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

Sepriano, A.R.

The *Gestalt* of Spondyloarthritis

From early recognition to long-term imaging outcomes

Alexandre Sepriano

2020

ISBN: 9789464190359

Alexandre Sepriano was supported by the *Fundação para a Ciência e Tecnologia* (FCT) grant SFRH/BD/108246/2015.

Cover design and thesis layout: GildePrint

Printing: Gildeprint - Enschede (www.gildeprint.nl).

Printing of this thesis was financially supported by the Dutch Arthritis Foundation (ReumaNederland) and UCB pharma B.V. which is gratefully acknowledged.

The *Gestalt* of Spondyloarthritis

From early recognition to long-term imaging outcomes

Proefschrift

Ter verkrijging van de graad van doctor aan de Universiteit Leiden, op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker, volgens besluit van het College voor Promoties te verdedigen op donderdag 19 November 2020 klokke 11:15 uur

door

Alexandre Rocha Sepriano

geboren te Lissabon, Portugal in 1985

Promotores:	Prof. dr. D.M.F.M. van der Heijde
	Prof. dr. R.B.M. Landewé (Universiteit van Amsterdam)
Copromotor:	Dr. M.S.A.C.O. Ramiro
Leden promotiecommissie:	Prof. dr. T.W.J. Huizinga
	Dr. F.A. van Gaalen
	Prof. dr. O.M. Dekkers
	Prof. dr. M. Boers (Universiteit van Amsterdam)
	Prof. dr. A.E.R.C.H. Boonen (Universiteit van Maastricht)

Medicine is a science of uncertainty and an art of probability

Sir William Osler

(1849-1919)

To Leonor and Carlos

To Cátia, Catarina and Beatriz

TABLE OF CONTENTS

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]

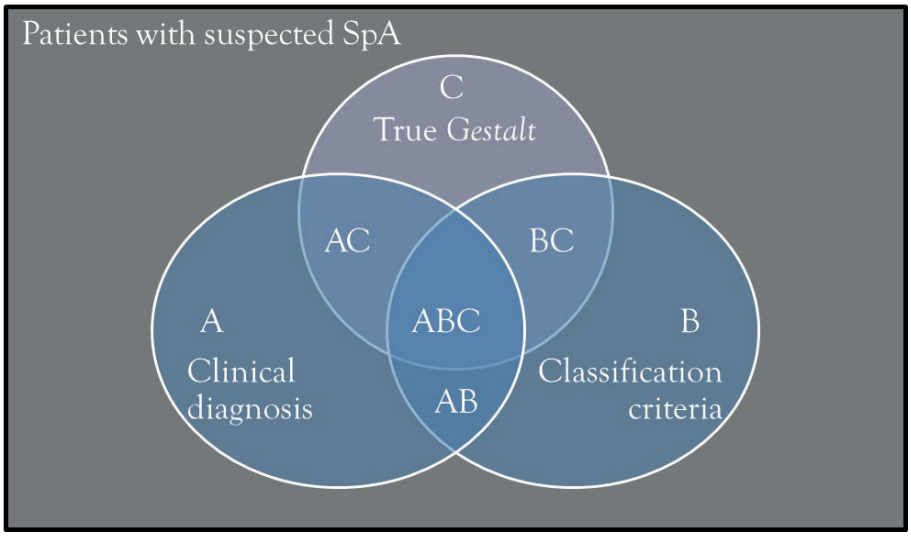


Figure 1. Relationship between clinical diagnosis (A), classification criteria (B) and the *Gestalt* (C) of axSpA in a cohort of patients with a suspected axSpA. The size of the circles and of their intersections do not necessarily represent the expected magnitude of the relationship between the three 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.

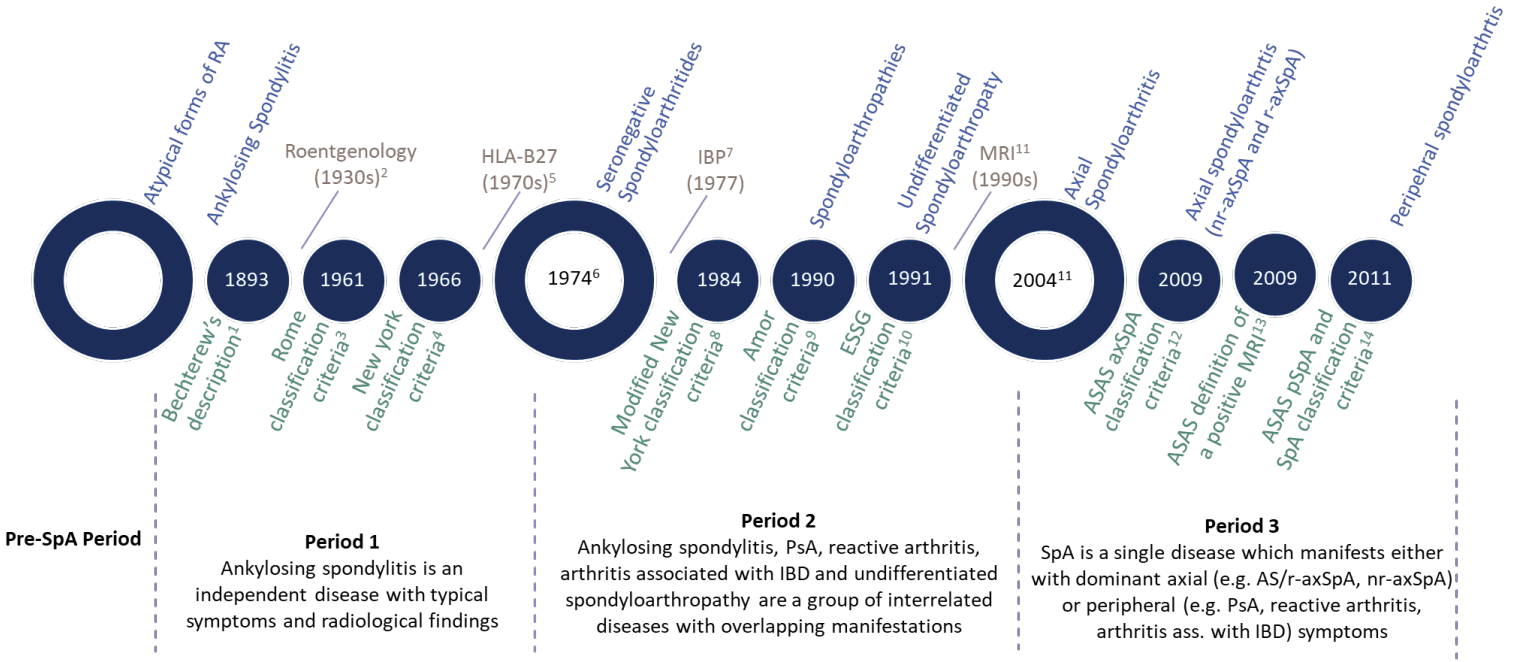


Figure 2. Evolution of the rheumatologist's perception for the *Gestalt* of spondyloarthritis translated into nomenclature and classification criteria in relation to paradigm-shift technological breakthroughs. **(1)** Bechterew W. Neurol Centralbl 1893;12:426–34. **(2)** Forestier J. Rheumatism 1964;20:28-34. **(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.

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.

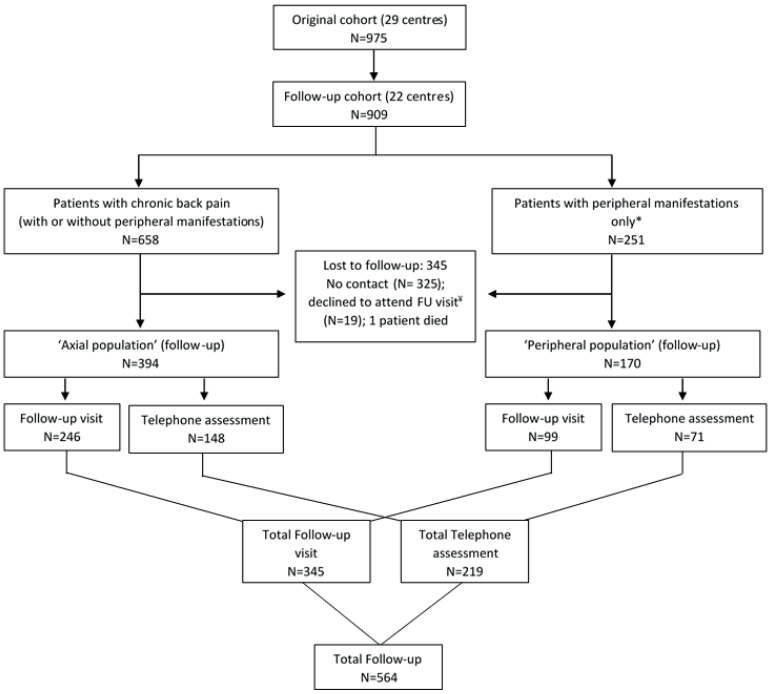


Figure 1. Follow-up of the Assessment of SpondyloArthritis international Society (ASAS) Cohort. *Patients with undiagnosed peripheral arthritis, and/or enthesitis, and/or dactylitis and absence of current back pain. ¥ And did not provide information via telephone. FU, follow-up.

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).

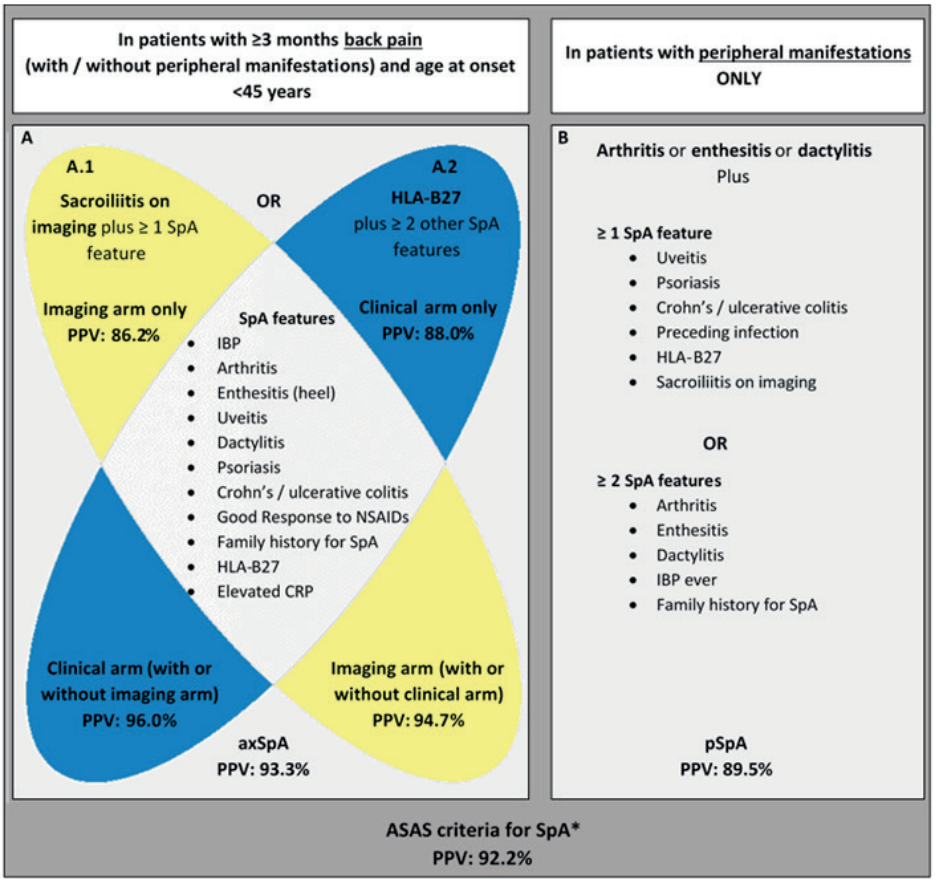


Figure 2. Predictive validity of the Assessment of SpondyloArthritis international Society (ASAS) classification criteria. *Combination of: **(A)** ASAS criteria for axial spondyloarthritis (axSpA) applied in patients with predominant back pain, with or without peripheral manifestations (N=394) and **(B)** 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.

Chapter 3

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

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

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

ABSTRACT

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

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

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

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

INTRODUCTION

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

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

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

METHODS

Literature search

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

Study selection, data extraction and assessment of risk of bias

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

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

Data analysis

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

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

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

RESULTS

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

Study populations

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

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

Table 1. Main study characteristics

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

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

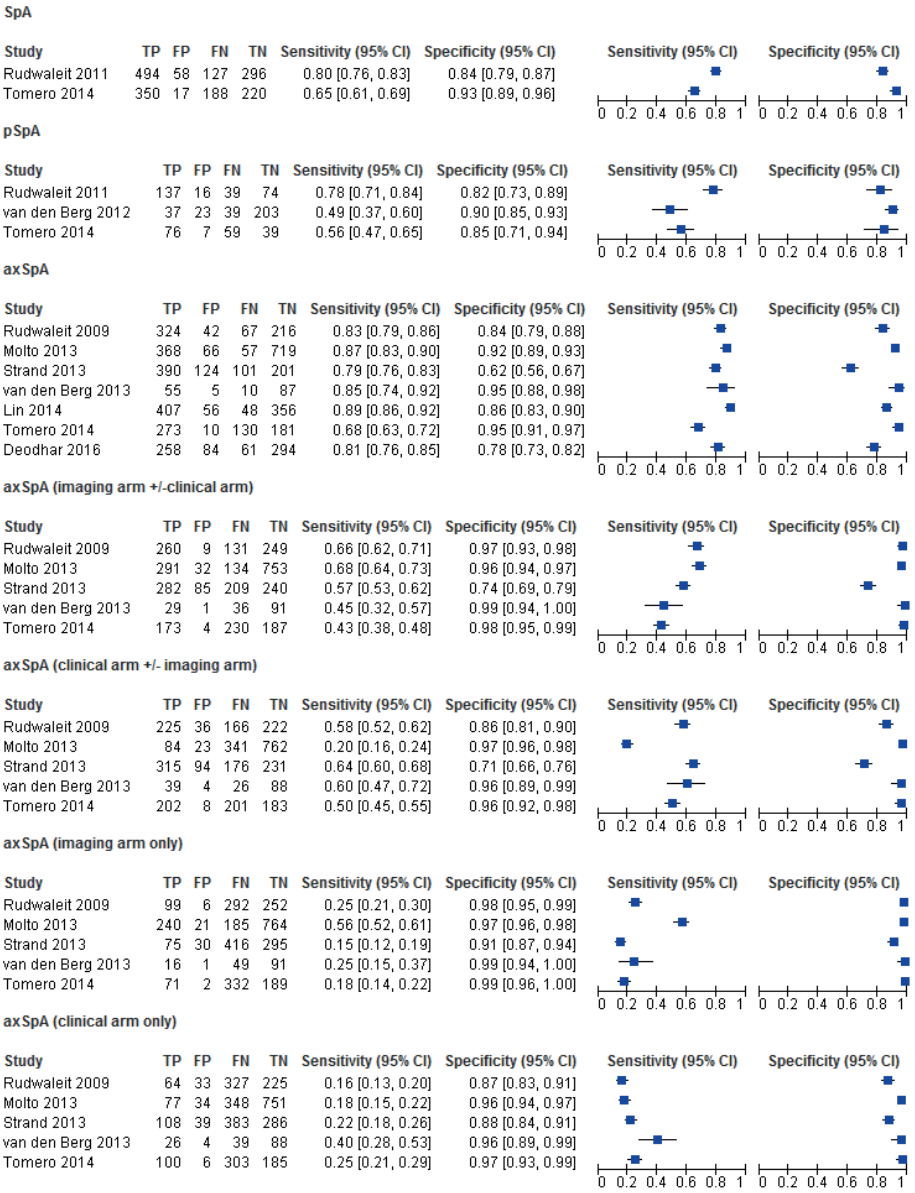


Figure 1. Performance of the ASAS SpA classification criteria across studies. ASAS, Assessment of SpondyloArthritis international Society; axSpA, axial spondyloarthritis; pSpA, peripheral spondyloarthritis; CI, confidence interval; TP, true positives, FP, false positives; FN, false negatives; TN, true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

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

Performance of the ASAS SpA classification criteria

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

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

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

Sensitivity analyses

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

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

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

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

DISCUSSION

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

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

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

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

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

SUPPLEMENTARY DATA

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

REFERENCES

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

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.

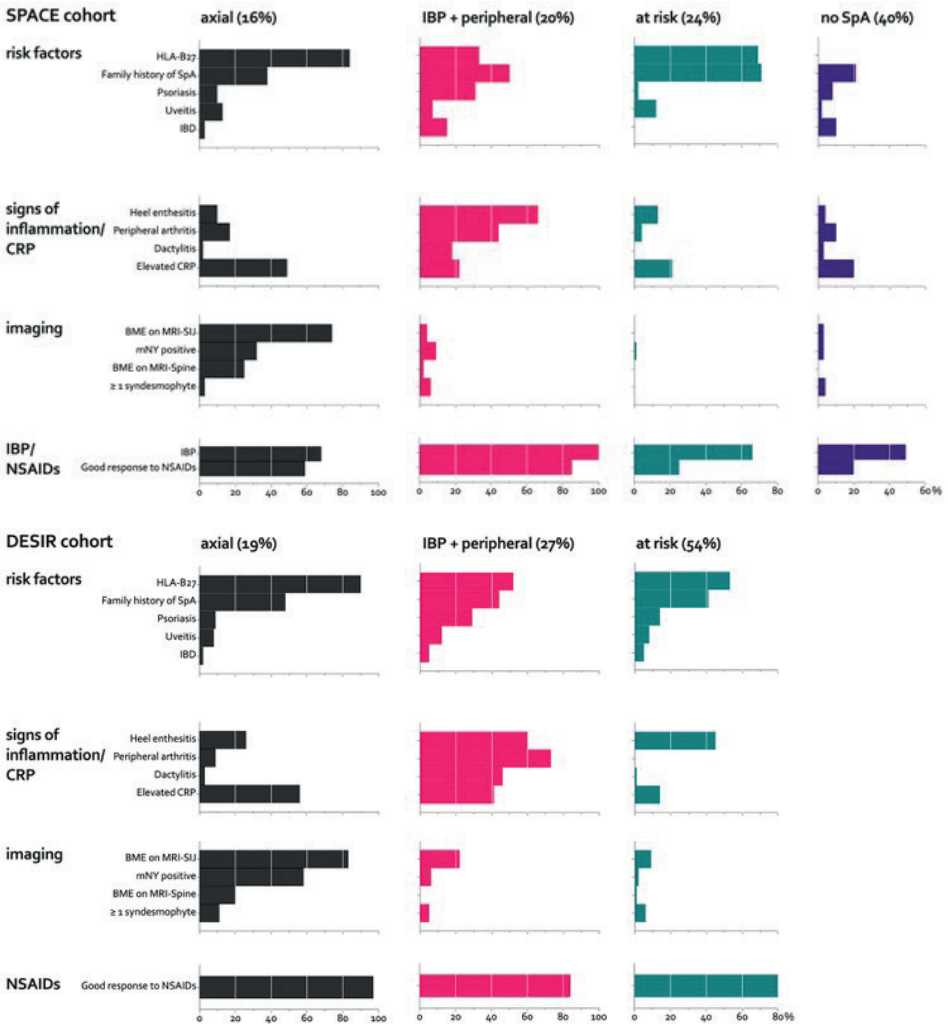


Figure 1. The Gestalt of axial SpA. Distribution of the probabilities of each feature according to the final LCA model in SPACE and DESIR. CRP, C-reactive protein; NSAIDs, nonsteroidal anti-inflammatory drugs; mNY, modified New York criteria; MRI, magnetic resonance imaging; SIJ, sacroiliac joints; BME, 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).

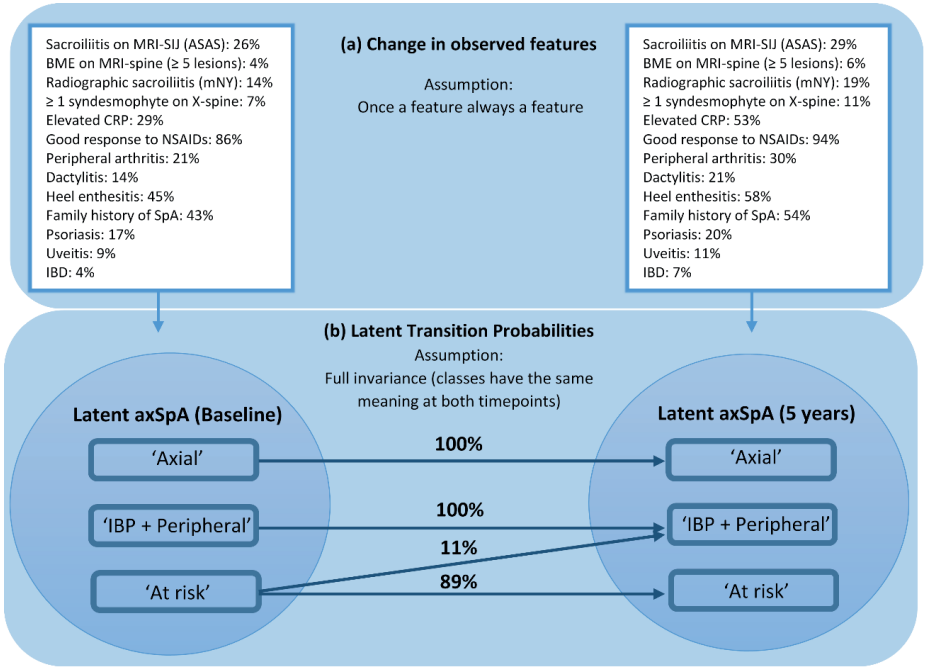


Figure 2. Final latent transition analysis (LTA) model (with full invariance*) in DESIR (N=576). (a) squares refer to observed (i.e. measurable) variables and (b) Circles to latent (i.e. unobserved variables). Arrows: latent transition analysis models the change in observed features (a) to estimate the 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.

Chapter 5

Five-year follow-up of radiographic sacroiliitis: progression as well as improvement?

Alexandre Sepriano, Martin Rudwaleit, Joachim Sieper, Rosaline van den Berg, Robert Landewé, Désirée van der Heijde.

Ann Rheum Dis. 2016 Jun;75(6):1262-3

Determining the presence of radiographic sacroiliitis is a key feature in the diagnostic process of radiographic axial spondyloarthritis (r-axSpA), synonymous to ankylosing spondylitis according to the modified New York criteria (mNY).[1] Its presence is considered prognostically relevant and paves the way for treatment with biological drugs.[2] Multiread and multireader exercises have proven that radiographic sacroiliitis is an ambiguous finding, as reflected by large inter-reader and intrareader variability.[3, 4]

Determining *progression* of radiographic sacroiliitis, which marks the arbitrary but irreversible change from non-radiographic axSpA (nr-axSpA) to r-axSpA, is even more ambiguous. The mNY lack sensitivity-to-change in this slowly progressing condition, and it is conceivable that *regression* of radiographic sacroiliitis is very rare if not impossible.[5] Previous studies addressing progression from nr-axSpA to r-axSpA have ignored regression and have only interpreted progression.[6] However, from a methodological perspective, bi-directional change cannot be ignored.

The aim of this study was therefore to assess positive and negative changes on plain pelvic radiographs (X-SI) over time in the Assessment of SpondyloArthritis international Society (ASAS)-cohort, in which X-SI judgements have been provided by single local readers from many centres worldwide.

In the ASAS cohort, 975 patients with either chronic back pain (>3 months, onset <45 years) of unknown origin or undiagnosed peripheral symptoms were assessed at baseline.[7, 8] Of these, 564 patients were reassessed after a mean follow-up of 4.4 years (range: 1.9-6.8). Patients with paired X-SI available (at baseline and follow-up) were included and judgements of the local observer (rheumatologist/radiologist) at both time points (either by the same or other reader) were analysed. Positive cases were defined as definite radiographic sacroiliitis according to the mNY.

In total, 357 patients had paired X-SI available. Of these, 17.4% (62/357) fulfilled the criteria for r-axSpA (table 1). At follow-up this proportion has raised to 22.4% (80/357) suggesting a net-progression of 5%. Cross-tabulation, however, revealed that more than half (36/62) considered mNY-positive at baseline were assessed mNY-negative at follow-up (table 2). If true, this would mean that radiographic sacroiliitis would have regressed in 58% of the cases. Conversely, only 54/295 patients (18.3%) became positive at follow-up.

It is very difficult to interpret these data, since progression, regression and measurement error (leading to spurious change) cannot be disentangled. Under the untenable assumption of 'no true change', the kappa statistic would yield a very poor figure of 0.21 (only marginally better than chance-agreement), which would make it useless from a diagnostic perspective.

If only positive change (progression) is valued and negative change is ignored, one would disregard measurement error and spuriously attribute part of the observed positive change to real progression.

The most likely explanation of our strange and extreme observation is that subtle radiographic progression (the signal) – if truly present – cannot be reliably distinguished from measurement error (the noise). These sobering data clearly illustrates that more research is needed in visualising progression in axSpA. Imaging modalities other than radiographs should be evaluated in future such as MRI and low-dose CT.

Table 1. Baseline characteristics of patients with baseline and follow-up pelvic radiographs

	Patients with paired radiographs (N=357)
Age, years (mean, SD)	33.8 (10.8)
Age at onset of back pain, years (mean, SD)	26.2 (8.8)
Male gender, n (%)	171 (47.9)
Number of SpA features* (mean, SD)	2.5 (1.4)
Definite radiographic sacroiliitis (mNY), n (%)	62 (17.4)
Active inflammation of SIJ¥, MRI, n (%) (n=223)	112 (50.2)
HLA-B27 positivity, n (%)	174 (48.7)
Elevated CRP, n (%)	135 (37.8)
IBP (According to experts definition), n (%)	178 (49.9)
Peripheral arthritis past or present, n (%)	193 (54.1)
Heel enthesitis past or present, n (%)	79 (22.1)
Uveitis past or present, n (%)	32 (9.0)
Dactylitis past or present, n (%)	39 (10.9)
Psoriasis past or present, n (%)	27 (7.6)
IBD past or present, n (%)	14 (3.9)
Good response to NSAIDs, n (%)	126 (35.4)
Family history of SpA, n (%)	79 (22.1)
Preceding infection, n (%)	11 (3.1)
Schober's test (cm), mean (SD) (n=354)	4.4 (2.5)
Chest expansion (cm), mean (SD) (n=351)	5.6 (5.7)
Active inflammation of the spine¥, MRI, n (%) (n=110)	29 (26.4)

* Features included: Inflammatory back pain (IBP), arthritis, heel enthesitis, dactylitis, uveitis, psoriasis, inflammatory bowel disease (IBD), good response to NSAIDs, family history of spondyloarthritis, elevated CRP. ¥ Presence or absence of typical signs of active inflammation independent of formal criteria. SIJ, sacroiliac joints; SpA, spondyloarthritis; mNY, modified New York criteria; MRI, magnetic resonance imaging.

Table 2. Radiographic sacroiliitis according to the modified New York criteria at baseline and at follow-up (on average 4.4 years)

	Follow-up radiograph		
Baseline radiograph	Positive	Negative	**Total**
Positive	26	36	**62**
Negative	54	241	**295**
Total	**80**	**277**	**357**
PPV (%)	41.9		
NPV (%)	81.7		

PPV, positive predictive value; NPV, negative predictive value.

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. van den Berg R, Stanislawska-Biernat E, van der Heijde DM. Comparison of recommendations for the use of anti-tumour necrosis factor therapy in ankylosing spondylitis in 23 countries worldwide. Rheumatology (Oxford). 2011 Dec; 50(12):2270-2277.
3. 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.
4. 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.
5. Sieper J, van der Heijde D. Review: Nonradiographic axial spondyloarthritis: new definition of an old disease? Arthritis Rheum. 2013 Mar; 65(3):543-551.
6. Poddubnyy D, Sieper J. Radiographic progression in ankylosing spondylitis/axial spondyloarthritis: how fast and how clinically meaningful? Curr Opin Rheumatol. 2012 Jul; 24(4):363-369.
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.

Chapter 6

Percentage of progressors in imaging: can we ignore regressors?

Alexandre Sepriano, Sofia Ramiro, Robert Landewé, Maxime Dougados, Désirée van der Heijde

RMD Open. 2019 Feb 26;5(1):e000848

ABSTRACT

Stopping or preventing structural progression is a goal common to all Inflammatory rheumatic diseases. Imaging may capture structural progression across diseases but is susceptible to measurement error. Progression can be analysed as a continuous change score over time (e.g. mean change of the van der Heijde-modified Sharp score), or as a binary change score (e.g. percentage of progressors according to the modified New York criteria). Here, we argue that the former takes measurement error into account while the latter ignores it, which may lead to spurious conclusions. We will argue that assumptions underlying commonly used binary definitions of progression are false and we propose a method that incorporates (inevitable) measurement error.

VIEWPOINT

Inflammatory rheumatic musculoskeletal diseases (RMDs), such as rheumatoid arthritis (RA) and spondyloarthritis (SpA), typically cause irreversible joint damage over time, particularly if left untreated. Recent landmark therapeutic advancements suggest modifying the destructive course of a disease is possible, but still much needs to be done in this regard.[1, 2] In order to capture treatment effects in joint damage progression, valid outcome measures are warranted, as prescribed by regulatory agencies worldwide.[3-6]

Conventional radiography is the standard modality for capturing and quantifying progression of structural damage in RMDs. Although we focus on conventional radiography as an example, the issues we address here apply similarly to all imaging modalities assessing structural damage. Equally important as the imaging modality itself is the analytical method used to quantify progression. For example, radiographic progression can be analysed as an averaged continuous change score (e.g. mean change of the van der Heijde-modified Sharp score (SvdH) over time; or the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) over time), or as a binary change score (e.g. percentage of 'progressors' according to the modified New York criteria (mNY)). Another way of presenting a binary change score is dichotomising a continuous change score (e.g. SvdH ≥ 5 vs <5; or mSASSS ≥ 2 vs <2). The quantification of radiographic progression, like outcome assessments in general and other imaging methods more specifically, is susceptible to measurement error. Here we will demonstrate that researchers using continuous change scores will implicitly take measurement error into account, while researchers using binary change scores will frequently omit measurement error.

We make a plea that measurement error (or: noise) should not be ignored when interpreting imaging studies. The 'signal-to-noise' ratio analogy has been recently proposed to better explain the fallacies of ignoring measurement error.[7] Here, this analogy will be used to argue the false assumptions underlying commonly used binary definitions of progression. The 'signal-to-noise' concept incorporates two types of information: (1) 'true change' ('signal'); and (2) error change ('noise'). The larger the measurement error the harder to capture the 'signal' and in some cases, disentangling the 'signal' from the 'noise' can be particularly challenging. Sources of 'noise' in reading radiographs are plenty and widely recognised (e.g. technical, intra- and inter-reader variability). To improve the 'signal-to-noise' ratio (the higher the better), investigators have been implementing strategies to reduce the denominator (i.e. 'noise') by, for instance, combining judgements from ≥ 2 trained central readers. Nevertheless, these (methodological) strategies cannot fully eliminate the undesired 'noise'. Thus, here we discuss how appropriate analytical

choices can further contribute to handle 'noise' in imaging assessment, ultimately contributing to its reduction.

