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

Sepriano, A.R.

Chapter 1

General Introduction

THE *GESTALT* OF SPONDYLOARTHRITIS

Spondyloarthritis (SpA) is an 'umbrella' term currently used to describe a group of clinical entities with common clinical, laboratory and imaging features.[1] These entities are grouped into two main phenotypical *patterns*: patients with predominant involvement of the axial skeleton are labelled as axial SpA (axSpA); and patients with predominant peripheral manifestations as peripheral SpA (pSpA). Patients with axSpA with evidence of radiographic sacroiliitis, as defined by the modified New York criteria (mNY),[2] are considered as radiographic axSpA (r-axSpA), and those without as non-radiographic axSpA (nr-axSpA). This description embodies the current rheumatologists' perception of the *Gestalt* of SpA.

Gestalt is a German word, mostly used in the field of psychology to explain how human beings build meaningful perceptions from surrounding stimuli.[3] According to this concept, the 'true *Gestalt'* of an entity (e.g. SpA) is more than the combination of its parts (e.g. SpA features) and is, therefore, *unmeasurable* by definition. The rheumatologist's perception of the *Gestalt* of SpA has changed substantially over the years, which, to some extent, influenced the development of the various SpA classification criteria (see below). Assuming that the disease itself remained the same, this change implies that the rheumatologist's perception of the disease (the 'perceived Gestalt') did not always overlap with the 'true *Gestalt*'. Figure 1 graphically represents the theoretical relationships across the concepts of the 'true *Gestalt*' of SpA, the clinical diagnosis (i.e. rheumatologist's perception of the *Gestalt*) and the classification criteria. Although an approximately equal degree of overlap is seen, this is likely an over-simplification of the truth. Here is the *conundrum*: how to judge the 'correctness' of the rheumatologist's perception (i.e. diagnosis) against the 'true *Gestalt*' if the latter is impossible to measure by definition? In this thesis we will attempt to address this fundamental question. We start by evaluating the evolution of the interaction between rheumatologist's perception of the *Gestalt* of SpA and its classification criteria as summarised in Figure 2 and briefly described below.

The first in-depth clinical descriptions of SpA appeared in the medical literature by the end of the 19th century. Wladimir von Bechterew's classical description was pivotal in defining r-axSpA as a clinical entity, independent of rheumatoid arthritis (RA).[4] This entity was, back then, named ankylosing spondylitis and was in some countries also known as Bechterew's disease. With the advent of **roentgenology**, in the 1920s,[5] and its subsequent application in the evaluation of patients with r-axSpA,[6] features such as radiographic 'sacroiliitis' and syndesmophytes were observed for the first time. This technological breakthrough expanded the clinicians' perception of the disease and started **'Period one'** in our theoretical timeline of *the 'history of SpA'* (Figure 2). Not, surprisingly, in the 1960s-80s imaging findings were awarded a very prominent place in the first sets of classification criteria for r-axSpA (Table 1).[7, 8] The recognition that r-axSpA patients mostly present with CBP with inflammatory characteristics led to the proposal of the modified New York criteria (mNY) which 'survived' until today.[2, 9]

In 1974, Moll and Wright published one of the most influential manuscripts in the field of SpA, in which they proposed that r-axSpA, psoriatic arthritis (PsA), reactive arthritis, arthritis associated with inflammatory bowel disease (IBD) and juvenile SpA are diseases with common features and as such should be considered together as a group.[10] This group was appropriately coined with the name *seronegative spondyloarthritis* (or spondyloarthropathies) to highlight the weak association with rheumatoid factor (RF) and a predilection for the involvement of the axial

skeleton. Contributing to inform this innovative *clustering* was another scientific breakthrough, this time in the field of genetics. Researchers recognised that **HLA-B27** positivity occurred more frequently within this nosologic group than in other diseases.[11] Studies on the role of infection and the involvement of the gut in triggering spondyloarthritis also played a role.[12]

