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

Sepriano, A.R.

Chapter 4

What is Axial Spondyloarthritis? A latent class and transition analysis in the SPACE and DESIR cohorts

Alexandre Sepriano, Sofia Ramiro, Désirée van der Heijde, Floris van Gaalen, Pierre Hoonhout, Anna Moltó, Alain Saraux, Roberta Ramonda, Maxime Dougados, Robert Landewé

Ann Rheum Dis. 2020 Mar;79(3):324-331

ABSTRACT

Objectives: To gain expert-judgement-free insight into the *Gestalt* of axial spondyloarthritis (axSpA), by investigating its 'latent constructs' and to test how well these latent constructs fit the Assessment of SpondyloArthritis international Society (ASAS) classification criteria.

Methods: Two independent cohorts of patients with early onset chronic back pain (SPondyloArthritis Caught Early (SPACE)) or inflammatory back pain (IBP) (DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR)) were analysed. Latent class analysis (LCA) was used to estimate the (unobserved) potential classes underlying axSpA. The best LCA model groups patients into clinically meaningful classes with best fit. Each class was labelled based on most prominent features. Percentage fulfilment of ASAS axSpA, peripheral SpA (pSpA) (ignoring IBP) or both classification criteria was calculated. Five-year data from DESIR were used to perform latent transition analysis (LTA) to examine if patients change classes over time.

Results: SPACE (n=465) yielded four discernible classes: 'axial' with highest likelihood of abnormal imaging and HLA-B27 positivity; 'IBP+peripheral' with 100% IBP and dominant peripheral symptoms; 'at risk' with positive family history and HLA-B27 and 'no SpA' with low likelihood for each SpA feature. LCA in DESIR (n=576) yielded similar classes, except for the 'no-SpA'. The ASAS axSpA criteria captured almost all (SPACE: 98%; DESIR: 93%) 'axial' patients, but the 'IBP+peripheral' class was only captured well by combining the axSpA and pSpA criteria (SPACE: 78%; DESIR: 89%). Only 4% of 'no SpA' patients fulfilled the axSpA criteria in SPACE. LTA suggested that 5-year transitions across classes were unlikely (11%).

Conclusion: The *Gestalt* of axSpA comprises three discernible entities, only appropriately captured by combining the ASAS axSpA and pSpA classification criteria. It is questionable whether some patients with 'axSpA at risk' will ever develop axSpA.

INTRODUCTION

Spondyloarthritis (SpA) encompasses heterogeneous entities with common clinical, laboratory and imaging features. The full spectrum of SpA includes patients with dominant axial symptoms (axial SpA (axSpA)) and patients with dominant peripheral symptoms (peripheral SpA (pSpA)).[1] The term axSpA aggregates patients with radiographic axSpA (r-axSpA; also known as ankylosing spondylitis) and non-radiographic axSpA (nr-axSpA), differing only by the presence of radiographic sacroiliitis in the former, as defined by the modified New York (mNY) criteria.[2]

Axial SpA is a syndrome described by classification criteria that supposedly best reflect its inherently unmeasurable 'latent' construct (*Gestalt*). The Assessment of SpondyloArthritis international Society (ASAS) criteria for axSpA have been developed to classify both r-axSpA and nr-axSpA. In the absence of a 'gold standard', expert opinion has been used as an external 'anchor' to develop and validate classification criteria.[3-5] The ASAS criteria outperform other criteria,[6] meaning that they contain several elements that experts consider relevant for their 'latent' picture of axSpA.

While such an approach for developing classification criteria has been pursued by default in rheumatology, it has a fundamental limitation that may jeopardise their construct- and content validity: circularity. If criteria are developed against expert opinion, and the expert finds certain characteristics [e.g. inflammation on magnetic resonance imaging (MRI) of the sacroiliac joints (SIJ)] more important than others, such characteristics may be awarded a too prominent place in the criteria. Subsequent cross-validation against an expert diagnosis may produce results driven by experts' beliefs rather than on an objective presence of axSpA. The *axiom* that 'early (diagnosis and treatment) is always better', a dominant view in modern rheumatology, may have contributed to rheumatologists' beliefs and as such trickled down into the ASAS criteria, designed to better capture patients with early disease. When classification criteria are (mis)used in a diagnostic context, overdiagnosis, followed by overtreatment, is a logical consequence.[7]

A more circularity-free determination of the *Gestalt* of axSpA is lacking in the literature, which hampers the study of the side effects of overdiagnosis and overtreatment. Here we propose to evaluate the *Gestalt* of axSpA using an analytical approach that excludes the rheumatologist's diagnostic opinion. Our aims were twofold: i. to gain an expert-judgement-free insight, into the concept of axSpA, by investigating its 'latent constructs'; and ii. to evaluate how well the ASAS SpA classification criteria capture these 'latent constructs'.

METHODS

Patients and study design

Baseline data from the SPondyloArthritis Caught Early (SPACE) cohort and baseline and 5-year data from the DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR) were used. Both cohorts have been previously described in detail.[6, 8] Briefly, in SPACE (ongoing multinational cohort), consecutive patients aged ≥ 16 years with chronic back pain (≥3 months, ≤2 years and onset <45 years) are included. In DESIR, consecutive patients aged 18-50 with inflammatory back pain (IBP) (>3 months but <3 years), and for whom the treating

rheumatologist considers the symptoms suggestive of axSpA (level of confidence (LoC) ≥5, scale 0-10), were included. Databases were locked in October 2017 (SPACE) and June 2016 (DESIR).

SpA features

The following features were collected in each cohort: HLA-B27, elevated C reactive protein (CRP) (≥6 mg/L), family history of SpA (ASAS definition),[5] good response to nonsteroidal anti-inflammatory drugs (NSAIDs), peripheral arthritis, heel enthesitis, dactylitis, psoriasis, inflammatory bowel disease, acute anterior uveitis, and IBP.

