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

Sepriano, A.R.

Chapter 2

Predictive validity of the ASAS classification criteria for axial and peripheral spondyloarthritis after follow-up in the ASAS cohort: a final analysis

Alexandre Sepriano, Robert Landewé, Désirée van der Heijde, Joachim Sieper, Nurullah Akkoc, Jan Brandt, Jürgen Braun, Eduardo Collantes-Estevez, Maxime Dougados, Oliver Fitzgerald, Feng Huang, Jieruo Gu, Yesim Kirazli, Walter P Maksymowych, Helena Marzo-Ortega, Ignazio Olivieri, Salih Ozgocmen, Euthalia Roussou, Salvatore Scarpato, Inge J Sørensen, Rafael Valle-Oñate, Filip Van den Bosch, Irene van der Horst-Bruinsma, Ulrich Weber, James Wei, Martin Rudwaleit, on behalf of ASAS

Ann Rheum Dis. 2016 Jun;75(6):1034-42

ABSTRACT

Objective: To establish the predictive validity of the Assessment of SpondyloArthritis international Society (ASAS) spondyloarthritis (SpA) classification criteria.

Methods: 22 centres (N=909 patients) from the initial 29 ASAS centres (N=975) participated in the ASAS-cohort follow-up study. Patients had either chronic (>3 months) back pain of unknown origin and age of onset below 45 years (N=658) or peripheral arthritis and/or enthesitis and/or dactylitis (N=251). At follow-up, information was obtained at a clinic visit or by telephone. The positive predictive value (PPV) of the baseline classification by the ASAS criteria was calculated using rheumatologist's diagnosis at follow-up as external standard.

Results: In total, 564 patients were assessed at follow-up (345 visits; 219 telephone) with a mean follow-up of 4.4 years (range: 1.9; 6.8) and 70.2% received a SpA diagnosis by the rheumatologist. 335 patients fulfilled the axial SpA (axSpA) or peripheral SpA (pSpA) criteria at baseline and of these, 309 were diagnosed SpA after follow-up (PPV SpA criteria: 92.2%). The PPV of the axSpA and pSpA criteria was 93.3% and 89.5% respectively. The PPV for the 'clinical arm only' was 88.0% and for the 'clinical arm' ± 'imaging arm' 96.0%, for the 'imaging arm only' 86.2% and for the 'imaging arm' +/- 'clinical arm' 94.7%. A series of sensitivity analyses yielded similar results (range: 85.1–98.2%).

Conclusions: The PPV of the axSpA and pSpA criteria to forecast an expert's diagnosis of 'SpA' after more than 4 years is excellent. The 'imaging arm' and 'clinical arm' of the axSpA criteria have similar predictive validity and are truly complementary.

INTRODUCTION

The term spondyloarthritis (SpA) encompasses a group of chronic rheumatic diseases sharing common clinical, genetic and imaging features. SpA patients can be divided (with some overlap) according to their clinical presentation into axial SpA (axSpA), for those with predominantly axial symptoms, and peripheral SpA (pSpA) if peripheral manifestations dominate the clinical picture.

It has become evident that the requirement for the presence of radiographic sacroiliitis, as defined by the modified New York criteria (mNY),[1] leads to a delayed diagnosis of axSpA.[2, 3] Magnetic resonance imaging (MRI) has been proven to detect inflammation in the sacroiliac joints (SIJ) early in the disease course, far before structural changes are seen in radiographs.[4, 5] These findings have initiated the aggregation of patients with non-radiographic (nr-axSpA) and radiographic axial SpA (r-axSpA – also known as ankylosing spondylitis) patients, under one 'umbrella' term being axSpA. The Assessment of SpondyloArthritis international Society (ASAS) has published criteria for axSpA and pSpA.[6-8]

Since their release, the ASAS criteria have been implemented worldwide. In the original validation studies,[7, 8] the new ASAS criteria proved to reflect the current perception of what 'SpA looks like' ('gestalt') better than the European Spondyloarthropathy Study Group[9] (ESSG) and Amor[10] criteria when tested against the expert's diagnosis. After that, the ASAS axSpA criteria,[11-13] the pSpA[14] criteria and the entire set[15, 16] have consistently shown good criterion and construct validity.

However, it has been argued that the ASAS axSpA criteria are too loose and include patients without SpA (mislabelling)[17]: Patients with nr-axSpA are more often women and have lower C-reactive protein (CRP) levels when compared with patients with r-axSpA.[18-20] Recent studies have suggested that the 'clinical arm' could drive such differences.[11, 21] However, the same studies have also shown that patients classified by the 'imaging arm' and 'clinical arm' are similar regarding the presence of SpA features and burden of clinical symptoms. Moreover, it has been hypothesised that the male gender is a risk factor for the development of radiographic damage,[2] and it has been shown that the elevated CRP drives progression to r-axSpA,[22] thereby explaining, at least partially, these differences in the nr-axSpA subpopulation.

While previous validation studies have shown high specificity of the ASAS criteria, mostly in cross-sectional analyses (except for one follow-up study in a Chinese population[12]), these studies do not give resolution with regard to predictive validity: will patients with a classification of axSpA still be considered as having a diagnosis of SpA after some years.

A similar question pertains to the pSpA criteria. Some claim that an entry symptom of arthritis may easily include patients with other forms of early arthritis,[23] and that the entry symptom of 'enthesitis' may evoke confusion with non-inflammatory diseases.[24]

Hence, it had been upfront decided that patients from the validation cohort would be reassessed after 5 years. Therefore, the aim of this study was to establish the predictive validity of an ASAS classification - either as axSpA (also split by imaging and clinical arm) or pSpA - by comparing such a classification with the final diagnosis after follow-up in the original ASAS-cohort.