We have used data from a recently published study from the DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR) cohort[8] to better illustrate the concept of 'signal-to-noise' ratio with a particularly challenging case. In our example, damage occurring in the sacroiliac joints (SIJs) over 5 years was evaluated in patients with axial SpA (axSpA), according to the mNY scoring system[9]. This scoring system has clearly been shown to be unreliable (much 'noise'), especially if scores from only one (untrained) reader are used.[10-12] We reduced the 'noise' by having baseline and 5-year films per patient scored by three trained central readers obtained independently, and used blinded chronological order to ensure unbiased measurement error in two directions (i.e. the readers did not know which is the baseline and which is the 5-year film when scoring the pair). Each reader reported a binary score (mNY-positive vs mNY-negative) and a (semi) continuous grade (range: 0-8; both SIJ together) per time-point. The final mNY binary status score was defined by the agreement of at least two of the three readers, and the continuous grade by the average of the three independent scores. The binary change scores can take 3 possible values (-1, 0, +1). For instance, if a patient is mNY-positive at baseline and negative at 5-years the binary change score is -1 (negative change or 'improvement'). Similarly, the continuous change score can also be positive and negative (range: -8 to +8), where a negative value means the mNY grade at 5-years is smaller than the grade at baseline. The resulting change scores are shown here in a way that makes measurement error better visible: (1) for the binary change score we show the crosstabulation between baseline and 5-year combined scores (table 1), and report positive change (i.e. worsenings; +1) and negative change (i.e. 'improvements'; -1); and (2) for the continuous change score we report a cumulative probability plot (figure 1), that (by default) also shows positive and negative change and, additionally, we overlay the binary changes in the plot to facilitate comparison. These data are used here as the 'common ground' from which we explore the assumptions of commonly used binary definitions of progression, and finally to propose an assumption-free approach. This is all under the assumption that structural damage is irreversible (which might not necessarily apply in all settings) and therefore improvements should be judged as measurement error.

Table 1. Change in the mNY status in patients with axSpA after 5 years in the DESIR cohort[8]

5-years / Baseline	mNY Positive	mNY Negative	**Total**
mNY Positive	59	3	**62**
mNY Negative	24	330	**354**
Total	**83**	**333**	**416**

mNY, radiographic sacroiliitis according to the modified New York criteria (agreement between ≥ 2 out of 3 trained central readers blinded to time-order); axSpA, axial spondyloarthritis.

Crude progression

At baseline, 62 (15%) of the 416 patients were classified as mNY-positive. Of the 354 mNY-negative patients at baseline, 24 changed into mNY-positive after 5 years (positive change or worsening; +1). Most studies would have only reported these 24 cases (6.8% (24/354)) as those who had progressed from mNY-negative to mNY-positive.[13-16]. But this rate is spuriously high for two reasons: First, it implies that the baseline reading is true and free of measurement error

(bias); second, it assumes that a change in the unexpected opposite direction (negative change or 'improvement'; -1) can be ignored. Since radiographic readings are not free of measurement error and readers are not aware of which film pertains to baseline and follow-up, such an approach does not provide a valid representation of the truth. Also when analyzing the data, one must consider the different possible scenarios, in this case meaning that 'improvement' or negative change, though less expected or warranted, can also happen, particularly due to measurement error. This method to measure progression does not accommodate this reality.

Conditional net progression

Recently, researchers from the DESIR and the German Spondyloarthritis Inception Cohort reported progression of radiographic sacroiliitis at 2 years.[17, 18] They acknowledged that a robust estimation of progression must not ignore the measured negative changes. Table 1 shows how this principle worked out: Positive changes ('worsening' in 24 of the 354 formerly mNY-negative patients (6.8%); '+1 change') and negative changes ('improvement' in 3 of the 62 formerly mNY-positive patients (4.8%); '-1 change') were seen, and 'net progression' was obtained by calculating the difference between both rates (2%). While this approach differs from the 'crude method' by acknowledging the relevance of negative changes, the 'net progression' rate of 2% is still conditional on the baseline classification status assumed to be free of bias. In other words, it implicitly assumes that 'worsening' can only happen in patients who are mNY-negative at baseline and 'improvement' only in mNY-positive patients. Since readers are not aware which film is the baseline film (scores had been obtained in pairs with full blinding of time order) this assumption does not hold.

Assumption-free net progression

We therefore propose an assumption-free method to analyse structural damage progression.[8] In principle, both 'positive changes' ('+1 change') and 'negative changes' ('-1 change') are 'allowed' and scores of individual patients are not interpreted as 'true progression' or 'noise'. Under the premise of reading with concealed (blinded) time order, measurement error ('noise') presumably occurs symmetrically. This means: it will affect scores with similar likelihood in both directions since readers are not aware of which image pertains to baseline and which to follow-up, as has been worked out by us previously for progression in RA.[19] So, with the 'assumption-free' method, the overall improvement contains (in theory) both 'true improvement' (i.e. repair) as well as measurement error. Similarly, worsening also includes 'true worsening' (i.e. progression) and measurement error. However, in a setting of irreversible damage, it is not unreasonable to expect that measurement error (rather than repair) largely dominates improvements. Still the direction and magnitude of residual bias (driven by bidirectional measurement error) is difficult to know with certainty for binary outcomes. Notwithstanding with the proposed method measurement error at least is incorporated and not ignored as done thus far.

With the 'assumption-free' method, if 'true progression' is present over-and-above measurement error, it will become obvious as a positive change when all zero changes, positive changes and negative changes occurring in the *entire* population are summed together. The area under the curve (AUC) of the probability plots (positive area minus negative area) provides the mean continuous change score taking measurement error into account since it incorporates, by

default, both positive (>0, i.e. corresponding to '+1 change') and negative (<0, i.e. corresponding to '-1 change') changes (figure 1). In our example the overall mean change-score (+0.20 [SD: 0.55]) can be obtained by the subtraction of the mean status score at baseline (1.40 [SD: 1.68]) from the mean status score at 5 years (1.60 [SD: 1.83]). Another way of getting the average continuous change score, is by summing all positive change-scores (+106.67 N=136 within the positive AUC), all negative changes scores (-24.67; N=53 within the negative AUC) and all no-changes (0; N=227) and divide the result by the total number of patients [(106.67 + (-24.67) + 0)/416= +0.20]. Thus, on average, the continuous change score is positive (+0.20) since positive change scores outweigh the negative change scores but, importantly, both are included in the calculation. The binary 'assumption-free' net progression is analytically similar, also capturing measurement error appropriately. However, measurement error is neglected by the first two definitions of binary change. If positive binary changes are scored +1, negative changes are scored -1, and no changes are scored zero, the total change is the sum of all +1 scores, -1 scores and zeros scores, divided by the total number of observations, and expressed as a percentage [(24 + (-3) + 0)/416 =5%]. Similar to the average continuous change score above (+0.20) an overall positive percentage implies that, at the group level, there is more progression than measurement error. By doing so we get an 'assumption free' net progression of +5% and not of +2% (as the conditional net progression).

Of note, the estimated progression is an averaged estimate which aims to approximate 'true progression' at the group level (i.e. beyond measurement error) but does not translate to individual patients. So, it becomes impossible to declare a patient as a 'progressor', as is often done in the context of clinical trials. Similarly, we estimate 5% progression from mNY-negative to mNY-positive after 5 years in the population of DESIR patients, and not 21 progressors out of 416.

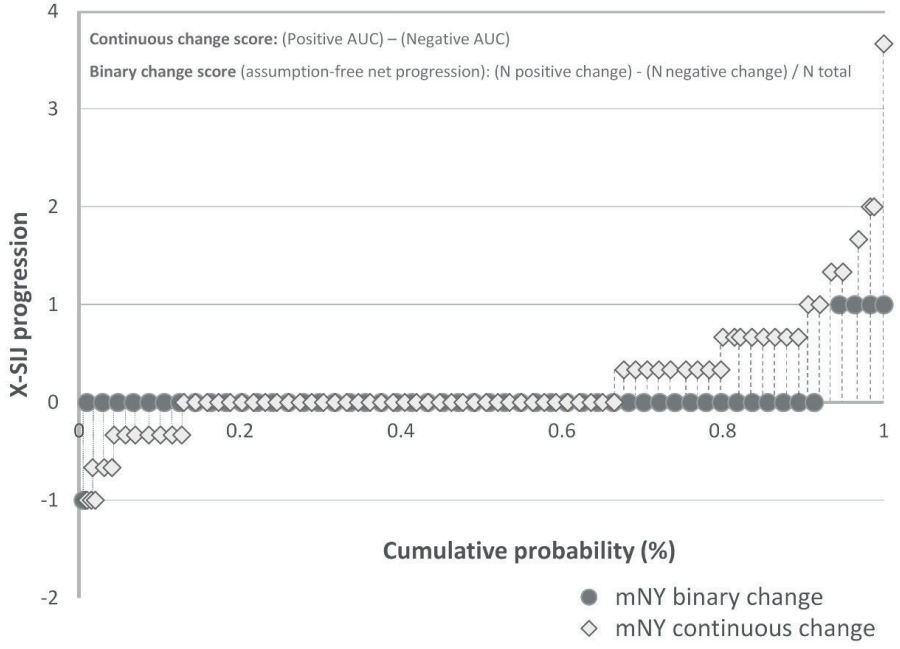


Figure 1. Cumulative probability plot. Structural progression in radiographs of the sacroiliac joints (X-SIJ) according to the modified New York criteria (mNY), measured as binary change (possible values: +1, 0, -1) and continuous grade change (range of possible values: -8; 8). Each datapoint represents either binary (black circles) or continuous (open diamonds) progression from one unique patient (selection from the total sample to increase readability but covering the full range of observed values). Positive AUC: Dashed lines; Negative AUC: dotted lines. AUC, area under the curve; N, number of patients.

Proposed method for future research

In summary, three methods to approximate binary progression to 'true progression' that are in use have been discussed here: (1) 'crude progression'; (2) 'conditional net progression'; and (3) 'assumption-free net progression'. This ´assumption-free net progression' yields the least biased estimates since it gives most credit to measurement error (i.e. always includes error without a prior assumption on the imaging modality ability to reliably capture change or on the baseline status score). Obviously, decreasing bias carries many benefits such as the better detection of treatment effects in randomised trials. Thus, we propose that this method will be applied in future studies with binary imaging outcomes. Importantly, this method applies to both continuous outcome measures that are dichotomised (e.g. SvdH ≥ 5 vs <5; or mSASSS ≥ 2 vs <2) as well to dichotomous measures by nature (e.g. mNY-positive vs mNY-negative),[9, 20] but should be used with caution since it implies that outcomes are irreversible (mainly structural damage), and are evaluated over not too long periods, as 'true repair' cannot be excluded with longer follow-up. A better understanding of what structural repair means (and importantly how to define it) is still a major unmet need in the field of rheumatology. Further studies are necessary to better understand 'negative changes' in settings other than irreversible damage and how 'true improvements' (i.e. repair) possibly contribute to the overall net progression. However, since the proposed 'assumption-free' method, different to what has been done so far, implies full disclosure of the bidirectional change (e.g. as a 2:2 table used in this viewpoint), together with the overall figure of 'net progression', it can facilitate research pursuing a consensual definition of 'repair' by acknowledging and, importantly, making 'negative change' more visible. This includes subtle distinctions between, for instance, spontaneous repair and repair driven by interventions which might reflect different pathophysiological pathways. Understanding these differences will allow a better interpretation of the treatment effects of drugs targeting specific pathways and how the 'assumption free' method captures these effects.

While we have used the example of radiographs in axSpA, the application of assumption-free net progression extends to all examples in rheumatology where imaging scores on structural damage are obtained under blinded conditions, and likely goes beyond. The example of axSpA should here be merely seen as an example of a methodological issue that we would welcome researchers to incorporate in their analysis of radiographic progression, independently of the disease being investigated. Too often we think that measurement error is not a big issue, while it is really there but often only not quantified.

References

1. Smolen JS, Landewé R, Bijlsma J, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. Ann Rheum Dis. 2017 Jun; 76(6):960-977.
2. van der Heijde D, Ramiro S, Landewé R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis. 2017 Jun; 76(6):978-991.
3. European Medicines Agency. Concept paper on clinical investigation of medicinal products for the treatment of axial spondyloarthritis. http://www.ema.europa.eu/docs/en_GB/docum

ent_library/Scientific_guideline/2015/04/WC500185187.pdf.

4. Guideline on clinical investigation of medicinal products for the treatment of rheumatoid arthritis. http://http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2018/01/WC500241042.pdf.
5. Clinical Trial Imaging Endpoint Process Standards Guidance for Industry. https://www.fda.gov/downloads/drugs/guidances/ucm268555.pdf.
6. Boers M, Brooks P, Strand CV, et al. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol. 1998 Feb; 25(2):198-199.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. Schattenkirchner M, Kruger K. Natural course and prognosis of HLA-B27-positive oligoarthritis. Clin Rheumatol. 1987 Sep; 6 Suppl 2:83-86.
14. 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.
15. 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.
16. Sampaio-Barros PD, Bortoluzzo AB, Conde RA, et al. Undifferentiated spondyloarthritis: a longterm followup. J Rheumatol. 2010 Jun; 37(6):1195-1199.
17. 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.
18. 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.
19. Landewé R, van der Heijde D. Radiographic progression depicted by probability plots: presenting data with optimal use of individual values. Arthritis Rheum. 2004 Mar; 50(3):699-706.
20. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol. 2000 Jan; 27(1):261-263.

Chapter 7

Sacroiliac radiographic progression in recent onset axial spondyloarthritis: the 5-year data of the DESIR cohort

Maxime Dougados, Alexandre Sepriano, Anna Moltó, Sofia Ramiro, Manouk de Hooge, Rosaline van den Berg, Victoria Navarro-Compán, Christophe Demattei, Robert Landewé, Désirée van der Heijde

Ann Rheum Dis. 2017 Nov;76(11):1823-1828

ABSTRACT

Objective: To estimate sacroiliac joint radiographic (X-SIJ) progression in patients with axial spondyloarthritis (axSpA) and to evaluate the effects of inflammation on MRI (MRI-SIJ) on X-SIJ progression.

Methods: X-SIJ and MRI-SIJ at baseline and after 2 and 5 years in patients with recent onset axSpA from the DESIR-cohort were scored by three central readers. Progression was defined as (1) the shift from non-radiographic (nr) to radiographic (r) sacroiliitis (by modified New York (mNY) criteria) or alternative criteria (2) a change of at least one grade or (3) a change of at least one grade but ignoring a change from grade 0 to 1. The effects of baseline inflammation on MRI-SIJ on 5-year X-SIJ damage (mNY) were tested by generalised estimating equations.

Results: In 416 patients with pairs of baseline and 5-year X-SIJ present, net progression occurred in 5.1% (1), 13.0% (2) and 10.3% (3) respectively, regarding a shift from nr- to r-axSpA (1), a change of at least one grade (2) or a change of at least one grade but ignoring a change from grade 0 to 1 (3). Baseline MRI-SIJ predicted structural damage after 5 years in human leukocyte antigen-B27 (HLA-B27) positive (OR 5.39 (95% CI: 3.25 – 8.94)) and in HLA-B27 negative (OR 2.16 (95% CI: 1.04 – 4.51)) patients.

Conclusions: Five-year progression of X-SIJ damage in patients with recent onset axSpA is limited but present beyond measurement error. Baseline MRI-SIJ inflammation drives 5-year radiographic changes.

INTRODUCTION

Axial spondyloarthritis (axSpA) comprises two subcategories based on the presence of structural changes in the sacroiliac joints (SIJs): radiographic (r)-axSpA and non-radiographic (nr)-axSpA). R-axSpA implies the fulfilment of the modified New York criteria (mNY).[1, 2]

Information about the natural course of radiographic sacroiliitis and factors that contribute to it is scarce.[3] Prospective cohorts should give resolution and long-term follow-up of patients with recent-onset disease is mandatory to 'capture' meaningful progression. Inherently, such studies face the risk of loss to-follow up and attrition bias.

DESIR (acronym in French for outcome of recent onset spondyloarthritis) is a prospective cohort of patients with recent onset axial spondyloarthritis (NCT01648907). With this study we address the primary objectives of DESIR, formulated as follows: (1) what proportion of patients switches from nr- to r-axSpA after 5 years?; (2) how sensitive are different outcome measures for radiographic damage of SIJ (X-SIJ) to change?; (3) does inflammation on magnetic resonance imaging of the SIJ (MRI-SIJ) lead to structural damage on X-SIJ after 5 years?

METHODS

Patients

The DESIR cohort has been previously described.[4] Briefly, consecutive patients (aged 18-50 from 25 centers in France) with inflammatory back pain[5, 6] and a duration ≥3 months but <3 years were included if the treating rheumatologist considered the symptoms suggestive of axSpA (a score ≥5 on a scale from 0 to 10, in which 0 was 'not suggestive' and 10 'very suggestive'). Between December 2007 and April 2010, 708 patients were included.

The study was conducted according to good clinical practice guidelines and was approved by the appropriate local medical ethical committees. A detailed description of the study protocol is available at the DESIR website (http://www.lacohortedesir.fr/desir-in-english/). The research proposal for this particular analysis was approved by the scientific committee of the DESIR cohort.

Clinical data

By using a standardized case report form (CRF) information was collected with questionnaires, physical examination, on-going treatments and laboratory tests according to the DESIR protocol. The database used for this analysis was locked in June 2016.

At baseline, age, gender, smoking status, HLA-B27 and duration of axial symptoms had been collected. At baseline, every 6 months during the first 2 years of follow-up, and annually thereafter the following parameters had been collected: Bath Ankylosing Spondylitis Disease Activity Index (BASDAI),[7]Bath Ankylosing Spondylitis Functional Index,[8] C-reactive protein (CRP), treatment including non-steroidal anti-inflammatory drugs (NSAID) by the Assessment of Spondyloarthritis International Society (ASAS)-NSAID score and tumour necrosis factor inhibitors (TNFi).[9]

Pelvic radiographs

Pelvic-radiographs collected at baseline, 2 years and 5 years of follow-up were evaluated in one session independently by three central readers (MdH, VNC and RvdB). Readers were blinded for time order and clinical information. Each reader evaluated each SIJ according to the mNY-grading-method (0: normal; 1: suspicious changes; 2: minimal abnormalities; 3: unequivocal abnormalities and 4: severe abnormalities (complete ankylosis)).[10]

Pelvic MRI

MRI-SIJ collected at baseline, 2 years and 5 years of follow-up were evaluated in one session independently by three central readers (MdH, VNC and MvL). Readers were blinded for time order and clinical information. MRI-SIJ was considered positive if bone marrow edema (BME) lesions highly suggestive of SpA were present (either one BME-lesion on ≥2 consecutive slices or several BME lesions on one slice).[11] An MRI-SIJ was considered positive if at least two out of three readers judged positivity. MRI-SIJ and X-SIJ were scored entirely independently.

Sample size calculation

The sample size calculation was based on an estimated prevalence of radiographic damage between 70% and 90% at year 5 irrespective of the baseline status. Moreover, we estimated the prevalence of inflammation on MRI-SIJ at baseline between 30% and 50%.[12, 13]

The number of patients was calculated based on a relative risk of 2-3 to observe radiographic damage at year 5 in case of a baseline MRI-SIJ inflammation. For a 5% bilateral alpha risk, a 90% power, and the different assumptions including an attrition rate between 15% and 20%, the number of required patients ranged from 685 to 768, and 700 was the chosen number.

Statistical analysis

SIJ radiographic progression

The 5-year X-SIJ progression was assessed in patients in whom baseline and year-5 X-SIJ were present (completers' population). Assessed were: (A) switch from nr-axSpA at baseline to r-axSpA (mNY score) at 5 years; (B) worsening of at least one grade in at least one SIJ; (C) worsening of at least one grade in at least one SIJ, but with a 5-year grade of at least 2 in the worsened joint; and (D) change in the total mNY score (expressed as a continuous variable) with a range from 0 to 8 (4 grades per SIJ).

In order to give sufficient credit to measurement error, we determined the proportion of 'progressors' (% of patients with worsening) as well as the proportion of 'regressors' (% of patients with improvement). Improvement was defined per outcome measure: (A) switching from r-axSpA at baseline to nr-axSpA at 5 years; (B) reduction of at least one grade in at least one SIJ; and (C) reduction of at least one grade in at least one SIJ with a baseline score of at least 2 in the improved joint. In addition, 'net' percentage of progression was defined as the number

of ‘progressors’ minus the number of ‘regressors’ (numerator) divided by the total number of the study population (denominator), and was analysed in the entire population and clinically relevant subgroups.

Sensitivity analyses that addressed the impact of missing data were performed in patients with a baseline and at least one post-baseline radiograph available (‘intention-to-follow’ population) using two imputation techniques: (1) last observation carried forward (LOCF) and (2) linear extrapolation (LE).

The continuous SIJ score (total scores of left plus right SIJ (ranging from 0 to 8) was the mean score of the 3 readers; for the binary definitions a change was considered present if at least two out of the three readers agreed.

Effect of baseline MRI-SIJ inflammation on the 5-year X-SIJ damage

The association between baseline MRI-SIJ inflammation and 5-year X-SIJ damage (primary outcome) was analysed by three different models: 1: binomial multivariable generalised estimating equations (GEEs) on the individual readers’ scores (1-level GEE model); 2: ‘traditional’ multivariable logistic regression on the aggregated (two out of three reader consensus scores for MRI and SIJ) X-SIJ progression scores; 3: a true longitudinal (2-level) multivariable GEE with time-lagged autoregressive variables (as in Ramiro *et al*).[14] The logistic regression models were also fit after multiple imputations with chained equations (MICE) in the ‘intention-to-follow’ population.

Potential baseline-confounders for the association of interest were selected based on their clinical relevance (gender, symptom duration, CRP, BASDAI, smoking status and treatment with NSAIDs). Statistical interactions between MRI-SIJ inflammation and baseline variables were excluded first, and if relevant ($p<0.15$ for the interaction term) the model was fitted per stratum.

RESULTS

Patients and study course

Pelvic radiographs were available for 685 of the 708 patients at baseline. Of the 685 patients with baseline X-SIJ, 519 and 416 patients had X-SIJ, from all readers, after 2 and 5 years respectively (completer’s population). A postbaseline X-SIJ (either at year 2 or 5) was available for 557 patients (intention to follow population). A baseline MRI-SIJ was available for 679 patients.

Table 1 summarises the baseline characteristics for patients with complete 5-year pelvic radiograph data and those without.

Table 1. Baseline characteristics according to the availability of complete 5-year radiographic data of the sacroiliac joints

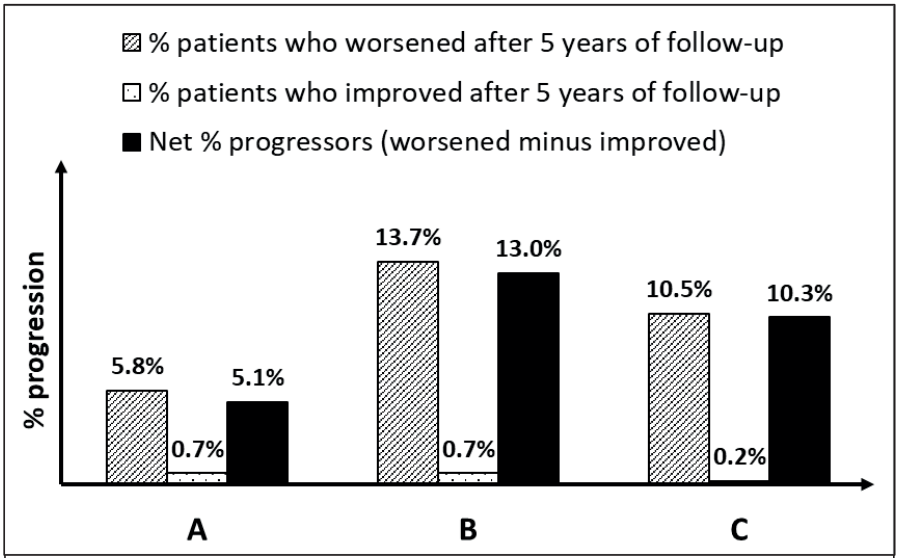
Characteristics	Status at year 5		All patients
	Completers¥	Non completers	
Number of patients	417	291	**708**
Age (mean, SD)	34.1 (8.6)	33.2 (8.6)	33.7 (8.6)
Symptom duration (years), (mean, SD)	1.5 (0.9) (n=416)	1.5 (0.8) (n=291)	1.5 (0.9) (n=707)
Male gender (%)	198 (47.5)	129 (44.3)	327(46.2)
HLA-B27 positivity (%)	267 (64.0) (n=417)	143 (49.3) (n=290)	410 (58.0) (n=707)
X-SIJ structural damage* (mNY) (%)	62 (14.9) (n=416)	29 (10.8) (n=268)	92 (13.5) (n=684)
MRI-SIJ-inflammation*‡ (%)	113 (28.1) (n=402)	67 (24.2) (n=277)	180 (26.5) (n=679)
Abnormal CRP† (%)	126 (31.5) (n=400)	78 (27.4) (n=285)	204 (29.8) (n=685)
BASDAI (0-10, mean, SD)	4.34 (1.99) (n=416)	4.65 (2.01) (n=288)	4.47 (2.00) (n=704)
ASDAS (mean, SD)	2.6 (1.0) (n=395)	2.6 (0.9) (n=281)	2.6 (1.0) (n=676)
BASFI (0-10, mean, SD)	2.92 (2.24) (n=413)	3.23 (2.32) (n=288)	3.04 (2.28) (n=701)

*According to the '2 out of 3' definition: agreement of at least 2 out of the 3 readers – if 2 readers disagree and the third reading is missing the combined score is set as missing (1 case for X-SIJ); ‡Presence of bone marrow edema according to the ASAS criteria at MRI-SIJ; †≥6 mg/L; X-SIJ: radiograph of the sacroiliac joints; ¥ patients with both baseline and 5-year X-SIJ available; mNY: modified New York criteria; MRI-SIJ, magnetic resonance imaging of the sacroiliac joints; CRP: c reactive protein; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; ASDAS: Ankylosing Spondylitis Disease Activity Score; BASFI: Bath Ankylosing Spondylitis Functional Index.

Radiographic progression after 5 years of follow-up

At baseline, the mNY criteria were fulfilled by 62/416 (14.9%; according to two out of three readers) of the patients in the completers' population. After 5 years, this proportion has increased to 20.0% in the completers' population and to 18.0% and 17.7% in the 'intention-to-follow' population (n=557), after LOCF and LE, respectively. A statistically significant worsening of the mean (SD) SIJ score was found in all scenarios (from 1.41 (1.68) to 1.60 (1.83) (Δ:0.19 (0.55); p<0.001) in the completers' population and from 1.32 (1.65) to 1.49 (1.81) (Δ:0.17 (0.59);p<0.001)(LOCF) or from 1.33 (1.65) to 1.50 (1.84) (Δ:0.17 (0.61);p<0.001)(LE) in the 'intention to follow' population).

Figure 1 summarises the observed changes in the binary outcome measures in the completers' population, in terms of '% worsened', '% improved'; and 'net % progression' (online supplementary figures S1 and S2 provide the same information for the 'intention-to-follow' population after LOCF and LE, yielding similar results).



A = Switch from nr to r-axSpA according to the mNY criteria (worsened) *minus* switch from r to nr-axSpA (N=416)
B = Change in at least one grade in at least one SIJ (N=408)
C = Change in at least one grade in at least one SIJ and a final (at year 5) absolute value of at least 2 in the worsened joint (worsened) *minus* change in at least one grade in at least 1 SIJ and a baseline (year 0) absolute value of at least 2 in the improved joint (N=408)

Figure 1. Changes in different binary SIJ-Plain X-ray outcome measures (completers' population). nr-axSpA, radiographic axial spondyloarthritis; r-axSpA, radiographic axial spondyloarthritis; SIJ, sacroiliac joint.

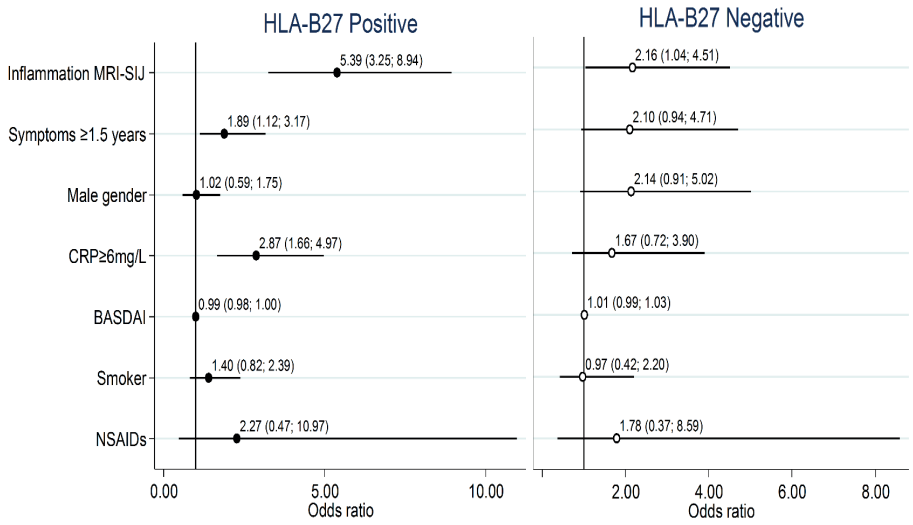


Figure 2. Effect of inflammation on MRI-SIJ on being mNY-positive after 5-years irrespective of baseline mNY status stratified according to the HLA-B27 status at baseline (1-level binomial multivariable GEE). Interaction between inflammation on MRI-SIJ and HLA-B27 at baseline: p=0.033. MRI-SIJ, magnetic resonance imaging of the sacroiliac joints; CRP: c reactive protein.

Effects of MRI-SIJ inflammation on X-SI damage

Figure 2 shows the effect of baseline MRI-SIJ inflammation on 5-year SIJ-damage according to the mNY criteria, stratified for HLA-B27 (interaction: p=0.033). Baseline MRI-SIJ inflammation was associated with radiographic damage after 5 years in HLA-B27 positive patients (OR 5.39 (95% CI: 3.25–8.94)) as well as HLA-B27 negative patients (OR 2.16 (95% CI: 1.04–4.51)). The association between baseline MRI-inflammation and 5-year SIJ-damage was consistently found, regardless of the analytical method and the definition of SIJ-progression (table 2).

Radiographic progression across clinically relevant subgroups

Figure 3 shows the 'net' progression from nr-axSpA to r-axSpA in different subgroups of patients according to relevant clinical characteristics and the interaction with HLA-B27.

HLA-B27-positive nr-axSpA-patients with a positive MRI-SIJ and CRP had a likelihood of 'net' progression of at least one grade of the X-SIJ mNY score that was more than twice as high as r-axSpA patients with similar baseline features (see online supplementary figures S3 and S4).

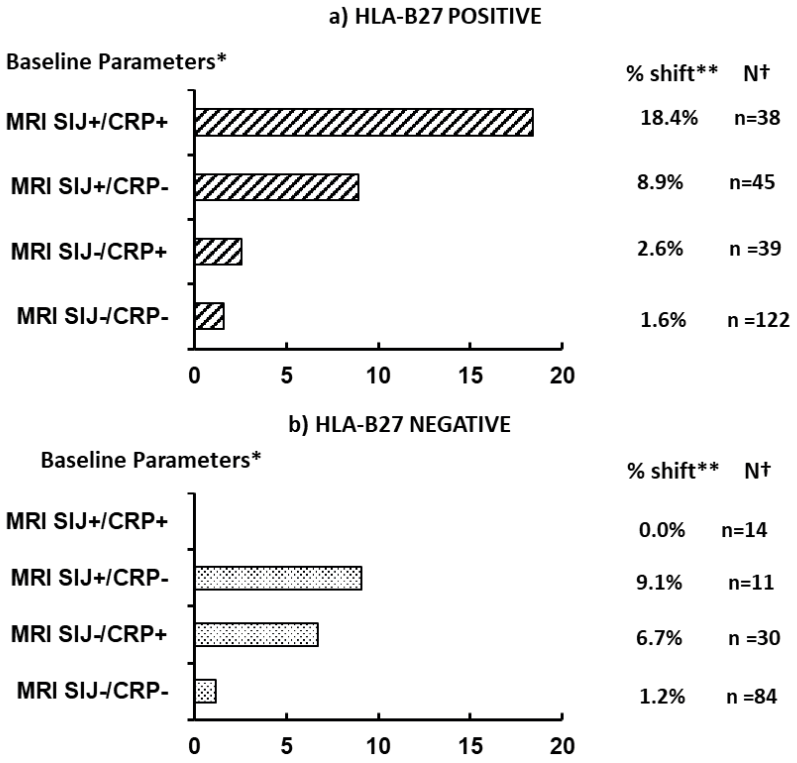


Figure 3. Net progression from nr-axSpA to r-axSpA according to baseline objective inflammatory markers and stratified on HLA-B27 status. BMO, bone marrow oedema; CRP, C reactive protein; MRISIJ, MRI of the sacroiliac joints; nr-axSpA, non-radiographic axial spondyloarthritis; r-axSpA, radiographic axial spondyloarthritis.