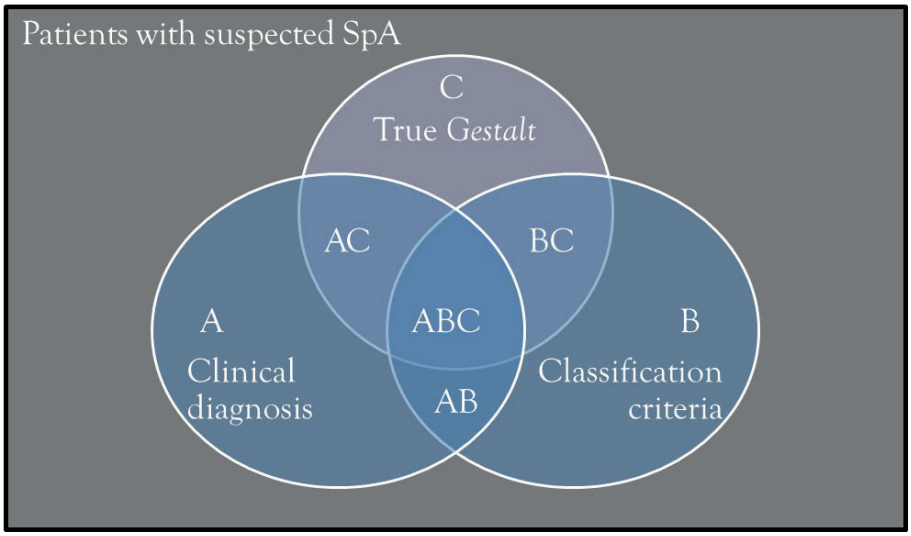

Figure 1. Relationship between clinical diagnosis (A), classification criteria (B) and the *Gestalt* (C) of axSpA in a cohort of patients with a suspected axSpA. The size of the circles and of their intersections do not necessarily represent the expected magnitude of the relationship between the three concepts. Interactions: 'AC', 'true SpA' phenotype recognised by the rheumatologist but not captured by the criteria; 'BC': 'true SpA' phenotype captured by the criteria but not recognised by the rheumatologist; 'AB', phenotype recognised by the rheumatologist and captured by the criteria but not representing 'true SpA' (misclassification and misdiagnosis); 'ABC': 'true SpA' phenotype recognised by the rheumatologist and captured by the criteria. 'A alone', a phenotype recognised only by the rheumatologist (wrong diagnosis); 'B alone': a phenotype captured only by criteria (misclassification): 'C alone': residual 'true SpA phenotype' intangible to rheumatologists and to the criteria they developed.

The change-of-paradigm proposal by Moll and Wright, undoubtedly changed the clinician's perception of SpA and marks the start of **'Period two'** in our timeline. Grouping together 'different' diseases, in theory, facilitates studies aiming at better understanding it. However, such studies need the proper 'tool' to guarantee that a homogeneous group of patients is included. While some of the diseases within the seronegative SpA concept had already their own classification criteria (e.g. r-axSpA, PsA, reactive arthritis), experts recognised that some patients with early and often milder forms did not classify as SpA even though they were perceived by the experts as having a *Gestalt* of SpA. This unmet need was addressed in the early 1990's with the development of the Amor and the European Spondyloarthropathy Study Group (ESSG) classification criteria.[13, 14] The Amor/ESSG expanded the range of manifestations allowing classification (Table 1). In addition, the term 'undifferentiated SpA' was coined to describe above-mentioned patients who fulfilled the ESSG classification criteria but did not fall within one of the major disease entities. The name of the disease was also changed. With such a wide spectrum of manifestations the term 'seronegative' became less relevant and was therefore abandoned. If we would build our Figure 1 based on the knowledge available when the mNY were developed and compare it with one based on knowledge present at the time of the Amor/ESSG criteria, an increase in the 'AC', and consequently, the 'BC' interaction would be evident. Obviously, this 'phenotypical expansion' is only apparent in retrospect.