At baseline, SpA features were considered positive if 'ever present' (i.e. any time in the past and/or baseline) in both cohorts, except dactylitis (available only as 'current' in SPACE). In DESIR, data on SpA features were also collected every 6 months up to 2 years and yearly thereafter up to 5 years. Change in time-varying features was defined as 'once-a-feature-always-a-feature (OFAF)': patients positive at baseline remained positive at 5-years, even if becoming negative or missing in between; patients negative at baseline, remained negative at follow-up if no switch to positive or if missing in between. A feature changed to positive if appearing anytime during follow-up.

Radiographs and MRIs of the SIJ (X-SIJ; MRI-SIJ) and spine (X-Spine; MRI-Spine) were obtained at baseline in both cohorts, and at 2 and 5 years in DESIR. Each image was independently scored, by three trained central readers in each cohort, blinded to chronology, clinical data and to the results of other modalities. Four binary imaging features, defined by agreement between ≥2 out of 3 readers, were assessed: inflammation on MRI-SIJ (ASAS definition);[9, 10] bone marrow edema (BME) on MRI-Spine (≥5 lesions);[11] definite structural damage in X-SIJ according to the mNY criteria;[2] and ≥1 syndesmophyte in X-spine.[12]

Statistical analysis

Latent class analysis (LCA) was performed with baseline data of each cohort separately, including patients with complete data on all features. LCA unmasks a 'latent' (i.e. unobserved) construct (here: *Gestalt* of axSpA) by splitting patients into mutually exclusive classes based on the covariance of observed SpA features. Extensive evidence supports the superiority of LCA in identifying latent data structures, compared with other clustering methods.[13-15] SpA features (15 variables in SPACE; 14 in DESIR (excluding IBP)) were selected 'a priori' based on content knowledge without predefined weights.

A detailed description LCA and how it can be used to identify the latent classes of the *Gestalt* of axSpA is provided in online supplementary text S1. Briefly, the number of classes was increased, one-by-one, until the best model was found, defined by: best goodness of fit assessed by Akaike's information criterion, Bayesian information criterion (BIC), sample-sized adjusted BIC (aBIC), entropy, likelihood ratio-test (comparing the model with the one with n-1 classes); and by clinically recognisable patterns within each class (i.e. a statistical criterion alone does not suffice). The classes of the final model were interpreted according to the probability of each feature and labelled as a clinically recognisable entity. Features were defined as: across-class

dominant (highest probability across classes); within-class dominant (probability >50% within each class); and not dominant across or within classes.

Maximum likelihood estimates were used to classify individual patients based on their posterior probability of class membership. This allowed us to describe the classes including also variables not used in the models and to evaluate the percentage of patients within each class fulfilling the ASAS axSpA, pSpA (ignoring IBP) and the SpA criteria (i.e. combination of either axSpA or pSpA criteria) at baseline.

To address between-cohort differences in study design, a sensitivity analysis was performed in SPACE: only in patients with a rheumatologist's diagnosis with LoC ≥ 5 (similar to DESIR).

Latent transition analysis (LTA) was used to estimate the likelihood of change across classes after 5 years in DESIR.[16] LTA includes the same patients and variables as in LCA. The number of classes best fitting the baseline and 5-year LCA formed the basis of the LTA model. Classes at baseline and follow-up can be assumed as: having the same meaning (full invariance); different meaning (full non-invariance); or the same meaning for some and different for others (partial invariance). The final LTA model has the number of classes at baseline and 5-year and class-(in)variance that best fits the data provided it is clinically meaningful.

LCA was performed in Stata V.15.1. LTA was performed in MPlus V.7.

RESULTS

Baseline characteristics

In total, 465 patients from SPACE and 576 from DESIR were included. In SPACE, included patients were more likely to be HLA-B27 positive (37% *vs* 57%) and less likely to have BME on MRI-SIJ (14% *vs* 30%) than those excluded (N=283). No differences were seen in DESIR (excluded: N=132) (online supplementary tables S1 and S2). Baseline characteristics of the included patients from both cohorts are shown in table 1. Patients from DESIR had, on average, more SpA features compared with those from SPACE, including peripheral features (e.g. heel enthesis 45% *vs* 20%) and axial imaging abnormalities (e.g. sacroiliitis on MRI-SIJ 27% *vs* 14%).

Table 1. Baseline patient characteristics in the SPACE and DESIR cohorts

	SPACE (N=465)	DESIR (N=576)
Age at baseline (years)	31 (8)	33 (8)
Male gender	161 (35)	269 (47)
Symptom duration (years)	1.8 (2.3)	1.5 (0.8)
ASAS axSpA criteria	172 (37)	358 (62)
axSpA according to Rheumatologist*	136 (30)	269 (47)
ASAS pSpA criteria	182 (39)	320 (56)
ASAS SpA criteria†	249 (54)	443 (77)
Sacroiliitis on MRI-SIJ (ASAS)	64 (14)	153 (27)
BME on MRI-spine (≥ 5 lesions)	21 (5)	25 (4)
Radiographic sacroiliitis (mNY)	38 (8)	78 (14)
≥ 1 syndesmophyte on X-spine	15 (3)	39 (7)
Elevated CRP (≥6 mg/L)	118 (25)	169 (29)
Good response to NSAIDs ever	189 (41)	491 (85)
Peripheral arthritis ever	76 (16)	122 (21)
Dactylitis ever	23 (5)	78 (14)
Heel enthesitis ever	91 (20)	261 (45)
HLA-B27	172 (37)	345 (60)
Family history of SpA	194 (42)	250 (43)
Psoriasis ever	54 (12)	99 (17)
Uveitis ever	33 (7)	52 (9)
Inflammatory Bowel Disease ever	35 (8)	25 (4)
Current arthritis / any enthesitis / dactylitis	317 (68)	398 (69)
Inflammatory back pain	308 (66)	576 (100)
Number of SpA features (0-9)‡	2 (1)	3 (1)

