for the 'imaging arm' and 'clinical arm' of the axSpA criteria, the pSpA criteria and the SpA criteria could not be performed.

Table 2. Results of the meta-analysis (pooled estimates)

	N patients (studies)	LR + (95% CI)	LR – (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
ASAS SpA criteria¥	1,750 (2 studies[2,11])	6.3 (3.2; 12.4)	0.31 (0.13; 0.70)	0.73 (0.47; 0.89)	0.88 (0.81; 0.93)
ASAS pSpA criteria*	749 (3 studies[2,7,11])	4.7 (3.5; 6.3)	0.43 (0.30; 0.62)	0.62 (0.47; 0.76)	0.87 (0.81; 0.91)
ASAS axSpA criteria*	4,990 (7 studies[1,8-13])	6.2 (3.7; 10.5)	0.20 (0.16; 0.27)	0.82 (0.77; 0.86)	0.87 (0.78; 0.92)
axSpA criteria* (imaging arm +/-clinical arm)	3,426 (5 studies[1,8-11])	13.6 (4.8; 38.7)	0.45 (0.37; 0.56)	0.57 (0.47; 0.66)	0.96 (0.88; 0.99)
axSpA criteria* (clinical arm +/- imaging arm)	3,426 (5 studies[1,8-11])	6.0 (2.9; 12.4)	0.56 (0.43; 0.72)	0.49 (0.34; 0.64)	0.92 (0.82; 0.96)
axSpA criteria* (imaging arm only)	3,426 (5 studies[1,8-11])	9.6 (4.4; 20.7)	0.76 (0.64; 0.90)	0.26 (0.16; 0.40)	0.97 (0.94; 0.99)
axSpA criteria* (clinical arm only)	3,426 (5 studies[1,8-11])	3.6 (2.0; 6.4)	0.83 (0.75; 0.91)	0.23 (0.17; 0.29)	0.94 (0.90; 0.96)

* Bivariate random-effects generalised mixed model. ¥ Univariate random-effects logistic regression. ASAS, Assessment of SpondyloArthritis international Society; axSpA, axial spondyloarthritis; pSpA, peripheral spondyloarthritis; CI, confidence interval; LR: likelihood ratio.

DISCUSSION

Pooled data from eight cohorts (including more than 5,500 patients) confirm the good performance of the various ASAS SpA classification criteria as tested against the rheumatologist's diagnosis. This review confirms that splitting the 'arms' of the axSpA criteria results in loosing sensitivity while retaining specificity, which indicates that the full set of axSpA criteria is the preferred set.

While the pooled specificity for both the axSpA criteria and pSpA criteria was similarly high (87% for both), the pooled sensitivity for the pSpA criteria was much lower than that for the axSpA criteria (62% vs 82%). This difference may be explained by restrictive inclusion criteria. Unlike the ASAS cohort the Early Arthritis Clinic cohort only included patients with arthritis, and not those with dactylitis only or enthesitis only.[7] Similar 'restrictions' were seen in the ESPERANZA-cohort.[11] The low sensitivity found in these studies suggests that both enthesitis and dactylitis are considered by the rheumatologists as fitting the pattern of pSpA, which adds to the credibility of the ASAS pSpA criteria (that include these presentations).

Sensitivity analyses have shown the 'robustness' of the axSpA criteria when applied in different settings (hospital and community), in patients with short (< 2 years) and long (≥2 years) symptom duration and in different populations.