METHODS

Study design

The ASAS-cohort is an international, multicentre, prospective study. From November 2005 to January 2009, rheumatologists from 29 ASAS centres worldwide have included 975 consecutive patients who first presented for diagnostic work-up. To be included, eligible patients had to fulfil one of two criteria: (1) 'axial population': chronic (>3 months) back pain of unknown origin (no definite diagnosis) with an age of onset below 45 years, with or without peripheral symptoms; (2) 'peripheral population': patients with peripheral arthritis and/or enthesitis and/or dactylitis and the absence of current back pain with suspicion of SpA but no definitive diagnosis.[7, 8]

All patients were assessed at baseline and after a mean follow-up of 4.4 years (range: 1.9-6.8). Of the 29 original ASAS centres, 22 participated in the follow-up corresponding to 909 of the original 975 patients. At follow-up, these patients were contacted to assess their willingness to attend the follow-up visit. A total of 345/909 physically attended the follow-up visit and 219 provided only information via telephone (figure 1). Of the 22 participating centres, 10 had ≥75% patients with follow-up data available (N=291), while 12 had <75% (N=273).

The current Good Clinical Practice guidelines were followed and the study has been approved by the local ethics committees. All patients provided written informed consent at the baseline visit that also included the follow-up visit.

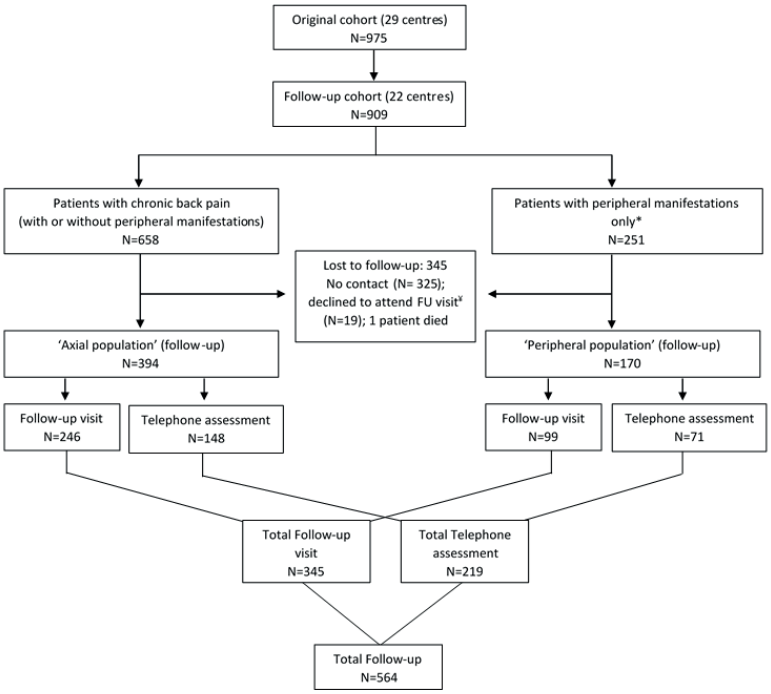

Figure 1. Follow-up of the Assessment of SpondyloArthritis international Society (ASAS) Cohort. *Patients with undiagnosed peripheral arthritis, and/or enthesitis, and/or dactylitis and absence of current back pain. ¥ And did not provide information via telephone. FU, follow-up.

Data collection

Clinical, laboratory and imaging data were collected for all patients at baseline. The same assessments (except for HLA-B27 typing) were also performed at follow-up for patients attending the follow-up visit. For these patients, the rheumatologist provided a diagnosis at both time-points (not necessarily the same clinician). Patients assessed by telephone at follow-up had also received a diagnosis by the rheumatologist at baseline, while the follow-up diagnosis was self-reported: Patients were asked whether during follow-up they had received a diagnosis that was different from the diagnosis based on the first study visit. Details on the methods used for data collection were previously published and were similar for both the 'axial population' and 'peripheral population'.[7, 8] A summary of these methods is provided in the online supplementary material -appendix 1.

Statistical analysis

All patients with follow-up data available were considered in the analysis (N=564). The rheumatologist's diagnosis (SpA *vs* no-SpA) at follow-up was used as external reference (combining the follow-up visit and telephone diagnosis), against which the baseline ASAS-classification was tested. The rheumatologists did not have access to the patients' baseline classification status according to the ASAS criteria. Missing values for baseline SpA features were interpreted as being absent. For patients assessed at follow-up, the level of confidence about the diagnosis was recorded on a numerical rating scale from 0 (not confident at all) to 10 (very confident).

The predictive validity of the baseline ASAS-classification for axSpA and pSpA was analysed in terms of positive predictive value (PPV) and negative predictive value (NPV). Similarly, the entire set was assessed combining the axSpA criteria (applied in patients with predominant back pain with/without peripheral manifestations) with the pSpA criteria (applied in patients with currently exclusive peripheral manifestations). The 'imaging arm' and the 'clinical arm' of the axSpA criteria were analysed separately using two approaches: (1) considering all patients who fulfil each arm irrespective of fulfilment of the other; and (2) considering patients who fulfil one arm exclusively.

In addition, the ASAS criteria predictive validity was assessed separately for countries with a low *versus* high background prevalence of HLA-B27 (median prevalence used as cut-off).

Sensitivity analyses

Three sensitivity analyses were performed to assess the possible effects of the following on the predictive validity results: (1) missing baseline data, (2) telephone *versus* physical visit, and (3) completeness of reassessed patients per centre. First, an analysis was performed on patients with complete data on all SpA features at baseline (N=345)]; Second, an analysis only on patients who physically attended the follow-up visit (N=345) was done. By chance the same number of patients, but different patients (n=345), were included in these analyses; finally, a '≥75%

complete follow-up analysis was done, including only patients from centres with high levels of follow-up participation (N=291).