Values are mean (standard deviation) for continuous variables or number (%) for binary variables. SpA features are positive if 'ever present' (any time in the past and/or baseline); * Clinical diagnosis of axSpA at baseline with a level of confidence >7; Missing data SPACE: axSpA according to Rheumatologist (n=454); Symptom duration (N=461); missing data DESIR: axSpA according to Rheumatologist (N=576); † fulfilment of either ASAS axSpA or ASAS pSpA classification criteria; ‡ peripheral arthritis, heel enthesitis, dactylitis, psoriasis, uveitis, inflammatory bowel disease, good response to NSAIDs, elevated CRP and family history of SpA.; SD, standard deviation; ASAS, Assessment of SpondyloArthritis international Society; axSpA, axial spondyloarthritis; pSpA, peripheral spondyloarthritis; CRP, C-reactive protein; NSAIDs, nonsteroidal anti-inflammatory drugs; mNY, modified New York criteria; MRI, magnetic resonance imaging; SIJ, sacroiliac joints; BME, bone marrow edema; X-spine, radiograph of the spine.

Latent class analysis in SPACE and DESIR

A 4-class (SPACE) and a 3-class (DESIR) LCA-model fitted the data best (table 2). The additional class in the 5-class (SPACE) and 4-class (DESIR) models, with worse model fit, did not yield a clinically recognisable pattern (online supplementary tables S3, S4 and S5).

The final LCA models are shown in Table 2. In SPACE, class 1 was characterised by highest likelihood (i.e. across-class dominance) of lesions present on axial imaging, elevation of CRP and HLA-B27-positivity, and was labelled as 'axial'. Class 2, was labelled 'IBP+peripheral', given the 100% likelihood of IBP and across-class dominance of peripheral features. Class 3 had across-class dominance of positive family history (71%) and within-class dominance of HLA-B27 positivity (69%) and IBP (66%) but low likelihood of other features and was labelled as 'at risk'. Class 4 was labelled 'no SpA' given the very low likelihood for each SpA feature.

Table 2. Final latent class analysis (LCA) models in SPACE (N=465) and DESIR (N=576) in probability scale (range: 0-1)

	SPACE				DESIR			
	Class 1 ('axial') (P*=16%)	Class 2 ('IBP+peripheral') (P*=20%)	Class 3 ('At risk') (P*=24%)	Class 4 ('no SpA') (P*=40%)	Class 1 ('axial') (P*=19%)	Class 2 ('IBP + peripheral') (P*=27%)	Class 3 ('at risk') (P*=54%)	Class 4 'no SpA'†
Inflammation on MRI-SIJ (ASAS)	0.74	0.04	0.00	0.03	0.83	0.22	0.09	
BME on MRI-Spine (≥ 5 lesions)	0.25	0.02	0.00	0.00	0.20	0.00	0.01	
Radiographic sacroiliitis (mNY)	0.32	0.09	0.01	0.03	0.58	0.06	0.02	
≥ 1 syndesmophyte on X-spine	0.03	0.06	0.00	0.04	0.11	0.05	0.06	
Elevated CRP (≥ 6 mg/dL)	0.49	0.22	0.21	0.20	0.56	0.41	0.14	
Good response to NSAIDs (ever)	0.59	0.85	0.25	0.20	0.97	0.84	0.82	
Peripheral arthritis (ever)	0.17	0.44	0.04	0.10	0.09	0.73	0.00	
Dactylitis (ever)	0.02	0.18	0.00	0.03	0.03	0.46	0.01	
Heel enthesitis (ever)	0.10	0.66	0.13	0.04	0.26	0.60	0.45	
HLA-B27	0.84	0.33	0.69	0.00	0.90	0.52	0.53	
Family history of SpA	0.38	0.50	0.71	0.21	0.48	0.44	0.41	
Psoriasis (ever)	0.10	0.31	0.02	0.08	0.09	0.29	0.14	
Uveitis (ever)	0.13	0.07	0.12	0.02	0.08	0.12	0.08	
IBD (ever)	0.03	0.15	0.00	0.10	0.02	0.05	0.05	
Inflammatory back pain	0.68	1.00	0.66	0.49	NA	NA	NA	

The table displays the main results of the LCA separately in each cohort. Values are the conditional probability for each SpA feature positivity within each latent class (range: 0-1).
* Probability of the latent class. † 'No SpA' latent class absent in DESIR; in DESIR all included patients have a high likelihood of axSpA. Heatmap legend: **Red:** highlights dominant features across latent classes; **yellow:** highlights dominant features (probability >50%) within each class but not across classes; **blank:** not dominant neither across nor within classes. SpA features are positive if 'ever present' (any time in the past and/or baseline); CI, confidence interval; ASAS, Assessment of SpondyloArthritis international Society; CRP, C-reactive protein; NSAIDs, nonsteroidal anti-inflammatory drugs; mNY, modified New York criteria; MRI, magnetic resonance imaging; SIJ, sacroiliac joints; BME, bone marrow edema; X-spine, radiograph of the spine; IBD, inflammatory bowel disease. NA, not applicable.