Table 2. Sensitivity analyses: effect of baseline MRI-SIJ inflammation on the different SIJ radiographic progression definitions, irrespective of baseline mNY status and using different analytical approaches

	Main effect aOR (95% CI)	**HLA-B27 positive** aOR (95% CI)	**HLA-B27-negative** aOR (95% CI)	**p-value interaction**
Outcome: mNY-positive				
Logistic regression*	NA	**9.26 (4.32; 19.86)** (N=247)	**3.79 (1.01; 14.28)** (N=143)	**0.106**
Logistic regression after MI†	**6.64 (3.67; 12.00)** (N=557)	NA	NA	NS
1-level GEE‡	NA	**5.39 (3.25; 8.94)** (N=248)	**2.16 (1.04; 4.51)** (N=143)	**0.033**
2-level GEE (longitudinal)¥	**2.42 (1.01; 5.78)** (N=493)	NA	NA	NS
Outcome: 1-grade progression				
Logistic regression*	**2.33 (1.21; 4.49)** (N=373)	NA	NA	NS
Logistic regression after MI†	**2.35 (1.13; 4.86)** (N=557)	NA	NA	NS
1-level GEE‡	**1.74 (1.05; 2.88)** (N=381)	NA	NA	NS
2-level GEE (longitudinal)¥	**1.90 (1.16; 3.13)** (N=486)	NA	NA	NS
Outcome: 1-grade progression + follow-up grade ≥2				
Logistic regression*	**3.45 (1.65; 7.23)** (N=373)	NA	NA	NS
Logistic regression after MI†	**3.47 (1.60; 7.54)** (N=557)	NA	NA	NS
1-level GEE‡	**1.82 (1.02; 3.27)** (N=381)	NA	NA	NS
2-level GEE (longitudinal)¥	**1.87 (1.04; 3.36)** (N=486)	NA	NA	NS

*Association between baseline MRI-SIJ inflammation and the X-SIJ score at year 5 with both variables according to the '2 out of 3' definition; N=patients with X-SIJ score available at year 5 and complete data on all covariates at baseline.

†Association between baseline MRI-SIJ inflammation and the X-SIJ score at year 5 both variables according to the '2 out of 3' definition, after multiple imputation; N= patients with X-SIJ available at baseline and in at least one postbaseline visit and complete data on all covariates at baseline.

‡Association between baseline MRI-SIJ inflammation and the X-SIJ score at year 5 incorporating measurements from all readers at baseline for MRI-SIJ and year 5 for the X-SIJ score and taking into account the within-reader correlation; N=patients with at least one baseline MRI-SIJ/5-year X-SIJ pair (i.e., at the same time points available) and complete data on all covariates at baseline. §Longitudinal association between MRI-SIJ inflammation and X-SIJ score (all measurements from all readers for both modalities) over the 5-year follow-up with time-lagged models and first-order autoregression, taking into account the within-reader and within-patient correlation for the repeated measurements; N=patients with at least one X-SIJ/ MRI-SIJ pair and complete data on all covariates for the available pairs.

aOR, adjusted OR (adjusted for: symptom duration, gender, CRP, BASDAI, smoking status, treatment with NSAIDs and treatment with TNFi for longitudinal models); BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; CRP, C reactive protein; GEE, generalised estimating equations; MI, multiple imputation; mNY, modified New York criteria; MRI-SIJ, MRI of the sacroiliac joints; NSAIDs, non-steroidal anti-inflammatory drugs; NA, not applicable—the main effect of MRI-SIJ inflammation on the different outcomes is only shown if the interaction with HLA-B27 is not significant (p≥0.15); NS, not significant; otherwise the effect of MRI-SIJ in each strata of HLA-B27 is shown; TNFi, tumour necrosis factor inhibitors.

DISCUSSION

The main findings of this 5-year follow-up study can be summarised as follows: (1) 5-year radiographic SIJ progression is statistically significant but of limited magnitude; (2) strategically chosen definitions of radiographic progression may be more sensitive to change over time than the rigid (binary) mNY based definition; and 3) inflammation on MRI-SIJ is highly predictive of a structural radiographic SIJ progression. Moreover, these data provide meaningful information for the clinician who likes to determine the risk of progression in an individual patient, using baseline parameters such as HLA-B27 positivity, radiographic structural damage, MRI-SIJ inflammation, and abnormal CRP.

In order to properly interpret the rate of progression of SIJ damage that we found in this study, two quantities have to be considered: (A) the proportion of patients with radiographic SIJ damage at baseline; and (B) the proportion of patients that change from nr-axSpA to r-axSpA over time.

Observed radiographic SIJ-damage in the DESIR cohort (15%) is in accordance with what has been found before, in light of the relatively short duration of the symptoms (between 3 months and 3 years).[15-17] These data suggest that structural damage can already be found very early in the disease.

Longitudinal studies that allow a proper evaluation of change from nr-axSpA to r-axSpA are scarce: Sampaio-Barros et al. found a 10% progression rate over 2 years in one study[18] [18] and a 24% progression rate over 10 years in another study.[19] However, only the researchers of the GESPIC cohort realised that a proper progression estimate should aggregate worsening as well as improvement, and reported progression in 9% after two years.[17]

The mNY criteria that quantify radiographic damage in SIJ have been proposed several decades ago for classifying a particular patient at a particular point of time. These inherently binary criteria (mNY+ or mNY-) were not intended to evaluate the natural course of the disease. Adaptations thereof may be more sensitive to change and simpler to interpret: our continuous score modification (a score from 0 to 8 based on the ordinal scale of mNY-grading) is more sensitive but harder to interpret to the data-analyst and the clinician. The statistician will worry about the handling of a semi-quantitative variable as if it were a continuous one and will argue the seemingly similar distance between different grades. Moreover, a continuous score is simply the sum of the scores obtained in two SIJs, as if they were independent. A simpler means to express progression to the clinician is to define progression as a change of at least 1 grade in at least 1 SIJ. This proposal has been used for the first time by the GESPIC researchers.[16] Since we felt that a change between grade 0 and grade 1 (and vice versa) is not clinically relevant, we proposed a third definition by ignoring a change from 0 to 1.[3] Our study has confirmed that the sensitivity to change of this adjusted definition is better than the one based on the mNY criteria.

The main weakness of these X-SIJ-based definitions is likely the poor interobserver reliability: the assessment of radiographic damage in the SIJ according to the binary mNY criteria is particularly susceptible to measurement error.[20] While trained central readers have shown better reliability than single (local) readers, a combined-score by our three central readers (‘2

out of 3' score) is still fallible in terms of measurement error, as is suggested by the finding of 'improvement' of SIJ-damage under fully blinded conditions in a significant proportion of patients.

This means that measurement-error (i.e. scoring-variability) must be taken into account when analysing X-SIJ-progression. We have addressed this in two ways: first, our analysis was assumption free. We allowed 'positive change' as well as 'negative change' to occur without labelling this as 'true progression' or 'noise'. We analysed to what extent 5-year SIJ structural damage was driven by baseline inflammation on MRI-SIJ, and we could confirm a positive association: more MRI-inflammation at baseline leads to a higher 5-year SIJ score. In addition, we have used an analytical approach that most efficiently captures all the available information in the model, which adds to precision. In fact, our main analysis (the 1-level GEE) was more precise (narrower CI) than the 'traditional' logistic regression.

This cohort study in early axSpA reiterates the importance of BME on MRI-SIJ as a predisposing factor for developing radiographic sacroiliitis 5 years later.[3, 20] Of note, HLA-B27 was an effect modifier: patients carrying this genetic (risk) marker had a larger effect of MRI-inflammation on radiographic damage than those not carrying this marker. This disparate effect suggests HLA-B27 is a critical factor for the severity of axSpA.[21, 22]

Our data suggest that a proper risk estimation in individual patients is within our scope: an nr-axSpA patient that is HLA-B27-negative, has a normal CRP and a negative MRI-SIJ has a likelihood of only 1.2% to progress to r-axSpA. In contrast, this likelihood is 18.4% if the patient is HLAB27-positive, the CRP is increased and the MRI-SIJ shows BME.

Further studies are required to better estimate the X-SIJ progression in axSpA and to better understand the role of inflammation on this progression.

SUPPLEMENTARY DATA

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

REFERENCES

1. Dougados M, Baeten D. Spondyloarthritis. Lancet. 2011 Jun 18; 377(9783):2127-2137.
2. Sieper J, Braun J, Dougados M, et al. Axial spondyloarthritis. Nat Rev Dis Primers. 2015 Jul 9; 1:15013.
3. 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.
4. Dougados M, Etcheto A, Molto A, et al. Clinical presentation of patients suffering from recent onset chronic inflammatory back pain suggestive of spondyloarthritis: The DESIR cohort. Joint Bone Spine. 2015 Oct; 82(5):345-351.
5. 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.
6. 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.
7. Garrett S, Jenkinson T, Kennedy LG, et al. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol. 1994 Dec; 21(12):2286-2291.
8. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the

Bath Ankylosing Spondylitis Functional Index. J Rheumatol. 1994 Dec; 21(12):2281-2285.

9. Dougados M, Simon P, Braun J, et al. ASAS recommendations for collecting, analysing and reporting NSAID intake in clinical trials/epidemiological studies in axial spondyloarthritis. Ann Rheum Dis. 2011 Feb; 70(2):249-251.
10. 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.
11. 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.
12. Heuft-Dorenbosch L, Weijers R, Landewé R, et al. Magnetic resonance imaging changes of sacroiliac joints in patients with recent-onset inflammatory back pain: inter-reader reliability and prevalence of abnormalities. Arthritis Res Ther. 2006; 8(1):R11.
13. 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.
14. 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.
15. Poddubnyy D, Brandt H, Vahldiek J, et al. The frequency of non-radiographic axial spondyloarthritis in relation to symptom duration in patients referred because of chronic back pain: results from the Berlin early spondyloarthritis clinic. Ann Rheum Dis. 2012 Dec; 71(12):1998-2001.
16. Said-Nahal R, Miceli-Richard C, Berthelot JM, et al. The familial form of spondylarthropathy: a clinical study of 115 multiplex families. Groupe Francais d'Etude Genetique des Spondylarthropathies. Arthritis Rheum. 2000 Jun; 43(6):1356-1365.
17. 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.
18. 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.
19. Sampaio-Barros PD, Bortoluzzo AB, Conde RA, et al. Undifferentiated spondyloarthritis: a longterm followup. J Rheumatol. 2010 Jun; 37(6):1195-1199.
20. 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.
21. Bennett AN, McGonagle D, O'Connor P, et al. Severity of baseline magnetic resonance imaging-evident sacroiliitis and HLA-B27 status in early inflammatory back pain predict radiographically evident ankylosing spondylitis at eight years. Arthritis Rheum. 2008 Nov; 58(11):3413-3418.
22. Chung HY, Machado P, van der Heijde D, et al. HLA-B27 positive patients differ from HLA-B27 negative patients in clinical presentation and imaging: results from the DESIR cohort of patients with recent onset axial spondyloarthritis. Ann Rheum Dis. 2011 Nov; 70(11):1930-1936.

Chapter 8

Is active sacroiliitis on MRI associated with radiographic damage in axial spondyloarthritis? Real-life data from the ASAS and DESIR cohorts

Alexandre Sepriano, Sofia Ramiro, Robert Landewé, Maxime Dougados, Désirée van der Heijde, Martin Rudwaleit

Rheumatology (Oxford). 2019 May 1;58(5):798-802

ABSTRACT

Objectives: To assess any association between bone marrow edema on MRI of the sacroiliac joints (MRI-SIJ) according to local readings in daily practice and the development of structural damage on radiographs of the SIJ (X-SIJ) in axial spondyloarthritis (axSpA).

Methods: Patients with axSpA from the Assessment of the SpondyloArthritis international Society (ASAS) and DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR) multicentre cohorts were included. MRI-SIJ and X-SIJ were obtained at baseline, and X-SIJ at follow-up after a mean 4.6 years (ASAS) and 5.1 years (DESIR). All images were scored by local readers. Structural damage in the X-SIJ was defined according to the modified New York criteria. The percentage of structural net progression (number of 'progressors' minus the number of 'regressors' divided by the total number of patients) was assessed and the effect of bone marrow edema on MRI-SIJ on X-SIJ damage evaluated by multivariable logistic regression.

Results: In total, 125 (ASAS-cohort) and 415 (DESIR-cohort) patients had baseline MRI-SIJ and complete X-SIJ data available. According to local readings, progression and 'improvement' in X-SIJ was seen in both the ASAS- and DESIR-cohort, yielding a net progression that was higher in the former than in the latter (19.2% and 6.3%). In multivariable analysis, baseline bone marrow edema on MRI-SIJ was strongly associated with X-SIJ structural progression in both ASAS (odds ratio=3.2 [95% CI: 1.3; 7.9]), and DESIR (odds ratio=7.6 [95% CI: 4.3; 13.2]).

Conclusion: Inflammation on MRI-SIJ is associated with future radiographic progression according to local readings despite an expected increased imprecision invoked by local readings.

INTRODUCTION

Axial spondyloarthritis (axSpA) is a term used to describe patients with SpA with predominant axial manifestations including those with (radiographic axSpA; r-axSpA) and without (non-radiographic axSpA; nr-axSpA) evidence of radiographic damage at the sacroiliac joint (SIJ) level (according to the modified New York criteria; mNY).[1]

Over the years, several studies have been assessing the rate of progression from nr-axSpA to r-axSpA (i.e. from mNY-negative to mNY-positive).[2-8] Overall, progression is known to be a slow process in axSpA, but some features have been shown to associate with an increase in SIJ damage accrual, especially objective inflammatory markers, such as elevated CRP and presence of inflammation at the local level as measured by subchondral bone marrow edema (BME) on MRI of the SIJ (MR-SIJ).[2, 3, 5, 9, 10]

To partially control for the well-known limitations of the mNY method (i.e. poor reliability due to substantial interobserver variation) and to arrive at the most reliable and unbiased progression rate, researchers have been relying on scores provided by trained central readers (often more than one) when assessing SIJ radiographic progression and predictors thereof.[11, 12] Indeed, central reading (especially when more than one reader contributes with scores) has been shown to increase the chances of finding subtle associations.[13] On the other hand, central reading findings are not easy to transfer to clinicians' daily clinical practice where central imaging interpretation is not available.

The effect of BME on MRI-SIJ on SIJ radiographic progression using imaging data provided by (untrained) local readers, has not been tested thus far. Therefore, the question remains whether the practicing clinician can use the imaging data available in daily clinical practice, though possibly less reliable, to make prognostic decisions when confronted with a positive MRI-SIJ, as suggested by studies with dedicated central reading procedures.

We aimed to test the possible effect of MRI-SIJ inflammation on structural damage in radiographs of the SIJ (X-SIJ), when both are assessed by local readers as in daily clinical practice.

METHODS

Patients and study design

Patients with axSpA according to their treating rheumatologist from the Assessment of the SpondyloArthritis international Society (ASAS) cohort (clinicaltrials.gov ID: NCT00328068) and from the DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR) cohort (clinicaltrials.gov ID: NCT01648907), with baseline MRI-SIJ and complete (i.e. baseline and follow-up) X-SIJ data available were included. Details on the inclusion criteria of the above-mentioned cohorts have been previously reported.[14, 15] Importantly, they differ in the duration of symptoms allowed for inclusion, which was not restricted in the ASAS cohort, but was limited to 3 years in DESIR.

Both studies were conducted according to good clinical practice guidelines and were approved by the appropriate local medical ethical committees. Written informed consent was obtained from participating patients before inclusion.

Data collection

Information on age, symptom duration (in years), gender, HLA-B27 status (positive/negative) and on CRP (mg/L) was collected at baseline in both cohorts. In addition, in DESIR, data on disease activity (BASDAI), smoking status (smoker/non-smoker) and treatment with non-steroidal anti-inflammatory drugs (yes/no) was also collected. MRI-SIJ and X-SIJ were obtained at baseline, and X-SIJ at follow-up (ASAS: mean (S.D.) 4.6 (0.8) years; DESIR: 5.1 (0.2) years) and evaluated by a local reader (i.e. rheumatologist and/or radiologist). Images were taken unblinded to other imaging information and clinical characteristics. Readers had the option to view the baseline image when scoring the follow-up image. BME at MRI-SIJ was assessed either without a formal definition (i.e. according to the reader overall judgement; ASAS-cohort) or according to the ASAS definition (DESIR-cohort) as present/absent.[16, 17] Structural damage in the X-SIJ was defined according to the mNY criteria (positive/negative).[18]

Statistical analysis

The percentage of structural net progression was defined as the number of 'progressors' (change from mNY-negative to mNY-positive) minus the number of 'regressors' (change from mNY-positive to mNY-negative) divided by the total number of patients. Net progression was assessed separately in the entire population of each cohort and in subgroups according to the CRP and BME status at baseline. The effect of baseline MRI-SIJ BME on X-SIJ damage at follow-up was evaluated in two types of logistic regression models adjusted for potential baseline confounders selected *a priori* based on clinical grounds: i) including only variables common to both cohorts (i.e. gender, HLA-B27, CRP, symptom duration); and ii Including all common variables plus the ones only available in DESIR (i.e. BASDAI, smoking status and treatment with non-steroidal anti-inflammatory drugs). All models were fit including all axSpA patients irrespective of the mNY status at baseline.

RESULTS

In total, 125 (out of 445) and 415 (out of 708) axSpA patients were included from the ASAS and DESIR cohorts, respectively. Patients that were included were more likely to be HLA-B27 positive and to have radiographic sacroiliitis and BME on MRI-SIJ at baseline than those that did not, in the DESIR cohort but were similar in the ASAS cohort (Supplementary Table S1 and S2 available at *Rheumatology* online).

Included patients from the ASAS cohort had longer mean symptom duration (6.7 vs 1.5 years) and were also more likely to be HLA-B27 positive (70% vs 64%), to have BME on MRI-SIJ (66% vs 40%) and elevated CRP (38% vs 30%) at baseline as compared with patients from the DESIR cohort.

Radiographic progression

From the total 125 patients in the ASAS cohort, 35 (28%) changed from mNY-negative to mNY-positive (positive change) after a mean of 4.6 years, while 11 (8.8%) changed in the opposite direction (negative change), resulting in a net percentage of progression of 19.2%. In DESIR, positive change occurred in 49 (11.8%) out of the total 415 patients after a mean of 5.1 years; and negative change in 23 (5.5%), yielding a net progression of 6.3%. In Fig. 1, net progression is shown in subgroups of patients according to the presence of objective signs of inflammation at baseline. In both cohorts, progression was much higher if BME on MRI-SIJ was present regardless of CRP elevation.

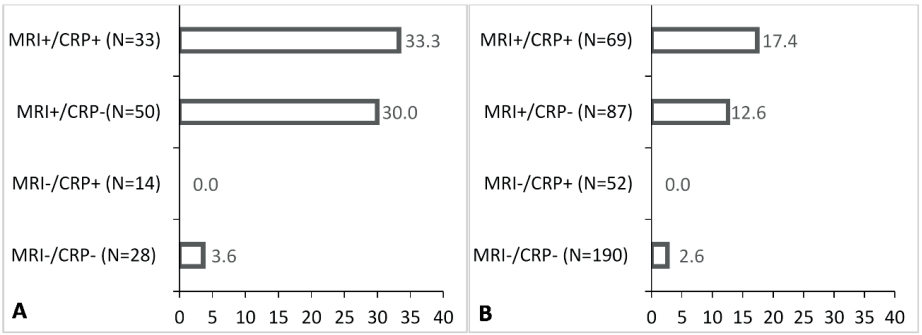


Figure 1. Net progression from mNY-negative to mNY-positive according to baseline objective inflammatory markers.
(A) ASAS cohort: N=125; (B) DESIR: N=398 (17 patients miss baseline CRP). MRI, magnetic resonance imaging; CRP, c reactive protein; mNY, modified New York criteria; SIJ, sacroiliac joints.

Effect of MIR-SIJ inflammation on X-SIJ progression

In the multivariable analysis (including only variables common to both cohorts), BME on MRI-SIJ was found to be an independent predictor of the development of radiographic damage both in the ASAS (odds ratio=3.2 [95% CI: 1.3; 7.9]), and DESIR (odds ratio=7.6 [95% CI: 4.3; 13.2]) cohort (Table 1). The results were similar also in the model adjusted for variables only available in DESIR (odds ratio=6.6 [95% CI: 3.7; 11.6]).

Table 1. Effect of inflammation on MRI-SIJ at baseline on the development of X-SIJ structural damage at follow-up

Predictor \ Outcome	mNY aOR (95% CI)
Sacroiliitis on MRI-SIJ (ASAS-cohort) (N=125)	**3.2 (1.3; 7.9) ***
Sacroiliitis on MRI-SIJ (DESIR-cohort) (N=398)	**7.6 (4.3; 13.2) ***

* Adjusted for gender, HLA-B27, CRP, symptom duration. MRI-SIJ, magnetic resonance imaging of the sacroiliac joints; X-SIJ, radiograph of the SIJ; mNY, modified New York criteria; c reactive protein; aOR, adjusted odds ratio.

DISCUSSION

In this study we analysed data from two independent multicentre cohorts conducted in daily practice with readings of MRI-SIJ and X-SIJ performed by local rheumatologists or radiologists. We have shown axSpA patients with inflammation on MRI-SIJ at baseline were 3-7-fold more likely to develop radiographic damage after 4.6-5.1 years. We were able to find this relationship despite the fact that local readers may not be necessarily well trained, and that the scores are usually based on one reader only, two factors that increase variation in scores. On the other hand, local readers were unblinded to the time order of images, which may increase precision, but also has the risk of expectation bias.

The assessment of SIJ radiographic progression based on the mNY grading system is challenging. Researchers have been implementing strategies to handle the well-known poor reliability of this method.[11, 12] The use of scores from at least one trained central reader being one of the most common.[2, 3, 5] Central reading reduces (but does not eliminate) the 'noise' and increases the likelihood of capturing true progression (i.e. the signal). Although the 'noise' is expectedly bidirectional (if readers are blinded to time-order) it is not unreasonable to assume that it explains the captured 'improvements' of structural damage.[13]

The above-mentioned concept of 'signal' to 'noise' ratio remained overlooked for several years. Only recently, researchers from the German Spondyloarthritis Inception Cohort acknowledged that 'improvements' should not be ignored when calculating progression.[5] In this cohort 3 (1.4% of the total) axSpA patients that were mNY-positive at baseline 'improved' after 2 years (i.e. became mNY-negative). Also, in DESIR, 'improvements' were seen in 7 patients (1.6% of the total) after 2 years and in 3 (0.7%) after five years with central reading.[2, 3] These studies, even within the same cohort, differ from each other in the method to obtain the scores (e.g. how to combine data from different readers) as well in the method to calculate 'net progression', but they all unequivocally show that improvements (i.e. noise) can still be seen even with central reading.

Thus, it is not surprising that when relying on local (untrained) readers, as in the current study, figures for improvements and potentially for worsening were even higher compared with studies with central readings. 'Improvements' were seen in 9% and 6% of all axSpA patients from the ASAS and DESIR cohorts, respectively, even though readers had the possibility to access the baseline scoring when judging the follow-up images. Yet, it was neither assessed in ASAS nor in DESIR, whether the same or different readers scored baseline and follow-up images, and whether or not readers reviewed baseline images (MRI/radiographs) at the time of scoring follow-up radiographs. After taking measurement error into account, the 'net progression' was higher in the ASAS (19%) than in the DESIR cohort (6%), which may be partly explained by prognostic dissimilarities between the two populations (i.e. patients from the ASAS cohort had higher likelihood of features known to associate with structural progression: e.g. elevated CRP and HLA-B27 positivity).[2, 5] Overall, it would be expected that a low signal/noise ratio of scoring radiographs could compromise the ability to detect significant associations, especially, because the predictor of interest (i.e. BME on MRI-SIJ) is not free of measurement error either, though to a lesser extent compared to radiographs.[17] Notwithstanding, and despite all the

noise, it is remarkable that inflammation on MRI-SIJ is still clearly associated with the development of radiographic damage in both cohorts (with different populations – adding to external validity).

The results from this study should be interpreted with some caution. Unblinded readings as done in daily practice may lead to a higher rate of progression (expectation bias). However, given the rather high rate of 'improvements' of X-SIJ at follow-up, unblinded readings appeared not to be a major confounder in this regard in both cohorts. Moreover, the association between baseline BME on MRI-SIJ and the later development of radiographic damage was found at the group level. This means that, on average, patients with BME on MRI are 3-7 times more likely to develop structural damage in a setting that the rheumatologist encounters in daily practice. However, our data do not support (and we do not claim) that finding inflammation on MRI-SIJ at the individual-patient level implies definite progression in that individual patient. Of note, this limitation applies in the same way to studies with central reading assessments.

In summary, our data from the two multicentre cohorts show, for the first time, that at the group level SIJ inflammation on MRI is associated with the later development of structural progression in radiographs according to local readings in clinical practice.

SUPPLEMENTARY DATA

Supplementary data are published online on the website of Rheumatology (Oxford)

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. 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.
3. 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.
4. 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.
5. 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.
6. Sampaio-Barros PD, Bortoluzzo AB, Conde RA, et al. Undifferentiated spondyloarthritis: a longterm followup. J Rheumatol. 2010 Jun; 37(6):1195-1199.
7. 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.
8. Schattenkirchner M, Kruger K. Natural course and prognosis of HLA-B27-positive oligoarthritis. Clin Rheumatol. 1987 Sep; 6 Suppl 2:83-86.
9. Bennett AN, McGonagle D, O'Connor P, et al. Severity of baseline magnetic resonance imaging-evident sacroiliitis and HLA-B27 status in early inflammatory back pain predict radiographically evident ankylosing spondylitis at eight years. Arthritis Rheum. 2008 Nov; 58(11):3413-3418.
10. Gong Y, Zheng N, Chen SB, et al. Ten years' experience with needle biopsy in the early diagnosis of sacroiliitis. Arthritis Rheum. 2012 May; 64(5):1399-1406.
11. 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.

12. 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.
13. 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.
14. Dougados M, Etcheto A, Molto A, et al. Clinical presentation of patients suffering from recent onset chronic inflammatory back pain suggestive of spondyloarthritis: The DESIR cohort. Joint Bone Spine. 2015 Oct; 82(5):345-351.
15. 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.
16. 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.
17. 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.
18. 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.

Chapter 9

Inflammation of the sacroiliac joints and spine on MRI predicts structural changes on MRI in axial spondyloarthritis: 5-year data from DESIR

Alexandre Sepriano, Sofia Ramiro, Robert Landewé, Pascal Claudepierre, Daniel Wendling, Maxime Dougados, Désirée van der Heijde

Accepted for publication, Arthritis Care & Research

ABSTRACT

Objective: To test the impact of inflammation on MRI-structural changes occurring in the sacroiliac joints (SIJ) and the spine.

Methods: Patients with early axSpA from the DESIR cohort were included. MRIs of the SIJ (MRI-SIJ) and spine (MRI-spine), obtained at baseline, 2 and 5 years, were scored by 3 central readers. Inflammation and structural damage on MRI-SIJ/MRI-spine were defined by the agreement of ≥2 of 3 readers (binary outcomes), and by the average of 3 readers (continuous outcomes). The effect of inflammation (MRI-SIJ/MRI-spine) on damage (MRI-SIJ/MRI-spine, respectively) was evaluated in two models: i. Baseline prediction model: effect of baseline inflammation on damage assessed at 5-year; and ii. Longitudinal model: effect of inflammation on structural damage assessed during 5 years.

Results: 202 patients were included. Both the presence of bone marrow edema (BME) on MRI-SIJ and on MRI-spine at baseline were predictive of 5-year damage (≥3 fatty lesions) on MRI-SIJ [OR=4.2 (95% CI: 2.4; 7.3)] and MRI-spine [OR=10.7 (95% CI: 2.4; 49.0)], respectively, when adjusted for CRP. The association was also confirmed in longitudinal models (when adjusted for ASDAS) both in the SIJ [OR=5.1 (95% CI: 2.7; 9.6)] and spine [OR=15.6 (95% CI: 4.8; 50.3)]. Analysis of other structural outcomes (i.e. erosions) on MRI-SIJ yielded similar results. In the spine, a significant association was found for fatty lesions but not for erosions and bone spurs, which occurred infrequently over time.

Conclusion: We found a predictive and longitudinal association between MRI-inflammation and several types of MRI-structural damage in patients with early axSpA which adds to the proof for a causal relationship.

INTRODUCTION

Axial spondyloarthritis (axSpA) is a disease predominantly characterized by involvement of the axial skeleton. Axial involvement often translates into imaging abnormalities, which usually represent either an underlying inflammatory or structural lesion. Magnetic resonance imaging of the sacroiliac joints (MRI-SIJ) and spine (MRI-spine) is a modality to detect, quantify and evaluate (change of) axial inflammation in axSpA. Thus far, conventional radiographs have been prescribed for assessing progression of structural damage in clinical practice and research.

Patients with axSpA experience varying levels of radiographic progression (e.g. the occurrence of radiographic 'sacroiliitis' and new syndesmophytes).[1-4] Identifying patients with a higher likelihood of damage accrual is key to tailor treatment strategies early in the disease course. Elevation of C-reactive protein (CRP), disease activity as measured with the Ankylosing Spondylitis Disease Activity Score (ASDAS) and bone marrow edema (BME) on MRI-SIJ or MRI-spine have been shown to associate with increased probability of structural progression on conventional radiographs.[3, 5-12] However evidence is scarce in early disease and mostly limited to studies on which structural damage was measured with conventional radiographs.

The Interpretation of data stemming from the above-mentioned studies may be jeopardized by limitations of the instruments used to measure structural progression, especially at the SIJ level. It is well established that radiographic sacroiliitis defined by the mNY criteria is poorly reliable.[13-15] Investigators have been implementing strategies to improve the 'signal-to-noise' ratio by, for instance, combining judgments from ≥2 trained central readers.[3] Still, these strategies cannot fully eliminate the 'noise'.

In recent years there has been a growing interest in evaluating axial damage with other imaging modalities, such as MRI. Definitions for individual lesions (e.g. fatty lesions, erosions) have been proposed and composite scores validated.[16-19] Although MRI-detected lesions, as any outcome measure, are far from being error-free, available literature shows higher reliability for MRI-SIJ compared to pelvic radiographs in detecting structural lesions.[20] A better 'signal-to-noise' ratio, in theory, improves the ability to detect change and predictors thereof, especially in early disease where, at the group level, damage is known to be limited and to progress slowly.[3, 21]

Thus far, no study has assessed the effect of inflammation on structural damage evaluated on MRI. We aimed to test the effect of inflammation on several types of structural lesions both assessed by MRI and at the level of the SIJ and the spine in patients with early axSpA.