Data analysis was performed using STATA V.12.1.

RESULTS

Baseline characteristics

Table 1 describes the baseline characteristics comparing patients with/without follow-up data available, and comparing patients assessed at the follow-up visit or by telephone. These groups were globally comparable.

At the end of follow-up 396 (70.2%) patients were diagnosed as SpA (257 (64.9%) in the follow-up visit group and 139 (35.1%) in the telephone group), while 168 (29.8%) received either another diagnosis or no diagnosis at all. Among the 'axial population' 280 (71.1%) were diagnosed as axSpA, while among the 'peripheral population' 116 (68.2%) got a diagnosis of pSpA. Table 2 shows the baseline characteristics of all patients with SpA and split for axSpA and pSpA. Additional information on baseline characteristics is provided in online supplementary tables S1 and S2.

Change in diagnosis and symptoms from baseline to follow-up

Among the 394 patients from the 'axial population', the baseline diagnosis was changed in 37 (30/246 (12.2%) in the follow-up visit group and 7/148 (4.7%) in the telephone group). Of these 394 patients, 246 were assessed at the follow-up visit (figure 1) providing information on the predominance of manifestations. The majority (185; 75.2%) maintained the same symptomatic pattern they had at baseline (i.e. back pain +/- peripheral manifestations), with few presenting with only peripheral symptoms (15; 6.1%) and 46 (18.7%) becoming asymptomatic. The majority of these asymptomatic patients were treated during follow-up (41; 89.1%) and half (23; 50.0%) were still receiving medication at the follow-up visit (NSAIDs: 10 (43.5%); methotrexate: 2 (8.7%); tumour necrosis factor inhibitors (TNFi): 6 (26.1%); and 5 (21.7%) different combinations).

Of the 170 patients from the 'peripheral population', 19 (11.1%) had their diagnosis changed between baseline and follow-up [18/99 (18.2%) in the follow-up visit group and 1/71 (1.4%) in the telephone group]. Of these 170 patients, 99 were assessed at the follow-up visit and only 31 (31.3%) maintained exclusive peripheral symptoms at follow-up, while 37 (37.4%) developed back pain and 31 (31.3%) became asymptomatic. Similar to the 'axial population', also the majority of asymptomatic patients (22; 71.0%) were treated during follow-up, and 16 (51.6%) still needed treatment at the follow-up visit (NSAIDs: 7 (43.8%); methotrexate: 1 (6.3%); TNFi: 3 (18.8%); and 5 (31.3%) different combinations).

In total, 77 (22.3%) patients were asymptomatic at the follow-up visit. On the other hand, 109 (31.6%) patients developed at least 1 new SpA feature compared with baseline.

Table 1. Baseline characteristics comparing patients with and without follow-up data available and comparing patients assessed by telephone to those with follow-up visit

	Follow-up data availability				Follow-up assessment			
					Follow-up visit (N=345)		Telephone assessment (N=219)	
	All (N=909)	No data available (N=345)	Data available (N=564)	p-value*	Axial (N=246)	Peripheral (N=99)	Axial (N=148)	Peripheral (N=71)
Age (years) at baseline, mean (SD)	34.1 (11.5)	36.1 (11.6)	32.8 (11.2)	<0.001	33.2 (10.9)	34.5 (10.5)	30.9 (11.4)	33.3 (12.4)
Age (years) at onset of back pain£, mean (SD)	26.3 (9.1)	27.0 (9.5)	25.8 (8.8)	0.085	26.0 (8.8)	NA	25.4 (8.9)	NA
Onset of back pain before 40 years£, n (%)	597 (90.7)	230 (87.1)	367 (93.2)	0.009	230 (93.5)	NA	137 (92.6)	NA
Duration of back pain in years£, mean (SD)	7.4 (9.3)	8.9 (10.3)	6.3 (8.3)	<0.001	7.0 (8.9)	NA	5.2 (7.2)	NA
Male gender, n (%)	432 (47.6)	144 (41.9)	288 (51.1)	0.007	111 (45.1)	55 (55.6)	73 (49.3)	24 (33.8)
SpA clinical diagnosis at baseline, n (%)	574 (63.2)	176 (51.0)	398 (70.6)	<0.001	185 (75.2)	72 (72.7)	96 (64.9)	45 (63.4)
ASAS criteria for SpA†, n (%)	506 (55.7)	171 (49.6)	335 (59.4)	0.004	NA	NA	NA	NA
ASAS criteria for axSpA£, n (%)	367 (55.8)	127 (48.1)	240 (60.9)	0.001	158 (64.2)	NA	82 (55.4)	NA
ASAS criteria for pSpA‡, n (%)	139 (55.4)	44 (54.3)	95 (55.9)	0.816	NA	58 (58.6)	NA	37 (52.1)
Number of SpA featuresΩ, mean (SD)	2.4 (1.4)	2.2 (1.4)	2.5 (1.4)	0.020	2.6 (1.5)	2.3 (1.1)	2.4 (1.6)	2.2 (1.0)
Presence of 2 or more SpA featuresΩ, n (%)	657 (72.3)	240 (69.6)	417 (73.9)	0.153	183 (74.4)	74 (74.8)	104 (70.3)	56 (78.9)
Definite radiographic sacroiliitis¥£, n (%)	122 (18.6) (N=657)	48 (18.2) (N=264)	74 (18.8) (N=392)	0.834	53 (21.5)	NA	21 (14.3)	NA
Active inflammation of sacroiliac joints, MRI£, n (%)	208 (41.0) (N=507)	57 (28.9) (N=197)	151 (48.7) (N=310)	<0.001	101 (51.3) (N=197)	NA	50 (44.3) (N=113)	NA
HLA - B27, n (%)	407 (46.2) (N=881)	132 (39.5) (N=334)	275 (50.3) (N=547)	0.002	137 (56.2) (N=244)	32 (33.3) (N=96)	80 (57.1) (N=140)	26 (38.8) (N=67)
Elevated CRP, n (%)	345 (38.0)	116 (33.6)	229 (40.6)	0.035	80 (32.5)	51 (51.5)	57 (38.5)	41 (57.8)
IBP (according to experts definition)£, n (%)	415 (63.1)	148 (56.1)	267 (67.8)	0.002	175 (71.1)	NA	92 (62.2)	NA
Peripheral arthritis past or present, n (%)	454 (49.9)	155 (44.9)	299 (53.0)	0.018	95 (38.6)	89 (89.9)	49 (33.1)	66 (93.0)
Enthesitis past or present, n (%)	376 (41.4)	130 (37.7)	246 (43.6)	0.078	101 (41.1)	49 (49.5)	61 (41.2)	35 (49.3)
Uveitis past or present, n (%)	69 (7.6)	30 (8.7)	39 (6.9)	0.325	29 (11.8)	3 (3.0)	6 (4.1)	1 (1.4)
Dactylitis past or present, n (%)	81 (8.9)	26 (7.5)	55 (9.8)	0.255	15 (6.1)	22 (22.2)	7 (4.7)	11 (15.5)
Psoriasis past or present, n (%)	68 (7.5)	32 (9.3)	36 (6.4)	0.108	17 (6.9)	10 (10.1)	6 (4.1)	3 (4.2)
IBD past or present, n (%)	23 (2.5)	7 (2.0)	16 (2.8)	0.452	9 (3.7)	3 (3.0)	2 (1.4)	2 (2.8)
Active inflammation of the spine, MRI£, n (%)	46 (18.3) (N=251)	11 (11.1) (N=99)	35 (23.0) (N=152)	0.017	26 (26.3) (N=99)	NA	9 (17.0) (N=53)	NA