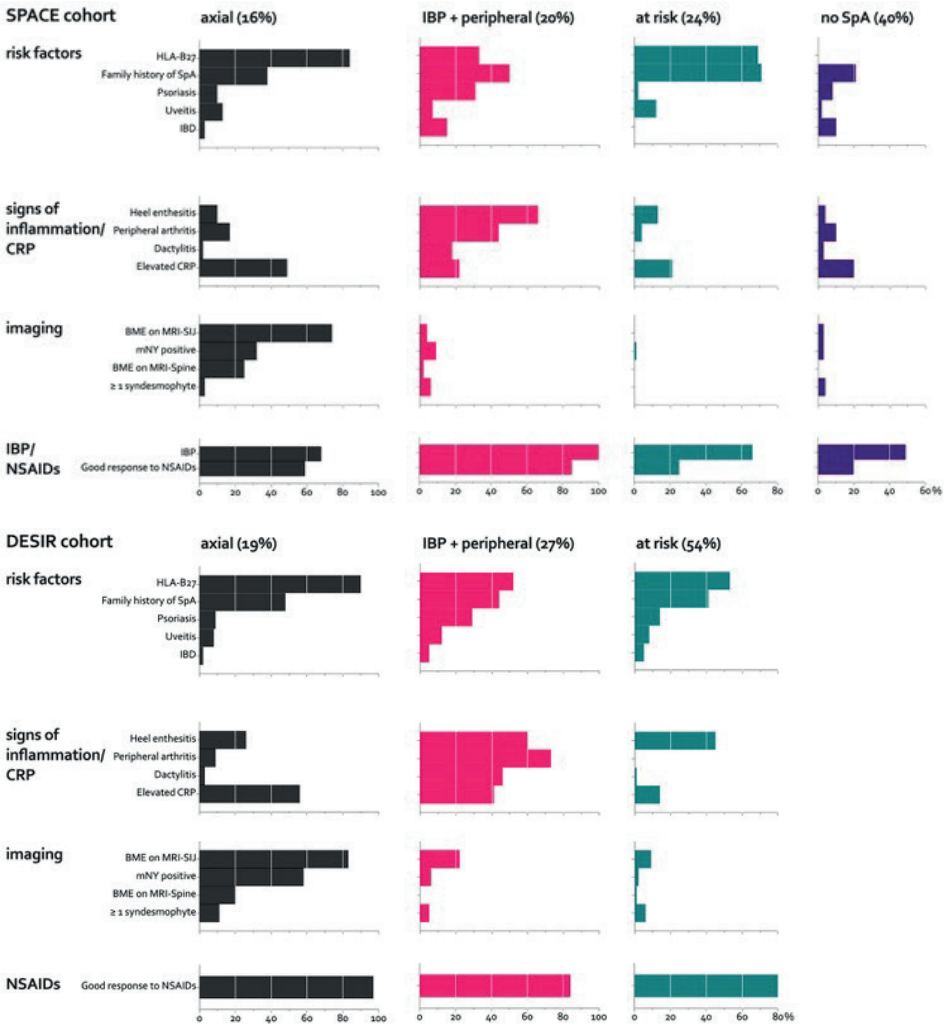

Figure 1. The Gestalt of axial SpA. Distribution of the probabilities of each feature according to the final LCA model in SPACE and DESIR. CRP, C-reactive protein; NSAIDs, nonsteroidal anti-inflammatory drugs; mNY, modified New York criteria; MRI, magnetic resonance imaging; SIJ, sacroiliac joints; BME, bone marrow edema; X-spine, radiograph of the spine; IBP, inflammatory back pain; IBD, inflammatory bowel disease.

The LCA analysis in DESIR yielded the same latent classes, except 'no SpA', and an overlapping pattern of dominance: amongst 42 possible comparisons (14 features (excluding IBP) multiplied by 3 classes (excluding 'no SpA')), in 37 (88%) the dominance-pattern was similar to SPACE (Table 2). Figure 1 graphically displays the between-cohort similarities, and also the phenotypical differences between the 'axial' and 'IBP+peripheral' classes which overlap with the 'at risk class' only partially, and even less with the 'no SpA' class.

The LCA model in SPACE, in patients with a rheumatologist's diagnosis of axSpA (LoC ≥5) (N=202) yielded the same classes as the main model, except 'no SpA' i.e. similar to DESIR ('axial':29%; 'IBP + peripheral':33%; 'at risk':38%; online supplementary table S6).