METHODS

Patients and study design

Five-year data from patients with early axSpA from the DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR) cohort have been used.[22] Patients had to have ≥2 consecutive MRI images (either of the SIJ or spine) during the 5-year follow-up to be included. The database used for the current analysis was locked on the 20th of June 2016. The study was conducted according to Good Clinical Practice guidelines and was approved by the appropriate

local ethics committees. Written informed consent had been obtained from participating patients before inclusion.

Imaging scoring procedures

MRI-SIJ and MRI-spine were performed at baseline for all patients. By protocol, at two and five years of follow-up MRIs were only performed in participating centres in Paris (n=9 out of the 25 participating centers). Each image was independently scored by 3 trained central readers blinded to chronology and clinical data. MRI-SIJ and MRI-spine were performed on a 1-1.5T scanner providing T1-weighted Turbo Spin-Echo (T1-w) and Short Tau Inversion Recovery (STIR) sequences. Scanning was performed in a coronal oblique plane for SIJ and in a sagittal plane for spine, with a slice thickness of 4mm. A detailed description of the MRI protocol in DESIR has been previously reported.[23, 24]

Structural damage on MRI

The Spondyloarthritis Research Consortium of Canada (SPARCC) MRI-SIJ Structural score by Weber *et al* was used to define individual structural lesions on MRI-SIJ.[18] In the absence of a formal definition for structural damage on MRI-SIJ, we considered 3 definitions previously shown most discriminatory between axSpA and no axSpA: ≥5 fatty lesions and/or erosions; ≥3 erosions; and ≥3 fatty lesions.[25] Continuous structural lesions on MRI-SIJ were defined as number of erosions, number of fatty lesions (both range: 0-40), number of fatty lesions and/or erosions (range: 0-80), and as the total number of lesions including fatty lesions, erosions, partial ankylosis / total ankylosis with the addition of sclerosis (not in the original score) (range: 0-144).

Structural lesions on MRI-Spine were scored according to the Canada–Denmark (CANDEN) method, modified to include only corner lesions.[16, 17] Similar to MRI-SIJ, in absence of a formal definition, we defined structural damage on MRI-spine as ≥5 fatty lesions, which has been previously shown highly specific for axSpA.[25, 26] In addition, we also considered ≥5 fatty lesions and/or erosions; ≥3 erosions; ≥3 fatty lesions; and ≥3 bone spurs. The total number of fatty lesions, erosions, bone spurs (range: 0-92; for each), fatty lesions and/or erosions (range: 0-184) and the total number of structural lesions (fatty lesions, erosions, bone spurs, including also ankylosis; range: 0-322) was assessed, as continuous structural outcomes.

Inflammation on MRI

Inflammation on MRI-SIJ was assessed using the Assessment of SpondyloArthritis international Society (ASAS)-definition (positive/negative) and the SPARCC-score (range: 0-72).[27-29] BME on MRI-Spine was defined according to the ASAS definition (≥3 vertebral corner lesions; positive/negative).[30] In addition, a cut-off of at least 5 lesions was assessed, as it has been shown to be highly specific of axSpA.[25] The total spine SPARCC score was used as a continuous inflammatory outcome (range: 0-414).[31]

The interreader reliability of the MRI scores used in this study has been reported elsewhere.[32]

Statistical analysis

Structural progression of binary scores was assessed in clinically relevant subgroups according to the CRP and BME status at baseline, and defined by the agreement of ≥2 out of 3 readers as the percentage of net progression: the number of 'progressors' (change from negative to positive) minus the number of 'regressors' (change from positive to negative) divided by the total number of patients, a method previously described in detail.[33]

The effect of inflammation, both on MRI-SIJ and MRI-spine, on structural outcomes, again both on MRI-SIJ and MRI-spine, respectively, was evaluated by two types of generalized estimating equations (GEE) models: i. a baseline model: effect of baseline inflammation on 5 years structural damage incorporating measurements from all readers (1-level GEE model adjusted for reader); and ii. A longitudinal model: effect of BME at *t* on structural outcomes at *t+1* over 5 years (longitudinal time-lagged 2-level GEE models with auto-regression). Binary variables of Inflammation (i.e. BME) were modelled using binary damage outcomes (binomial GEE), while continuous variables of inflammation (i.e. SPARCC) were modelled using continuous outcomes of damage (linear GEE).

Table 1. Baseline patient and disease characteristics comparing patients with MRI available in ≥2 consecutive (included) visits to those without (excluded)

	MRI on ≥2 consecutive visits (N=202)	MRI on <2 consecutive visits (N=60)
Age at baseline (years)	34 (9)	33 (8)
Male gender	96 (48)	27 (45)
Symptom duration (years)	2 (1)	1 (1)
HLA-B27	125 (62)	32 (53)
ASAS axSpA criteria	133 (66)	35 (60)
Sacroiliitis on MRI-SIJ† (ASAS)	58 (29)	15 (28)
BME on MRI-Spine† (ASAS)	14 (7)	3 (6)
≥ 5 BME lesions on MRI-spine	10 (5)	2 (4)
Radiographic sacroiliitis† (mNY)	25 (13)	8 (14)
≥ 3 fatty lesions on MRI-SIJ	23 (12)	7 (14)
≥ 3 erosions on MRI-SIJ	29 (15)	9 (17)
≥ 3 fatty lesions on MRI-spine	3 (2)	0 (0)
≥ 3 erosions on MRI-spine	0 (0)	0 (0)
≥ 3 bone spurs on MRI-spine	0 (0)	0 (0)
BASDAI (0-10)	4 (2)	47 (21)
ASDAS-CRP	3 (1)	3 (1)
Elevated CRP (≥6 mg/L)	52 (27)	12 (21)
BASFI£ (0-10)	3 (2)	33 (28)
Treatment with NSAIDs	192 (95)	57 (95)
Treatment with TNFi	0 (0)	0 (0)

Values are mean (SD) for continuous variables and number (percentage) for dichotomous variables. * Independent samples t-test for continuous and Chi2 for dichotomous variables; † agreement between 2 out of 3 readers; <5% missing data: mNY, BME on MRI-spine (ASAS), ≥ 5 BME lesions on MRI-spine, ≥ 3 fatty lesions on MRI-spine, ≥ 3 erosions on MRI-spine, ≥ 3 bone spurs on MRI-spine, ASDAS, CRP; <1% missing data: sacroiliitis on MRI-SIJ, ≥ 3 fatty lesions on MRI-SIJ, ≥ 3 erosions on MRI-SIJ, BASDAI, BASFI. MRI, magnetic resonance imaging; SIJ, sacroiliac joints; ASAS, Assessment of SpondyloArthritis international Society; mNY, radiographic sacroiliitis according to the modified New York criteria; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; ASDAS, Ankylosing Spondylitis Disease Activity Score; CRP, C-reactive protein; BASFI, Bath Ankylosing Spondylitis Functional Index; NSAID, non-steroidal anti-inflammatory drugs; TNFi, tumor necrosis factor inhibitors; NA, not applicable

The final multivariable models included variables that were found to confound the association of interest (i.e. that importantly changed the effect of inflammation on structural outcomes). The following variables were tested as possible confounders: age (in years), gender (male vs female), HLA-B27 (positive vs negative), smoking status (smoker vs non-smoker), CRP (mg/L), Bath Ankylosing Spondylitis Activity Index (BASDAI), Ankylosing Spondylitis Disease Activity Score (ASDAS) (BASDAI plus CRP and ASDAS tested in separate models to avoid collinearity), treatment with nonsteroidal anti-inflammatory drugs (NSAIDs) (yes/no) and tumor necrosis factor inhibitors (TNFi) (yes/no). Variables with a potential to change over time were modelled as such (i.e. all the above except gender and HLA-B27) in the longitudinal models

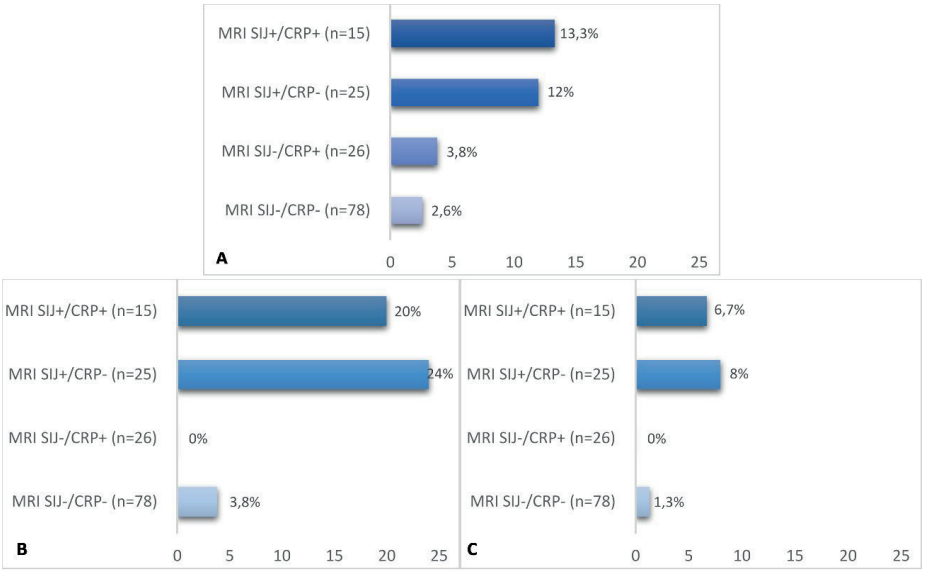


Figure 1. Net progression from MRI-SIJ without structural lesions (MRI-SIJ-STR negative) to MRI-SIJ with structural lesions (MRI-SIJ-STR positive) defined by **(A) ≥5 fatty lesions and/or erosions, (B) ≥3 fatty lesions and (C) ≥3 erosions**, according to baseline objective inflammatory markers (MRI-SIJ inflammation and CRP); MRI-SIJ+: Presence of bone marrow edema on MRI-SIJ according to the ASAS definition, CRP+: CRP ≥6 mg/l at baseline. Net progression from MRI-SIJ-STR negative to MRI-SIJ-STR positive at year 5: number of 'progressors' minus the number of 'regressors' divided by the total number of patients in each category (N=144; MRI-SIJ available both at baseline and year 5 and CRP available at baseline). MRI, magnetic resonance imaging; SIJ, sacroiliac joints; STR, structural; CRP, C-reactive protein.

RESULTS

Baseline characteristics

Of the total 708 patients from DESIR, 262 could have imaging at follow-up according to the protocol and 202 had at least 2 consecutive visits with data available either on MRI-SIJ or MRI-Spine (196 had both modalities, 3 had MRI-SIJ only and 3 had MRI-Spine only) and were therefore included. No significant baseline differences were found between patients included and not included in this study (Table 1). The presence of BME at baseline was more frequent in the SIJ (29%) than in the spine [7% (ASAS definition); 5% for ≥5 BME lesions]. Likewise, structural

damage was higher in the SIJ (e.g. ≥3 fatty lesions on MRI-SIJ: 12%) than in the spine (e.g. ≥3 fatty lesions on MRI-spine: 2%).

Table 2. Effect of MRI inflammation on MRI structural damage in the SIJ (multivariable models)

Binary scores	≥5 fatty lesions/erosions OR (95% CI)	≥3 fatty lesions OR (95% CI)	≥3 erosions OR (95% CI)
BME at baseline† (N=144-151)	**5.6 (3.1; 10.0)***	**4.2 (2.4; 7.3)***	**4.1 (2.1; 7.8)**
BME over 5 years‡ (N=197-199)	**7.7 (4.5; 13.4)¥**	**5.1 (2.7; 9.6)¥**	**3.2 (1.9; 5.3)**
Continuous scores	Fatty lesions/erosions β (95% CI)	Fatty lesions β (95% CI)	Erosions β (95% CI)
SPARCC at baseline† (N=144-151)	**0.23 (0.15; 0.31)***	**0.12 (0.05; 0.19)***	**0.12 (0.06; 0.18)**
SPARCC over 5 years‡ (N=197-199)	**0.13 (0.07; 0.19)¥**	**0.10 (0.04; 0.16)¥**	**0.04 (0.01; 0.06)**

† Multilevel GEE models: effect of inflammation at baseline on the outcome at 5 years taking the scores from the individual readers into account, ‡ longitudinal multilevel time-lagged GEE models with autoregression (i.e. effect of inflammation at *t* on the outcome at *t+1* adjusted for the outcome at t, taking the scores from the individual readers into account); * Adjusted for CRP at baseline; ¥ Adjusted for time-lagged ASDAS-CRP. BME, bone marrow edema according to the ASAS definition (positive/negative); MRI-SIJ, magnetic resonance of the sacroiliac joints; SPARCC, spondyloarthritis research consortium of Canada; ASDAS, Ankylosing Spondylitis Disease Activity Score; CRP, C-reactive protein.

Structural progression according to the presence of objective inflammation at baseline

In total, 155 patients had complete MRI data at baseline and 5 years (141 both modalities, 10 MRI-SIJ only and 4 MRI-Spine only). Net progression, defined by ≥ 5 fatty lesions and/or erosions, ≥3 fatty lesions and ≥3 erosions on MRI-SIJ, according to baseline objective inflammatory markers is shown in Figure 1. Patients with BME on MRI-SIJ present at baseline had higher net progression rates compared to those that were BME negative for all outcomes, irrespective of the CRP status (range if BME positive: 7% to 24%; range if BME is negative: 0% to 4%). On MRI-spine overall net progression was -0.7% both for ≥5 fatty lesions and/or erosions and for ≥5 fatty lesions; 0.7% for ≥3 fatty lesions and 0% for ≥3 erosions and for ≥3 bone spurs. These low numbers precluded further analysis according to the presence of inflammatory markers at baseline.

Effect of inflammation on structural progression (multivariable models)

Sacroiliac joints

The presence of BME on MRI-SIJ at baseline was predictive of the development of fatty lesions and erosions on MRI-SIJ 5 years later for all binary definitions [range odds ratio (OR): 4.1-5.6], after adjustment for CRP at baseline (Table 2). Similar results were found in the longitudinal models (after adjustment for ASDAS). On average, patients with BME on MRI-SIJ had a 5 times higher likelihood of having at least 3 fatty lesions in the subsequent visit as compared to those without BME [OR (95% CI): 5.1 (2.7; 9.6)] (Figure 2). The association between the continuous SPARCC score on MRI-SIJ and the various continuous structural outcomes was also always statistically significant, and present in both models.

Spine

Testing the association of interest on MRI-spine was hampered by low number of lesions, leading to imprecise estimates and, for some outcomes (i.e. ≥3 erosions and ≥5 fatty lesions/erosions), precluded the estimation of the effect (Table 3). Only the association between BME and ≥3 fatty lesions was statistically significant. The presence of baseline BME (ASAS definition) on MRI-spine was positively associated with ≥ 3 fatty lesions at 5 years on MRI-spine [OR (95% CI): 10.7 (2.4; 15.6)]. This effect was also positive in the longitudinal model [OR (95% CI): 15.6 (4.8; 50.3)] (Figure 2). As in MRI-SIJ, CRP (baseline models) and ASDAS (longitudinal models) have been found to confound the association of interest. Testing the effect of ≥5 BME lesions yielded similar results, but with wider 95% CI (Online Supplementary Table S1). For continuous variables a positive association could be found for fatty lesions alone or in combination with erosions, but not for erosions alone and bone spurs, both in baseline and longitudinal models.

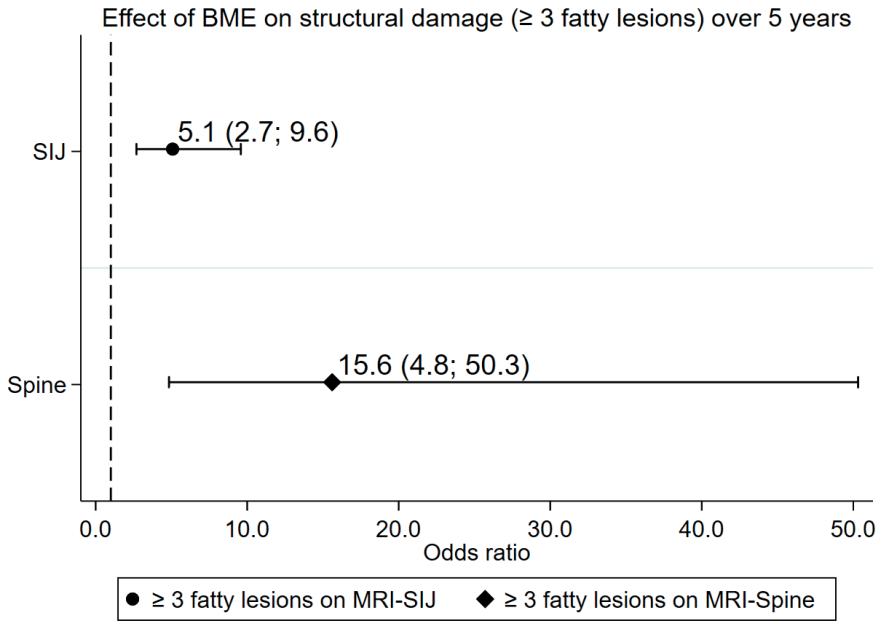


Figure 2. Effect of BME (according to the ASAS definition) on structural damage (defined as ≥3 fatty lesions) both in the SIJ and spine (longitudinal time-lagged models with autoregression). BME, bone marrow edema; ASAS, Assessment of SpondyloArthritis international Society; BME, bone marrow edema; SIJ, sacroiliac joints; MRI, magnetic resonance imaging.

Table 3. Effect of MRI inflammation on MRI structural damage in the spine (multivariable models)

Binary scores	≥5 fatty lesions/erosions	≥ 5 fatty lesions	≥3 fatty lesions	≥3 erosions	≥3 bone spurs
BME at baseline† (N=139)	¥	¥	**10.7 (2.4; 49.0)***	¥	3.2 (0.4; 27.8)*
BME over 5 years‡ (N=197)	¥	0.9 (0.8; 1.2)£	**15.6 (4.8; 50.3)£**	¥	2.8 (0.8; 9.6)£
Continuous scores	Fatty lesions/erosions	Fatty lesions		Erosions	Bone spurs
SPARCC at baseline† (N=139-145)	**0.10 (0.01; 0.18)***	**0.08 (0.02; 0.14)**		0.02 (0.00; 0.03)†	0.01 (-0.01; 0.03)†
SPARCC over 5 years‡ (N=197)	**0.06 (0.02; 0.11)£**	**0.07 (0.02; 0.11)£**		0.00 (-0.01; 0.01)£	0.01 (0.00; 0.02)

† Multilevel GEE model (i.e. effect of inflammation at baseline on the outcome at 5 years taking the scores from the individual readers into account), ‡ longitudinal multilevel time-lagged GEE models with autoregression (i.e. effect of inflammation at t on the outcome at *t+1* adjusted for the outcome *at t*, taking the scores from the individual readers into account); * Adjusted for CRP at baseline; £ Adjusted for time-varying lagged ASDAS-CRP. BME, bone marrow edema according to the ASAS definition (≥3 lesions; positive/negative); MRI-Spine, magnetic resonance of the spine; ¥ model fails to converge due to low number of events. SPARCC, spondyloarthritis research consortium of Canada; ASDAS, Ankylosing Spondylitis Disease Activity Score; CRP, C-reactive protein; ASAS, Assessment of SpondyloArthritis international Society.

DISCUSSION

In this prospective observational cohort study, we have shown that axial inflammation detected on MRI predicts subsequent development of structural lesions (especially fatty lesions) also on MRI over 5 years in patients with early axSpA. This effect is independent of systemic inflammation and is seen both at the SIJ and spinal level but is measured more precisely in the SIJ where damage prevails in early disease. Our results add to the existing evidence by showing that the association between axial inflammation and some lesions reflecting structural damage can be measured with MRI in patients with early axSpA.

In the current study we have demonstrated an association between local inflammation and structural damage both measured on MRI in patients with early axSpA. Involvement of the axial skeleton in axSpA usually starts at the SIJ level.[21, 34, 35] In line with the literature, we found that 6 times more patients showed structural damage (e.g. ≥ 3 fatty lesions) on MRI-SIJ (12%) than on MRI-spine (2%) at baseline. Consequently, the longitudinal association between BME and structural damage (e.g. ≥3 fatty lesions) on MRI-SIJ [OR 5.1 (95% CI: 2.7; 9.6)] was found with a substantially higher precision (narrower confidence intervals) compared to the same effect in the spine [OR: 15.6 (95% CI 4.8; 50.3)]. Although it may seem that the effect of inflammation on damage is stronger on the spine than on the SIJ (OR: 16 vs 5), this is not necessarily the case. It is well-known that imprecise estimates tend to overestimate effect-sizes.[36]

Evidence that inflammation on MRI drives structural damage in early axSpA is relevant to the practicing rheumatologist since it argues in favor of its use for prognostic stratification. In addition, if inflammation drives damage, it is logical to expect that interventions targeting the former will prevent, or at least retard, the latter. However, thus far, trial data do not support this claim.[37] The complex, and yet not fully understood, pathophysiology of new bone formation in axSpA may, at least in part, explain this disappointing result. For instance, it has been shown that systemic inflammation, measured by ASDAS, predicts spinal radiographic progression in radiographic axSpA (r-axSpA).[6, 8] However, progression was still found in patients with inactive disease. Similarly, in another study, inflammation at the vertebral unit level increased the likelihood of the formation of a new syndesmophyte in the same location 2 years later, but most new syndesmophytes appeared in vertebral units without signs of inflammation.[12] These data highlight the relevance of inflammation in driving structural progression but also suggest that other mechanisms may play a role.

However, biology cannot fully explain the failure of anti-inflammatory drugs in modifying the effect of inflammation on structural damage. Outcome measures (lack of) sensitivity to change, has also been previously proposed as a likely explanation.[38] If an intervention truly prevents further damage by reducing inflammation (or by any other means), low sensitivity to change of the outcome measure may prevent that such effect becomes evident (e.g. no significant difference between active drug and placebo). Thus far, progression of structural damage has been mostly measured in conventional radiographs, with mSASSS and the mNY grading system as the most often used outcomes in the spine and SIJ, respectively. However, both the mSASSS and the mNY have low sensitivity to change and assessing radiographic progression with the latter is further challenged by its poor reliability.[3, 14, 15, 39] It remains to be proven that structural lesions detected on MRI are more sensitive to change than those on radiographs.

However, our study supports that different lesions may yield different results. For instance, compared with erosions or bony spurs, fatty lesions were more prevalent in our early axSpA population, especially in the SIJ leading to more precise estimates. Thus, our data may inform future research aiming at clarifying whether MRI is valid alternative to conventional radiography in detecting structural treatment effects in patients with axSpA.

Our study is not without limitations. First, Inflammatory and structural lesions, per patient, were read together by the same reader, which may obviously result in overestimating the association between both. This contrasts with other studies where inflammation and damage were blindly measured with different imaging modalities. However, it should be stressed that readers were still blinded to time-order. That is, they did not know if a certain lesion (e.g. BME) pertained to a baseline or to a follow-up image. Thus, 'causality by reading' though not impossible, is unlikely to fully explain the impressive associations found in our study. Second, the lack of an association between vertebral corner inflammation on MRI-spine and erosions and bone spurs, should be interpreted with caution. Even though a 'true' lack of association cannot be ruled out, as mentioned above, this may be also due to low statistical power driven by low number of these lesions in the spine. The role of inflammation on sites other than vertebral corners for the progression of spinal damage should be addressed in future studies.

In summary, we have shown that local inflammation is associated with development of structural damage (e.g. fatty lesions), both measured with MRI, over 5 years in the SIJ and spine in early axSpA. This association is detected with more precision on the SIJ where structural damage prevails, compared to the spine, in early disease. These findings support the concept that MRI is a valid alternative to conventional radiographs in detecting the structural consequences of axial inflammation in patients with early axSpA.

SUPPLEMENTARY DATA

Supplementary data are published online on the website of Arthritis Care & Research

REFERENCES

1. 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.
2. Poddubnyy D, Brandt H, Vahldiek J, et al. The frequency of non-radiographic axial spondyloarthritis in relation to symptom duration in patients referred because of chronic back pain: results from the Berlin early spondyloarthritis clinic. Ann Rheum Dis. 2012 Dec; 71(12):1998-2001.
3. 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.
4. 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.
5. Baraliakos X, Listing J, Rudwaleit M, et al. The relationship between inflammation and new bone formation in patients with ankylosing spondylitis. Arthritis Res Ther. 2008; 10(5):R104.
6. 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.

7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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 Aug 17.
20. 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.
21. 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 Jul 23.
22. Dougados M, Etcheto A, Molto A, et al. Clinical presentation of patients suffering from recent onset chronic inflammatory back pain suggestive of spondyloarthritis: The DESIR cohort. Joint Bone Spine. 2015 Oct; 82(5):345-351.
23. de Hooge M, Pialat JB, Reijnierse M, et al. Assessment of typical SpA lesions on MRI of the spine: do local readers and central readers agree in the DESIR-cohort at baseline? Clin Rheumatol. 2017 Jul; 36(7):1551-1559.
24. Jacquemin C, Rubio Vargas R, van den Berg R, et al. What is the reliability of non-trained investigators in recognising structural MRI lesions of sacroiliac joints in patients with recent inflammatory back pain? Results of the DESIR cohort. RMD Open. 2016; 2(2):e000303.
25. 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.
26. Bennett AN, Rehman A, Hensor EM, et al. The fatty Romanus lesion: a non-inflammatory spinal MRI lesion specific for axial spondyloarthropathy. Ann Rheum Dis. 2010 May; 69(5):891-894.
27. 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.
28. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for

assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Oct 15; 53(5):703-709.

29. 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.
30. Hermann KG, Baraliakos X, van der Heijde DM, et al. Descriptions of spinal MRI lesions and definition of a positive MRI of the spine in axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI study group. Ann Rheum Dis. 2012 Aug; 71(8):1278-1288.
31. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Aug 15; 53(4):502-509.
32. Madari Q, Sepriano A, Ramiro S, et al. 5-year follow-up of spinal and sacroiliac MRI abnormalities in early axial spondyloarthritis: data from the DESIR cohort. RMD Open. 2020 Feb; 6(1).
33. Sepriano A, Ramiro S, Landewé R, et al. Percentage of progressors in imaging: can we ignore regressors? RMD Open. 2019; 5(1):e000848.
34. 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.
35. Wanders AJ, Landewé RB, Spoorenberg A, et al. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the available methods based on the Outcome Measures in Rheumatology Clinical Trials filter. Arthritis Rheum. 2004 Aug; 50(8):2622-2632.
36. Van Calster B, Steyerberg EW, Collins GS, et al. Consequences of relying on statistical significance: Some illustrations. Eur J Clin Invest. 2018 May; 48(5):e12912.
37. 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.
38. Maksymowych WP. The role of imaging in the diagnosis and management of axial spondyloarthritis. Nat Rev Rheumatol. 2019 Oct 7.
39. Ramiro S, van der Heijde D, Sepriano A, et al. Spinal radiographic progression in early axial Spondyloarthritis: Five-year results from the DESIR cohort. Arthritis Care Res (Hoboken). 2018 Oct 24.

Chapter 10

Integrated longitudinal analysis does not compromise precision and reduces bias in the study of imaging outcomes: A comparative 5-year analysis in the DESIR cohort

Alexandre Sepriano, Sofia Ramiro, Désirée van der Heijde, Maxime Dougados, Pascal Claudepierre, Antoine Feydy, Monique Reijnierse, Damien Loeuille, Robert Landewé

Semin Arthritis Rheum. 2020 Mar 3. pii: S0049-0172(20)30070-6

ABSTRACT

Objective: To assess if an integrated longitudinal analysis using all available imaging data affects the precision of estimates of change in patients with axial spondyloarthritis (axSpA), with completers analysis as reference standard.

Methods: Patients from the DESIR cohort fulfilling the ASAS axSpA criteria were included. Radiographs and MRIs of the sacroiliac joints and spine were obtained at baseline, 1, 2 and 5 years. Each image was scored by 2 or 3 readers in 3 'reading-waves' (or campaigns). Each outcome was analysed: i. According to a 'combination algorithm' (e.g. '2 out of 3' for binary scores); and ii. Per reader. Change over time was analysed with generalised estimating equations by 3 approaches: (a)'integrated-analysis' (all patients with ≥1 score from ≥1 reader from all waves); (b1)Completers-only analysis (patients with 5-year follow-up, using scores from individual readers); (b2)Completers analysis using a 'combination algorithm' (as (b1) but with combined scores). Approaches (b1) and (b2) were considered the 'reference'.

Results: In total, 413 patients were included. The 'integrated analysis' was more inclusive with similar levels of precision of the change estimates as compared to both completers analyses. In fact, for low-incident outcomes (e.g. % mNY-positive over 5-years), an increased incidence was 'captured', with more precision, by the 'integrated analysis' compared to the completers analysis with combined scores (% change/year (95%CI): 1.1 (0.7; 1.5) vs 1.2 (0.5; 1.8), respectively).

Conclusion: An efficient and entirely assumption-free 'integrated analysis' does not jeopardise precision of the estimates of change in imaging parameters and may yield increased statistical power for detecting changes with low incidence.

INTRODUCTION

Axial spondyloarthritis (axSpA) is a chronic inflammatory rheumatic disease that primarily affects the axial skeleton. Patients with axSpA show, in different degrees, inflammatory and structural (osteoproliferative and/or osteodestructive) changes in the sacroiliac joints (SIJs) and spine. However, the complex relationship between these abnormalities, including their sequence, frequency and rate of change over time, is not yet well known.[1]

Axial pathological lesions in axSpA can be detected and quantified by the available imaging techniques, including both inflammatory (magnetic resonance imaging; MRI) and structural changes (both radiographs and MRI), and several scores have been developed for this purpose.[2-5] The role of imaging to assess axial inflammatory activity and structural damage over time in axSpA has been assessed in previous studies, but these are few,[6-8] rendering the appropriate use of imaging in the monitoring of axSpA yet to be defined.[9]

To clarify this role, long-term data is needed. However, collection and analysis of such data pose some methodological challenges, including loss to follow-up that often jeopardises the interpretation of findings. The Interpretation may further be challenged by the fact that different readers may have contributed to obtaining scores, in multiple 'reading-waves'. A common approach is to choose a convenient read wave, to only evaluate patients with complete follow-up (completers analysis) and to aggregate scores of individual readers into some algorithm (e.g. agreement ≥ 2 out of 3 readers). Such approaches are not assumption-free, may cause non-random data loss (bias by study completion), and may as such yield biased estimates and loss of external validity.

An alternative method has been previously proposed to analyse long-term imaging data in patients with rheumatoid arthritis (RA) using all available information provided by all readers in different 'reading-waves' in an assumption-free manner (a so called 'integrated analysis').[10] Our aim was to investigate if the use of the 'integrated analysis' affects the precision of estimates for imaging outcomes in patients with axSpA, with a conventional completers analysis as reference standard.

METHODS

Patients and study design

Five-year follow-up data of patients with inflammatory back pain (≥ 3 months but <3 years), and with symptoms suggestive of axSpA according to the treating rheumatologist from the DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR) cohort (clinicaltrials.gov ID: NCT01648907) were used.[11] In addition, patients had to fulfil the Assessment of SpondyloArthritis international Society (ASAS) axSpA criteria and to have at least one radiograph and/or MRI reading available during the 5-year follow-up. The database used for the current analysis was locked on 20th of June 2016.

The study was conducted according to Good-Clinical-Practice-guidelines and was approved by the appropriate local medical ethical committees. Written informed consent was obtained from participating patients before inclusion.