* Chi-square test for categorical variables and the independent samples t-test for continuous variables. £ Only applicable in patients from the 'axial population' (N=658 at baseline and N=394 at follow-up). † Combination of ASAS criteria for axSpA and criteria for pSpA. ‡ Only applicable in patients from the 'peripheral population' (N=251 at baseline and N=170 at follow-up). Ω Features included: Inflammatory back pain (IBP) according to experts' definition, arthritis (ever), heel enthesitis (ever), dactylitis (ever), uveitis (ever), psoriasis (ever), inflammatory bowel disease (IBD) (ever), good response to NSAIDs, family history of spondyloarthritis (SpA), elevated CRP. ¥ ≥ grade 2 bilateral or ≥ grade 3 unilateral. axSpA, axial spondyloarthritis; pSpA, peripheral spondyloarthritis; IBD, inflammatory bowel disease; NA, not applicable; MRI, magnetic resonance imaging. Although imaging of the axial skeleton was performed.

Table 2. Baseline characteristics comparing patients with SpA and no-SpA according to the rheumatologist's follow-up diagnosis in each study population

	All (N=564)		Axial Population (N=394)		Peripheral Population (N=170)	
	SpA (N=396)	No-SpA (N=168)	axSpA (N=280)	No-SpA (N=114)	pSpA (N=116)	No-SpA (N=54)
Age (years) at baseline, mean (SD)	31.2 (11.1)	36.7 (10.5)	30.7 (10.9)	36.1 (10.8)	32.1 (11.6)	37.9 (9.6)
Age (years) at onset of back pain[£], mean	25.0 (8.5)	27.7 (9.3)	25.0 (8.5)	27.7 (9.3)	NA	NA
Onset of back pain before 40 years[£], n	265 (94.6)	102 (89.5)	265 (94.6)	102 (89.5)	NA	NA
Duration of back pain in years[£], mean (SD)	5.7 (7.2)	7.9 (10.5)	5.7 (7.2)	7.9 (10.5)	NA	NA
Male gender, n (%)	224 (56.6)	64 (38.1)	147 (52.5)	39 (34.2)	77 (66.4)	25 (46.3)
Number of SpA features[*], mean (SD)	2.9 (1.3)	1.5 (1.1)	3.1 (1.4)	1.3 (1.1)	2.5 (1.0)	1.8 (0.8)
Presence of 2 or more SpA features[*], n (%)	339 (85.6)	78 (46.4)	241 (86.1)	46 (40.4)	98 (84.5)	32 (59.3)
Definite radiographic sacroiliitis[¥£], n (%)	70 (25.0) (N=280)	4 (3.5) (N=114)	70 (25.0) (N=280)	4 (3.5) (N=114)	NA	NA
Active inflammation of sacroiliac joints, MRI[£], n (%)	141 (63.8) (N=221)	10 (11.2) (N=89)	141 (63.8) (N=221)	10 (11.2) (N=89)	NA	NA
HLA - B27, n (%)	247 (63.5) (N=389)	28 (17.7) (N=158)	191 (70.0) (N=273)	26 (23.4) (N=111)	56 (48.3) (N=116)	2 (4.3) (N=47)
Elevated CRP, n (%)	192 (48.5)	37 (22.2)	123 (43.9)	14 (12.3)	69 (59.5)	23 (42.6)
IBP (according to experts definition[†£]), n (%)	224 (80.0)	43 (37.7)	224 (80.0)	43 (37.7)	NA	NA
Peripheral arthritis past or present, n (%)	231 (58.3)	68 (40.5)	125 (44.6)	19 (16.7)	106 (91.4)	49 (90.7)
Enthesitis past or present, n (%)	206 (52.0)	40 (23.8)	136 (48.6)	26 (22.8)	70 (60.3)	14 (25.9)
Uveitis past or present, n (%)	34 (8.6)	5 (3.0)	30 (10.7)	5 (4.4)	4 (3.5)	0 (0.0)
Dactylitis past or present, n (%)	47 (11.9)	8 (4.8)	21 (7.5)	1 (0.9)	26 (22.4)	7 (13.0)
Psoriasis past or present, n (%)	29 (7.3)	7 (4.2)	17 (6.1)	6 (5.3)	12 (10.3)	1 (1.9)
IBD past or present, n (%)	16 (4.0)	0 (0.0)	11 (3.9)	0 (0.0)	5 (4.3)	0 (0.0)
Active inflammation of the spine, MRI[£], n (%)	32 (34.8) (N=92)	3 (5.0) (N=60)	32 (34.8) (N=92)	3 (5.0) (N=60)	NA	NA