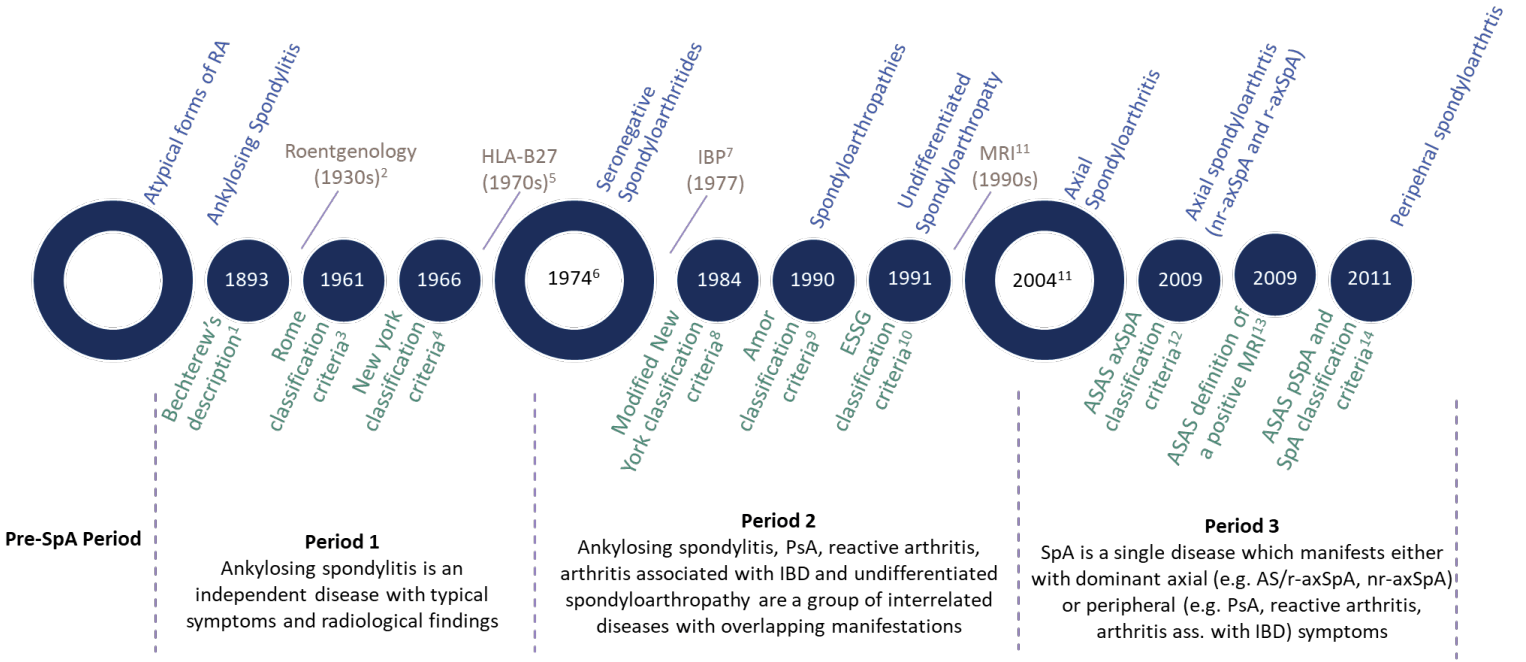

Figure 2. Evolution of the rheumatologist's perception for the *Gestalt* of spondyloarthritis translated into nomenclature and classification criteria in relation to paradigm-shift technological breakthroughs. **(1)** Bechterew W. Neurol Centralbl 1893;12:426–34. **(2)** Forestier J. Rheumatism 1964;20:28-34. **(3)** Kellgren JH, et al. Oxford: Blackwell 1963; 326-7. **(4)** Bennet PH, et al. Amsterdam: Exerpta Medica Foundation 1968:456-7. **(5)** Schlosstein L, et al. N Engl J Med 1973;288:704–6. **(6)** Moll JM, et al. Medicine (Baltimore) 1974;53:343–64. **(7)** Calin A, et al. JAMA 1977;237:2613-14. **(8)** Van der Linden S, et al. Arthritis Rheum 1984;27:241-9. **(9)** Amor B, et al. Rev Rhum Osteoartic 1990;57:85-9; **(10)** Dougados M, et al. Arthritis Rheum 1991;34(1):1218-27. **(11)** Braun J, et al. Arthritis Rheum 1994;37:1039–45. **(11)** Rudwaleit M, et al. Ann Rheum Dis. 2004;63(5):535-43. **(12)** Rudwaleit M, et al. Ann Rheum Dis 2009;68:777-83. **(13)** Rudwaleit M, et al. Ann Rheum Dis 2009;68:1520-7. (14) Rudwaleit M, et al. Ann Rheum Dis 2011;70(1):25-31. RA, rheumatoid arthritis; HLA-B27, human leukocyte antigen B27; IBP, inflammatory back pain; PsA, psoriatic arthritis; IBD, inflammatory bowel disease; ESSG, European Spondyloarthropathy Study Group; MRI, magnetic resonance imaging; ASAS, Assessment of SpondyloArthritis international Society; SpA, spondyloarthritis; axSpA, axial spondyloarthritis; nr-axSpA, non-radiographic axSpA; r-axSpA, radiographic axSpA; AS, ankylosing spondylitis; pSpA, peripheral spondyloarthritis.

Since the release of the Amor and ESSG criteria, new pieces of evidence had emerged that had, once again, changed our perception of the *Gestalt* of SpA and ultimately led to a new revision of the classification criteria. Neither the Amor nor the ESSG criteria distinguish between patients with predominant axial- and predominant peripheral-patterns. The relevance of such distinction was supported by studies showing that patients respond differently to treatment depending on the pattern.[15] Also, new evidence showed that not all SpA patients with predominant axial involvement develop pathological changes on pelvic radiographs (mNY-positive).[16, 17] A technological breakthrough was important here again, which marked the beginning of **'Period three'**. When **Magnetic resonance imaging** (MRI) became available, researchers found that inflammation was often seen on MRI of the sacroiliac joints (MRI-SIJ) long before changes appeared on pelvic radiographs.[18-20] Objective MRI inflammation could therefore be used to identify patients with SpA already early in the disease course. Such evidence prompted the Assessment of SpondyloArthritis international Society (ASAS) experts to develop classification criteria for patients with predominant axial involvement, with (r-axSpA) and without (nr-axSpA) radiographic sacroiliitis,[21] and for patients with predominant peripheral involvement that -if combined- would enclose the entire new perception of the *Gestalt* of SpA according to experts.[22] In our Figure 1, nr-axSpA would further expand the 'AC' and 'BC' interactions.

In the original validation studies, the ASAS SpA classification criteria proved to reflect the current perception of 'Gestalt' better than the ESSG and Amor criteria when tested against the expert's diagnosis. Since their release, the ASAS axSpA criteria,[21, 23-28] the pSpA criteria,[22, 28, 29] and the entire set,[22, 28] have consistently shown good criterion and construct validity against this 'external anchor'. However, it has been argued that the ASAS criteria are too loose and evoke confusion with patients with non-inflammatory disease (AB-misclassification in Figure 1).[30] Those who classify as nr-axSpA, especially in the absence of sacroiliitis on MRI-SIJ (the so-called 'clinical arm') have been the major source of criticism. Patients with nr-axSpA are more often female and less likely to show elevation of CRP than those with r-axSpA. [31-33] Despite similar burden of disease between nr-axSpA and r-axSpA, such differences are often pointed out as proof of mislabelling.[24, 34] Thus far, most studies testing the validity of the ASAS SpA criteria were cross-sectional. Arguably, testing their performance against an expert diagnosis made after a period of follow-up may yield more robust conclusions (predictive validity). However, the best that these studies can do is to inform us on how well the criteria 'capture' the expert's perception of the Gestalt of SpA (Area 'A' in Figure 1). This is not detrimental *per se*, provided that this perception is a good reflection of the 'true Gestalt' (Area 'C'). The inherent problem, though, is *circularity*: ASAS experts have developed the ASAS SpA criteria which were subsequently cross-validated against an expert's diagnosis. It has been argued that such circular reasoning may have contributed to develop criteria that are driven by experts' beliefs rather than by an objective presence of axSpA.[35, 36] An analysis of the *Gestalt* of SpA independent of the expert opinion, can contribute to clarify whether that was the case or not. However, such an analysis is lacking in the literature.