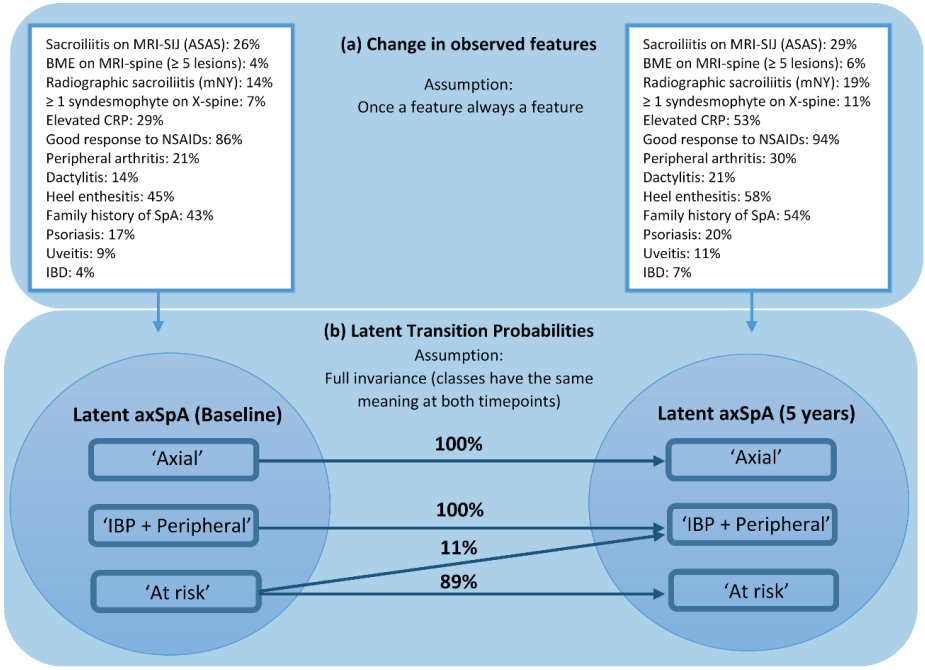

Figure 2. Final latent transition analysis (LTA) model (with full invariance*) in DESIR (N=576). (a) squares refer to observed (i.e. measurable) variables and (b) Circles to latent (i.e. unobserved variables). Arrows: latent transition analysis models the change in observed features (a) to estimate the latent (b) transition probabilities between classes from baseline to 5-years. LTA, latent transition analysis; ASAS, Assessment of SpondyloArthritis international Society; CRP, C reactive protein; NSAIDs, nonsteroidal anti-inflammatory drug; mNY, modified New York criteria; MRI, magnetic resonance imaging; SIJ, sacroiliac joints; BME, bone marrow edema; X-spine, radiograph of the spine; IBD, inflammatory bowel disease; axSpA, axial spondyloarthritis. * Selection of final LTA model according to goodness of fit detailed in online supplementary table S9 and full final model in online supplementary table S10.

Latent Transition Analysis in DESIR

Of the 576 patients in DESIR 500 (87%) completed the 5-year follow-up. The change in SpA- and imaging-features between baseline and 5 years is shown in Figure 2a. Because of how SpA features were defined (OFAF), all increased in prevalence over time, but changes were more pronounced with peripheral (e.g. peripheral arthritis: 21% to 30%) than with imaging features (e.g. BME on MRI-SIJ:26% to 29%).

Similar to baseline LCA, a 3-class model at 5 years best fitted the data (online supplementary table S7 and S8). Accordingly, an LTA model with 3 classes at both timepoints was fit. Although the model fit (online supplementary table S9) was better with partial invariance, the resulting model did not yield a clinically recognisable pattern (data not shown), so the simplest assumption (full invariance) was taken to define the final LTA model (Figure 2b and online

supplementary table S10). LTA revealed a 0% probability of switch from the 'axial' and 'IBP+peripheral' to another class. 'at risk' patients at baseline had 11% likelihood to change to 'IBP+ peripheral' over 5 years.

Observed characteristics and fulfilment of the ASAS classification criteria

The patterns of observed characteristics per latent class in SPACE and DESIR were, expectedly, similar to the model-based estimates (table 3). In addition, across-class dominance of males in the 'axial' class (SPACE: 66%; DESIR: 73%), and current arthritis/enthesitis/dactylitis (i.e. entry criterion for pSpA criteria) in the 'IBP+peripheral class (SPACE: 87%; DESIR: 88%) were observed.

The ASAS axSpA criteria captured almost all patients from the 'axial' class in SPACE (63/64; 98%). This percentage was much lower with 'IBP+Peripheral' (41/92; 49%), and missed patients were most often female (78%), positive for current arthritis/enthesitis/dactylitis (92%) and HLA-B27 and MRI-SIJ/mNY negative. The pSpA criteria captured 67% of the 'IBP+peripheral' patients and this figure was 78% when the axSpA and pSpA criteria were combined. Fifty-nine (60%) patients from the 'at risk' class fulfilled the axSpA criteria (58/59=98% fulfilling the 'clinical arm only'). Among the 58 fulfilling the 'clinical arm only', family history of SpA (75%) and IBP (85%) were the most common features. Only nine patients (4%) from the 'no SpA' class fulfilled the axSpA criteria, all of which captured by the imaging arm only (78% positive for IBP or good response to NSAIDs). Results were similar in DESIR, except that the percentage of 'at risk' patients fulfilling the 'clinical arm only' was somewhat lower (148/177=84%).

Table 3. Baseline observed patient- and disease-characteristics per latent class in SPACE and DESIR

	SPACE				DESIR		
	'Axial' (N=64)	'IBP+Peripheral' (N=92)	'At risk' (N=99)	'No SpA' (N=210)	'Axial' (N=110)	IBP+Peripheral' (N=137)	'At risk' (N=329)
Clinical and demographic							
Age at baseline (years)	30	32	30	31	31	33	34
Male gender	***66***	38	32	25	***73***	43	40
Symptom duration (years)	1	2	2	2	2	1	2
Imaging							
Inflammation on MRI-SIJ (ASAS)	***86***	3	0	2	***88***	22	8
BME on MRI-Spine (≥ 5 lesions)	**28**	2	0	1	**20**	0	1
Radiographic sacroiliitis (mNY)	**34**	11	1	2	***59***	5	2
≥ 1 syndesmophyte on X-spine	3	7	0	3	11	6	6
SpA features							
Elevated CRP (≥ 6 mg/dL)	**50**	22	24	20	***56***	39	16
Good response to NSAIDs (ever)	*59*	***89***	32	18	***97***	*84*	82
Peripheral arthritis (ever)	17	**47**	3	9	7	***83***	0
Dactylitis (ever)	2	**19**	0	2	3	***55***	0
Heel enthesitis (ever)	9	***72***	12	3	24	***59***	47
Current arthritis/ enthesitis/ dactylitis	48	***87***	*64*	*68*	*56*	***88***	*65*
HLA-B27	***86***	36	***85***	0	***93***	*52*	*52*
Family history of SpA	38	*52*	***72***	24	47	43	42
Psoriasis (ever)	9	**35**	1	7	7	**29**	16
Uveitis (ever)	13	7	**15**	2	8	**12**	8
IBD (ever)	3	**15**	0	9	1	5	5
Inflammatory back pain	*67*	***100***	*69*	50	100*	100*	100*
Number of SpA features (0-9)†	3	**5**	2	1	3	**4**	2
ASAS classification criteria							
ASAS axSpA criteria	**98**	45	60	4	**93**	58	54
ASAS pSpA criteria	48	**70**	56	15	56	**82**	44
ASAS SpA criteria‡	**98**	78	79	17	**99**	89	64