*Features included: Inflammatory back pain (IBP) according to experts definition, arthritis (ever), heel enthesitis (ever), dactylitis (ever), uveitis (ever), psoriasis (ever), inflammatory bowel disease (IBD) (ever), good response to NSAIDs, family history of spondyloarthritis (SpA), elevated CRP. ¥ ≥ grade 2 bilateral or ≥ grade 3 unilateral. £ Only applicable in patients from the axial population. axSpA, axial spondyloarthritis; pSpA, peripheral spondyloarthritis; IBD, inflammatory bowel disease; NA, not applicable; MRI, magnetic resonance imaging. Although imaging of the axial skeleton was performed in some patients from the 'peripheral population', the significant amount of missing data precludes unbiased proportions to be calculated.

Predictive validity of the ASAS SpA classification criteria

The predictive validity of the ASAS SpA classification criteria is presented in table 3 and figure 2. Of the 564 patients with follow-up assessment, 335 had fulfilled the axSpA or pSpA criteria at baseline, and 229 had not. Of these 335 patients, 309 were diagnosed as SpA at follow-up (PPV: 92.2%). Of the 229 patients not fulfilling ASAS criteria at baseline, 142 were indeed considered having no or another diagnosis than SpA (NPV: 62.0%), but 87 received a diagnosis of SpA at follow-up. The PPV of the axSpA and pSpA criteria was 93.3% and 89.5% respectively.

Table 3. Predictive validity of the ASAS classification criteria, by testing the classification at baseline against the rheumatologist's diagnosis at follow-up (on average 4.4 years)

Criteria	Predictive values		Classification at baseline	Rheumatologist's diagnosis at follow-up		
	PPV (%)	NPV (%)		SpA	No-SpA	
SpA*	92.2	62.0	**Positive**	309	26	**335**
			Negative	87	142	**229**
				396	**168**	**564**
pSpA	89.5	58.7	**Positive**	85	10	**95**
			Negative	31	44	**75**
				116	**54**	**170**
axSpA	93.3	63.6	**Positive**	224	16	**240**
			Negative	56	98	**154**
				280	**114**	**394**
axSpA: Imaging arm (with/without clinical arm)	94.7	51.0	**Positive**	180	10	**190**
			Negative	100	104	**204**
				280	**114**	**394**
axSpA: Clinical arm (with/without imaging arm)	96.0	48.9	**Positive**	168	7	**175**
			Negative	112	107	**219**
				280	**114**	**394**
axSpA: Imaging arm only	86.2	31.9	**Positive**	56	9	**65**
			Negative	224	105	**329**
				280	**114**	**394**
axSpA: Clinical arm only	88.0	31.4	**Positive**	44	6	**50**
			Negative	236	108	**344**
				280	**114**	**394**

*Combination of ASAS criteria for axSpA (in patients with predominant back pain with or without peripheral manifestations) and criteria for pSpA for patients with peripheral manifestations only. axSpA, axial spondyloarthritis; pSpA, peripheral spondyloarthritis; PPV, positive predictive value; NPV, negative predictive value.

The PPV of the ASAS SpA criteria did not differ when applied in patients from countries with high *versus* low background HLA-B27 prevalence (91.2% and 92.7% respectively; online supplementary material -appendix 3).

The sensitivity analyses yielded a PPV of the ASAS SpA (range: 92.6%-95.1%), axSpA (range: 93.4%-95.1%) and pSpA (range: 87.9%; 95.7%) criteria similar to the main analysis (table 4). Comparable results were found for the 'imaging arm' (range: 94.5%-96.5%) and 'clinical arm' (range: 96.4%-98.2%); and also considering those fulfilling the 'imaging arm' only (range: 85.1%-86.7%) and 'clinical arm' only (range: 87.9%-92.9%) (see online supplementary material - appendix 4).