Table 1. History of classification criteria for spondylarthritis

Period 1 (1983 – 1974)			Period 2 (1974– 2004)		Period 3 (2004– present day)		
Rome criteria (1961)	New York criteria (1966)	Modified NY criteria (1984)	Amor criteria (1990)	ESSG criteria (1991)	ASAS criteria for SpA		
				Main criteria	Patients with ≥3 months of BP and age onset <45 years		Patients with peripheral symptoms only
				IBP (Calin) Synovitis	Sacroiliitis imaging HLA-B27		Arthritis or enthesitis or dactylitis
Clinical criteria (n=5)	**Clinical criteria (n=3)**	**Clinical criteria (n=3)**	**Clinical and lab criteria (n=11)**	**Clinical criteria (n=6)**	**SpA features (n=10)**		**SpA features (set 1; n=6)**
Spinal pain	**Spinal pain**	**Spinal pain**	**Spinal and buttock pain**	**Buttock pain**	**Spinal pain**		**Extra-articular features**
LBP and stiffness >3M, not relieved by rest	History/presence of pain at dorso-lumbar junction or LS	LBP and stiffness for > 3M which improves with exercise, but is not relieved by rest.	Pain at night (spine); or morning stiffness (1)	Altern. buttock pain	IBP (expert's definition)		Uveitis
Pain and stiffness in the TS	**Limited mobility**	**Limited mobility**	Any gluteal (buttock) pain; or alternating buttock pain (2)	**Peripheral features**	**Peripheral features**		Psoriasis
Limited mobility	Limitation of motion LS in all three planes*	Limitation of motion of the LS in both the sagittal and frontal planes.	**Peripheral features**	Enthesitis (heel)	Arthritis		IBD
Limited motion in the LS	Limitation of chest expansion to 2.5 cm, measured at the fourth intercostal space	Limitation of chest expansion relative to normal values corrected for age and sex.	Asymmetric oligoarthritis (2)	**Extra-articular features**	Enthesitis (heel)		**Previous infection¥**
Limited chest expansion			Dactylitis (2)	Psoriasis	Dactylitis		Preceding infection
Uveitis			Enthesitis (heel) (2)	IBD	**Extra-articular features**		**Genetical background**
History or evidence of iritis or its sequelae			**Extra-articular features**	**Previous infection¥**	Uveitis		HLA-B27
			Uveitis (2)	Urethritis, cervicitis or acute diarrhea	Psoriasis		**Imaging criterion**
			Pso, balanitis or IBD (2)	**Genetical background**	IBD		**Sacroiliitis on imaging**
			Previous infection¥	Family history£	**Genetical background**		**SpA features (set 2; n=5)**
			Urethritis/cervicitis (1)		Family history		**Spinal pain**
			Acute diarrhea (1)		HLA-B27		IBP
			Genetical background		**Response to NSAIDs**		**Peripheral features**
			HLA-B27/family history‡ (2)		Good response to NSAIDs		Arthritis
			Response to NSAIDs Rapid (<48 h) good response to NSAIDs (2)		**Other**		Enthesitis
					Elevated CRP		Dactylitis
							Genetical background
Radiological criterion	**Radiological criteria†**	**Radiological criteria†**	**Radiological criteria†**	**Radiological criteria†**			Family history
Bilateral sacroiliac changes characteristic of AS	(a) Bilateral sacroiliitis: grade 3-4	(a) Bilateral sacroiliitis: grade 2-4	Sacroiliitis (bilateral grade 2 or unilateral grade 3): 2	Bilateral sacroiliitis: grade 2-4 or unilateral sacroiliitis: grade 3-4			
	(b) Unilateral sacroiliitis: grade 3-4; or bilateral sacroiliitis: grade 2	(b) unilateral sacroiliitis: grade 3-4					
Positive classification	**Positive classification**	**Positive classification**	**Positive classification**	**Positive classification**	**Positive classification**		**Positive classification**
Bilateral sacroiliitis plus ≥1/5 clinical criteria	**Definite AS**	**Definite AS**	Score ≥ 6 out of 23 possible points	IBP or synovitis plus ≥1 of the 7 additional criteria	Sacroiliitis on imaging plus ≥ 1 SpA feature; or HLA-B27 plus ≥ 2 SpA features		Arthritis or enthesitis or dactylitis plus ≥ 1 SpA feature from set 1, or ≥2 SpA features from set 2
	Grade 3-4 bilateral sacroiliitis plus ≥1 clinical criterion; or grade 3-4 unilateral or grade 2 bilateral sacroiliitis plus clinical criterion 2 or with both clinical criteria 1 and 3	Bilateral sacroiliitis: grade 2-4 or unilateral sacroiliitis: grade 3-4 plus ≥1 clinical criterion					

* Anterior flexion, lateral flexion and extension; † Grade 0: normal; grade 1: suspicious changes; Grade 2: Minimal abnormality – small localized areas with erosion or sclerosis without alterations on joint width; grade 3: unequivocal abnormality – moderate or advanced sacroiliitis with ≥ 1 of the following: erosions, sclerosis, widening, narrowing or partial ankylosis; grade 4: severe abnormality – total ankylosis; ¥ ≤ 1 month of arthritis onset; ‡ AS, ReA, uveitis, psoriasis or IBD; £ Presence in first-degree or second-degree relatives of any of the following: AS, psoriasis, acute uveitis, ReA, IBD; LBP, low back pain; M, months; AS, ankylosing spondylitis; NY, New York; TS, thoracic spine; LS, lumbar spine; Lab, laboratory; ESSG, European Spondyloarthropathy Study Group; NSAIDs, Nonsteroidal anti-inflammatory drugs; IBP, inflammatory back pain; MRI, magnetic resonance imaging; ASAS, Assessment of SpondyloArthritis international Society; SpA, spondyloarthritis; CRP, C reactive protein; IBD, inflammatory bowel disease; HLA-B27, human leukocyte antigen B27; ReA, reactive arthritis.