Imaging scoring procedures

Radiographs and MRIs of the SIJ (X-SIJ; MRI-SIJ) and spine (X-Spine; MRI-Spine) were obtained at baseline, 1, 2 and 5 years. Radiographs were performed in all centers (N=25) and in all time-points. MRIs were performed at baseline in all centers and, by protocol, follow-up MRIs were only performed in centers in Paris (N=9). Each image was independently scored, in 3 separate 'reading-waves' (or campaigns) by trained central readers, blinded to clinical data and to the results of other imaging modalities and without known chronology. In wave 1, baseline images were scored by 2 readers and 1 adjudicator (in case of disagreement). In wave 2, images from baseline, 1 and 2 years were also scored by 2 readers and one adjudicator. In wave 3, images from baseline, 2 and 5 years were scored by 3 central readers. The readers and adjudicators varied across modalities and waves (Online Supplementary Table S1).

SIJ imaging outcomes

Inflammation on MRI-SIJ was assessed according to the ASAS definition (positive/negative) and by the Spondyloarthritis Research Consortium of Canada (SPARCC) score (range: 0-72).[2, 3, 12] The adapted SPARCC MRI-SIJ Structural score by Webers *et al* was used to define individual structural lesions on MRI-SIJ (fatty lesions, erosions, sclerosis, partial ankylosis and total ankylosis).[13] In the absence of a formal definition of a positive structural MRI-SIJ, we considered three definitions that have been shown to be the most discriminatory in early axSpA: ≥5 fatty lesions and/or erosions; ≥3 erosions; and ≥3 fatty lesions.[14] Continuous structural lesions on MRI-SIJ were defined as number of fatty lesions and/or erosions (range: 0-80), number of erosions (range: 0-40), number of fatty lesions (range: 0-40) and total number of lesions (range: 0-144). Structural lesions on X-SIJ were assessed according to the mNY-grading method as a continuous variable (range: 0-8) and as mNY positive/negative.[15] Two binary definitions of X-SIJ structural damage were also assessed: worsening of ≥ 1 grade in ≥ 1 SIJ (yes/no); and worsening of ≥ 1 grade in ≥1 SIJ, with grade ≥ 2 in the worsened joint at 5 years (yes/no).[16]

Spine imaging outcomes

Bone marrow edema (BME) on MRI-Spine was defined according to the ASAS definition (≥3 corner lesions; yes/no).[17] In addition, a cut-off of 5 lesions was also assessed, as it has been shown to be highly specific of axSpA.[14] The spine SPARCC score (range: 0-414) and spine Berlin score (range: 0-69) were used as continuous inflammatory outcomes.[4, 18] Structural lesions on MRI-Spine were scored according to the Canada–Denmark (CANDEN) method.[5] As for MRI-SIJ, in the absence of a formal definition, we defined structural damage as ≥5 fatty lesions, since this cut-off has been shown to be highly specific for axSpA.[14] The total number of structural lesions (fatty lesions, erosions, bone spurs, ankylosis; range: 0-322) was also assessed. Structural lesions on X-Spine were assessed as the presence of ≥1 syndesmophyte (yes/no) and by the modified Stoke Ankylosing Spondylitis Spine Score (mSASSS).[19]

Statistical analysis

Each outcome was analysed by generalised estimating equations (GEE) models with an exchangeable 'working' correlation structure, taking into account the repeated scores over time. The parameter estimate for 'time', as the main variable of interest in the models, can be interpreted as the absolute change of the score per year for continuous outcomes; and as the change per year in the percentage of positive cases for binary outcomes. Each outcome was analysed per patient and per time-point in two ways: i. according to a 'combination algorithm'; and ii. per individual reader. For the algorithm, the combined score for binary (yes/no) outcomes in waves 1 and 2 resulted from the agreement of 2 readers and, in case of disagreement, involves the adjudicator score. Binary outcomes in wave 3 were scored by the agreement of ≥ 2 out of 3 readers. The combined scores for continuous outcomes were defined as the mean of the available scores.

The change per year was estimated with three analytical-methods: (a) 'integrated-analysis', including all patients with ≥1 available score from ≥1 reader from all 'reading-waves' (reader and the wave added to the models to adjust for higher levels of correlation); (b1) completers only analysis, including only patients with complete 5-year follow-up, using scores from individual readers from wave 3 (adjusted for reader); and (b2) aggregated completers analysis, using a combination algorithm (as (b1) but with combined scores, thus without reader adjustment). Both completers analysis (b1 and b2) were used as the 'reference' against which the 'integrated analysis' was compared.

Goodness-of-fit statistics (quasi-likelihood under the independence model criterion; QIC), were used to get an impression on how much of the outcome variability is explained by each model. Different transformations of time were tested to assess which yielded the lowest QIC (better fit). A non-linear model was chosen if best fitting the data, and if the non-linear factor (e.g. quadratic term) added to the model was statistically significant ($p<0.05$).

RESULTS

Change of inflammatory and structural lesions over time

In total, 413 patients were included and 366 completed the 5-year follow-up. The mean (SD) symptom duration was 1.6 (0.9) years; 52% were males and 89% HLA-B27% positive (Online Supplementary Table S2).

The estimated change over time of the SIJ imaging outcomes, with the 'integrated analysis' is shown in Fig. 1 (spine outcomes: Online Supplementary Fig. S1). Inflammation on MRI-SIJ was detected in a large proportion of patients at baseline [estimated % (95%CI): 43 (38; 47)] and significantly decreased over time, especially during the first 2 years, i.e. following a quadratic distribution (QIC linear model: 8726; QIC quadratic model: 8710; quadratic term p-value: 0.028). On the contrary, structural damage on MRI-SIJ and X-SIJ significantly increased over time. For instance, we found an increase of 1.1% per year in the percentage of patients being mNY-positive over a time span of 5 years. In general, spine abnormalities were scarce at baseline and remained low over time.

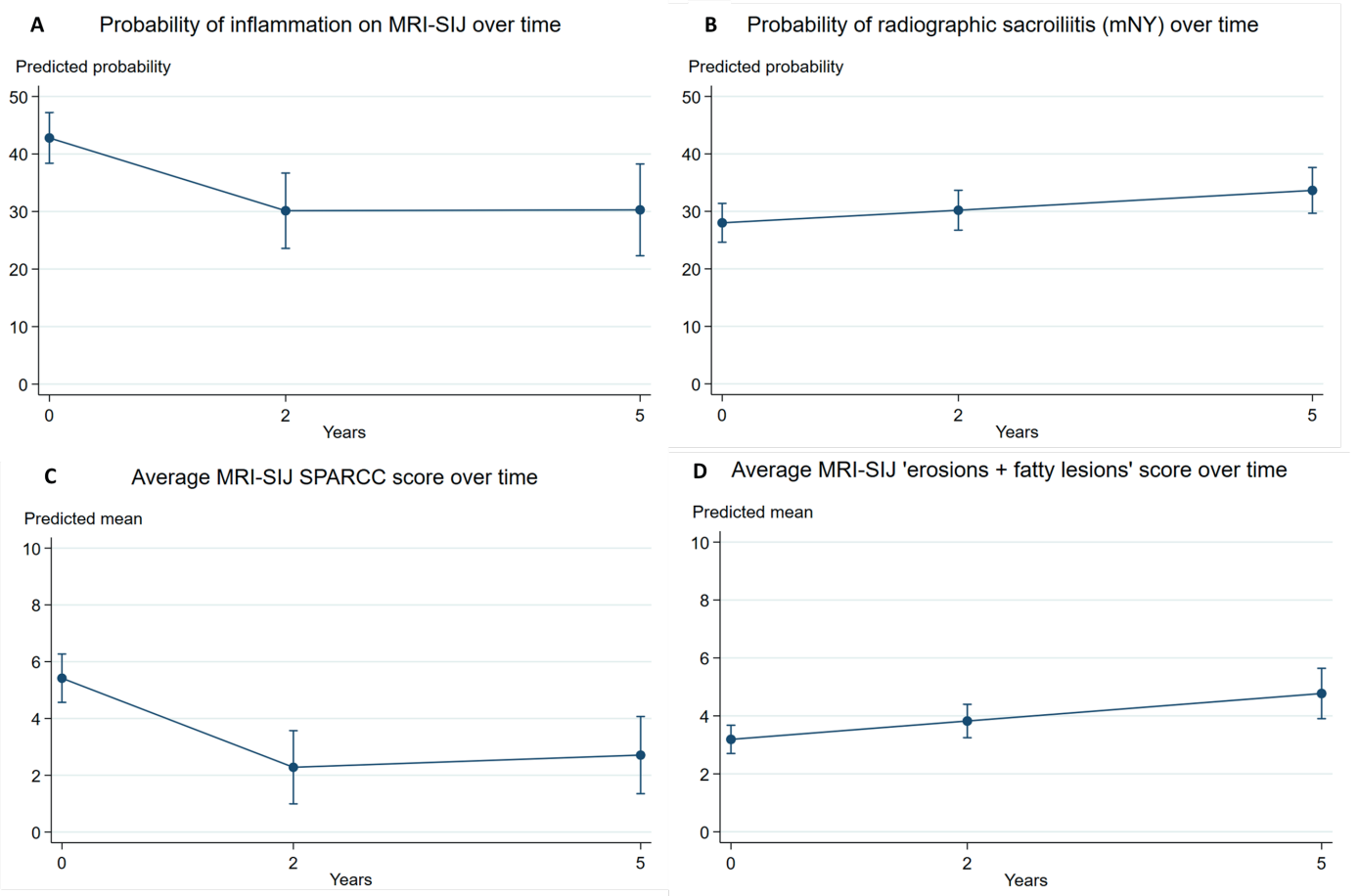


Figure 1. Estimated change of sacroiliac joints outcomes over 5 years ('Integrated analysis'). Point estimates: probability of each binary outcome or mean of each continuous outcome in each time point. Error bars: 95% confidence intervals. **Panel A**: There is an average yearly 7.4% (95% CI: -11.7; -3.1) reduction in the probability of having BME on MRI-SIJ (ASAS definition), following a quadratic distribution (quadratic term: OR=1.06 (95% CI 1.01; 1.11); p-value=0.028), which means that every year there is a decrease of 0.06 odds in the rate of decrease of BME positivity. **Panel B:** There is an average yearly 1.1% (95% CI: 0.7; 1.5) increase in the probability of radiographic sacroiliitis (mNY definition), following a linear distribution (quadratic term not significant). **Panel C:** there is an average yearly decrease of 1.7 (95% CI: -2.6; -0.9) units in the SPARCC score over 5 years, following a quadratic distribution (quadratic term: β=0.34 (95% CI 0.14; 0.55); p-value=0.001), which means that every year the rate of decrease reduces by 0.34 units. **Panel D:** There is an average yearly 0.32 (95% CI: 0.2; 0.5) units increase in the total number of erosions and fatty lesions over time, following a linear distribution (quadratic term not significant). MRI, magnetic resonance imaging; SIJ, sacroiliac joints; BME, bone marrow edema; mNY, modified New York criteria; SPARCC, spondyloarthritis research consortium of Canada; CI, confidence interval.

Comparison of different analytical methods to capture change

The estimated change over time for binary and continuous imaging outcomes by the three analytical approaches is shown in Tables 1 and 2, respectively. The 'integrated analysis' (method a) was more inclusive compared to the completers analysis with individual readers' scores (method b1) and completers analysis with combined scores (method b2), both for binary ((a): N=360-411 vs (b1 and b2): N=313-364) and continuous outcomes ((a): N=399-411 vs (b1): 342-364 and (b2): 338-364).

The decrease of MRI-SIJ detected inflammation was captured by all analytical methods with similar precision both for the binary ASAS definition of a positive MRI-SIJ and the continuous SPARCC score (negative coefficients with similar 95%CI excluding zero). Similar findings were also seen for MRI-SIJ structural changes, but in the opposite direction (positive coefficients with similar 95%CI excluding zero). Of note, the subtle increase in binary X-SIJ structural lesions was detected with more precision by the 'integrated analysis' as compared to both completers analysis [e.g. worsening of ≥ 1 grade in ≥1 SIJ with a grade ≥2 in the worsened joint at 5 years: (a): 1.76 (1.06; 2.46) vs (b1): 1.55 (0.78; 2.32) and (b2): 2.05 (0.81; 3.28), respectively].

All analytical methods were unable to detect a significant change for both inflammatory and structural lesions in the spine, except for the formation of new syndesmophytes, captured with similar precision by the three approaches (% change/year (95% CI): (a): 0.84 (0.46; 1.22) vs (b1): 0.48 (0.16; 0.80) vs (b2): 0.50 (0.10; 0.91)).

DISCUSSION

In this 5-year longitudinal study in patients with early axSpA, we tested a new approach to analyse imaging outcomes over time as compared to the 'traditional' completers analysis. We have shown that, by applying the 'integrated analysis', we can efficiently use all available data in an entirely assumption-free manner without compromising precision, and it may even yield increased statistical power for detecting low incident abnormalities. In addition, the 'integrated analysis' may, to some extent, protect against attrition bias and avoid bias by 'convenient choices'.

A previous post-hoc analysis of two randomised trials in patients with RA has also shown the robustness of the 'integrated analysis' as compared to a completers analysis.[10] Here we report, for the first time, the application of this innovative analytical method to observational data and in patients with early axSpA. We 'challenged' this technique with several imaging scores and have shown that the precision of the estimates of change was similar to the one obtained by the completers analysis, or even better: in case of outcomes with a low incidence.

Table 1. Change per year in the percentage of positive cases for binary imaging outcomes over 5-years of follow-up, according to 3 different analytical methods, in early axSpA patients fulfilling the ASAS axSpA criteria

	Integrated analysis (a)*	**Completers analysis with individual readers scores** (b1)†	**Completers analysis with combined scores for readers** (b2)‡
Imaging outcomes	**% change per year (95% CI) (N=360-411)**	**% change per year (95% CI) (N=313-364)**	**% change per year (95% CI) (N=313-364)**
SACROILIAC JOINTS			
Inflammatory lesions (MRI-SIJ)			
Sacroiliitis (ASAS criteria)[2]	**-7.35 (-11.65; -3.05)£**	**-5.40 (-8.87; -1.92)£**	**-3.13 (-5.09; -1.18)**
Structural lesions (MRI-SIJ)[13]			
≥ 5 fatty lesion and / or erosions	**4.41 (2.30; 6.53)£**	**3.17 (1.49; 4.85)£**	**2.12 (0.97; 3.27)**
≥ 3 erosions	0.25 (-0.67; 1.17)	0.28 (-0.58; 1.13)	0.10 (-1.30; 1.49)
≥ 3 fatty lesions	**4.68 (2.68; 6.67)£**	**3.30 (1.73; 4.86)£**	**2.03 (1.02; 3.04)**
Structural lesions (X-SIJ)			
mNY dichotomous	**1.10 (0.67; 1.53)**	**0.87 (0.48; 1.26)**	**1.18 (0.54; 1.81)**
mNY 1-grade change[16]	**2.18 (1.40; 2.96)**	**2.03 (1.16; 2.89)**	**2.30 (0.88; 3.71)**
mNY 1-grade change and value ≥ 2[16]	**1.76 (1.06; 2.46)**	**1.55 (0.78; 2.32)**	**2.05 (0.81; 3.28)**
SPINE			
Inflammatory lesions (MRI-Spine)			
BME: ≥ 3 lesions (ASAS criteria)[17]	-0.82 (-2.31; 0.67)	-0.44 (-1.39; 0.51)	0.14 (-0.88; 1.17)
BME: ≥ 5 lesions (ASAS criteria)[14]	-0.72 (-2.20; 0.76)	-0.30 (-1.26; 0.65)	-0.33 (-1.41; 0.76)
Structural lesions (MRI-Spine)			
≥ 5 fatty lesions[14]	-0.22 (-0.85; 0.41)	-0.12 (-0.45; 0.20)	¥
Structural lesions (X-Spine)			
≥ 1 syndesmophyte	**0.84 (0.46; 1.22)**	**0.48 (0.16; 0.80)**	**0.50 (0.10; 0.91)**

*Analysis taking into account the 3 different reading campaigns, i.e. waves, and the different readers from all waves; 3-level generalised estimating equations (GEE) models, taking into account the within-patient correlation for the repeated measures and adjusting for the reader and wave; † Data from one reading wave only (wave 3) and taking the different readers (n=3 per modality) into account; 2-level GEE, taking into account the within-patient correlation for the repeated measures and adjusting for the reader; ‡ Data from one reading wave only (wave 3) and using combined scores calculated from the individual readers (n=3) scores; 1-level GEE, taking into account the within-patient correlation for the repeated measures of the combined scores (i.e. '2 out of 3'); £Quadratic transformation; ¥ No convergence achieved: only 5 events during follow-up.

axSpA, axial spondyloarthritis; MRI-SIJ, magnetic resonance imaging of the sacroiliac joints; X-SIJ, radiograph of the sacroiliac joints; ASAS, Assessment of SpondyloArthritis international Society; mNY, radiographic sacroiliitis according to the modified New York criteria; MRI-spine, MRI of the spine; X-spine, radiograph of the spine; SPARCC, Spondyloarthritis Research Consortium of Canada score; mSASSS: modified Stoke Ankylosing Spondylitis Spine Score; GEE: generalised estimating equations.

Table 2. Yearly progression rate of continuous imaging outcomes over 5-years of follow-up, according to 3 different analytical methods, in early axSpA patients from the DESIR-cohort who fulfil the ASAS axSpA classification criteria

	Integrated analysis (a)*	**Completers analysis with individual readers scores**(b1)	**Completers analysis with combined scores for readers** (b2)
Imaging outcomes	**units change per year (95% CI) (N=399-411)**	**units change per year (95% CI) (N=342-364)**	**units change per year (95% CI) (N=338-364)**
SACROILIAC JOINTS			
Inflammatory lesions (MRI-SIJ)			
SPARCC SIJ score (0-72)[3]	**-1.74 (-2.57; -0.90)£**	**-1.02 (-1.57; -0.46) £**	**-1.03 (-1.60; -0.47) £**
Structural lesions (MRI-SIJ)[13]			
Number of fatty lesions /erosions (0-80)	**0.32 (0.18; 0.45)**	**0.51 (0.28; 0.74) £**	**0.28 (0.16; 0.40)**
Number of erosions (0-40)	0.05 (-0.03; 0.12)	0.04 (-0.02; 0.10)	0.03 (-0.03; 0.10)
Number of fatty lesions (0-40)	**0.27 (0.16; 0.38)**	**0.45 (0.25; 0.65) £**	**0.25 (0.15; 0.35)**
Total structural lesions†† (0-144)	**0.39 (0.24; 0.54)**	**0.37 (0.23; 0.50)**	**0.37 (0.23; 0.50)**
Structural lesions (X-SIJ)			
mNY continuous grade (0-8)	**0.05 (0.03; 0.07)**	**0.04 (0.03; 0.06)**	**0.04 (0.03; 0.06)**
SPINE			
Inflammatory lesions (MRI-Spine)			
SPARCC Spine score (0-414)[4]	-0.21 (-0.54; 0.12)	-0.14 (-0.37; 0.10)	-0.15 (-0.39; 0.10)
Berlin Spine score (0-69)[18]	-0.11 (-0.25; 0.02)	-0.05 (-0.13; 0.03)	-0.05 (-0.14; 0.03)
Structural lesions (MRI-Spine)			
Total structural lesions** (0-322)[20]	0.02 (-0.01; 0.05)	0.03 (-0.0003; 0.06)	0.03 (-0.01; 0.06)
Structural lesions (X-Spine)			
mSASSS score (0-72)	**0.09 (0.04; 0.14)**	**0.07 (0.03; 0.11)**	**0.06 (0.02; 0.10)**

*Analysis taking into account the 3 different reading campaigns, i.e. waves, and the different readers from all waves; 3-level generalised estimating equations (GEE) models, taking into account the within-patient correlation for the repeated measures and adjusting for the reader and wave; † Data from one reading wave only (wave 3) and taking the different readers (n=3 per modality) into account; 2-level GEE, taking into account the within-patient correlation for the repeated measures and adjusting for the reader; ‡ Data from one reading wave only (wave 3) and using combined scores calculated from the individual readers (n=3) scores; 1-level GEE, taking into account the within-patient correlation for the repeated measures of the combined scores (i.e. Mean of 3 readers); £ Quadratic transformation; †† fatty lesions, erosions, sclerosis, partial ankylosis, total ankylosis; ** fatty lesions, erosions, bone spurs, ankylosis;

axSpA, axial spondyloarthritis; MRI-SIJ, magnetic resonance imaging of the sacroiliac joints; X-SIJ, radiograph of the sacroiliac joints; ASAS, Assessment of SpondyloArthritis international Society; mNY, radiographic sacroiliitis according to the modified New York criteria; MRI-spine, MRI of the spine; X-spine, radiograph of the spine; SPARCC, Spondyloarthritis Research Consortium of Canada score; mSASSS: modified Stoke Ankylosing Spondylitis Spine Score; CD score, Canada-Denmark score; GEE: generalised estimating equations.

The largely overlapping precision suggests that both analytical approaches can be applied when analyzing change over time in imaging outcomes. However, our results argue in favour of using the 'integrated analysis' for several reasons. First, with this method, we included all patients with at least one score in at least one time point who would, otherwise, be excluded from a completers analysis. Thus, to some extent, it may deal better with possible bias by attrition – a common problem of long-term cohorts. Second, this technique directly handles data from

different readers and 'reading-waves', with no need for 'combined scores' (e.g. 2 out of 3), which are not without assumptions and prone to bias. The 'trade off' is adding some variability ('noise') to the estimates, which may lead to a lower precision (i.e. wider 95% CI). But that is not what we have found. Arguably, by including all scoring data without 'hidden' assumptions, we may better approximate the 'true' point-estimates (the 'signal'). In fact, despite similar levels of precision, differences in the point-estimates were found between methods. Third, integrated analysis increases statistical power to detect subtle changes, which is of particular interest when assessing structural damage in patients with early disease as shown here. Taken all together, the 'integrated analysis' increases external validity without compromising (or even improving) internal validity.

In addition, the integrated analysis 'increases the sample size without increasing the number of patients. This means: the number of available scores for analysis is not only determined by the number of patients but also by the number visits, the number of readers and the number of 'reading-waves'. Obviously, these multiple observations per patient cannot be interpreted as independent observations. Each time point is clustered within patient, each patient is clustered within reader, and each reader is clustered within the 'reading-wave'. Ignoring the lack of independency between observations would result in an artificially narrow 95% CI. This is why we have applied GEE models, which appropriately deals with correlated data.[21, 22]

In summary, here we describe the 'integrated analysis', a novel and sophisticated analytical method that may be used in future studies focusing on imaging, including those dealing with the assessment of treatment effects on imaging outcomes. This approach may be of special interest in studies with long-term follow-up, and/or when the outcomes are expected to occur infrequently over time.

SUPPLEMENTARY DATA

Supplementary data are published online on the website of Seminars in Arthritis and Rheumatism

REFERENCES

1. Baraliakos X, Braun J. Opinion: Perspectives on imaging in axial spondyloarthritis. Nat Rev Rheumatol. 2013 Aug; 9(8):498-502.
2. 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.
3. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Oct 15; 53(5):703-709.
4. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Aug 15; 53(4):502-509.
5. 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.
6. Navarro-Compan V, Ramiro S, Landewé R, et al. Disease activity is longitudinally related to sacroiliac inflammation on MRI in male patients with axial spondyloarthritis: 2-years of the DESIR cohort. Ann Rheum Dis. 2016 May; 75(5):874-878.

7. 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.
8. Ramiro S, van der Heijde D, Sepriano A, et al. Spinal radiographic progression in early axial Spondyloarthritis: Five-year results from the DESIR cohort. Arthritis Care Res (Hoboken). 2019 Dec;71(12):1678-1684.
9. van der Heijde D, Ramiro S, Landewe R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis. 2017 Jun; 76(6):978-991.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. Hermann KG, Baraliakos X, van der Heijde DM, et al. Descriptions of spinal MRI lesions and definition of a positive MRI of the spine in axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI study group. Ann Rheum Dis. 2012 Aug; 71(8):1278-1288.
18. Lukas C, Braun J, van der Heijde D, et al. Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: a multireader experiment. J Rheumatol. 2007 Apr; 34(4):862-870.
19. 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.
20. 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.
21. Hanley JA, Negassa A, Edwardes MD, et al. Statistical analysis of correlated data using generalised estimating equations: an orientation. Am J Epidemiol. 2003 Feb 15; 157(4):364-375.
22. 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.

Chapter 11

Which imaging outcomes for axSpA are most sensitive to change? A 5-Year analysis of The DESIR Cohort

Alexandre Sepriano, Sofia Ramiro, Désirée van der Heijde, Maxime Dougados, Pascal Claudepierre, Antoine Feydy, Monique Reijnierse, Damien Loeuille, Robert Landewé

Accepted for publication, Arthritis Care & Research

ABSTRACT

Objective: To compare the sensitivity to change of different imaging scoring methods in patients with early axial spondyloarthritis (axSpA).

Methods: Patients from the DESIR cohort fulfilling the ASAS axSpA criteria were included. Radiographs and MRI of the sacroiliac joints (SIJ) and spine were obtained at baseline, 1, 2 and 5 years. Each image was scored by 2 or 3 readers in 3 separate 'reading-waves'. The rate of change of outcomes measuring spinal and SIJ inflammation (e.g. SPARCC score) and structural damage on MRI (e.g. ≥3 fatty lesions) and radiographs (e.g. mNY grading) was assessed using multilevel generalized estimating equations (GEE) models (taking all readers and waves into account). To allow comparisons across outcomes, rates were standardized (difference between the individual's value and the population mean divided by the standard deviation).

Results: In total, 345 patients were included. Inflammation on MRI-SIJ (standardized rate range: -0.278; -0.441) was more sensitive to change compared to spinal inflammation (range: -0.030; -0.055). Structural damage in the SIJ showed a higher standardized rate of change on MRI-SIJ (range: 0.015-0.274) compared to X-SIJ (range: 0.043-0.126). MRI-SIJ damage defined by ≥3 fatty lesions showed the highest sensitivity to change (0.274). Spinal structural damage slowly progressed over time with no meaningful difference between radiographic (range: 0.037-0.043) and MRI structural outcomes (range: 0.008-0.027).

Conclusion: Structural damage assessed in pelvic radiographs has low sensitivity to change, while fatty lesions detected on MRI-SIJ are a promising alternative. In contrast, MRI-spine is not better than X-spine in detecting structural changes in early axSpA patients.

INTRODUCTION

Several imaging outcomes have been developed to assess inflammation and structural damage over time in patients with axial spondyloarthris (axSpA). A recent systematic literature review (SLR) informing the EULAR recommendations for the use of imaging in the diagnosis and management of SpA in clinical practice identified several studies testing the utility of magnetic resonance imaging (MRI) and radiographs of the sacroiliac joints (SIJ) and spine on monitoring disease activity and structural damage over time.[1] However, these studies mostly assessed only one score each, and focused on comparing imaging to clinical measures of disease activity, disability and mobility, which means they mostly addressed their validity.

In addition to validity, in order to prioritize imaging outcomes measuring similar aspects of the disease (i.e. inflammation or structural damage), the other aspects of the Outcome Measures in Rheumatology (OMERACT) filter, namely discrimination (sensitivity to change and reliability) and feasibility should also be taken into account.[2] However, direct comparisons of the discriminative ability and feasibility of imaging outcomes in axSpA have been seldom performed, and almost only in later phases of the disease (radiographic axSpA; r-axSpA).[3-5] An exception to this, is the comparison of the different spinal radiographic scoring methods performed in the DESIR cohort and previously reported by us.[6]

A better understanding on which imaging findings (reflecting inflammation or structural damage), imaging modality (MRI or radiographs) and anatomical location (SIJ or spine) are most informative to monitor axial changes in the entire spectrum of axSpA (also including non-radiographic axSpA; nr-axSpA) over time is still a major unmet need. We aimed to compare the sensitivity to change of different MRI and radiographic scoring methods in patients with early axSpA.

METHODS

Patients and study design

Five-year data from patients with early axSpA from the DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR) cohort have been used (clinicaltrials.gov ID: NCT01648907).[7] Patients had to fulfill the Assessment of SpondyloArthritis international Society (ASAS) axSpA criteria and to have ≥1 radiograph and/or MRI reading available during the 5-year follow-up to be included in the current study. The database used for the current analysis was locked on 20th of June 2016. The study was approved by the appropriate local medical ethical committees. All patients signed the informed consent upon participation.

Imaging scoring procedures

Radiographs and MRIs of the SIJ (X-SIJ; MRI-SIJ) and spine (X-Spine; MRI-Spine) were obtained at baseline, 1, 2 and 5 years. Each image was independently scored, in 3 ‘reading-waves’ by trained central readers, blinded to chronology, clinical data and to the results of other imaging modalities. In wave 1 baseline images were scored by two readers and one adjudicator (in case of disagreement). In wave 2, images from baseline, 1 and 2 years were also scored by 2 readers and one adjudicator. In wave 3, images from baseline, 2 and 5 years were scored by 3 central

readers. Readers and adjudicators varied across modalities and waves (Online Supplementary Table S1).[8] By protocol, radiographs have been performed in all 25 participating centers at each time point, but MRIs were only performed in all centers at baseline, while MRIs at 1, 2 and 5 years were only obtained in 9 centers from Paris.

Inflammation outcomes

Inflammation on MRI-SIJ was assessed using the ASAS definition (positive/negative) and the Spondyloarthritis Research Consortium of Canada (SPARCC) score (range: 0-72).[9-11]

Bone marrow edema (BME) on MRI-Spine was defined according to the ASAS definition (≥3 vertebral corner lesions; yes/no).[12] In addition, a cut-off of 5 vertebral corner BME lesions (typical of axSpA and present in ≥2 consecutive slices) was also assessed, according to the Canada-Denmark method, as it has been shown to be highly specific of axSpA.[13] The total spine SPARCC (range: 0-414) and Berlin (range: 0-69) scores were used as continuous inflammatory outcomes.[3, 14]

Structural outcomes

Structural damage on X-SIJ was assessed according to the modified New York (mNY) system as continuous (range: 0-8) and as a binary (positive / negative) score.[15] Two additional binary definitions were assessed: worsening of ≥1 grade in ≥1 SIJ (yes/no); and worsening of ≥1 grade in ≥1 SIJ, with a 5-year grade ≥2 in the worsened joint (yes/no).[16]

An adaptation of the MRI-SIJ Structural score by Weber *et al*, previously described by us,[17] was used to define individual structural lesions on MRI-SIJ.[18] In summary, fatty lesions, erosions and ankylosis/partial ankylosis are scored as originally described. Sclerosis was added. Fatty lesions, erosions and sclerosis were marked as present if seen on ≥2 consecutive slices (maximum 5 lesions in 6 slices per each of the 8 quadrants in both SIJs). Ankylosis or partial ankylosis was considered present if seen on a single slice. Partial ankylosis and ankylosis cannot occur simultaneously in a quadrant, and ankylosis always involves two quadrants; therefore, the corresponding scoring range is 0–24.In the absence of a formal definition of presence of structural damage on MRI-SIJ, we considered 3 definitions previously shown most discriminatory in early axSpA: ≥5 fatty lesions and/or erosions; ≥3 erosions; and ≥ 3 fatty lesions.[13] Continuous structural lesions on MRI-SIJ were defined as number of fatty lesions and/or erosions (range: 0-80), number of erosions (range: 0-40), number of fatty lesions (range: 0-40) and total number of lesions with (range: 0-144) and without (range: 0-104) sclerosis.

Structural lesions on X-Spine were assessed as the presence of ≥1 syndesmophyte (yes/no), and by the modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS; range: 0-72).[19]

Structural lesions on MRI-Spine were scored according to the Canada–Denmark (CANDEN) method.[20, 21] In the absence of a formal definition, we define structural damage as ≥5 fatty lesions, also previously shown to be highly specific for axSpA.[13] The total number of structural lesions (fatty lesions, erosions, bone spurs, ankylosis; range: 0-322) was assessed, as well as the total number of fatty lesions, erosions and bone spurs (range: 0-92; for all).