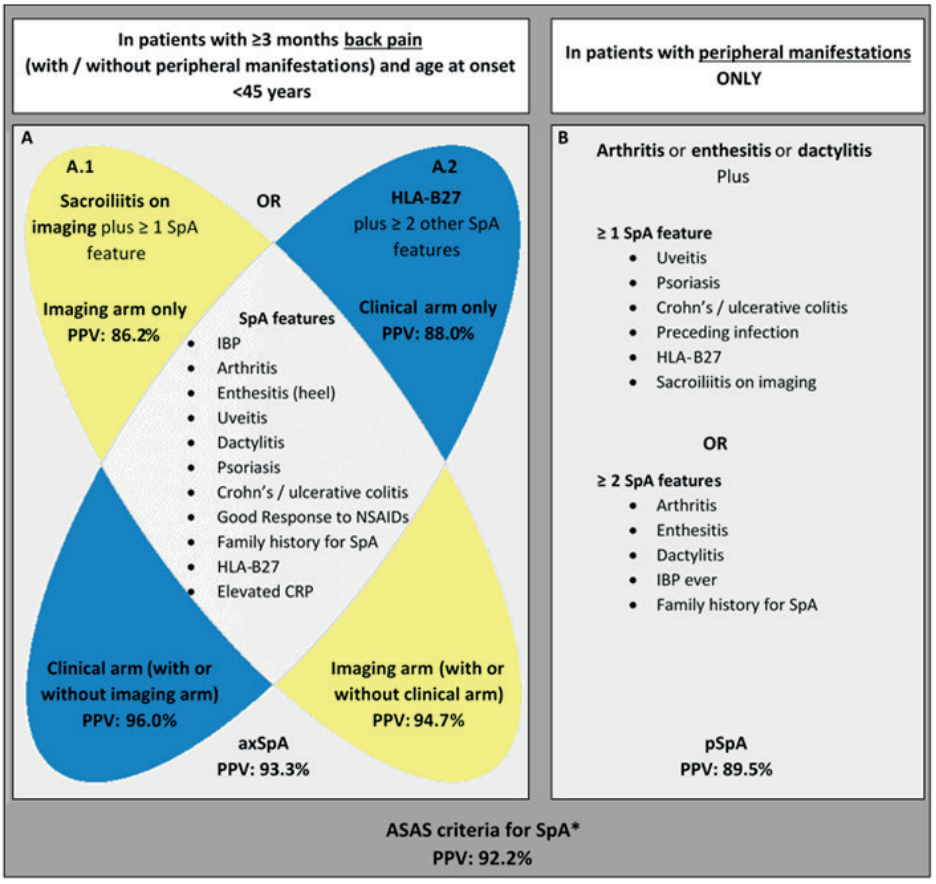

Figure 2. Predictive validity of the Assessment of SpondyloArthritis international Society (ASAS) classification criteria. *Combination of: **(A)** ASAS criteria for axial spondyloarthritis (axSpA) applied in patients with predominant back pain, with or without peripheral manifestations (N=394) and **(B)** criteria for peripheral spondyloarthritis (pSpA) applied in patients with peripheral manifestations only (N=170). The positive predictive value (PPV) of The **(A.1)** 'imaging arm' and the **(A.2)** 'clinical arm' of the axSpA criteria are shown considering all patients that fulfil each arm irrespective of fulfilment of the other (bottom of each ellipsis) and considering patients that fulfil one arm exclusively (top of each ellipsis). IBP, inflammatory back pain; CRP, C reactive protein.

Imaging arm of the axSpA criteria

Among the 240 patients classified positive according to the axSpA criteria at baseline, 190 (79.2%) had sacroiliitis on imaging (radiograph and/or MRI), hence fulfilling the 'imaging arm'

(irrespective of fulfilment of the 'clinical arm'). Remarkably, when imaging was positive, almost all patients were classified positive (190/193: 98.4%) by the axSpA criteria at baseline and almost all received a SpA diagnosis at follow-up (PPV: 94.7%). The PPV was similarly high comparing patients with only radiographic sacroiliitis (n=42; PPV: 97.6%), only sacroiliitis on MRI (n=117; PPV: 94.9%) and with both (n=31; PPV: 90.3%).

Similarly, patients fulfilling the 'imaging arm' only (thus excluding patients who also fulfil the 'clinical arm') had a high probability (PPV: 86.2%) of being diagnosed axSpA after more than 4 years (mean (SD) level of confidence: 8.6 (1.5)).

Table 4. Sensitivity analyses.

	Main analysis¥ (N=564)		Complete cases† (N=345)		FU visit£ (N=345)		≥ 75% FU‡ (N=291)	
ASAS Criteria	**PPV (%)**	**NPV (%)**	**PPV (%)**	**NPV (%)**	**PPV (%)**	**NPV (%)**	**PPV (%)**	**NPV (%)**
SpA*	92.2	62.0	93.6	63.5	92.6	55.8	95.1	64.8
pSpA	89.5	58.7	95.7	31.3	87.9	48.8	94.9	48.4
axSpA	93.3	63.6	93.4	68.2	94.3	59.1	95.1	73.7
axSpA: Imaging arm (with/without clinical arm)	94.7	51.0	94.5	55.2	94.6	46.6	96.5	52.3
axSpA: Clinical arm (with/without imaging arm)	96.0	48.9	96.4	49.7	98.2	43.7	97.4	52.9
axSpA: Imaging arm only	86.2	31.9	86.0	32.1	85.1	27.1	86.7	26.3
axSpA: Clinical arm only	88.0	31.4	87.9	31.0	92.9	27.1	89.7	26.7

*Combination of ASAS criteria for axSpA (in patients with predominant back pain with or without peripheral manifestations) and criteria for pSpA for patients with peripheral manifestations only. ¥ All patients with follow-up data available (N=564). £ Only patients with complete information regarding all SpA features at baseline (N=345). † Only patients with follow-up visit (N=345). ‡ Only patients from centers with ≥ 75% complete follow-up data (N=291). axSpA, axial spondyloarthritis; pSpA, peripheral spondyloarthritis; PPV, positive predictive value; NPV, negative predictive value; FU, follow-up. See online supplementary file 1 - table S1, S2 and S3 for raw data regarding all sensitivity analyses.