IMAGING OUTCOMES: CHALLENGES AND OPPORTUNITIES

For several years, conventional radiography has been the imaging modality of choice to assess progression of axial structural damage in axSpA. The modified stoke ankylosing spondylitis spinal score (mSASSS) is the most sensitive to change and valid scoring method to quantify progression in spinal radiographs.[37, 38] Despite large inter-patient variability, it is estimated that on average, patients with r-axSpA show an increase of approximately 2 mSASSS units every 2 years.[39] Lacking better alternatives, the mNY grading system has been the most often used score to measure progression in SIJs. According to this score, definite damage is defined as the presence of bilateral grade 2 or unilateral grade 3 or 4 'sacroiliitis' ('mNY-positive').[2] It is estimated that approximately 10% of patients with axSpA progress from mNY-negative to mNY-positive over 2 years.[40] However, the mNY was not originally developed as an outcome measure but rather for classification purposes (see above). Also, unlike mSASSS, the mNY suffers from poor reliability, even when assessed by trained readers, which may have led to biased estimates of progression.[16, 17] An analogy with the concept of 'signal-to-noise ratio' in electronics has been recently proposed to illustrate the problem.[41] This ratio incorporates two types of information: 'true change' ('signal') and error change ('noise'). The larger the measurement error, the harder to capture the 'signal'. Approaches to optimize the detection of the 'signal' have been proposed, such as protocolled imaging acquisition, combining scores from multiple trained central readers and scoring with concealed time order. However, these strategies cannot fully eliminate the 'noise'. Thus, the 'true' rate of structural progression at the SIJs remains uncertain.

Researchers have not only pursued accurate measurements of radiographic progression but also predictors thereof. Over the last decade, a lot of focus was on whether inflammation drives structural spinal damage in axSpA.[42-47] These efforts have yielded a reasonably solid base of evidence supporting such a claim. For example, it has been shown that one unit increase in ASDAS, a validated measure of systemic inflammation in axSpA, results in an increase of 0.72 mSASSS units 2 years later.[43] Also, bone marrow edema (a lesion reflecting an underlying inflammatory process) in vertebral corners, visualized on MRI of the spine, predicts the subsequent development of syndesmophytes at the same location.[47] At the SIJ level this association is far less well studied.[20, 48] It should be noted, however, that in both above-mentioned studies the magnitude of the effect was rather low and, importantly, new bone formation still occurred in the absence of inflammation. This suggests that damage accrual in axSpA is only partially dependent on inflammation-driven processes, which may explain why it has been difficult to demonstrate that anti-inflammatory drugs (e.g. TNF inhibitors) halt, or at least retard, bone formation.

Measuring the subtle association between inflammation and damage with more precision (narrower confidence intervals) may increase the statistical power to unveil the structural effects of therapeutical intervention, that many experts believe must be present, within the relatively short 'window' and the small sample size of randomised clinical trials. It has been suggested that this ambitious goal can be achieved with other imaging modalities than conventional radiographs.[49] Even though mSASSS is the spinal radiographic outcome measure of choice with highest sensitivity to change, at least 2 years are needed for a meaningful change to be detected and for the subtle association between inflammation and damage to become

apparent.[39] In pelvic radiographs, low reliability of the mNY grading further challenges the detection of meaningful associations.[16, 17] Considering such limitations with conventional radiography, in recent years there has been a growing interest in evaluating axial damage with MRI. T1-weighted (T1W) sequences on MRI have been shown to accurately detect chronic changes in the spine and SIJ in patients with axSpA. These include fatty lesions, erosions, bone spurs, sclerosis and ankylosis. In addition, scores combining individual structural lesions on MRI have been validated and can thus be used in clinical studies.[19, 50-52] However, thus far, no study has evaluated whether inflammation seen on MRI predicts structural progression also assessed on MRI and, importantly, it remains to be proven if MRI outcomes are truly more sensitive to change than radiographic outcomes as they are thought to be.

OVERARCHING AIMS OF THIS THESIS

- To test the longitudinal validity of the ASAS SpA classification criteria against an expert clinical diagnosis;
- To gain better insight into the *Gestalt* of axSpA, independent of expert judgement, and to evaluate how the ASAS SpA criteria 'capture' this *Gestalt*;
- To identify and address the challenges in measuring and analysing structural damage progression at the SIJ and spinal level and its relationship with inflammation.

To address the aims of this thesis, data from 3 cohorts were used.