Values are means for continuous variables or percentages for binary variables. * by study design all patients in DESIR have IBP; † peripheral arthritis, heel enthesitis, dactylitis, psoriasis, uveitis, inflammatory bowel disease, good response to NSAIDs, elevated CRP and family history of SpA; ‡ fulfillment of either ASAS axSpA or ASAS pSpA classification criteria. Values in bold highlight discriminant features across latent classes. Values in italic highlight dominant features (probability >50%) within each class. SpA are positive if 'ever present' (any time in the past and/or baseline); CI, confidence interval; ASAS, Assessment of SpondyloArthritis international Society; CRP, C reactive protein; NSAIDs, nonsteroidal anti-inflammatory drug; mNY, modified New York criteria; MRI, magnetic resonance imaging; SIJ, sacroiliac joints; BME, bone marrow edema; X-spine, radiograph of the spine; IBD, inflammatory bowel disease; axSpA, axial spondyloarthritis; pSpA, peripheral spondyloarthritis.

DISCUSSION

Using a data-driven approach, we identified three separate clinical entities, remarkably stable over time, together forming the *Gestalt* of axSpA, in two independent cohorts, that we labelled 'Pure axial SpA' ('axial'), 'Axial SpA with peripheral signs' ('IBP+peripheral') and 'Axial SpA at risk' ('at risk'). In SPACE, a cohort that includes back pain patients without axSpA, these three axSpA classes decently discerned themselves from a fourth labelled as 'no SpA'. This adds to the credibility of our data, since the absence of 'no SpA' in DESIR was expected based on enrolment criteria. The ASAS axSpA classification criteria captured almost entirely the 'axial' class but missed several patients from the 'IBP+peripheral' class: The latter is better captured when combining the axSpA and pSpA criteria, suggesting a larger overlap between axSpA and pSpA than previously thought, when the ASAS criteria were developed. Taken together, at the group level these results confirm the robustness of the classification criteria. The 'at risk' class is an entity characterised by the presence of presumed risk factors for axSpA but the absence of objective clinical signs. While these patients often fulfil the ASAS axSpA classification criteria, it is likely that some do not actually have or will ever develop axSpA. Overdiagnosis of axSpA in the 50% of patients in this class is likely if classification criteria are ticked for diagnosis.

A diagnosis of axSpA is challenging and should rely on thorough knowledge and recognition of 'the appropriate pattern'.[17, 18] The rheumatologists' perception of the 'SpA-pattern' evolved over the last 40 years as a result of efforts by the international rheumatology community. Initially only r-axSpA (ankylosing spondylitis) was recognised and classified by the mNY-criteria.[2] In the 70s-80s Moll and Wright defined SpA as a group of entities with common features,[19] and the Amor and the European Spondyloarthropathy Study Group (ESSG) classification criteria were proposed.[20, 21] Both criteria-sets capture the broader 'SpA-pattern' by combining axial and peripheral features and do not distinguish between patients with dominant axial- and dominant peripheral patterns. Since then, evidence has emerged supporting that patients with the axial and peripheral pattern may respond differently to treatment,[22, 23] and that not all patients with axSpA will develop sacroiliitis on pelvic radiographs (mNY-positive). When they do, this is frequently a late and unreliable finding and often preceded by sacroiliitis on MRI-SIJ for many years.[24-31] Such evidence prompted ASAS-experts to develop classification criteria for patients with predominant axial involvement,[5] also capturing those that are mNY-negative (nr-axSpA) as axSpA, and for patients with predominant peripheral involvement that -if combined- enclose the entire *Gestalt* of SpA according to experts.[4]

The ASAS axSpA and pSpA classification criteria were validated against an external 'gold standard': expert opinion.[3-5] Extensive evidence supports that the ASAS criteria perform well against this anchor,[32] but misclassification remains a matter of intense debate.[33] It has been argued that expert opinion may have contributed to designing criteria that encompass circular reasoning,[34, 35] that is, features deemed important by experts, especially those that allow early detection (e.g. sacroiliitis on MRI), were awarded a too prominent place in criteria that were subsequently again validated by experts. However, whether or not circularity has played a decisive role remains unclear, since an expert judgement free assessment of the *Gestalt* of axSpA has not been pursued so far. This is exactly what we have done in this study.

Using LCA we could describe the *Gestalt* of axSpA without any pre-assumptions on the contribution ('weight') of each SpA feature. This was only possible because LCA, following selection of parameters for analysis, does not need interpretational input from experts, whose beliefs therefore do not influence the analysis. The only inevitable influence experts potentially had was deciding if the patient should be included in the cohort. One of the phenotypes that arose from this analytical framework was a syndrome characterised by a high likelihood of axial imaging abnormalities, HLA-B27 positivity and male dominance, which we have subsequently labelled as 'axial'. This phenotype closely resembles the rheumatologist's conventional clinical picture of axSpA. Of note, LCA did not distinguish nr-axSpA from r-axSpA, even after forcing one additional class to the model. This is in line with extensive evidence suggesting that the split of axSpA in nr-axSpA and r-axSpA is artificial and supports the view that both are part of the same disease spectrum.[1, 26, 36, 37]