Clinical arm of the axSpA criteria

The PPV of the 'clinical arm' (± 'imaging arm') was 96% and the majority of the 50 patients fulfilling the 'clinical arm' only at baseline were diagnosed as SpA at follow-up (PPV: 88.0%). Similar to the 'imaging arm' only, the follow-up diagnosis for these 50 patients was established

with high confidence (mean: 8.5 (SD: 1.5)) and was consistent with baseline diagnosis: of the 44 patients diagnosed as axSpA at follow-up, 38 (86.4%) had also received the same diagnosis at baseline.

Patients fulfilling the 'clinical arm' only had a mean of 3.4 (SD: 1.1) SpA features at baseline, and inflammatory back pain (IBP) (43; 86.0%) was most prevalent, followed by good response to NSAIDs (34; 68.0%), peripheral arthritis (23; 46.0%) and elevated CRP (20; 40%). The large majority (36; 72.0%) of these patients still had either axial or peripheral symptoms at the end of follow-up.

DISCUSSION

The long-term follow-up of the original ASAS-cohort provided an excellent predictive validity for the ASAS axSpA and pSpA classification criteria and for the combined set. In addition, patients fulfilling the 'clinical arm' had disease characteristics in accordance with the rheumatologists' perception of what 'SpA looks like' ('gestalt') resulting in a good predictive validity similar to that of the 'imaging arm'.

A previous report on the ASAS axSpA criteria predictive validity has shown similarly good results (PPV: 87.9%).[12] However, this study was limited to Chinese patients and had a short follow-up (2 years). Moreover, patients with r-axSpA and with predominantly peripheral manifestations were excluded limiting the study's external validity.

The current study is the first prospectively testing the entire set of the ASAS SpA criteria against the rheumatologist's diagnosis in a worldwide population over 4 years later. In fact, most of previous studies tested the ASAS criteria concurrent validity, where both the criteria and the ´external reference' (rheumatologist's diagnosis) were determined simultaneously. In the current study, the time-lag between the criteria application (baseline) and the rheumatologist's diagnosis (follow-up) allowed assessment of the criteria accuracy for predicting a diagnosis of SpA taking into account the disease course (predictive validity).

Several metrics are generally used to describe criteria performance, among which sensitivity and specificity are the most often reported. However, since these metrics are defined on the basis of subjects with or without the disease, they do not inform about the probability of having SpA once the criteria are applied (post-test probability).[25] This probability is given by the predictive values (both positive and negative), which, as stated above, are particularly informative when derived from longitudinal studies, such as the ASAS cohort.

The somewhat low NPV should be interpreted cautiously in the context of a longitudinal study, particularly in SpA, which exhibits often an evolving character with increasing number of manifestations over time. Indeed, during follow-up approximately one third of the patients developed at least one additional SpA feature, which may explain why some patients not captured by the ASAS criteria at baseline were regarded as SpA by the rheumatologist at follow-up. Thus, the NPV may reflect not only the number of patients with SpA that, at baseline, are not captured by the criteria, and also the natural course of the disease.

It has been argued that, when applied in clinical practice, the 'clinical arm only' carries the risk of misclassification.[17, 24] In that sense, it is a common belief that the 'clinical arm' adds sensitivity to the axSpA criteria, while compromising specificity. Our findings do not support these claims. On the contrary, we found similarly high PPVs for both arms of the axSpA criteria. Moreover, the additional patients captured by the 'clinical arm' showed a 'SpA-like' phenotype, which persisted over time, possibly explaining the consistency and the high level of confidence for the diagnosis of this subgroup. These data support the view that the 'clinical arm' comprises a group of patients who belong to the SpA spectrum as much as those fulfilling the 'imaging arm'. Thus, the 'clinical arm' is truly complementary and may be of particular use when imaging is not available.

A noteworthy finding in this study is the dominant place that sacroiliitis on MRI holds in the ASAS axSpA criteria. Remarkably, almost all patients who had sacroiliitis on imaging were classified 'positive' and most patients fulfilling the 'imaging arm' had only sacroiliitis on MRI (without radiographic sacroiliitis). The fact that most of these were indeed diagnosed as axSpA at follow-up (PPV: 94.9%) demonstrates how well the axSpA criteria reflect the rheumatologists' expectations on the ability of sacroiliitis on MRI to discriminate between patients with and without axSpA. However, it is important to highlight that sacroiliitis on MRI was at the basis of the nr-axSpA concept[18] and instigated the development of the ASAS axSpA criteria.[2] Hence, circularity in reasoning cannot be excluded, but is not necessarily detrimental as long as sacroiliitis on MRI truly reflects the disease consequences closely linked to their risk factors and pathophysiology as it is currently believed. More research is needed to clarify this issue.

The HLA-B27 prevalence in patients with pSpA was expectedly lower (48.3%) than in axSpA, but similar to what is known for pSpA and also found in another recent cohort (47.5%; Early Arthritis Clinic: EAC).[14] Despite this, the prevalence of pSpA in that cohort was much lower (3.8%) when compared with the current study (68%). Importantly, the pSpA criteria discriminated well between pSpA and no-SpA (PPV: 89.5%), even with similar proportions of peripheral arthritis in both groups (91.4% *vs* 90.7%). However, there was a significant difference in the proportion of enthesitis (60.3% vs 25.9%), which was infrequent in the EAC cohort (17.1%), possibly reflecting different inclusion criteria. This may, at least in part, explain the pSpA prevalence disparity between the two cohorts and stresses the central role of enthesitis in the disease. Thus, the allowance of enthesitis as an entry feature yields more pSpA cases without increased risk of mislabelling, as previously suggested.