The **ASAS cohort** is an international, multicentre, prospective study. From November 2005 until January 2009, rheumatologists from 29 ASAS centres worldwide have included 975 consecutive patients who first presented for diagnostic work-up. To be included, patients had to have no definitive diagnosis and to fulfil one of two criteria: i) chronic (>3 months) back pain of unknown origin (no definite diagnosis) with an age of onset below 45 years, with or without peripheral symptoms; ii) patients with peripheral arthritis and/or enthesitis and/or dactylitis and absence of current back pain with suspicion of SpA.[21, 22] Of the 29 original ASAS centres, 22 participated in the follow-up assessment (mean 4.4 years) corresponding to 909 of the original 975 patients. In total, 564 patients had a follow-up assessment with 345 physically attended the follow-up visit and 219 provided only information via telephone. Data from the ASAS cohort is presented in **chapters 2, 5 and 8**.

The **SPACE cohort** is an ongoing observational study initiated in 2009 at the Leiden University Medical Center (LUMC, the Netherlands).[24] Patients aged ≥ 16 years with chronic back pain (CBP; ≥3 months, ≤2 years and onset <45 years), of unknown origin, referred to the rheumatology outpatient clinic were included. The presence of other painful conditions not associated with axSpA that could interfere with the evaluation of disease activity led to exclusion. Patients were recruited from multiple rheumatology centres in Europe; the Netherlands (Leiden, Amsterdam and Gouda), Norway (Oslo), Italy (Padova) and Sweden (Göteborg, Malmö, Falun, Skövde, Västerås, Huddinge, Stockholm). A detailed description of the SPACE cohort has been published elsewhere.[24] Data from the SPACE cohort is presented in **chapter 4**.

DESIR is a longitudinal inception cohort for which the inclusion period was between December 2007 and April 2010 in 25 participating centres in France.[53] Patients between 18 and 50 years old with IBP according to the Calin[9] or Berlin[54] criteria (≥3 months, ≤3 years) were included. Moreover, the symptoms had to be suggestive of axSpA according to the treating rheumatologist expressed in a score of five or more on a numerical rating scale from zero to ten. Main exclusion criteria were: the presence of a clearly defined spinal disease; history of treatment with any biological drug; and corticosteroid intake of a dose higher than 10 mg prednisone per day prior to baseline. Data from the DESIR cohort is presented in **chapters 4 and 6 to 10**.

THESIS OVERVIEW

The first aim of this thesis is to evaluate the longitudinal validity of the ASAS SpA classification criteria tested against the rheumatologist's diagnosis. In the original validation studies, the ASAS axSpA, pSpA and SpA (axSpA and pSpA combined) criteria had shown good specificity against this concurrent 'external reference' (84%, 82%, and 84%, respectively).[21, 22] Despite that, as mentioned above, some argue that the criteria are too lenient and lead to 'mislabelling'. Most criticism pertains to patients who classify without definite damage on pelvic radiographs (nr-axSpA).[30] Similar concerns apply to the pSpA criteria.[55] Specificity, tells us how likely it is for a patient to classify as negative if he/she does not have a clinical diagnosis of SpA. Arguably, a more relevant question would be: how likely is it for a patient to receive a clinical diagnosis of SpA if he/she classifies as positive? The mathematical representation of the latter is the positive predictive value and is better appreciated if the clinical diagnosis is made after a certain period of follow-up. In **Chapter 2**, we used the follow-up data from the ASAS cohort to test the predictive validity of the SpA classification criteria. In **Chapter 3**, we have systematically reviewed studies that have challenged the ASAS SpA classification criteria by reproducing the original validation exercise in different populations than the one from the ASAS cohort and combined their data to report pooled metrics of performance.

In the study in chapter 2, like in the literature reviewed in chapter 3, the ASAS classification criteria were tested against the expert diagnosis as 'external-standard'. This is a common approach in rheumatology which is, however, not without limitations. Arguably, *circular reasoning*, is the most important one, that may ultimately jeopardise the construct and content validity of the ASAS criteria. Circularity occurs when experts find certain characteristics more important than others, and such characteristics are awarded a too prominent place in the criteria. Subsequent cross-validation against an expert diagnosis may produce results driven by experts' beliefs rather than by an objective presence of axSpA. These beliefs, as mentioned above, are volatile and have changed considerably over the years. In '21st century rheumatology' early detection of SpA is a priority, so it is not surprising that features such as BME on MRI-SIJ became prominent in the ASAS axSpA criteria. But the question remains to what extent these beliefs reflect the 'true Gestalt' of axSpA. Circularity leads to mislabelling ('AB' interaction in Figure 1) if the overlap between the expert perception of the *Gestalt* and the 'true *Gestalt*' ('AC') is too narrow. In such scenario, 'overdiagnosis', and consequently 'overtreatment', occurs if classification criteria are wrongly applied for diagnostic purposes. To find resolution on whether circularity played a role in the development of the ASAS axSpA criteria, an analysis independent

of expert opinion is needed. In **Chapter 4**, we address the second aim of this thesis, by using an analytical technique that circumvents expert opinion (latent class analysis) to determine, in a circularity-free manner, the *Gestalt* of axSpA and to evaluate how the ASAS classification criteria capture its latent constructs.

Definite damage at the SIJ level is defined according to the mNY grading system, as the presence of bilateral grade 2 or unilateral grade 3 or 4 'sacroiliitis' ('mNY-positive').[2] As mentioned above, this definition was originally proposed to classify patients with r-axSpA. With the recognition that patients with axSpA are often mNY-negative (nr-axSpA), especially early in the disease, the use of the mNY scoring as an outcome measure gained popularity (i.e. change from mNY-negative to mNY-positive).[56-59] However, progression measured by the mNY in pelvic radiographs has been found rather unreliable.[16, 17] In **Chapter 5**, we add to this evidence by comparing two films read several years apart, by local untrained readers, among patients with suspected SpA from the ASAS cohort. Findings from this study address the third aim of the thesis and led us to propose, in **Chapter 6**, a new analytical approach, the so-called 'assumption-free net progression', which we argue best handles measurement error in the context of a binary judgement, such as a change from mNY-negative to mNY-positive.