However, the 'axial' class is only one part of the *Gestalt* of axSpA: We identified a separate phenotype, defined by the presence of IBP (100%) in close conjunction with peripheral signs and symptoms ('IBP+peripheral'). These axSpA patients (mostly female) had back pain but were unlikely to be positive for sacroiliitis on imaging and HLA-B27. Thus, these patients rather fulfilled the pSpA than the axSpA classification criteria since the latter require either positive imaging ('imaging arm') or HLA-B27 ('clinical arm'). Formally, the ASAS pSpA criteria could not have been applied, since all patients had IBP.[4] We ignored this rule to better understand the possible overlap between *SpA* with predominantly peripheral features (original 'target' of the pSpA criteria) and *axSpA* with peripheral signs (the entity described here). The high percentage of 'IBP+Peripheral' patients fulfilling the pSpA criteria argues in favour of a significant overlap. This is in line with another study in DESIR, in which a different analytical approach (cluster analysis) was pursued that, unlike LCA, assumes an *a priori* presence of subgroups.[38] Taken all together, our findings undermine the current stand that either sacroiliitis on imaging or presence of HLA-B27 is mandatory to classify patients as axSpA. Several (female) patients presenting with IBP and concomitant peripheral manifestations but without manifest sacroiliitis or HLAB27 are not recognised as axSpA and therefore not included in axSpA trials. These patients have consistently shown to have significant burden of disease.[38-41] Whether or not these patients truly have inflammatory SpA or rather a chronic pain syndrome is a question that cannot be resolved by this analysis.

A third phenotype we identified is based on the presence of risk factors for axSpA (i.e. positive family history and HLA-B27) in association with IBP and only sporadically other SpA features. We have labelled this phenotype axSpA 'at risk'. Here 'at risk' means that patients present with features suggestive of axSpA, but such a diagnosis is not beyond any doubt. In other words, the 'at risk' class implies a higher level of uncertainty (grey zone) than the other classes, such as the 'axial' and the 'IBP+peripheral' classes. Too often, when dealing with uncertain or difficult cases clinicians apply classification criteria to inform binary diagnostic judgements (e.g. axSpA vs no axSpA) that do not allow grey zones. In addition, the anchoring features of this class (i.e. family history and HLA-B27) have shown redundancy,[42] but yet count separately for classification, which may contribute to overcalling axSpA when the ASAS axSpA criteria are wrongly used for diagnostic purposes. The high likelihood of IBP in these patients does not further help in discriminating SpA and no-SpA, since it also occurs in half of the patients of the 'No SpA' class. This is in line with recent data suggesting that specificity of IBP is lower than previously

thought.[43, 44] Although a longer follow-up may reveal more across-class switches over time, the low likelihood of 'at risk' patients to switch to a more profound phenotype within 5 years adds to the notion that 'At risk' patients may not have 'real' axSpA and will most often also not develop it later. A logical consequence would be to refrain from treating them as if they really have axSpA and from including these 'at risk' patients in axSpA trials which is indeed done as in addition to fulfilment of the ASAS criteria objective signs of inflammation are required.

In summary, we identified three latent phenotypes of the *Gestalt* of axSpA with a method that largely circumvents the circularity by expert opinion. 'Pure axial SpA' is the 'classical' phenotype of axSpA. 'axSpA with peripheral signs' is a recognisable phenotype in the spectrum of patients presenting with chronic back pain, best captured by the pSpA criteria suggesting that the overlap between axSpA and pSpA is larger than anticipated. The 'at Risk' class is the least well-defined of all entities and may encompass individuals at risk of axSpA, but without fully established disease, and also individuals that do not have SpA or will ever develop it. Studies addressing the prognosis of these subphenotypes, especially that of the 'at Risk' class, should inform us better on the real outcome of axial SpA.

SUPPLEMENTARY DATA

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

REFERENCES

1. Sieper J, van der Heijde D. Review: Nonradiographic axial spondyloarthritis: new definition of an old disease? Arthritis Rheum. 2013 Mar; 65(3):543-551.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. Landewé RBM. Overdiagnosis and overtreatment in rheumatology: a little caution is in order. Ann Rheum Dis. 2018 Oct; 77(10):1394-1396.
8. Dougados M, d'Agostino MA, Benessiano J, et al. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine. 2011 Dec; 78(6):598-603.
9. Lambert RG, Bakker PA, van der Heijde D, et al. Defining active sacroiliitis on MRI for classification of axial spondyloarthritis: update by the ASAS MRI working group. Ann Rheum Dis. 2016 Nov; 75(11):1958-1963.
10. Rudwaleit M, Jurik AG, Hermann KG, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. Ann Rheum Dis. 2009 Oct; 68(10):1520-1527.
11. de Hooge M, van den Berg R, Navarro-Compan V, et al. Patients with chronic back pain of short

duration from the SPACE cohort: which MRI structural lesions in the sacroiliac joints and inflammatory and structural lesions in the spine are most specific for axial spondyloarthritis? Ann Rheum Dis. 2016 Jul; 75(7):1308-1314.