This study has a number of limitations. The most relevant one is the high number of patients without follow-up data. Attrition unfortunately is common in long-term follow-up studies, especially if there is no regular protocol with assessments between the baseline and follow-up visit. Understandably, patients who complied with a follow-up visit had more active sacroiliitis on MRI at baseline, deemed to be associated with 'worse prognosis'. Hence, it could be expected that, if 'good prognosis' patients have preferentially dropped out, the performance of the criteria in centres with high participation rates (≥75% complete data) would be worse than in centres with low participation rates. However, this was not the case and argues against 'channelling bias' causing a spuriously high PPV. Finally, patients with less definite ('equivocal') diagnoses at baseline were not more likely to be lost to follow-up either since the level of

diagnostic confidence was almost identical in patients with follow-up (mean (SD): 8.3 (1.5)) compared with those lost to follow-up (8.2 (1.5)).

Missing data on MRI are another potential limitation. However, missing data are common in observational cohorts, as they reflect clinical practice, where clinicians must make decisions (on diagnosis) even without complete information. It is plausible to assume that, in such a scenario, missing information can best be considered negative. Nevertheless, it is always possible that patients diagnosed as no-SpA at baseline are more likely to have missing data, which would decrease their likelihood of fulfilling the criteria. Under that scenario, an analysis of patients with complete information only would yield worse PPVs, but that was not what we found.

Another limitation of this study is the self-reported diagnosis in some patients. However, the predictive values of the ASAS criteria in all patients *versus* patients who presented physically at a follow-up visit were similar, which adds to the credibility of the self-reported diagnosis provided by telephone.

In conclusion, and keeping in mind how the above-mentioned constraints were handled in the analysis, the ASAS SpA criteria have proven to accurately discriminate between patients with and without the disease when applied in patients with similar symptoms. Therefore, the ASAS criteria are valid for selecting patients for clinical and therapeutic trials and, especially when applied in settings similar to the ASAS cohort, they may guide rheumatologists in establishing a proper diagnosis.

SUPPLEMENTARY DATA

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

REFERENCES

1. 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 Apr; 27(4):361-368.
2. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new criteria? Arthritis Rheum. 2005 Apr; 52(4):1000-1008.
3. Mau W, Zeidler H, Mau R, et al. Clinical features and prognosis of patients with possible ankylosing spondylitis. Results of a 10-year followup. J Rheumatol. 1988 Jul; 15(7):1109-1114.
4. Braun J, Baraliakos X. Imaging of axial spondyloarthritis including ankylosing spondylitis. Ann Rheum Dis. 2011 Mar; 70 Suppl 1:i97-103.
5. Weber U, Lambert RG, Ostergaard M, et al. The diagnostic utility of magnetic resonance imaging in spondylarthritis: an international multicenter evaluation of one hundred eighty-seven subjects. Arthritis Rheum. 2010 Oct; 62(10):3048-3058.
6. Rudwaleit M, Landewé R, van der Heijde D, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. Ann Rheum Dis. 2009 Jun; 68(6):770-776.
7. 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.
8. 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.
9. Dougados M, van der Linden S, Juhlin R, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of

spondylarthropathy. Arthritis Rheum. 1991 Oct; 34(10):1218-1227.

10. Amor B, Dougados M, Mijiyawa M. [Criteria of the classification of spondylarthropathies]. Rev Rhum Mal Osteoartic. 1990 Feb; 57(2):85-89.
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). 2015 Jul; 54(7):1336.
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. 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.
14. 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.
15. Cheung PP, Paternotte S, Burki V, et al. Performance of the assessment in Spondyloarthritis International Society classification for axial and peripheral spondyloarthritis in an established clinical cohort: comparison with criteria sets of Amor and the European Spondylarthropathy Study Group. J Rheumatol. 2012 Apr; 39(4):816-821.
16. 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.
17. Robinson PC, Wordsworth BP, Reveille JD, et al. Axial spondyloarthritis: a new disease entity, not necessarily early ankylosing spondylitis. Ann Rheum Dis. 2013 Feb; 72(2):162-164.
18. Rudwaleit M, Haibel H, Baraliakos X, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. Arthritis Rheum. 2009 Mar; 60(3):717-727.
19. Ciurea A, Scherer A, Exer P, et al. Tumor necrosis factor alpha inhibition in radiographic and nonradiographic axial spondyloarthritis: results from a large observational cohort. Arthritis Rheum. 2013 Dec; 65(12):3096-3106.
20. Kiltz U, Baraliakos X, Karakostas P, et al. Do patients with non-radiographic axial spondylarthritis differ from patients with ankylosing spondylitis? Arthritis Care Res (Hoboken). 2012 Sep; 64(9):1415-1422.
21. Molto A, Paternotte S, van der Heijde D, et al. Evaluation of the validity of the different arms of the ASAS set of criteria for axial spondyloarthritis and description of the different imaging abnormalities suggestive of spondyloarthritis: data from the DESIR cohort. Ann Rheum Dis. 2015 Apr; 74(4):746-751.
22. Poddubnyy D, Rudwaleit M, Haibel H, et al. Rates and predictors of radiographic sacroiliitis progression over 2 years in patients with axial spondyloarthritis. Ann Rheum Dis. 2011 Aug; 70(8):1369-1374.
23. Gladman DD. Editorial: What is peripheral spondyloarthritis? Arthritis Rheumatol. 2015 Apr; 67(4):865-868.
24. Garg N, van den Bosch F, Deodhar A. The concept of spondyloarthritis: where are we now? Best Pract Res Clin Rheumatol. 2014 Oct; 28(5):663-672.
25. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994 Jul 9; 309(6947):102.