Patients with axSpA experience varying levels of radiographic progression, both at the spinal and at the SIJ level. Inflammation has been shown to predict damage accrual at the spinal level, but similar evidence at the SIJ, especially in early disease is mostly absent.[42-47] In **Chapter 7**, we evaluate if objective inflammation on MRI-SIJ is associated with radiographic progression at the SIJ level in patients with early axSpA. Using data from three trained central readers, different definitions of progression were tested, including the change from mNY-negative to mNY-positive, and applying the method described in Chapter 6. In **Chapter 8**, we perform a similar analysis using scores yielded by local untrained readers from DESIR and also from the ASAS cohort in order to understand whether inflammation in MRI-SIJ, as available in clinical practice, can effectively be used for prognostic stratification. Given the limitations of radiographs in the assessment of progression of structural damage, MRI has been proposed as an alternative imaging modality. In fact, current evidence supports the view that MRI is able to detect structural damage with higher reliability than conventional radiographs at the SIJ level, thus arguing in favour of its use in studies evaluating structural progression and predictors thereof.[36] In **Chapter 9** we test the longitudinal effect of inflammation on MRI of the SIJ and spine on the subsequent development of structural damage also measured on MRI over five years of follow-up.

Collection and analysis of long-term imaging data pose important methodological challenges. The longer the study the higher the likelihood of loss to follow-up (right censoring bias). Also, different readers may contribute to obtaining scores, in multiple 'reading-waves' over time. Common solutions include analyzing completers only, to choose a convenient read wave, and to aggregate scores of individual readers into some algorithm. Such approaches are not assumption-free and may as such yield biased estimates. In **Chapter 10** we investigated if a technique that makes use of all available information provided by all readers in different 'reading-waves' in an assumption-free manner (a so called 'integrated analysis'),[60] will affect the precision of parameter estimates for imaging outcomes in patients with axSpA, with a conventional completers analysis as reference standard. In **Chapter 11** we applied this method

to compare the sensitivity to change of various inflammatory and structural outcomes measured in MRIs and conventional radiographs of the SIJ and spine performed in patients with early axSpA.

A summary and general discussion about the findings of this thesis is presented in **Chapter 12**. A summary of this thesis in Dutch is provided in **Chapter 13**.

References

1. van Tubergen A. The changing clinical picture and epidemiology of spondyloarthritis. Nat Rev Rheumatol. 2015 Feb; 11(2):110-118.
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. Arnheim R. The Gestalt Theory of Expression. Psychological Review. 1949; 56:156-171.
4. Bechterew W. Steifigkeit der Wirbelsaule und ihre Verkrummung alsbesondere Erkrankungsform. Neurol Centralbl. 1893; 12:426–434.
5. Forestier J, Jacqueline F, Rolesquerol J, eds. Ankylosing spondylitis: clinical considerations, roentgenology, pathologic anatomy, treatment. Springfield (IL): Charles C Thomas. 1956. (Translated by AU Desjardins.).
6. Forestier J. Ankylosing Spondylitis at the Beginning of the Century. Rheumatism. 1964 Apr; 20:28-34.
7. Kellegren JH, Jeffrey, M. R., AND Ball, J. (eds). The Epidemiology of Chronic Rheumatism. Blackwell, Oxford. 1963; 1:326-327.
8. Bennett PH, Burch TA. The epidemiological diagnosis of ankylosing spondylitis. In: Bennett PH, Wood PHN, eds. Proceedings of the 3rd international symposium of population studies of the rheumatic diseases. Amsterdam: Excerpta Medica Foundation. 1966:305–313.
9. Calin A, Porta J, Fries JF, et al. Clinical history as a screening test for ankylosing spondylitis. JAMA. 1977 Jun 13; 237(24):2613-2614.
10. 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.
11. Schlosstein L, Terasaki PI, Bluestone R, et al. High association of an HL-A antigen, W27, with ankylosing spondylitis. N Engl J Med. 1973 Apr 5; 288(14):704-706.
12. Sieper J, Braun J, Rudwaleit M, et al. Ankylosing spondylitis: an overview. Ann Rheum Dis. 2002 Dec; 61 Suppl 3:iii8-18.
13. Amor B, Dougados M, Mijiyawa M. [Criteria of the classification of spondylarthropathies]. Rev Rhum Mal Osteoartic. 1990 Feb; 57(2):85-89.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.