12. Creemers MC, Franssen MJ, van't Hof MA, et al. Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system. Ann Rheum Dis. 2005 Jan; 64(1):127-129.
13. Brusco MJ, Shireman E, Steinley D. A comparison of latent class, K-means, and K-median methods for clustering dichotomous data. Psychol Methods. 2017 Sep; 22(3):563-580.
14. Magidson J, Vermunt JK. Latent class models for clustering : a comparison with K-means. 2002; 2002.
15. Schreiber JB. Latent Class Analysis: An example for reporting results. Res Social Adm Pharm. 2017 Nov; 13(6):1196-1201.
16. Lanza ST, Patrick ME, Maggs JL. Latent Transition Analysis: Benefits of a Latent Variable Approach to Modeling Transitions in Substance Use. J Drug Issues. 2010; 40(1):93-120.
17. Rudwaleit M, Feldtkeller E, Sieper J. Easy assessment of axial spondyloarthritis (early ankylosing spondylitis) at the bedside. Ann Rheum Dis. 2006 Sep; 65(9):1251-1252.
18. Rudwaleit M, van der Heijde D, Khan MA, et al. How to diagnose axial spondyloarthritis early. Ann Rheum Dis. 2004 May; 63(5):535-543.
19. Moll JM, Haslock I, Macrae IF, et al. Associations between ankylosing spondylitis, psoriatic arthritis, Reiter's disease, the intestinal arthropathies, and Behcet's syndrome. Medicine (Baltimore). 1974 Sep; 53(5):343-364.
20. Amor B, Dougados M, Mijiyawa M. [Criteria of the classification of spondylarthropathies]. Rev Rhum Mal Osteoartic. 1990 Feb; 57(2):85-89.
21. 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.
22. Zochling J, van der Heijde D, Dougados M, et al. Current evidence for the management of ankylosing spondylitis: a systematic literature review for the ASAS/EULAR management recommendations in ankylosing spondylitis. Ann Rheum Dis. 2006 Apr; 65(4):423-432.
23. Gossec L, van der Heijde D, Melian A, et al. Efficacy of cyclo-oxygenase-2 inhibition by etoricoxib and naproxen on the axial manifestations of ankylosing spondylitis in the presence of peripheral arthritis. Ann Rheum Dis. 2005 Nov; 64(11):1563-1567.
24. 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.
25. Sepriano A, Rudwaleit M, Sieper J, et al. Five-year follow-up of radiographic sacroiliitis: progression as well as improvement? Ann Rheum Dis. 2016 Jun; 75(6):1262-1263.
26. van den Berg R, Lenczner G, Feydy A, et al. Agreement between clinical practice and trained central reading in reading of sacroiliac joints on plain pelvic radiographs. Results from the DESIR cohort. Arthritis Rheumatol. 2014 Sep; 66(9):2403-2411.
27. van Tubergen A, Heuft-Dorenbosch L, Schulpen G, et al. Radiographic assessment of sacroiliitis by radiologists and rheumatologists: does training improve quality? Ann Rheum Dis. 2003 Jun; 62(6):519-525.
28. 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.
29. Dougados M, Demattei C, van den Berg R, et al. Rate and Predisposing Factors for Sacroiliac Joint Radiographic Progression After a Two-Year Follow-up Period in Recent-Onset Spondyloarthritis. Arthritis Rheumatol. 2016 Aug; 68(8):1904-1913.
30. Dougados M, Sepriano A, Molto A, et al. Sacroiliac radiographic progression in recent onset axial spondyloarthritis: the 5-year data of the DESIR cohort. Ann Rheum Dis. 2017 Nov; 76(11):1823-1828.
31. Sepriano A, Ramiro S, Landewé R, et al. Is active sacroiliitis on MRI associated with radiographic damage in axial spondyloarthritis? Real-life data from the ASAS and DESIR cohorts. Rheumatology (Oxford). 2019 May 1;58(5):798-802.
32. Sepriano A, Rubio R, Ramiro S, et al. Performance of the ASAS classification criteria for axial and peripheral spondyloarthritis: a systematic literature review and meta-analysis. Ann Rheum Dis. 2017 May; 76(5):886-890.
33. Dubreuil M, Deodhar AA. Axial spondyloarthritis classification criteria: the debate continues. Curr Opin Rheumatol. 2017 Jul; 29(4):317-322.
34. Lukas C, Cyteval C, Dougados M, et al. MRI for diagnosis of axial spondyloarthritis: major advance with critical limitations 'Not everything that glisters is gold (standard)'. RMD Open. 2018; 4(1):e000586.
35. 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.
36. Baraliakos X, Braun J. Non-radiographic axial spondyloarthritis and ankylosing spondylitis: what are the similarities and differences? RMD Open. 2015; 1(Suppl 1):e000053.
37. Sieper J, Hu X, Black CM, et al. Systematic review of clinical, humanistic, and economic outcome

comparisons between radiographic and non-radiographic axial spondyloarthritis. Semin Arthritis Rheum. 2017 Jun; 46(6):746-753.

38. Costantino F, Aegerter P, Dougados M, et al. Two Phenotypes Are Identified by Cluster Analysis in Early Inflammatory Back Pain Suggestive of Spondyloarthritis: Results From the DESIR Cohort. Arthritis Rheumatol. 2016 Jul; 68(7):1660-1668.
39. de Winter JJ, Paramarta JE, de Jong HM, et al. Peripheral disease contributes significantly to the level of disease activity in axial spondyloarthritis. RMD Open. 2019 Jan 11;5(1):e000802.
40. Lopez-Medina C, Dougados M, Ruyssen-Witrand A, et al. Evaluation of concomitant peripheral arthritis in patients with recent onset axial spondyloarthritis: 5-year results from the DESIR cohort. Arthritis Res Ther. 2019 Jun 6; 21(1):139.
41. Lopez-Medina C, Molto A, Dougados M. Peripheral manifestations in spondyloarthritis and their effect: an ancillary analysis of the ASAS-COMOSPA study. J Rheumatol. 2020 Feb;47(2):211-217.
42. van Lunteren M, van der Heijde D, Sepriano A, et al. Is a positive family history of spondyloarthritis relevant for diagnosing axial spondyloarthritis once HLA-B27 status is known? Rheumatology (Oxford). 2019 Sep 1;58(9):1649-1654.
43. de Hooge M, van Gaalen FA, Renson T, et al. Low specificity but high sensitivity of inflammatory back pain criteria in rheumatology settings in Europe: confirmation of findings from a German cohort study. Ann Rheum Dis. 2019 Nov;78(11):1605-1606.
44. Poddubnyy D, Callhoff J, Spiller I, et al. Diagnostic accuracy of inflammatory back pain for axial spondyloarthritis in rheumatological care. RMD Open. 2018 Dec 5;4(2):e000825.