A detailed description of all scores is provided in Online Supplementary Tables S2-S10. The interreader reliability of the radiographic and MRI outcomes used in this study has been reported in detail elsewhere and is summarized in Online Supplementary Text S1.[6, 17]

Statistical analysis

The baseline value for each outcome was defined by a 'combination algorithm' of the scores from the 3 readers from wave 3 (agreement between ≥2 out of 3 for binary, and mean of 3 readers for continuous outcomes).

The rate of change of each outcome was analyzed by generalized estimating equations (GEE), with 'time' in years as the explanatory variable of interest. Each outcome was analyzed per patient, per time-point and per individual reader and the yearly rate of change estimated using the so-called 'integrated-analysis', including all patients with ≥1 score from ≥1 reader from ≥1 'reading-wave'. Different to traditional measures of sensitivity to change (e.g. Cohen's effect size), this method, which we have previously explained in detail,[8] appropriately handles the multilevel data structure of our data. All patients had to have ≥1 score from all outcomes, thus ensuring that the same patients are used across all analyses. All variables were standardized. A standardized variable (metric-free) was defined at the patient level as: difference between the individual's value and the population mean divided by the population standard deviation (SD). Each standardized variable has a mean of 0 and a variance of 1 and reads as the number of SD above (positive) or below (negative) the mean.

In addition, the relative standardized rate of change (i.e. the standardized yearly rate of change of an outcome divided by the corresponding rate of a reference imaging outcome) was calculated. For this calculation, a value > 1 means larger sensitivity and a value <1 lower sensitivity compared to the reference (the further away from 1 the larger the difference). Three types of references were defined: i. 'Inflammation common reference': comparing all inflammation outcomes to sacroiliitis on MRI-SIJ (ASAS definition); ii. 'Structural common reference': comparing all structural outcomes to sacroiliitis on X-SIJ (mNY); and iii. 'Modality reference': comparing outcomes to a reference within each modality and anatomical site.

Goodness-of-fit statistics (quasi-likelihood under the independence model criterion; QIC), were used to get an impression on how much of the outcome variability is explained by each model. Different transformations of time were tested to assess which yielded the lowest QIC (better fit). A non-linear model was chosen if best fitting the data, and if the non-linear factor (e.g. quadratic term) added to the model was significant ($p<0.05$). Stata V15.1 was used for the analyses.

RESULTS

Baseline characteristics

In total, 345 patients were included [mean (SD) symptom duration: 1.6 (0.9) years; 53% were males and 89% HLA-B27 positive; Table 1]. Baseline inflammation on MRI was more frequently present at the SIJ (active sacroiliitis: 39%) than at the spine level (BME ≥5 lesions: 6%) (Table 2).

Structural damage at baseline was limited in the SIJ (21% mNY positive) and even more in the spine (≥1 syndesmophyte: 6%) (Table 3).

Sensitivity to change of the different imaging outcomes

Inflammation on MRI-SIJ showed a higher sensitivity to change than on MRI-spine, the latter remaining essentially unchanged over time. This was true for the dichotomous ASAS MRI-SIJ score (standardized yearly rate of change -0.278) and especially for the continuous SPARCC score (standardized yearly rate of change -0.441), while the standardized yearly rates of change for MRI-spine ranged only between -0.030 and -0.055 (Table 2). The differences between SIJ and spine inflammation outcomes become especially evident with the relative standardized rate of change. Compared to the ASAS definition of a positive MRI-SIJ ('inflammation common reference'; i.e. value of 1) all inflammation outcomes in the spine were much less sensitive to change (range of relative standardized rates: 0.094; 0.531; i.e. all values far below 1).

Table 1. Patient- and disease-characteristics at baseline and during follow-up

	Baseline (N=345)	1 year (N=345)	2 years (N=342)	5 years (N=320)
Age at baseline (years), mean (SD)	31.0 (7.0)			
Male gender, n (%)	183 (53)			
Symptoms duration (years), mean (SD)	1.6 (0.9)			
Current smokers*, n (%)	135 (39)	127 (39)	118 (37)	92 (34)
HLA-B27, n (%)	307 (89)			
Radiographic sacroiliitis (mNY)**	73 (21)	NA	68 (23)	68 (27)
BASDAI*, mean (SD) (0-10)	4.1 (2.0)	3.2 (2.2)	3.1 (2.2)	2.9 (2.0)
ASDAS-CRP**, mean (SD)	2.6 (1.0)	2.1 (0.9)	2.0 (0.9)	2.0 (0.9)
Elevated CRP** (≥6 mg/L), n (%)	109 (33)	64 (20)	69 (22)	57 (22)
BASFI*, mean (SD) (0-10)	2.7 (2.2)	2.1 (2.1)	2.1 (2.2)	2.0 (2.0)
TNFi treatment**, n (%)	0 (0)	76 (24)	94 (29)	111 (42)
NSAID treatment*, n (%)	329 (95)	250 (77)	216 (68)	180 (66)

*Missing data <15% in each visit; ** Missing data <20% in each visit. BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; ASDAS, Ankylosing Spondylitis Disease Activity Score; CRP, C-reactive protein; BASFI, Bath Ankylosing Spondylitis Functional Index; TNFi, tumor necrosis factor inhibitors; NSAID, non-steroidal anti-inflammatory drugs; mNY, modified New York criteria (scored in wave 3); NA, not applicable (imaging in wave 3 is only scored at baseline, 2 and 5 years)

Structural damage in the SIJ increased over time but with a larger yearly rate on MRI-SIJ (standardized rate range: 0.015-0.274) compared to X-SIJ (standardized rate range: 0.043-0.126) (Table 3). Three or more fatty lesions on MRI-SIJ was the SIJ structural outcome with highest sensitivity to change (standardized rate: 0.274; relative rate of 6.227 comparing to mNY). On the contrary, ≥3 erosions on MRI-SIJ was the least sensitive (standardized rate: 0.015) of all SIJ structural outcomes (including both MRI-SIJ and X-SIJ). Importantly, ≥3 fatty lesions alone was

slightly more sensitive to change than combining fatty lesions with erosions, i.e. ≥ 5 fatty lesion and/or erosions (relative rate of 1.151 for the former compared to the latter).

Amongst the X-SIJ structural outcomes, worsening of ≥1 grade in ≥1 SIJ and worsening of ≥1 grade in ≥1 SIJ, with a 5-year grade ≥2 in the worsened joint were far more sensitive to change compared to the mNY binary definition as the 'modality reference' (relative rate: 2.864 and 2.705, respectively). Of note, the mNY continuous grading and the mNY binary score had comparable sensitivity to change (relative rate of the continuous vs the reference binary score = 0.977).

Overall, the standardized yearly rate of change of the spinal radiographic outcomes (range: 0.037-0.043) was higher as compared to MRI-Spine structural outcomes (range: 0.012-0.027) (Table 3), although all are relatively low. Amongst MRI-Spine outcomes, the total number of bone spurs was the outcome that most captured change (standardized rate: 0.027; and relative rate of 2.077 compared to ≥ 5 fatty lesions – the 'modality reference'). Yet, the best MRI-Spine outcome is still less sensitive to change as compared to X-spine outcomes, with a standardized rate of 0.037 for ≥ 1 syndesmophyte and of 0.043 for the continuous mSASSS.

Table 2. Baseline score and standardized yearly rate of change of inflammatory imaging outcomes over 5 years of follow-up in early axSpA patients fulfilling the ASAS axSpA classification criteria

Imaging outcomes	Baseline score* (N=334-344)	Standardized rate of change/year¥	Relative sRoC (Common Reference: ASAS MRI-SIJ)	Relative sRoC per modality & anatomical site
Inflammatory lesions (MRI-SIJ)[9-11]				
Sacroiliitis (ASAS criteria)	134 (39.2%)	-0.278£	1	1
SPARCC SIJ score (0-72)	4.7 (7.9)	-0.441£	1.586	1.586
Inflammatory lesions (MRI-Spine)[3,12-14]				
BME: ≥ 3 lesions	32 (9.4%)	-0.032	0.319	1
BME: ≥ 5 lesions	19 (5.6%)	-0.030	0.094	0.938
23 DVU SPARCC Spine score (0-414)	2.6 (7.7)	-0.050	0.531	1.563
Berlin Spine score (0-69)	0.9 (2.7)	-0.055	0.104	1.719

* Agreement of ≥2 out of 3 readers for binary variables and mean (SD) of 3 readers for continuous variables from wave 3; ¥ Estimated from a model where all independent variables (time, reader and wave) and the outcome are standardized; £ Quadratic transformation led to a better model goodness of fit (QIC: quasi-likelihood under the independence model criterion); ASAS, Assessment of SpondyloArthritis international Society; BME, bone marrow edema; MRI, magnetic resonance imaging; SIJ, sacroiliac joints; SPARCC, spondyloarthritis research consortium of Canada; DVU, discovertebral unit; sRoC, standardized rate of change.

Table 3. Baseline score and standardized yearly rate of change structural imaging outcomes over 5 years of follow-up in early axSpA patients fulfilling the ASAS axSpA classification criteria

Imaging outcomes	Baseline score* (N=313-344)	Standardized rate of change/year¥	Relative sRoC (Common Reference: mNY)	Relative sRoC per modality and anatomical site
Structural lesions (X-SIJ)[15, 16]				
mNY dichotomous	73 (21.2%)	0.044	1	1
mNY 1-grade change**	NA	0.126	2.864	2.864
mNY 1-grade change and value ≥ 2††	NA	0.119	2.705	2.705
mNY continuous grade (0-8)	1.7 (1.8)	0.043	0.977	0.977
Structural lesions (MRI-SIJ)[17]				
≥ 5 fatty lesion and/or erosions	66 (19.5%)	0.238£	5.409	1
≥ 3 erosions	60 (17.7%)	0.015	0.341	0.063
≥ 3 fatty lesions	56 (16.5%)	0.274£	6.227	1.151
Number of fatty lesions/erosions (0-80)	2.9 (4.9)	0.111	2.523	0.466
Number of erosions (0-40)	1.3 (2.2)	0.030	0.682	0.126
Number of fatty lesions (0-40)	1.5 (3.5)	0.140	3.182	0.588
Total structural lesions† (0-144)	3.4 (5.9)	0.115	2.614	0.483
Total structural lesions no sclerosis (0-104)	3.2 (5.8)	0.124	2.818	0.521
Structural lesions (X-Spine)[18]				
≥ 1 syndesmophyte	19 (5.5%)	0.037	0.841	1
mSASSS score (0-72)	0.3 (1.3)	0.043	0.977	1.162
Structural lesions (MRI-Spine)[19, 20]				
≥ 5 fatty lesions	5 (1.6%)	-0.013	0.295	1
Total structural lesions‡ (0-322)	0.4 (1.0)	0.016	0.364	1.231
Number of fatty lesions (0-92)	0.3 (0.8)	0.008	0.182	0.615
Number of corner erosions (0-92)	0.1 (0.2)	0.012	0.273	0.923
Number of corner bone spurs (0-92)	0.1 (0.3)	0.027	0.614	2.077

* Agreement of ≥2 out of 3 readers for binary variables and mean (SD) of 3 readers for continuous variables from wave 3; ¥ Estimated from a model where all independent variables (time, reader and wave) and the outcome are standardized; † fatty lesions, erosions, sclerosis, partial ankylosis/total ankylosis; ** Change of at least one grade in at least one sacroiliac joint (SIJ); †† Change of at least one grade in at least one SIJ, but with a 5-year grade ≥ 2 in the worsened joint; £ Quadratic transformation led to a better model goodness of fit (QIC: quasi-likelihood under the independence model criterion); NA, not applicable; ASAS, Assessment of SpondyloArthritis international Society; MRI, magnetic resonance imaging; X, radiograph; SIJ, sacroiliac joints; mSASSS, modified Stoke Ankylosing Spondylitis Spinal Score; sRoC, standardized rate of change.

DISCUSSION

In this prospective observational study, we have shown that, in patients with early axSpA, MRI outcomes of inflammation are more sensitive to change in the SIJ than in the spine. In addition, pelvic radiographs yield low sensitivity to change in detecting structural damage, while fatty lesions detected on MRI-SIJ emerges as a promising alternative. In contrast, MRI-spine is not better than X-spine in detecting structural changes in early axSpA patients.

In the current study, we directly compared, for the first time, inflammation outcomes on MRI-SIJ and MRI-spine and have shown that the former are more sensitive to change. Inflammation on MRI-spine remained low and essentially unchanged over a period of 5 years. Different from previous studies evaluating the sensitivity to change of imaging outcomes over shorter periods, we have applied an analytical technique ('integrated analysis') that we have previously shown to be robust for the evaluation of change over long periods of follow-up, especially with outcomes that are expected to occur infrequently over time.[8] Of note, combination algorithms (e.g. agreement between 2 out of 3 readers) are not needed when using this method. Instead

each individual reader score is analysed as it is in an assumption-free manner which, to some extent, handles across-reader variability.

The ASAS/OMERACT MRI working group has previously compared different (continuous) scores to quantify inflammation on MRI-SIJ.[22] In a multi-reader exercise the SPARCC method has been shown to be the most reliable and sensitive to change among patients with r-axSpA. The current study adds to this data by showing that both the continuous SPARCC score and the binary ASAS definition of a positive MRI-SIJ yield good sensitivity to change in the entire spectrum of axSpA (including nr-axSpA) during the early phases of the disease.

The same group performed a similar exercise for MRI-spine (also in r-axSpA).[3] This experiment has shown discrepant reliability results for the comparison between the 6- discovertebral unit (DVU) SPARCC score, the Ankylosing Spondylitis spine MRI activity (ASspiMRI-a) score and the Berlin method (SPARCC performed better when using the intraclass correlation coefficient but worse when using the smallest detectable change). All methods yielded excellent sensitivity to change according to the Guyatt's effect-size. Here, we compared the 23-DVU SPARCC to the Berlin method and 2 binary outcomes and found that all yield very poor sensitivity to change. Of note, these studies differ in several aspects, including the reading methods and population. In fact, our early axSpA population had lower baseline levels of inflammation compared to patients from the ASAS/OMERACT exercise (mean (SD) Berlin: 0.9 (2.7) vs 6 (9.0), respectively), which may hinder the detection of change, that we have shown before to be small in early axSpA.[17] Of note, in patients with nr-axSpA and high disease activity selected for RCTs, inflammation on MRI-spine performed well both in terms of sensitivity to change and in discriminating response between treatment arms.[23, 24] This confirms that the ability of the scoring methods to detect change is not only dependent on their intrinsic characteristics, but also on the population in which they are applied.

A recent study, also from DESIR, has shown that 'net' progression from mNY-negative to mNY-positive (i.e. considering measurement error) is very limited.[16] In the current study we have additionally shown that the change in the mNY (continuous) grading is as poorly sensitive to change as the mNY binary score (relative rate ≈ 1). On the other hand, the change in at least 1-grade in at least one SIJ, with or without considering the change between grade 0 and grade 1, perform better in detecting change.[16, 25]

Information on the sensitivity to change of MRI-SIJ structural outcomes is very scarce.[26] To the best of our knowledge, no previous formal comparison with X-SIJ scores has been performed thus far. We have found that ≥3 fatty lesions on MRI-SIJ largely outperform all X-SIJ outcomes. Erosions, however, performed poorly in this early population. Thus, our study yields encouraging data supporting MRI (in particular fatty lesions) as an alternative to radiographs in detecting change of structural damage at the SIJ. In contrast, in the spine, we found no evidence that MRI is better than radiographs in detecting change of structural damage. Despite the disappointing results with MRI, our results are in line with previous studies, showing that spinal radiographic progression can be detected even in early phases of the disease.[4, 27] A recent study has shown that low dose computerized tomography of the spine is more sensitive to detect new syndesmophytes than conventional radiographs promising to further expand our ability to detect change in axial damage.[28]

Our study has some limitations. First, not all available scoring systems were assessed. However, to the best of our knowledge, this is so far the largest direct comparison across scores, which includes those currently more often used in research and clinical practice. Second, we did not assess all domains of the OMERACT filter, namely validity, reliability and feasibility.[2] Thus, we cannot, and do not claim to, evoke superiority of one score over others based on our data alone. Instead, our results should be interpreted in light of the literature already informing on these aspects but falling short on direct comparisons of sensitivity to change. Third, the observed levels of inflammation, structural damage and changes over time are limited in this cohort, especially in the spine, which reduces the possibility to detect differences across methods. Finally, our data are limited to patients with early axSpA, thus our findings cannot be generalized to all patients with axSpA from clinical practice especially those with more advanced disease (i.e. with r-axSpA).

In conclusion, we have shown that MRI inflammation scores are more sensitive to change in the SIJ than in the spine. Also, X-SIJ structural outcomes are less sensitive to change compared to fatty lesions on MRI-SIJ. In contrast, MRI-spine is no better than X-spine in detecting structural changes in this early axSpA cohort. These data may help in prioritizing imaging scoring methods in subsequent observational or interventional studies in early axSpA.

SUPPLEMENTARY DATA

Supplementary data are published online on the website of Arthritis Care & Research

REFERENCES

1. Mandl P, Navarro-Compan V, Terslev L, et al. EULAR recommendations for the use of imaging in the diagnosis and management of spondyloarthritis in clinical practice. Ann Rheum Dis. 2015 Jul; 74(7):1327-1339.
2. Boers M, Kirwan JR, Wells G, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. J Clin Epidemiol. 2014 Jul; 67(7):745-753.
3. Lukas C, Braun J, van der Heijde D, et al. Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: a multireader experiment. J Rheumatol. 2007 Apr; 34(4):862-870.
4. 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.
5. Wanders AJ, Landewé RB, Spoorenberg A, et al. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the available methods based on the Outcome Measures in Rheumatology Clinical Trials filter. Arthritis Rheum. 2004 Aug; 50(8):2622-2632.
6. 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 Jul 23.
7. 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.
8. Sepriano A, , Ramiro S, van der Heijde D, Dougados M, Claudepierre P, Feydy A, Reijnierse M, Loeuille D, Landewé R. Integrated longitudinal analysis does not compromise precision and reduces bias in the study of imaging outcomes: A comparative 5-year analysis in the DESIR cohort. Semin Arthritis Rheum. 2020 Mar 3. pii: S0049-0172(20)30070-6.
9. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Oct 15; 53(5):703-709.
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. 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.

12. Hermann KG, Baraliakos X, van der Heijde DM, et al. Descriptions of spinal MRI lesions and definition of a positive MRI of the spine in axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI study group. Ann Rheum Dis. 2012 Aug; 71(8):1278-1288.

13. 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.

14. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Aug 15; 53(4):502-509.

15. 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.

16. 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 Jul 06.

17. Madari Q, Sepriano A, Ramiro S, et al. 5-year follow-up of spinal and sacroiliac MRI abnormalities in early axial spondyloarthritis: data from the DESIR cohort. RMD Open. 2020 Feb; 6(1).

18. 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.

19. 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.

20. 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.

21. 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.

22. van der Heijde DM, Landewé RB, Hermann KG, et al. Application of the OMERACT filter to scoring methods for magnetic resonance imaging of the sacroiliac joints and the spine. Recommendations for a research agenda at OMERACT 7. J Rheumatol. 2005 Oct; 32(10):2042-2047.

23. Maksymowych WP, Dougados M, van der Heijde D, et al. Clinical and MRI responses to etanercept in early non-radiographic axial spondyloarthritis: 48-week results from the EMBARK study. Ann Rheum Dis. 2016 Jul; 75(7):1328-1335.

24. Sieper J, van der Heijde D, Dougados M, et al. Efficacy and safety of adalimumab in patients with non-radiographic axial spondyloarthritis: results of a randomised placebo-controlled trial (ABILITY-1). Ann Rheum Dis. 2013 Jun; 72(6):815-822.

25. 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.

26. Baraliakos X, van der Heijde D, Braun J, et al. OMERACT magnetic resonance imaging initiative on structural and inflammatory lesions in ankylosing spondylitis--report of a special interest group at OMERACT 10 on sacroiliac joint and spine lesions. J Rheumatol. 2011 Sep; 38(9):2051-2054.

27. Cho SK, Sakai R, Nanki T, et al. A comparison of incidence and risk factors for serious adverse events in rheumatoid arthritis patients with etanercept or adalimumab in Korea and Japan. Modern Rheumatology. 2014 Jul; 24(4):572-579.

28. de Koning A, de Bruin F, van den Berg R, et al. Low-dose CT detects more progression of bone formation in comparison to conventional radiography in patients with ankylosing spondylitis: results from the SIAS cohort. Ann Rheum Dis. 2018 Feb; 77(2):293-299.

Chapter 12

Summary and conclusions

SUMMARY AND CONCLUSIONS

With the work presented in this thesis we aimed at contributing to a better understanding of the concept of spondyloarthritis (SpA) as well as to elucidate how imaging of the axial skeleton can, more efficiently, be used to monitor and predict disease progression over time. Our main contributions to the field were as follows: First, we have addressed the issue of misclassification by the Assessment of SpondyloArthritis international Society (ASAS) classification criteria for axial SpA (axSpA) and peripheral SpA (pSpA) by evaluating their longitudinal validity against the rheumatologist's perception of the *Gestalt* of the disease. Second, we have, for the first time, shed light on the 'latent' phenotypes underlying the *Gestalt* of axSpA determined independently of the rheumatologist's opinion. This unprecedented approach has allowed us to better understand whether circularity played a role in the development of the criteria and how well the modern perception of axSpA, and the criteria developed in light of such a perception, truly overlap with the 'true Gestalt'. Third, we have proposed analytical approaches to improve our ability to reliably detect change of imaging outcomes, as well as predictive factors thereof, by limiting underlying assumptions and by giving more credit to measurement error. Fourth, we have used these approaches to provide further insights to the link between inflammation and damage in axSpA as well as to determine which outcomes should be prioritized for the monitoring of patients in clinical practice and in subsequent observational or interventional studies in early axSpA.

The studies presented in this thesis were performed in three independent cohorts: The ASAS cohort,[1, 2] the Spondyloarthritis Caught Early (SPACE) cohort,[3] and the Devenir des Spondyloarthropathies Indifférenciées Récentes (DESIR) cohort.[4] The ASAS cohort is a multicentre, prospective study in which patients had to fulfil one of two criteria to be included: 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) peripheral arthritis and/or enthesitis and/or dactylitis in the absence of current back pain with suspicion of SpA but no definitive diagnosis. SPACE is an ongoing multinational cohort in which consecutive patients aged ≥ 16 years with chronic back pain (CBP; ≥3 months, ≤2 years and onset <45 years) are included. DESIR is a longitudinal prospective cohort that includes adults aged over 18 and less than 50 years from 25 regional centres in France. At inclusion, patients have inflammatory back pain (IBP) with more than 3 months and less than 3 years and symptoms suggestive of SpA according to the opinion of the local investigator (level of confidence >5, scale 0-10).

In this final chapter we will summarize the main findings of the studies presented in this thesis and we will also discuss future perspectives as well as a research agenda for the topics that we have studied.

Classification and *Gestalt* of spondyloarthritis

In this thesis, we addressed the issue of misclassification by the ASAS SpA classification criteria. We started in **Chapter 2,** by determining, in the original ASAS cohort, what is the likelihood for a patient who classifies as positive to receive a clinical diagnosis of SpA after follow-up (positive predictive value; PPV). We have compared the baseline classification status according to the ASAS axSpA (also 'imaging arm' and 'clinical arm' separately), pSpA and SpA (both combined) to the clinical diagnosis ('external reference') made by ASAS experts after a mean follow-up of 4.4 years. Several important conclusions could be drawn from this exercise which, among others, argue against misclassification: first, we found that the large majority of the patients who fulfilled either the axSpA or pSpA criteria at baseline were in fact diagnosed as SpA at follow-up (PPV: 92%), which adds to the validity of the ASAS SpA criteria as a whole. Second, the pSpA criteria discriminated well between a clinical diagnosis of pSpA and no-SpA (PPV: 90%), even with similar proportions of peripheral arthritis in both groups (91%). There was, however, a significant difference in the proportion of enthesitis (pSpA: 60% vs no pSpA: 26%), which highlights the central role of enthesitis in the disease. It also indicates that the allowance of enthesitis as an entry feature does not lead to mislabelling, as previously suggested. Third, the PPV was equally high for the 'imaging arm only' (86%) and the 'clinical arm only' (88%) separately, which argues against misclassification by the 'clinical arm' and supports the view that the 'clinical arm' comprises a group of patients that belong to the SpA *Gestalt* as much as those fulfilling the 'imaging arm'. Third, almost all patients who had sacroiliitis on imaging classified positive for axSpA (98%) with many of those in the 'imaging arm' (irrespective of the 'clinical arm') having only sacroiliitis on MRI (62%). Since most of these were indeed diagnosed as axSpA at follow-up (PPV: 95%), our data reflect the dominant place that sacroiliitis on MRI holds in the ASAS axSpA criteria and testify to the high diagnostic value attributed to this feature by the rheumatologists. Even though the study presented in chapter 2 has a number of noteworthy limitations (e.g. losses to follow-up, missing data), sensitivity analyses taught us that these had little impact in our PPV calculations.

In Chapter 2 we tested validity of the ASAS SpA criteria against the expert opinion in the ASAS cohort. In addition, the ASAS classification criteria have been further challenged around the world in different cohorts. Some of these cohorts differ in several aspects from the ASAS cohort, thus yielding unique insights into the criteria performance and applicability in a broad population of patients. In **Chapter 3** we performed a systematic literature review (SLR) of the published data pertaining to the performance of the ASAS classification criteria tested against the rheumatologist's diagnosis. In total, data from eight independent cohorts including more than 5,500 patients, was evaluated. In addition to the original studies for the development of the ASAS axSpA and pSpA criteria,[1, 2] 5 studies assessing the ASAS axSpA criteria,[3, 5-8] one study the pSpA criteria,[9] and one study the combined SpA criteria (providing separate data also for the axSpA and pSpA criteria) were also included.[10]

The pooled analysis revealed an excellent sensitivity and specificity of the ASAS SpA (axSpA and pSpA combined) criteria (73%; 88%; respectively). Good performance was also noted for the axSpA criteria (sens: 82%; spec: 87%), which was robust to variations in the setting (hospital vs community), in symptom duration (<2 years vs ≥2 years) and type of population ('restricted' vs 'original ASAS population') (sens range: 78-85%; spec range: 90-93%). Of note, splitting the

axSpA criteria into 'imaging arm only' and 'clinical arm only' compromised sensitivity (26% and 23%, respectively), but retained very high specificity (97%; 94%). This finding is aligned with Chapter 2 and further argues in favour of the combined use of both 'arms' to avoid missing axSpA patients. The finding of a higher positive likelihood ratio (LR+) for the 'imaging arm' (13.6) compared to the 'clinical arm' (6.0) again highlights the rheumatologist's reliance on positive imaging findings for making an axSpA diagnosis. Similar to the ASAS axSpA criteria, the pooled specificity of the pSpA criteria was excellent (87%). However, sensitivity was much lower (62%). The low pooled sensitivity of the pSpA criteria was driven by the two studies including patients only based on the presence of peripheral arthritis, which once again highlights the relevance of enthesitis, and dactylitis, and adds to the credibility of the ASAS pSpA criteria, that include these clinical presentations.

Both in the development and validation of the ASAS SpA classification criteria, expert opinion was used as an external 'anchor' in the absence of a 'true' gold-standard. This approach, however, entails one fundamental limitation which might compromise the criteria content validity: circularity.[11] However, circularity is not necessarily detrimental, provided the rheumatologist's perception is a good reflection of the 'true *Gestalt*'. The only way to verify this premise is to exclude the opinion of the rheumatologist from the analysis. In **Chapter 4**, we have used latent class analysis (LCA), to reveal the 'latent' (i.e. unobserved) *Gestalt* of axSpA (independent of expert judgement) by splitting patients from the SPACE and DESIR cohorts (analysed separately) into mutually exclusive classes (or phenotypes) based on the covariance of observed SpA features. SpA features were selected by us *a priori*, without any assumption on their relative value to the *Gestalt* of axSpA.

We identified three separate clinical entities, together forming the *Gestalt* of axSpA, in both cohorts. We labelled these 'Pure axial SpA' ('Axial'), 'Axial SpA with peripheral signs' ('IBP+Peripheral') and 'Axial SpA at risk' ('At Risk'). The 'Axial' class is characterised by a high likelihood of axial imaging abnormalities (e.g. 74% and 84% likelihood of inflammation on MRI-SIJ in SPACE and DESIR, respectively), HLA-B27 positivity and male dominance. This phenotype closely resembles the rheumatologist's conventional clinical picture of axSpA. Thus, it is not surprising that the ASAS axSpA classification criteria (developed by experts) captured almost entirely the 'Axial' class (98% in SPACE and 93% in DESIR). The 'IBP+Peripheral' phenotype is defined by the presence of IBP (100%) in conjunction with peripheral signs and symptoms. 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 (SPACE: 70%; DESIR: 82%) than the axSpA classification criteria (SPACE: 45%; DESIR: 58%). The 'At Risk' class is an entity characterised by the presence of presumed risk factors for axSpA (i.e. positive family history and HLA-B27) in association with IBP but only sporadically other SpA features. These patients often fulfil the ASAS axSpA classification criteria (SPACE: 60%; DESIR: 54%). It was remarkable that five-year transitions across classes were very unlikely.

Further discussion and future perspectives

The studies included in this thesis further testify to the good performance of the ASAS SpA classification criteria. Provided the criteria are applied in patients already with a clinical

diagnosis of SpA, as they are supposed to, the risk of misclassification is not higher than the risk with other diseases. For instance, the 2010 American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) classification criteria for rheumatoid arthritis (RA) were also developed to capture patients early in the disease course.[12] A meta-analysis evaluating their performance against several reference standards revealed an overall sensitivity of 82% and specificity of 61%.[13] The same compromise in specificity has not been observed in SpA, where we found a pooled sensitivity of 82% and specificity of 87% for the axSpA criteria. The fear that the 'clinical arm' would be responsible for misclassification does not find support in our data either. Rheumatologists from all over the world recognised patients fulfilling the 'clinical arm' within their *Gestalt* of SpA as much as those with evidence of sacroiliitis on imaging ('imaging arm').

Our data strengthen the view that non-radiographic axSpA (nr-axSpA) and radiographic axSpA (r-axSpA) are not separate entities as previously claimed, but rather part of the same disease continuum. The in-depth analysis into the *Gestalt* of axSpA in Chapter 4 clearly identified an 'Axial' phenotype without any distinction between nr-axSpA and r-axSpA. This expert judgement-free observation is in line with extensive evidence suggesting that a split of axSpA into nr-axSpA and r-axSpA is artificial.[14-16] More recently, a meta-analysis further demonstrated that both groups have mostly a similar disease presentation, as well as burden of disease.[17] Differences were also noted, but these should be interpreted thoughtfully. Patients with r-axSpA are most likely male, often smokers, have more often elevation of CRP, longer symptom duration and higher impairment of spinal mobility than patients with nr-axSpA. Available literature tells us that these translate into prognostic rather than diagnostic dissimilarities. In fact, these features are well known to associate with damage accrual in axSpA, thus obviously dominate the r-axSpA phenotype.[18-20] Similar poor prognostic 'funnelling' is seen in RA patients with erosive disease as compared to those without erosions.[12] However, it has never been argued that these two phenotypes in fact are two forms of the same disease.