23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. Landewé RB. Magnetic resonance imaging in the diagnosis of ankylosing spondylitis: be aware of gold standards and circularity. J Rheumatol. 2010 Mar; 37(3):477-478.
36. 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.
37. Ramiro S, Claudepierre P, Sepriano A, et al. Which scoring method depicts spinal radiographic damage in early axial spondyloarthritis best? Five-year results from the DESIR cohort. Rheumatology (Oxford). 2018 Nov 1;57(11):1991-2000.
38. Ramiro S, van Tubergen A, Stolwijk C, et al. Scoring radiographic progression in ankylosing spondylitis: should we use the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) or the Radiographic Ankylosing Spondylitis Spinal Score (RASSS)? Arthritis Res Ther. 2013 Jan 17; 15(1):R14.
39. Ramiro S, Stolwijk C, van Tubergen A, et al. Evolution of radiographic damage in ankylosing spondylitis: a 12 year prospective follow-up of the OASIS study. Ann Rheum Dis. 2015 Jan; 74(1):52-59.
40. Sieper J, van der Heijde D. Review: Nonradiographic axial spondyloarthritis: new definition of an old disease? Arthritis Rheum. 2013 Mar; 65(3):543-551.
41. Landewé RBM, van der Heijde D. "Big Data" in Rheumatology: Intelligent Data Modeling Improves the Quality of Imaging Data. Rheum Dis Clin North Am. 2018 May; 44(2):307-315.
42. Poddubnyy D, Protopopov M, Haibel H, et al. High disease activity according to the Ankylosing Spondylitis Disease Activity Score is associated with accelerated radiographic spinal progression in patients with early axial spondyloarthritis: results from the GErman SPondyloarthritis Inception Cohort. Ann Rheum Dis. 2016 Dec; 75(12):2114-2118.
43. Ramiro S, van der Heijde D, van Tubergen A, et al. Higher disease activity leads to more structural damage in the spine in ankylosing spondylitis: 12-

year longitudinal data from the OASIS cohort. Ann Rheum Dis. 2014 Aug; 73(8):1455-1461.

44. Machado PM, Baraliakos X, van der Heijde D, et al. MRI vertebral corner inflammation followed by fat deposition is the strongest contributor to the development of new bone at the same vertebral corner: a multilevel longitudinal analysis in patients with ankylosing spondylitis. Ann Rheum Dis. 2016 Aug; 75(8):1486-1493.
45. Maksymowych WP, Chiowchanwisawakit P, Clare T, et al. Inflammatory lesions of the spine on magnetic resonance imaging predict the development of new syndesmophytes in ankylosing spondylitis: evidence of a relationship between inflammation and new bone formation. Arthritis Rheum. 2009 Jan; 60(1):93-102.
46. Baraliakos X, Heldmann F, Callhoff J, et al. Which spinal lesions are associated with new bone formation in patients with ankylosing spondylitis treated with anti-TNF agents? A long-term observational study using MRI and conventional radiography. Ann Rheum Dis. 2014 Oct; 73(10):1819-1825.
47. van der Heijde D, Machado P, Braun J, et al. MRI inflammation at the vertebral unit only marginally predicts new syndesmophyte formation: a multilevel analysis in patients with ankylosing spondylitis. Ann Rheum Dis. 2012 Mar; 71(3):369-373.
48. 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.
49. van der Heijde D, Landewé R. Inhibition of spinal bone formation in AS: 10 years after comparing adalimumab to OASIS. Arthritis Res Ther. 2019 Nov 6; 21(1):225.
50. Ostergaard M, Maksymowych WP, Pedersen SJ, et al. Structural lesions detected by magnetic resonance imaging in the spine of patients with spondyloarthritis - Definitions, assessment system, and reference image set. Journal of Rheumatology. 2009; 36(SUPPL. 84):18-34.
51. Krabbe S, Sorensen IJ, Jensen B, et al. Inflammatory and structural changes in vertebral bodies and posterior elements of the spine in axial spondyloarthritis: construct validity, responsiveness and discriminatory ability of the anatomy-based CANDEN scoring system in a randomised placebo-controlled trial. RMD Open. 2018; 4(1):e000624.
52. Maksymowych WP, Lambert RG, Ostergaard M, et al. MRI lesions in the sacroiliac joints of patients with spondyloarthritis: an update of definitions and validation by the ASAS MRI working group. Ann Rheum Dis. 2019 Nov;78(11):1550-1558.
53. 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.
54. Rudwaleit M, Metter A, Listing J, et al. Inflammatory back pain in ankylosing spondylitis: a reassessment of the clinical history for application as classification and diagnostic criteria. Arthritis Rheum. 2006 Feb; 54(2):569-578.
55. Gladman DD. Editorial: What is peripheral spondyloarthritis? Arthritis Rheumatol. 2015 Apr; 67(4):865-868.
56. Schattenkirchner M, Kruger K. Natural course and prognosis of HLA-B27-positive oligoarthritis. Clin Rheumatol. 1987 Sep; 6 Suppl 2:83-86.
57. Mau W, Zeidler H, Mau R, et al. Outcome of possible ankylosing spondylitis in a 10 years' follow-up study. Clin Rheumatol. 1987 Sep; 6 Suppl 2:60-66.
58. Sampaio-Barros PD, Conde RA, Donadi EA, et al. Undifferentiated spondyloarthropathies in Brazilians: importance of HLA-B27 and the B7-CREG alleles in characterization and disease progression. J Rheumatol. 2003 Dec; 30(12):2632-2637.
59. Sampaio-Barros PD, Bortoluzzo AB, Conde RA, et al. Undifferentiated spondyloarthritis: a longterm followup. J Rheumatol. 2010 Jun; 37(6):1195-1199.
60. Landewé R, Ostergaard M, Keystone EC, et al. Analysis of integrated radiographic data from two long-term, open-label extension studies of adalimumab for the treatment of rheumatoid arthritis. Arthritis Care Res (Hoboken). 2015 Feb; 67(2):180-186.