Not surprisingly, the splitting of the axSpA criteria into two 'arms' compromised sensitivity, but retained (very high) specificity, if patients that fulfil each 'arm' irrespective of fulfilment of the other were considered, and if those that fulfil one 'arm' exclusively were analysed. The larger LR+ for the 'imaging arm' as compared with the 'clinical arm' reflects the rheumatologist's reliance on positive imaging findings. The prospective validation of the ASAS criteria against the rheumatologist's diagnosis after >4 years of follow-up in the ASAS-cohort has shown that both 'arms' still properly discriminate between axSpA and non-axSpA.[14] Another prospective study has also suggested the arms' low specificity when tested against radiographic sacroiliitis (modified New York criteria) after 8 years of follow-up ('imaging arm': 22%; 'clinical arm': 56%), but the setting in this study was a prognostic rather than a diagnostic setting and figures are difficult to interpret.[15]

In conclusion, the ASAS axSpA and pSpA criteria have shown to perform well in patients included in several cohorts all over the world, as assessed by rheumatologists. This review does not give resolution to the applicability of the ASAS classification criteria in primary care, since such a setting had not been tested. It is important to realise that the criteria's performance depends entirely on the prevalence of SpA in the underlying population (pretest likelihood).

SUPPLEMENTARY DATA

Supplementary data are published online on the website of the Annals of the Rheumatic Diseases

REFERENCES

1. Rudwaleit M, van der Heijde D, Landewé R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis. 2009 Jun; 68(6):777-783.
2. 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 Jan; 70(1):25-31.
3. Aggarwal R, Ringold S, Khanna D, et al. Distinctions between diagnostic and classification criteria? Arthritis Care Res (Hoboken). 2015 Jul; 67(7):891-897.
4. Sackett D WR, Rosenberg W, et al. Evidence-based medicine: how to practice and teach EBM. London: Churchill Livingstone, 1997.
5. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18; 155(8):529-536.
6. Takwoingi Y, Guo B, Riley RD, et al. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. Stat Methods Med Res. 2017 Aug; 26(4):1896-1911.
7. van den Berg R, van Gaalen F, van der Helm-van Mil A, et al. Performance of classification criteria for peripheral spondyloarthritis and psoriatic arthritis in the Leiden Early Arthritis cohort. Ann Rheum Dis. 2012 Aug; 71(8):1366-1369.
8. Molto A, Paternotte S, Comet D, et al. Performances of the Assessment of SpondyloArthritis International Society axial spondyloarthritis criteria for diagnostic and classification purposes in patients visiting a rheumatologist because of chronic back pain: results from a multicenter, cross-sectional study. Arthritis Care Res (Hoboken). 2013 Sep; 65(9):1472-1481.
9. 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). 2015 Jul; 54(7):1336.
10. Strand V, Rao SA, Shillington AC, et al. Prevalence of axial spondyloarthritis in United States rheumatology practices: Assessment of SpondyloArthritis International Society criteria versus rheumatology expert clinical diagnosis. Arthritis Care Res (Hoboken). 2013 Aug; 65(8):1299-1306.
11. Tomero E, Mulero J, de Miguel E, et al. Performance of the Assessment of Spondyloarthritis International Society criteria for the classification of spondyloarthritis in early spondyloarthritis clinics participating in the ESPERANZA programme. Rheumatology (Oxford). 2014 Feb; 53(2):353-360.
12. Lin Z, Liao Z, Huang J, et al. Evaluation of Assessment of Spondyloarthritis International Society classification criteria for axial spondyloarthritis in Chinese patients with chronic back pain: results of a 2-year follow-up study. Int J Rheum Dis. 2014 Sep; 17(7):782-789.
13. Deodhar A, Mease PJ, Reveille JD, et al. Frequency of Axial Spondyloarthritis Diagnosis Among Patients Seen by US Rheumatologists for Evaluation of Chronic Back Pain. Arthritis Rheumatol. 2016 Jul; 68(7):1669-1676.
14. Sepriano A, Landewé R, van der Heijde D, et al. Predictive validity of the ASAS classification criteria for axial and peripheral spondyloarthritis after follow-up in the ASAS cohort: a final analysis. Ann Rheum Dis. 2016 Jun; 75(6):1034-1042.
15. Aydin SZ, Maksymowych WP, Bennett AN, et al. Validation of the ASAS criteria and definition of a positive MRI of the sacroiliac joint in an inception cohort of axial spondyloarthritis followed up for 8 years. Ann Rheum Dis. 2012 Jan; 71(1):56-60.