In the introduction of this thesis we have used a Venn diagram (Figure 1) to illustrate the theoretical relationships of the concepts of the 'true *Gestalt*' of SpA ('C'), the rheumatologist's perception (diagnosis) of the *Gestalt* ('A') and the classification criteria ('B').This framework allows a critical discussion of the overlapping circles (misclassification and misdiagnosis) and is helpful for elucidating why inflammation on magnetic resonance imaging of the sacroiliac joints (MRI-SIJ) might have been awarded a too dominant place in the ASAS axSpA classification criteria due to circular reasoning.[11] Since sacroiliitis on MRI was at the basis of the nr-axSpA concept, which instigated the development of the ASAS axSpA criteria by experts, subsequent cross-validation against an expert's diagnosis may have resulted in classification criteria that reflect experts' beliefs ('A') rather than an objective presence of axSpA ('C'). Both in Chapter 2 and 3 we have indeed seen that the presence of inflammation on MRI-SIJ was almost synonymous to a clinical diagnosis in axSpA. This is in line with a recent study in the SPACE cohort in which patients with CBP received a diagnosis from a rheumatologist before and after the latter became aware of the result of imaging. Once known, a switch in diagnosis in 51% of the patients for whom the MRI-SIJ/pelvic radiograph result and the first diagnostic-judgement were incongruent was seen.[21] This is not so surprising though. Already in the original study on which the axSpA criteria were developed with 'paper patients', ASAS experts changed their diagnosis in 21% of the patients after knowing the result of the MRI-SIJ.[22]

Nonetheless, this evidence alone does not clarify whether the dominance of inflammation on MRI-SIJ is 'inappropriate' or just. In other words, it might be that such dominance according to the rheumatologist ('A') and translated into the axSpA criteria ('B') is coherent with the 'true *Gestalt*' of axSpA ('C'). The identification of the above-mentioned 'Axial' phenotype, with dominant imaging abnormalities (including inflammation on MRI-SIJ), independently of the rheumatologist's opinion, seems to suggest that. However, this phenotype, corresponds to less than 20% of the patients in SPACE and DESIR. This means that the phenotypical expansion of axSpA driven by MRI-SIJ, and reflected in the ASAS axSpA criteria, had indeed increased the 'AC' and 'BC' interactions (more 'true patients' diagnosed and classified), but at the cost of overlooking non-imaging-dominated phenotypes. The 'IBP+Peripheral', with female dominance, very low likelihood of abnormalities on axial imaging and a weak association with HLA-B27, is often seen by the expert clinician. However, this does not find reflection in the ASAS axSpA criteria, because these criteria either require positive imaging or HLA-B27-positivity to classify as positive. Thus, our data support that inappropriate circularity had indeed occurred when developing the ASAS axSpA classification criteria.

Our data help us to better appreciate the likely under-representation of the 'IBP+Peripheral' phenotype in previous studies (e.g. randomised clinical trials) in which the axSpA criteria were used for inclusion. However, a solution to the issue of over-valuing inflammation on MRI-SIJ and HLA-B27 positivity in the ASAS axSpA criteria is not straightforward. Even though the pSpA criteria perform reasonably well in capturing axSpA patients who have negative imaging and are HLA-B27 negative, they were developed to be applied in patients with exclusively peripheral manifestations, not in patients with current back pain as with the 'IBP+Peripheral' phenotype. More research is needed to better understand the overlap between axSpA and pSpA which is found to be greater than initially thought when the ASAS classification criteria were developed. One possible way forward is to better understand the 'cause' of IBP among female axSpA patients without imaging abnormalities. The lower than expected specificity of this feature needs also to be considered.[23, 24]

We live in the era of early diagnosis and early treatment. This modern paradigm, which undoubtedly brought many benefits to patients, also raises important challenges.[25] Axial SpA is difficult to diagnose and rheumatologists rely on pattern recognition for its identification, which is more than a simple sum of SpA features.[26] The SpA-pattern is less obvious in early disease when 'typical' features may still be absent which leads to uncertainty ('grey-zone'). The experienced clinician, will disentangle patients who do not have axSpA from those with the disease and appropriately handle those for whom either diagnosis is not beyond any doubt. However, others, when dealing with uncertain or difficult cases may be tempted to apply classification criteria to inform binary diagnostic judgements (e.g. axSpA vs no axSpA) that do not allow grey zones. We have shown that such clinicians are seriously in risk of 'overdiagnosing' and consequently 'overtreating' axSpA. We have labelled patients who drive this clinical conundrum, as 'At Risk', and for the first time provided a clear description of this entity. Presumed risk factors for axSpA (i.e. positive family history and HLA-B27) are the 'anchors', which often associate with IBP but only sporadically with other SpA features. It is easy to see why these patients often fulfil the ASAS axSpA criteria, especially the 'clinical arm' which require HLA-B27 to be positive in addition to two SpA features (family history and IBP). However, family history has been shown to be redundant when HLA-B27 is known.[27] Moreover, IBP is less

specific than initially thought which may contribute to overcalling axSpA when the ASAS axSpA criteria are wrongly used for diagnostic purposes. Of note, this is not a problem if the criteria are appropriately used only after a clinical diagnosis has been made, thus continuing efforts for education are key to avoid 'overdiagnosis' and 'overtreatment'.[28] Future studies should give resolution on the long-term outcomes of this and the previously described phenotypes of the *Gestalt* of axSpA.

Assessment of radiographic progression at the sacroiliac joints

Definite damage seen on the radiographs of the SIJ (X-SIJ) is defined according to the modified New York (mNY) grading system, as the presence of bilateral grade 2 or unilateral grade 3 or 4 'sacroiliitis' ('mNY-positive'), which is a key feature in the classification of r-axSpA.[29] However radiographic 'sacroiliitis' has been shown to be an unreliable finding, especially when assessed by untrained local readers.[30, 31] Determining the irreversible progression from mNY-negative (nr-axSpA) to mNY-positive (r-axSpA) is, arguably, even more ambiguous. Previous studies have focused only on positive change,[16] however, from a methodological perspective, bi-directional change, if present, cannot be ignored.

In **Chapter 5**, we compared two pelvic radiographs (X-SIJ) read several years apart (4.4 years on average) in patients with suspected SpA from the ASAS cohort in order to assess positive and negative change according to the mNY criteria. In total, 357 had paired X-SIJ available (at baseline and follow-up) read by the local observer. Of the 357 included patients, 17% (62/357) were mNY-positive at baseline. At follow-up this proportion increased to 22% (80/357). However, more than half (36/62) of those considered mNY-positive at baseline were assessed mNY-negative at follow-up. Assuming that structural damage in the SIJ is an inherently irreversible feature, and knowing that readers were aware of the correct time-order, these 'improvements' are very difficult to understand. The sobering truth is that progression, 'regression' and measurement error are not easy to disentangle in this setting. Thus, the question remained on what is the 'real' rate of radiographic progression at the SIJ level and how to handle measurement error on its calculation.

In previous studies, researchers have largely ignored or overlooked measurement error when reporting binary scores of progression, such as the change from mNY-negative to mNY-positive.[18, 19, 32-35] In **Chapter 6**, we undertook an analytical exercise that testifies to the truth of this statement and we made a plea for measurement error (or 'noise') not to be ignored when interpreting imaging studies.[36] We exposed the false assumptions underlying commonly used binary definitions of progression and proposed an analytical approach that we argue will best handle error. We evaluated the change between mNY-negative and mNY-positive after 5 years in the DESIR cohort, in which, contrary to the ASAS cohort, readers were blinded to time-order. In this setting, 'improvements' (i.e. change from mNY-positive to mNY-negative) should be judged as measurement error ('noise'). Each reader reported a binary score (mNY-positive vs mNY-negative) and the final status score was defined by the agreement of at least 2 of the 3 readers. The cross-tabulation between the baseline and 5-year reading, resulted in 3 possible change scores (mNY-positive to mNY-negative, no change, mNY-negative to mNY-positive).

At baseline, 62 (15%) of the 416 included patients were mNY-positive. Of the 354 mNY-negative patients at baseline, 24 (6.8%) changed into mNY-positive after 5 years. We labelled this figure as 'Crude progression', the simplest and most often used method to measure pelvic radiographic progression, which refers only to the rate of positive change. However, this method is spurious since it implies that the baseline reading is free of error and that a change in the opposite direction (here: 3/62: 4.8%) can be ignored. More recently, the method of 'Conditional net progression' has been proposed which gives credit to the rate of negative change by subtracting it from the rate of positive change (6.8%-4.8%: 2%).[18, 19] However, this method implicitly assumes that 'worsening' can only happen in patients who are mNY-negative at baseline and 'improvement' only in mNY-positive patients. Since readers are not aware which film is the baseline film, this assumption does not hold.

We therefore proposed a third method that we called (assumption-free) 'net progression' with which both 'positive change' and 'negative changes' are 'allowed' and scores of individual patients are not interpreted as 'true progression' or 'noise'. 'Net progression' is expressed as a percentage calculated as follows: number of positive changes minus number of negative changes divided by all patients [(24-3)/416=5%]. This calculation follows the same reasoning of the area under the curve (AUC) of probability plots (positive area minus negative area) that provides the mean continuous change score taking measurement error into account.[37] Thus, this ´net progression' yields the least biased estimates since it gives most credit to measurement error. That is, always includes error without a prior assumption on the imaging modality ability to reliably capture change.

Further discussion and future perspectives

Our data argue that, in a clinical practice setting, the arbitrary distinction between a mNY-negative and mNY-positive X-SIJ, is of little prognostic (but also diagnostic) utility. The same conclusion does not necessarily apply to its use in clinical research, where strategies to reduce measurement error can be implemented. Having films read by calibrated and trained central readers and the final scores determined by an 'agreement algorithm' (e.g. 2 out of 3) are some examples that reduce the 'noise'. However, even with such strategies, measurement error cannot be fully eliminated as shown by the occurrence of the unexpected improvements in Chapter 6. Thoughtful analytical approaches can help, as the one we propose (the so-called 'net progression'), to be used in clinical research to handle measurement error and to best estimate true progression of binary change-scores. Obviously, decreasing bias carries many benefits such as the better detection of treatment effects in randomised trials. Even though, we have used radiographic progression at the SIJ in axSpA to describe this method, its application extends to all examples where imaging scores on structural damage are obtained under blinded conditions.

Despite its merits, it should be noted, that this method implies that outcomes are irreversible (mainly structural damage), and are evaluated over short periods, as 'true repair' cannot be excluded with longer follow-up. Further studies should help us to understand the meaning of 'negative changes' in other settings than those with irreversible damage. They should also explain how 'true improvements' (i.e. repair) possibly contribute to the overall net progression. In addition, the 'assumption-free' method yields an average estimate of 'true progression' at the

group level (i.e. beyond measurement error) but does not translate to individual patients. So, it becomes impossible to declare an individual patient as a 'progressor'. Consequently, net progression is not to be used in prediction models aiming at determining factors associated with radiographic progression, such as inflammation on MRI.

Relationship between inflammation and structural damage

Patients with axSpA experience varying levels of radiographic progression, so identifying those with a higher likelihood of damage accrual is key for prognostic stratification. A significant effort has been put forward to the study of drivers of damage in axSpA, with inflammation receiving a large amount of attention by the international rheumatology community. At the time of the start of this thesis, there was already robust evidence supporting that inflammation associates with radiographic progression at the spinal level.[20, 38-42] It would be expected for the same association to be present at the SIJ level, however evidence supporting or rejecting this hypothesis was still scarce at that time.[18, 19]

In **Chapter 7**, we sought to determine whether inflammation on MRI-SIJ (ASAS definition) was associated with structural damage on X-SIJ (mNY grading) 5 years later in the DESIR cohort. As mentioned above, the estimated net progression from mNY-negative to positive was 5%. Other binary definitions of progression based on the grading of the SIJs proved to be more sensitive to change. Net progression was 13% for the change in at least one grade in at least one SIJ and 10% for change in at least one grade in at least one SIJ and a final absolute value of at least 2 in the worsened joint. Objective inflammatory markers (CRP and inflammation on MRI-SIJ) at baseline had a large impact on the likelihood of net progression especially among patients who were HLA-B27-positive. For instance, we found that patients who were HLA-B27-negative and who had a normal CRP and a negative MRI-SIJ had a likelihood of only 1% to (net) progress from mNY-negative to mNY-positive. In contrast, this likelihood was eighteen times higher (18%) if all three variables were positive.

We further tested whether inflammation on MRI-SIJ at baseline associated with subsequent radiographic progression at 5 years in multivariable models. Since the figure of net progression does not identify individual patients it could not be used in the models. Instead we defined our outcomes at 5 years irrespective of the scoring at baseline. For instance, the outcome was the mNY status (positive vs negative) at 5-years and not the change from mNY-negative to positive which would imply including in the analysis only those unreliably judged mNY-negative at baseline. This approach not only reduces bias but also increases the statistical power by including a larger number of patients. Two main conclusions could be drawn. First, we found that inflammation on MRI-SIJ was independently associated with damage at 5-years; second, that this association was modified by the baseline HLA-B27 status. That is, the effect of MRI-SIJ inflammation on mNY after 5 years was stronger in HLA-B27 positive patients [odds ratio (OR) 5.39 (95% CI: 3.25–8.94)] than in HLA-B27 negative patients [OR 2.16 (95% CI: 1.04–4.51)].

Although the results from this Chapter 7 are methodologically robust, they are also hard to translate to clinical practice where images are not read by multiple trained readers blinded to chronology. As demonstrated in Chapter 5, substantial 'noise' is expected in locally read films. Thus, it was unclear whether inflammation on MRI-SIJ as seen in clinical practice had the same

prognostic connotation as inflammation found with central reading. In **Chapter 8**, using locally read data from the ASAS and DESIR cohorts, we found that, despite all the 'noise', there was a clear prognostic value for objective inflammation: patients with a normal CRP and no inflammation on MRI-SIJ were unlikely to progress from mNY-negative to mNY-positive (ASAS: 4%; DESIR: 3%), whereas those who had both elevated CRP and inflammation on MRI-SIJ had very high probability to progress (ASAS: 33%; DESIR: 17%). In the multivariable analysis, inflammation on MRI-SIJ was found to be an independent predictor of the development of radiographic damage both in the ASAS (OR=3.2 [95% CI: 1.3; 7.9]), and DESIR (OR=7.6 [95% CI: 4.3; 13.2]) cohort. This study strongly argues in favour of the prognostic value of inflammation on MRI-SIJ as available in daily clinical practice.

Recently, there has been an increasing interest in the use of MRI not only to measure inflammation but also structural damage. Definitions of individual lesions (e.g. fatty lesions, erosions) have been proposed and composite scores validated.[43-46] In **Chapter 9** we tested, for the first time, the effect of inflammation on MRI of the SIJ and spine on the subsequent development of structural damage also measured on MRI over five years in the DESIR cohort. The presence of BME on MRI-SIJ at baseline was predictive of structural damage on MRI-SIJ 5 years later according to several binary definitions [range OR: 4.1-5.6]. Testing the association of interest on MRI-spine was challenged by low numbers of lesions, resulting in lower precision. Only the association between inflammation and ≥3 fatty lesions was statistically significant. In addition to the baseline models, we have shown that axial inflammation detected on MRI is longitudinally associated with subsequent development of structural damage also on MRI over 5 years (longitudinal models) both at the SIJ and at the spinal level. This study adds to the existing evidence by showing that the association between axial inflammation and structural damage can also be measured with MRI in patients with early axSpA.

Further discussion and future perspectives

Our findings add to the literature by showing that an association between inflammation and damage is seen at the SIJ level similar to what was previously shown for the spine. If fact, no matter how we look into it, an unequivocal positive association between these two types of lesions is always found: In early disease (Chapter 7 and 9) and in more established disease (Chapter 8); with local readings (Chapter 8) and with central readings (Chapter 7 and 9); with damage measured in conventional radiographs (Chapter 7 and 8) and in MRI (Chapter 9). Of note, we did not only find a predictive association between baseline inflammation and follow-up damage; We also found that having inflammation on MRI in one visit increased the likelihood of having structural damage in the subsequent visit up to 5 years of follow-up adjusting for the presence of damage at the first visit. These, so-called, time-lagged and 'autoregressive' models allow a more causal interpretation and add credibility to our findings.

It should be noted that all analyses were performed at the patient level. For instance, inflammation was said to be present on the SIJs according to the ASAS definition and damage if the mNY criteria for 'sacroiliitis' were met. Another interesting question would be to evaluate whether inflammation in one specific SIJ quadrant leads to subsequent damage at the same quadrant. Such an analysis likely yields further insights into the complex pathophysiology of

axSpA. For instance, it has been shown that inflammation at the vertebral unit level increased the likelihood of the formation of a new syndesmophyte in the same location 2 years later, but most new syndesmophytes appeared in vertebral units without signs of inflammation.[42] This remarkable finding suggests that unknown pathophysiological mechanisms may play a role in structural progression in axSpA. Some have argued that local injury and muscle dysfunction may be responsible for driving inflammation-independent damage.[47, 48] However, such mechanisms are not yet fully understood.

In Chapter 9 we found that inflammation on MRI of the SIJ and spine was associated with the subsequent formation of both fatty lesions and erosions. However, the interpretation of such association is not straightforward, because the underlying cause of these imaging findings remains to be clarified. One hypothesis defends that inflammation in axSpA fluctuates and that bone proliferation (e.g. formation of syndesmophytes in the spine and ankylosis of the SIJs) is a repair process that ignites only once inflammation subsides and is mediated by the formation of fatty lesions.[39, 42] On the other hand, if inflammation is persistent, repair is not possible and catabolic bone changes dominate, which in turn leads to bone destruction (e.g. erosions). Understanding the complex interplay between inflammation, bone formation and bone destruction in axSpA potentially has important therapeutical consequences. However, axSpA is a slowly progressive disease, thus long-term studies are needed to better understand the complex relationship between these abnormalities, including their sequence, frequency and rate of change over time. These studies pose some methodological challenges that we address in the following section.

Multilevel analysis of imaging data

Researchers designing long-term cohort studies, usually do not want to wait several years before their data can be analysed. A common practice is to 'split' the cohort into parts after a certain period of data collection.[36] For instance, patients included in the DESIR cohort are planned to be followed up to 10 years,[4] but it was already possible to use the available 5-year data to address several relevant research questions. In this setting imaging-data is usually read in several 'reading-waves'. In DESIR imaging data were, thus far, collected at baseline, 1 year, 2 years and 5 years and read by trained central readers in 3 'reading waves': in wave 1, baseline images were scored by 2 readers and 1 adjudicator (in case of disagreement); in wave 2, images from baseline, 1 and 2 years were also scored by 2 readers and one adjudicator; in wave 3, images from baseline, 2 and 5 years were scored by 3 central readers.

In previous chapters, we evaluated whether inflammation on MRI at baseline predicted subsequent structural lesions after 5 years in the DESIR cohort (baseline prediction models). Therefore, we have used data from the only 'reading-wave' that contains 5-year data: 'wave 3'. Although logical, this choice it is not without underlying assumptions that are often not fully appreciated. For instance, to be included in our baseline prediction models, patients had to have complete 5-year data ('completers analysis'), meaning that all those who had only follow-up imaging data at 2 years in wave 3 or images scored only in the other 2 waves were excluded (right censoring). In addition, yet another analytical choice had to be made. In wave 3 each of the 3 readers reported a score per patient, and these scores had to be somehow combined to

define the inflammation and damage variables. We have decided for the rule of 2 out of 3 for binary variables (e.g. inflammation present /absent according to the ASAS definition; or mNY positive/negative;) and the average of 3 for continuous variables (e.g. the SPARCC score; or the mNY grading). These combination algorithms are practical but are also not assumption-free and lead to loss of information.

We have already partially addressed these analytical problems in previous chapters. When we modelled baseline inflammation against the 5-year damage (both with combined scores), a 'traditional' logistic regression model would suffice. However, we have pursued two additional approaches. First, we have modelled the baseline inflammation against the 5-year structural outcomes using data from each individual reader separately. Since scores per reader are not independent, the assumption of independency of observations of logistic regression models does not hold. Therefore, we used generalised estimating equations (GEE), an analytical technique that takes correlated data into account. Our baseline predictive GEE models had one level of correlation (the reader), thus we called them 1-level GEE models. Still, these models leave out the 2-year visit data. So, we have used a so-called time-lagged longitudinal model that take all visits from wave 3 into account. That is, we have tested the association between baseline MRI-SIJ inflammation and 2-year damage and between 2-year inflammation and 5-year damage in a '2-level longitudinal GEE model'. The first level being the patient (repeated scores over time) and the second level the reader.

Despite the merits of the 1-level and 2-level GEE models, we are still disregarding a large number of scores yielded by the central readers and adjudicators from wave 1 and wave 2. In theory, including all data without aggregation-algorithms protects against bias, since the analyst does not need to intervene in data selection and computation. The 'trade off' is adding some variability ('noise') to the estimates, which may lead to a lower precision (i.e. wider 95% CI). Combining all available imaging-information has been previously shown to be a robust approach to analyse long-term imaging data in patients with RA using all available information.[49] In **Chapter 10** we investigated if an 'integrated analysis' affects the precision of estimates of change of imaging outcomes in patients with axSpA, with a conventional completers analysis as reference standard. To achieve that we had to consider one additional level of correlated data (the 'reading-wave' level). Thus, our data has 3 levels of correlation: Each visit is clustered within patient, each patient is clustered within reader, and each reader is clustered within the 'reading-wave'. To estimate the change over time while considering the various levels of correlation, 3-level GEE models were used. In these models, time was our independent variable of interest. The 'exchangeable' working correlation structure was found to best fit the data when taking the repeated scores over time per patient. The two 'higher' levels of correlation (reader and wave) were added as covariates to the model, an approach that has been previously proposed.[20]

We have challenged the 'integrated analysis' with a large number of continuous and binary outcomes reflecting: i. inflammation at the SIJ and spinal level (e.g. SPARCC score and the ASAS definition of inflammation at MRI-SIJ and MRI-spine);[50-54] ii. damage in spine radiographs (e.g. mSASSS and the presence of ≥ 1 syndesmophyte);[55] iii. damage on pelvic radiographs (e.g. mNY continuous grade and mNY-positivity);[29] iv. damage on MRI-SIJ (e.g. ≥ 3 fatty lesions);[45, 56] and v. damage on MRI-spine (e.g. ≥ 5 fatty lesions).[57] Each outcome was tested in a separate model. We did not focus on the point estimates of time, but rather on their

95% confidence intervals (CI). The narrower the 95%CI the higher the precision. We have compared the 'integrated analysis' with two types of completers analysis: i. a completers-only analysis, including only patients with complete 5-year follow-up, using scores from individual readers from wave 3 (adjusted for reader; 2-level); and aggregated completers analysis, using a combination algorithm (thus without reader adjustment; 1-level).

This analytical experiment proved the superiority of the 'integrated analysis' in comparison with both types of completers analysis in several ways. First, the 'integrated analysis' was more inclusive: out of 413 patients, the 'integrated analysis' models could include between 399 and 411 patients (depending on the outcome), whereas both 'completers analyses' included 364 at maximum. Second, we have proven that adding all data from individual readers and from all waves, without combination algorithms, does not affect the precision of the estimates of change. In fact, the 95% CI intervals were mostly similar across the three analytical approaches for most outcomes. Of note, the subtle increase in binary X-SIJ structural lesions (e.g. worsening of ≥ 1 grade in ≥1 SIJ with a grade ≥2 in the worsened joint at 5 years) was detected with more precision by the 'integrated analysis' analysis (95% CI: 1.06; 2.46) as compared to both completers analyses (2-level model 95% CI: 0.78; 2.32; 1-level model 95% CI: 0.81; 3.28). These data confirm that the 'integrated analysis' increases external validity (less patients excluded) without compromising (or even improving) internal validity.

We list above several of the imaging outcomes which have been developed and validated to assess inflammation and structural damage over time in patients with axSpA. However, direct comparisons of their sensitivity to change are mostly absent in the literature especially in early axSpA. [57-61] This knowledge gap precludes informed decisions on which outcome to prioritize in the follow-up and monitoring of patients with axSpA. Therefore, in **Chapter 11** we applied the 'integrated analysis' to study the sensitivity to change of the same scores used in chapter 10. Different to Chapter 10, however, here we focused on the point estimates of time (interpreted as change per year of the outcome) and not on the 95% CI. Since scores differ in the units of change and some are continuous while other binary, all variables were standardized (difference between the individual's value and the population mean divided by the population standard deviation [SD]) to allow comparisons. Each standardized variable has a mean of 0 and a variance of 1 and reads as the number of SD above (positive) or below (negative) the mean (standardized rate of change). We have compared the imaging outcomes sensitivity to change by calculating their relative standardized rate of change, i.e. the standardized yearly rate of change of an outcome divided by the corresponding rate of a reference imaging outcome.

We found that MRI outcomes of inflammation are more sensitive to change at the SIJ level than in the spine (e.g. range of relative standardized rate of change of spinal outcomes compared to the ASAS definition of a positive MRI-SIJ: 0.094; 0.531; i.e. all values far below 1). In addition, pelvic radiographs yield low sensitivity to change in detecting structural damage, while fatty lesions detected on MRI-SIJ emerged as a promising alternative (relative rate of change of ≥ 3 fatty lesions vs mNY: 6.2; i.e. far above 1). In contrast, MRI-spine (range rate of change: -0.013; 0.027) is not better than X-spine spine (range rate of change: 0.037; 0.043) in detecting structural changes in early axSpA patients.

Further discussion and future perspectives

The main strength of the proposed 'integrated analysis' is its ability to use all available imaging data in an assumption-free way with no intervention by the analyst. We have shown that this approach does not compromise precision, unlike what was expected since more data in principle implies more 'noise'. On the contrary, for outcomes that occurred infrequently over time, precision was even improved. Thus, this approach may be of special interest in studies with long-term follow-up, and/or when the outcomes are expected to occur infrequently over time. Our results are aligned with a previous study in RA, and may as well apply to other diseases. In fact, long-term studies evaluating imaging outcomes are highly relevant in the field of rheumatology. Over the years, several cohorts have been started to address some of the most fundamental and across-diseases long-lasting research questions: What is the natural history of the disease? Are we able to distinguish the patients who will follow a fairly benign path from those with a worse prognosis? Can we intervene in this process and ultimately drive lasting therapeutic benefits? The 'integrated analysis' may help researchers solving these questions in future studies focusing on imaging.

Arguably, by including all scoring data without 'hidden' assumptions, we may better approximate the 'true' point-estimates (the 'signal'). In fact, despite similar levels of precision, differences in the point-estimates were found between methods. That is, the estimated rates of change for the various imaging outcomes differed across analytical approaches. This might be, at least, in part explained by the fact that different patients are included depending on the method, with the 'integrated analysis' being the most inclusive. An alternative explanation might find ground on the old 'wisdom of the crowd' theory. An article published more than 30 years ago explains how this theory works by using a simple 'bean jar experiment'.[62] Briefly, a classroom of students is asked about the number of beans contained in a transparent jar. First, with no specific instructions each student yields an estimate and the average is calculated. Then the students are instructed about how to best estimate the number of beans and the exercise is repeated. Almost all students failed the exact number, but the average estimate came very close to the real number. Against expectations, however, the second average estimate (after instructions) was far less close to the exact number. The explanation for this counterintuitive finding is that the errors in each guess in the first exercise was independent from each other but became dependent in the second exercise when all students learned about the same instructions. This simple experiment tells us that combining multiple observations approximates the 'truth' provided the independence assumption holds. In theory, the larger the number of observations, the closer to the truth we will get.

The 'integrated analysis' uses a far greater number of observations than any of the 2 'completers analysis'. Let's use as starting point the 413 patients from DESIR who were included in the analysis of Chapter 10. For simplicity, let's assume that all 413 patients (p) complete the 5-year follow-up. These patients had at least one available score from at least one of the visits (t=4) read by at least one reader/adjudicator (r=3) from all available 'reading-waves' (w=3). In the 'completers analysis' using wave 3 only and aggregated outcomes (e.g. 2 out of 3) the statistical model could include at maximum 1,239 observations (413p * 3t *1r * 1w). This figure increases to 3,717 observations if we use individual-reader data instead of a combined outcome (413p * 3t *3r * 1w). Finally, with the 'integrated analysis' an impressive 14,868 observations can be

used at maximum (413p * 4t *3r * 3w). Obviously, the actual number of observations is variable depending on missing data, but in theory this approach can increase by almost 15 times the number of observations used to estimate our coefficient of interest. By applying a statistical technique that handles correlated data, we can then apply the principle of the 'wisdom of the crowd' and use all this information to approximate the 'true' estimate of change of each imaging outcome better than any of the 'completers analysis'.

The application of the 'integrated analysis' to compare the sensitivity to change of imaging outcomes yielded important insights, which may help in prioritizing imaging scoring methods in subsequent observational or interventional studies in early axSpA. Imaging abnormalities were found to be scarce and to hardly change over the period of 5 years at the spinal level regardless of the outcome and imaging modality. The opposite was observed at the SIJ level, which is aligned with the literature supporting that structural damage usually starts at the SIJ level and that the spine gets involved later on, and only in some of the patients.[63] We add to the literature by showing, for the first time, that MRI-SIJ outcomes of structural damage are more sensitive to change than the 'conventional' pelvic radiographic outcomes. This finding can be used to plan future studies aiming at studying progression of structural damage at the SIJ level, including those testing interventions aiming at disease modification.

Final comments

In this thesis we have pursued innovative analytical solutions for some of the most challenging questions in the field of SpA. We have gained better insights into the concept of axSpA by studying it independently of the rheumatologist's opinion. Our findings likely add knowledge to what axSpA really is. Future studies will learn us how much of these insights will translate into a better recognition of the disease in clinical practice and in better classifying them for research purposes. Since SpA is a slowly progressing disease, several years are needed to see meaningful changes in imaging abnormalities of the axial skeleton, which poses methodological challenges. We have shown that thoughtful analytical approaches, that make best use of imaging data, are helpful in better estimating progression, in unravelling its determinants and in clarify which outcomes are best to monitor disease. Efforts are made to further improve outcome measurement in axSpA, including the development of new imaging techniques, which can benefit from our proposed solutions to long-term imaging scoring. No question is too difficult when methodological rigor and creativity are put to work together:

Aut viam inveniam aut faciam

REFERENCES

1. Rudwaleit M, van der Heijde D, Landewé R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis. 2009 Jun; 68(6):777-783.
2. Rudwaleit M, van der Heijde D, Landewé R, et al. The Assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. Ann Rheum Dis. 2011 Jan; 70(1):25-31.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Aletaha D, Neogi T, Silman AJ, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis. 2010 Sep; 69(9):1580-1588.
13. Radner H, Neogi T, Smolen JS, et al. Performance of the 2010 ACR/EULAR classification criteria for rheumatoid arthritis: a systematic literature review. Ann Rheum Dis. 2014 Jan; 73(1):114-123.
14. Baraliakos X, Braun J. Non-radiographic axial spondyloarthritis and ankylosing spondylitis: what are the similarities and differences? RMD Open. 2015; 1(Suppl 1):e000053.
15. 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.
16. Sieper J, van der Heijde D. Review: Nonradiographic axial spondyloarthritis: new definition of an old disease? Arthritis Rheum. 2013 Mar; 65(3):543-551.
17. Lopez-Medina C, Ramiro S, van der Heijde D, et al. Characteristics and burden of disease in patients with radiographic and non-radiographic axial Spondyloarthritis: a comparison by systematic literature review and meta-analysis. RMD Open. 2019; 5(2):e001108.
18. 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.
19. 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.
20. 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.
21. Ez-Zaitouni Z, Landewé R, van Lunteren M, et al. Imaging of the sacroiliac joints is important for diagnosing early axial spondyloarthritis but not

all-decisive. Rheumatology (Oxford). 2018 Jul 1; 57(7):1173-1179.

22. 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.
23. 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.
24. Poddubnyy D, Callhoff J, Spiller I, et al. Diagnostic accuracy of inflammatory back pain for axial spondyloarthritis in rheumatological care. RMD Open. 2018; 4(2):e000825.
25. Landewé RBM. Overdiagnosis and overtreatment in rheumatology: a little caution is in order. Ann Rheum Dis. 2018 Oct; 77(10):1394-1396.
26. Ez-Zaitouni Z, Bakker PAC, van Lunteren M, et al. Presence of multiple spondyloarthritis (SpA) features is important but not sufficient for a diagnosis of axial spondyloarthritis: data from the SPondyloArthritis Caught Early (SPACE) cohort. Ann Rheum Dis. 2017 Jun; 76(6):1086-1092.
27. 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.
28. Rudwaleit M. Fibromyalgia is not axial spondyloarthritis. Rheumatology (Oxford). 2018 Sep 1; 57(9):1510-1512.
29. 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.
30. 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.
31. 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.
32. 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.
33. Sampaio-Barros PD, Bortoluzzo AB, Conde RA, et al. Undifferentiated spondyloarthritis: a longterm followup. J Rheumatol. 2010 Jun; 37(6):1195-1199.
34. 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.
35. Schattenkirchner M, Kruger K. Natural course and prognosis of HLA-B27-positive oligoarthritis. Clin Rheumatol. 1987 Sep; 6 Suppl 2:83-86.
36. 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.
37. Landewé R, van der Heijde D. Radiographic progression depicted by probability plots: presenting data with optimal use of individual values. Arthritis Rheum. 2004 Mar; 50(3):699-706.
38. 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.
39. 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.
40. 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.
41. 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.
42. 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.
43. 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.
44. 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.

45. 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.
46. 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.
47. Schett G, Stolina M, Dwyer D, et al. Tumor necrosis factor alpha and RANKL blockade cannot halt bony spur formation in experimental inflammatory arthritis. Arthritis Rheum. 2009 Sep; 60(9):2644-2654.
48. White A, Abbott H, Masi AT, et al. Biomechanical properties of low back myofascial tissue in younger adult ankylosing spondylitis patients and matched healthy control subjects. Clin Biomech (Bristol, Avon). 2018 Aug; 57:67-73.
49. 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.
50. Hermann KG, Baraliakos X, van der Heijde DM, et al. Descriptions of spinal MRI lesions and definition of a positive MRI of the spine in axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI study group. Ann Rheum Dis. 2012 Aug; 71(8):1278-1288.
51. 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.
52. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Aug 15; 53(4):502-509.
53. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Rheum. 2005 Oct 15; 53(5):703-709.
54. 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.
55. 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.
56. 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.
57. de Hooge M, Pialat JB, Reijnierse M, et al. Assessment of typical SpA lesions on MRI of the spine: do local readers and central readers agree in the DESIR-cohort at baseline? Clin Rheumatol. 2017 Jul; 36(7):1551-1559.
58. Lukas C, Braun J, van der Heijde D, et al. Scoring inflammatory activity of the spine by magnetic resonance imaging in ankylosing spondylitis: a multireader experiment. J Rheumatol. 2007 Apr; 34(4):862-870.
59. 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.
60. Wanders AJ, Landewé RB, Spoorenberg A, et al. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the available methods based on the Outcome Measures in Rheumatology Clinical Trials filter. Arthritis Rheum. 2004 Aug; 50(8):2622-2632.
61. 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.
62. Treynor JL. Market Efficiency and the Bean Jar Experiment. Financial Analysts Journal. 1987; 43(3):50-53.
63. 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.

Chapter 13

Samenvatting en conclusies

Samenvatting en conclusies

De term spondyloarthritis (SpA) wordt gebruikt om een groep van chronische reumatische aandoeningen op basis van ontstekingen te beschrijven die gemeenschappelijke kenmerken en symptomen hebben die als SpA-kenmerken worden aangeduid. Er is een sterke genetische overerving, vooral via de genetische marker HLA-B27 die vaker voorkomt in SpA dan in gezonde mensen of bij patiënten met andere ziekten. Patiënten met SpA die vooral klachten in de wervelkolom (bijv. chronische rugpijn) en in de bekkengewrichten (bijv. bilpijn) hebben, hebben axiale SpA. Bekkengewrichten worden ook sacroiliacale gewrichten (SI-gewrichten) genoemd. Patiënten met vooral klachten in de gewrichten van de ledematen (bijv. zwelling) of pijn in de omliggende weefsels zoals bijvoorbeeld peesaanhechtingen hebben perifere SpA. Axiale SpA bestaat uit twee vormen: radiografische axSpA die wordt gekenmerkt door schade op röntgenfoto's van de SI-gewrichten; en niet-radiografische axiale SpA zonder schade op röntgenfoto's van de SI-gewrichten. Deze syndromen worden beschreven aan de hand van classificatiecriteria (d.w.z. een lijst van tekenen en symptomen die een patiënt moet hebben om positief te worden beschouwd). Classificatiecriteria worden verondersteld het beste weer te geven wat de ziekte SpA is (de *'Gestalt').* Deze criteria worden gebruikt om patiënten in studies op te nemen die al een diagnose van SpA hebben gekregen van de reumatoloog.

Met het werk dat in dit proefschrift wordt gepresenteerd willen we bijdragen aan een betere kennis van SpA en duidelijk maken hoe beeldvorming van het axiale skelet (ruggengraat en sacroiliacale gewrichten) efficiënter kan worden gebruikt om patiënten in de loop van de tijd te monitoren. Onze belangrijkste bijdragen op dit gebied zijn de volgende: Ten eerste hebben we de kwestie van verkeerde classificatie aangepakt door de beoordeling van de classificatiecriteria van de SpondyloArthritis International Society (ASAS) voor axiale SpA en perifere SpA. Er is sprake van een verkeerde classificatie wanneer een patiënt voldoet aan de classificatiecriteria voor SpA, maar de reumatoloog de patiënt niet als zodanig diagnosticeert (en vice versa). Ten tweede hebben we voor het eerst de verschillende vormen van de *'Gestalt'* van axiale SpA (d.w.z. wat de ziekte werkelijk is) bepaald, onafhankelijk van de mening van de reumatoloog. Ten derde hebben we methoden voorgesteld om veranderingen beter op te kunnen sporen die door middel van beeldvorming (d.w.z. röntgenfoto's en door middel van magnetische resonantie) in de loop van de tijd worden opgespoord, evenals factoren die deze verandering bepalen. Ten vierde hebben we deze methoden gebruikt om meer inzicht te krijgen in het verband tussen ontsteking en schade in axiale SpA. Ten slotte hebben we bepaald welke beeldvormingsmethodes voorrang moeten krijgen bij het monitoren van patiënten in de klinische praktijk en in studies.

De studies die in dit proefschrift worden gepresenteerd zijn uitgevoerd in drie onafhankelijke cohorten: Het ASAS-cohort, het Spondyloarthritis Caught Early (SPACE) cohort en het Devenir des Spondyloarthropathies Indifférenciées Récentes (DESIR) cohort. Het ASAS-cohort is een studie waarbij patiënten aan één van de volgende twee criteria moesten voldoen: i) chronische rugpijn van onbekende oorsprong (geen definitieve diagnose) met een leeftijd van minder dan 45 jaar; ii) perifere artritis (gezwollen gewrichten) en/of enthesitis (pijn bij de aanhechting van pezen en banden) en/of dactylitis (gezwollen vinger/teen) bij afwezigheid van actuele rugpijn met vermoeden van SpA, maar geen definitieve diagnose. SPACE is een doorlopend cohort waarin opeenvolgende patiënten van ≥ 16 jaar met chronische rugpijn (≤2 jaar en beginnend <45 jaar) zijn opgenomen. DESIR is een cohort waarin volwassenen ouder dan 18 jaar en jonger dan 50 jaar uit 25 regionale centra in Frankrijk zijn opgenomen. Bij het begin hebben patiënten inflammatoire rugpijn (rugpijn die verbetert bij inspanning maar niet bij rust en geassocieerd is

met stijfheid) met meer dan 3 maanden en minder dan 3 jaar en symptomen die volgens de mening van de lokale behandelaar suggereren dat het om SpA gaat.

In dit laatste hoofdstuk zullen we de belangrijkste bevindingen van de in dit proefschrift gepresenteerde studies samenvatten.

Classificatie en *'Gestalt'* van spondyloarthritis

De studies in dit proefschrift, laten de goede prestaties van de ASAS SpA classificatiecriteria bij toetsing aan de diagnose van de reumatoloog zien. In hoofdstuk 2 wilden we weten wat de waarschijnlijkheid is dat een patiënt die aan het begin van het ASAS-cohort aan de ASAS-classificatiecriteria voldoet (d.w.z. die de volgens de criteria vereiste tekenen en symptomen heeft om als positief te worden beschouwd), 5 jaar later nog steeds geacht wordt de ziekte te hebben volgens de diagnose van de reumatoloog. We vonden dat de grote meerderheid (92%) van de patiënten die voldeden aan de axiale SpA of perifere SpA criteria bij het begin in feite gediagnosticeerd werden als SpA bij de follow-up. Deze waarschijnlijkheid was ook hoog voor patiënten die afzonderlijk aan de ASAS perifere SpA-criteria voldeden (90%). Patiënten kunnen voldoen aan de ASAS axiale SpA classificatiecriteria als ze sacroiliitis hebben op röntgenfoto's van het SI-gewricht en/of magnetische resonantie beeldvorming (MRI) van het SI-gewricht (gecombineerd is dat sacroiliitis op beeldvorming) plus ten minste één SpA-kenmerk (de zogenaamde 'beeldvormingsarm'); of als ze HLA-B27 positief zijn plus ten minste 2 extra SpA-kenmerken (de 'klinische arm'). Let wel, de kans op een diagnose door de reumatoloog op 5 jaar was even groot voor patiënten die aan de 'beeldvormingsarm' (86%) en de 'klinische arm' (88%) afzonderlijk voldeden aan het begin van het cohort, wat pleit tegen een verkeerde classificatie door de laatste, zoals eerder werd gesuggereerd. Tot slot waren bijna alle patiënten die een sacroiliitis hadden op de beeldvorming positief voor axiale SpA (98%), waarbij velen alleen sacroiliitis op MRI van het SI-gewricht hadden (62%). Aangezien de meeste van deze patiënten inderdaad bij de follow-up als axiale SpA werden gediagnosticeerd (95%), weerspiegelen onze gegevens de dominante plaats die sacroiliitis op MRI van het SI-gewricht inneemt in de criteria voor axiale SpA en de hoge diagnostische waarde die door de reumatologen aan dit kenmerk wordt toegekend.

De ASAS-classificatiecriteria zijn wereldwijd verder getest in verschillende cohorten. Sommige van deze cohorten verschillen in meerdere opzichten van het ASAS-cohort, waardoor unieke inzichten ontstaan in de toepasbaarheid van de criteria in een brede populatie van patiënten. In hoofdstuk 3 hebben we deze onderzoeken bekeken en die lieten een uitstekende sensitiviteit (sens; d.w.z. waarschijnlijkheid om SpA te hebben als de criteria positief zijn) en specificiteit (spec; d.w.z. waarschijnlijkheid om geen SpA te hebben als de criteria negatief zijn) zien van de ASAS SpA (axiale SpA en perifere SpA gecombineerd) criteria (73%; 88%; respectievelijk) . Goede prestaties werden ook gezien voor de axiale SpA-criteria (sens: 82%; spec: 87%), die robuust waren voor variaties in de studieopzet. Opmerkelijk is dat het splitsen van de axiale SpA-criteria in 'beeldvormingsarm' en 'klinische arm' de sensitiviteit in gevaar bracht (respectievelijk 26% en 23%), maar een zeer hoge specificiteit behield (97%; 94%). Deze bevinding is in lijn met hoofdstuk 2 en bewijst dat reumatologen van over de hele wereld patiënten van beide 'armen' herkennen als axiale SpA-patiënten, die daarom samen moeten worden gebruikt om te voorkomen dat er axiale SpA-patiënten ontbreken. De specificiteit van de pSpA-criteria was uitstekend (87%). De sensitiviteit was echter veel lager (62%), veroorzaakt door studies die enkel gebaseerd zijn op de aanwezigheid van gezwollen gewrichten, wat de relevantie van bijkomende

kenmerken benadrukt die zijn opgenomen in de ASAS perifere SpA criteria, namelijk enthesitis (pijn bij de aanhechting van pezen en banden) en/of dactylitis (gezwollen vinger/teen)

In de geneeskunde is de 'gouden standaard' de meest nauwkeurige test om te bepalen of een ziekte al dan niet aanwezig is. Net als in de vorige hoofdstukken zijn de classificatiecriteria van ASAS SpA ontwikkeld en gevalideerd aan de hand van het deskundig oordeel, dat wil zeggen het oordeel van de reumatoloog over de vraag of de patiënt de ziekte al dan niet had, als de 'gouden standaard'. Dit is een veel gebruikte benadering in de reumatologie, want in tegenstelling tot ziekten als diabetes of hypertensie, kunnen we SpA niet in één getal uitdrukken. Deze benadering brengt echter één fundamentele beperking met zich mee die afbreuk zou kunnen doen aan wat de criteria werkelijk meten: de cirkelredenering. Als criteria worden ontwikkeld aan de hand van de mening van een deskundige, en de deskundige vindt bepaalde kenmerken (bijv. sacroiliitis op de MRI) belangrijker dan andere, kunnen dergelijke kenmerken een te prominente plaats krijgen in de criteria. Latere testen met een deskundige diagnose kan leiden tot resultaten die eerder gebaseerd zijn op de overtuigingen van de deskundigen dan op een objectieve aanwezigheid van SpA. In feite hebben we gezien dat de aanwezigheid van sacroiliitis op de MRI bijna synoniem was aan een klinische diagnose van axSpA. Dat alleen maakt echter niet duidelijk of de dominantie van de ontsteking op de MRI correct is of juist niet. Met andere woorden, het kan zijn dat een dergelijke dominantie volgens de reumatoloog en vertaald naar de axiale SpA-criteria klopt met de ware *'Gestalt'* van axiale SpA. De enige manier om dit uitgangspunt te verifiëren is het uitsluiten van de mening van de reumatoloog bij de analyse. In hoofdstuk 4 hebben we een statistische techniek, namelijk latent class analysis (LCA), gebruikt om de *'Gestalt'* van axiale SpA te achterhalen, onafhankelijk van de mening van experts, door patiënten van SPACE en DESIR te verdelen in groepen met vergelijkbare kenmerken (fenotypen). We identificeerden drie afzonderlijke fenotypen, die we bestempelden als 'Pure axiale SpA' ('Axiale'), 'Axiale SpA met perifere tekens' ('IBP+Perifere', IBP=inflammatoire rugpijn) en 'Axiale SpA at risk' ('At Risk').

De 'Axiale' presentatie wordt gekenmerkt door een hoge waarschijnlijkheid van beeldvormingsafwijkingen in de wervelkolom en sacroiliacale gewrichten ('de as'), HLA-B27 positiviteit en mannelijke dominantie. Deze presentatie lijkt sterk op wat de meeste reumatologen denken dat axiale SpA is. Het is dan ook niet verwonderlijk dat de ASAS axiale SpA classificatiecriteria (ontwikkeld door deskundigen) bijna volledig de 'Axiale' patiënten omvatten (98% in SPACE en 93% in DESIR). Opgemerkt moet worden dat het 'axiale' fenotype geen onderscheid maakte tussen niet-radiografische axSpA en radiografische axSpA, wat aantoont dat dit geen afzonderlijke entiteiten zijn zoals eerder werd gedacht. De identificatie van de 'Axiale' presentatie, met dominante beeldvormingsafwijkingen (inclusief sacroiliitis op MRI), onafhankelijk van de mening van de reumatoloog, lijkt te suggereren dat dominante beeldvorming inderdaad in overeenstemming is met de *'Gestalt'* van axiale SpA. Deze presentatie komt echter overeen met minder dan 20% van de patiënten in SPACE en DESIR. De 'IBP+Perifere' presentatie wordt gedefinieerd door de aanwezigheid van IBP (100%) in combinatie met perifere tekenen en symptomen (bijv. gezwollen gewrichten). Deze axiale SpA-patiënten (meestal vrouwen) met een zeer lage kans op afwijkingen op beeldvorming en meestal HLA-B27 negatief worden vaak gezien door de deskundige clinicus in de klinische praktijk. Het is echter onwaarschijnlijk dat axiale SpA-patiënten met de 'IBP+Perifere'-presentatie positief zijn voor de ASAS-classificatiecriteria die vereisen dat aan HLA-B27-positiviteit of sacroiliitis op beeldvorming wordt voldaan. Deze bevinding ondersteunt het standpunt dat bij de ontwikkeling van de ASAS axiale SpA-indeling inderdaad sprake was van een ongewenste cirkelredenering.

Het moderne paradigma van vroegtijdige diagnose en vroegtijdige behandeling heeft ongetwijfeld veel voordelen voor de patiënten opgeleverd, maar brengt ook belangrijke uitdagingen met zich mee. Axiale SpA is moeilijk te diagnosticeren en reumatologen vertrouwen op patroonherkenning voor de identificatie ervan. Het SpA-patroon is minder voor de hand liggend bij een vroege ziekte wanneer 'typische' kenmerken nog ontbreken, wat tot onzekerheid leidt ('grijze zone'). De ervaren clinicus zal een onderscheid maken tussen patiënten die geen axiale SpA hebben en patiënten die de ziekte wel hebben en zal op passende wijze omgaan met degenen voor wie de diagnose nog niet duidelijk is. Anderen kunnen echter in de verleiding komen om bij de behandeling van onzekere of moeilijke gevallen classificatiecriteria toe te passen om diagnoses te stellen (bijvoorbeeld axiale SpA versus geen axiale SpA) die geen grijze zones toelaten. We hebben aangetoond dat dergelijke clinici een ernstig risico lopen op 'overdiagnose' en dus op 'overbehandeling' van personen met een onterechte diagnose van axiale SpA. Wij hebben de patiënten die het meeste risico lopen op een 'overdiagnose' gelabeld als 'At Risk' en deze patiënten voor het eerst beschreven. Sommige risicofactoren voor axiale SpA (namelijk het hebben van familieleden met de ziekte en HLA-B27) zijn de belangrijkste kenmerken die vaak in verband worden gebracht met inflammatoire rugpijn, maar slechts sporadisch met andere SpA-kenmerken. Het is gemakkelijk te begrijpen waarom deze patiënten vaak voldoen aan de ASAS axiale SpA criteria, met name de 'klinische arm' die vereist dat HLA-B27 positief is naast twee SpA kenmerken (familiegeschiedenis en inflammatoire rugpijn). Dit is echter geen probleem als de criteria pas na een klinische diagnose op de juiste wijze worden gebruikt, zodat voortdurende inspanningen op het gebied van educatie essentieel zijn om 'overdiagnose' en 'overbehandeling' te voorkomen.

Beoordeling van de radiografische progressie bij de sacro-iliacale gewrichten

Definitieve schade op röntgenfoto's van het SI-gewricht wordt volgens het gewijzigde classificatiesysteem van New York (mNY) gedefinieerd als de aanwezigheid van bilaterale graad 2 of eenzijdige graad 3 of 4 "sacroiliitis" ("mNY-positief"), wat een belangrijk kenmerk is van de indeling van radiografische axiale SpA. Het is echter gebleken dat radiografische 'sacroiliitis' een onbetrouwbare bevinding is, dat wil zeggen dat verschillende lezers het vaak niet eens zijn of een bepaalde patiënt mNY-negatief of positief is. Het bepalen van de progressie van mNY-negatief naar mNY-positief, in de loop van de tijd, is nog moeilijker. In hoofdstuk 5 hebben we twee röntgenfoto's van het SI-gewricht vergeleken, door ongetrainde lokale lezers, met enkele jaren ertussen bij patiënten met een verdenking op SpA uit het ASAS-cohort. Van de 357 geïncludeerde patiënten was 17% (62/357) mNY-positief bij het begin. Bij de follow-up steeg dit percentage tot 22% (80/357). Meer dan de helft (36/62) van de patiënten die bij het begin als mNY-positief werden beschouwd, werd echter bij de follow-up als mNY-negatief beoordeeld. Ervan uitgaande dat schade in de sacroiliacale gewrichten een inherent onomkeerbaar kenmerk is, en wetende dat de lezers zich bewust waren van de volgorde waarin de röntgenfoto's werden gemaakt, zijn deze 'verbeteringen' zeer moeilijk te begrijpen. Deze gegevens tonen aan dat in de klinische praktijk, zoals in het ASAS-cohort, het onderscheid tussen een mNY-negatieve en mNY-positieve röntgenfoto arbitrair is (te veel fout) en dus van weinig waarde is voor de reumatoloog.

Dezelfde conclusie geldt niet noodzakelijkerwijs voor het gebruik van röntgenfoto's van het SI-gewricht in klinisch onderzoek, waar strategieën om de fout te verminderen kunnen worden toegepast. Het laten lezen van films door getrainde lezers en het bepalen van de eindscores door een 'overeenkomst-algoritme' (bijv. 2 van de 3 lezers zijn het erover eens dat de patiënt mNY-

positief is) zijn enkele voorbeelden hiervan. Maar zelfs dergelijke strategieën kunnen de meetfout niet volledig elimineren. Daarom zijn geschikte methoden voor het berekenen van de progressie van cruciaal belang om de effecten van de meetfout tot een minimum te beperken. In hoofdstuk 6 hebben we de verandering tussen mNY-negatief en mNY-positief na 5 jaar in het DESIR-cohort geëvalueerd. In tegenstelling tot het ASAS-cohort waren de lezers niet op de hoogte van de volgorde in de tijd en werden de scores gedefinieerd door de '2 uit 3'-regel. Bij het begin waren 62 (15%) van de 416 opgenomen patiënten mNY-positief. Van de 354 mNY-negatieve patiënten bij het begin veranderden 24 (6,8%) na 5 jaar in mNY-positief. We bestempelden dit cijfer als 'Ruwe progressie'. Deze eenvoudige methode gaat ervan uit dat de nulmeting foutloos is en dat een verandering in de tegenovergestelde richting (hier: 3/62: 4,8%) kan worden genegeerd, wat niet juist is. Er is een andere methode voorgesteld die het percentage negatieve verandering in aanmerking neemt door het af te trekken van het percentage positieve verandering (6,8%-4,8%: 2%). Deze methode is echter ook niet geschikt omdat het impliceert dat 'verslechtering' alleen kan plaatsvinden bij patiënten die mNY-negatief zijn op de basislijn en 'verbetering' alleen bij mNY-positieve patiënten, ook al weten de lezers niet welke de eerste röntgenfoto was.

Daarom hebben we een derde methode voorgesteld die we 'netto progressie' noemden en waarmee zowel 'positieve verandering' als 'negatieve veranderingen' worden 'toegestaan'. De 'netto progressie' voor een groep patiënten wordt als volgt berekend: aantal positieve veranderingen min aantal negatieve veranderingen gedeeld door alle patiënten [(24-3)/416=5%]. De 'netto progressie' is de minst vertekenende methode, omdat deze het meeste rekening houdt met de meetfout. Hoewel we voor de beschrijving van deze methode gebruik hebben gemaakt van radiografische progressie bij de sacroiliacale gewrichten in axiale SpA, strekt de toepassing zich uit tot alle voorbeelden waarbij beeldvormingsscores op structurele schade worden verkregen onder geblindeerde omstandigheden. Er moet echter worden opgemerkt dat deze methode impliceert dat de resultaten onomkeerbaar zijn en over korte perioden worden geëvalueerd, aangezien echte negatieve verandering ('reparatie') niet kan worden uitgesloten met een langere follow-up. Verdere studies moeten ons helpen om de betekenis van 'negatieve veranderingen' in andere omgevingen dan die met onomkeerbare schade te begrijpen.

Relatie tussen ontsteking en structurele schade

Patiënten met axiale SpA laten radiografische progressie in verschillende mate zien. Bij sommige patiënten ontwikkelen zich snel structurele veranderingen, terwijl bij andere dat misschien nooit zal gebeuren. Het identificeren van degenen die meer kans hebben om schade te ontwikkelen (d.w.z. prognostische gelaagdheid), helpt de clinicus dus om beslissingen te nemen over de behandeling. Er is een aanzienlijke inspanning geleverd om de oorzaken van de schade in axSpA te bestuderen, waarbij de ontsteking veel aandacht krijgt. Ten tijde van het begin van dit proefschrift was er al solide bewijs dat de ontsteking leidt tot radiografische progressie ter hoogte van de wervelkolom. Hetzelfde zou verwacht kunnen worden bij de sacroiliacale gewrichten, maar het bewijs was op dat moment nog schaars.

In hoofdstuk 7 vonden we dat een ontsteking op MRI in het SI-gewricht de kans op schade aan de röntgenfoto's van het SI-gewricht 5 jaar later in het DESIR-cohort verhoogt, met beelden die door getrainde centrale lezers worden gelezen. C-reactief eiwit (CRP) is een andere objectieve marker van de ontsteking gemeten in bloedmonsters. CRP had ook een grote invloed op de

waarschijnlijkheid van progressie, vooral bij patiënten die HLA-B27-positief waren. Zo hadden patiënten die HLA-B27-negatief waren en een normale CRP en een negatieve MRI van het SI-gewricht een kans van slechts 1% om van mNY-negatief naar mNY-positief te gaan. Deze kans was daarentegen achttien keer zo groot (18%) als alle drie de variabelen positief waren.

Hoewel de resultaten van dit hoofdstuk 7 methodologisch robuust zijn, zijn ze moeilijk te vertalen naar de klinische praktijk waar beelden niet worden gelezen door meerdere getrainde lezers die zich niet bewust zijn van de chronologie van de beelden. Het was dus niet duidelijk of een ontsteking op de MRI van het SI-gewricht die in de klinische praktijk wordt gezien, de kans op schade vergroot, net zoals werd gevonden met getrainde lezers. In hoofdstuk 8, met behulp van lokaal gelezen gegevens van de ASAS- en DESIR-cohorten, vonden we dat, ondanks de grotere onbetrouwbaarheid, er inderdaad een duidelijke prognostische waarde was voor een objectieve ontsteking op de MRI in beide cohorten.

De laatste tijd is er een toenemende belangstelling voor het gebruik van MRI, niet alleen voor het meten van de ontsteking, maar ook voor structurele schade. In hoofdstuk 9 hebben we voor het eerst aangetoond dat de aanwezigheid van ontstekingen op MRI de ontwikkeling van verschillende structurele letsels op MRI 5 jaar later voorspelt (bv. vetstapeling en erosies). Onze resultaten zijn echter nog steeds moeilijk te begrijpen omdat de werkelijke betekenis van deze afwijkingen bij axiale SpA nog niet helemaal duidelijk is. Een hypothese stelt dat de ontsteking in axiale SpA schommelt en dat de botproliferatie (bv. vorming van verbening in de wervelkolom) een herstelproces is dat pas begint wanneer de ontsteking verdwijnt en vetstapeling (d.w.z. abnormale vervanging van normaal bot door vet) een tussenstap is. Aan de andere kant, als de ontsteking aanhoudt, is herstel niet mogelijk en overheerst botdestructie (bijv. erosies). Inzicht in de complexe relatie tussen ontsteking, botvorming en botdestructie in axiale SpA heeft mogelijk belangrijke therapeutische gevolgen. Aangezien axiale SpA echter een langzaam voortschrijdende ziekte is, zijn langetermijnstudies nodig waarbij zich een aantal uitdagingen aandienen die we in de volgende paragraaf behandelen.

Multilevel-analyse van beeldvormingsgegevens

Onderzoekers die langetermijnstudies ontwerpen, willen meestal niet enkele jaren wachten voordat hun gegevens kunnen worden geanalyseerd. Een gangbare praktijk is om de studie al na een bepaalde periode van gegevensverzameling te analyseren. Zo is het de bedoeling dat patiënten die in het DESIR-cohort zijn opgenomen tot 10 jaar worden opgevolgd, maar het was al mogelijk om de beschikbare gegevens voor 5 jaar te gebruiken om verschillende onderzoeksvragen te behandelen. In deze setting worden de beelden meestal in 'leesrondes' gelezen. Dat wil zeggen dat bij elke analyse alle op dat moment beschikbare beeldvormingsgegevens (bijv. begin en één vervolgbezoek) worden gescoord. Het proces wordt dan in elke analyse herhaald als er meer gegevens worden verzameld. In DESIR werden tot nu toe beeldvormingsgegevens verzameld bij het begin, na 1 jaar, 2 jaar en 5 jaar en gelezen door getrainde centrale lezers in 3 opeenvolgende 'leesrondes'.

In deze 'leesrondes' worden grote hoeveelheden gegevens gegenereerd. Meestal selecteren onderzoekers de gegevens die het meest geschikt zijn voor hun analyse. In de vorige hoofdstukken wilden we bijvoorbeeld de 5-jarige beeldvormingsgegevens in DESIR analyseren, dus we gebruikten alleen 'ronde 3', de enige met 5-jaars gegevens. Dergelijke keuzes kunnen echter leiden tot vertekening en verlies van informatie (wat te doen met de gegevens van de andere rondes?). In theorie beschermt het opnemen van alle gegevens zonder keuzes tegen

vertekening, aangezien de onderzoeker niet hoeft in te grijpen in de selectie en berekening van de gegevens. Het nadeel is het toevoegen van enige variabiliteit, wat kan leiden tot een lagere nauwkeurigheid.

Het combineren van alle beschikbare beeldvormingsinformatie is eerder aangetoond als een robuuste aanpak voor het analyseren van lange termijn beeldvormingsgegevens bij patiënten met reumatoïde artritis: een zogenaamde 'geïntegreerde analyse'. In hoofdstuk 10 hebben we aangetoond dat deze methode de snelheid van verandering van de beeldvormingsresultaten van patiënten met axiale SpA kan bepalen, zonder dat dit ten koste gaat van de nauwkeurigheid. Integendeel, voor uitkomsten die in de loop van de tijd zelden voorkwamen, werd de precisie zelfs verbeterd. Deze aanpak kan dus van bijzonder belang zijn bij studies met een lange-termijn follow-up, en/of wanneer de uitkomsten naar verwachting in de loop van de tijd niet vaak zullen voorkomen. Onze resultaten zijn gebaseerd op het bovengenoemde onderzoek naar reumatische artritis, en kunnen net zo goed van toepassing zijn op andere ziekten. In feite zijn lange termijn studies die de resultaten van de beeldvorming evalueren zeer relevant op het gebied van de reumatologie. In de loop der jaren zijn verschillende cohorten gestart om enkele van de meest fundamentele en voor vele aandoeningen toepasbare onderzoeksvragen aan te pakken: Wat is het natuurlijk beloop van de ziekte? Zijn we in staat om de patiënten die een vrij goedaardig beloop zullen hebben te onderscheiden van de patiënten met een slechtere prognose? Kunnen we in dit proces ingrijpen en uiteindelijk duurzame therapeutische voordelen bewerkstelligen? De 'geïntegreerde analyse' kan onderzoekers helpen deze vragen op te lossen in toekomstige studies die zich richten op beeldvorming.

Er zijn verschillende beeldvormingsuitkomsten ontwikkeld om de ontsteking en de schade bij patiënten met axiale SpA te beoordelen. Vergelijkingen van hun vermogen om veranderingen over de tijd te meten (d.w.z. hun gevoeligheid voor verandering) zijn echter schaars, zodat het onduidelijk blijft welke uitkomstmaten het beste zijn om patiënten met axiale SpA op te volgen en te monitoren. Daarom hebben we in hoofdstuk 11 de gevoeligheid voor verandering van een aantal scores bestudeerd. We vonden dat beeldvormingsafwijkingen schaars waren en nauwelijks veranderden in de periode van 5 jaar op het niveau van de wervelkolom, ongeacht de uitkomstmaat en het type beeldvorming. Het tegenovergestelde werd waargenomen op het niveau van de sacroiliacale gewrichten, hetgeen verwacht kon worden omdat de schade meestal begint bij de sacroiliacale gewrichten waarbij de wervelkolom later betrokken raakt, en dat laatste zelfs slechts bij een deel van de patiënten. We hebben ook voor het eerst aangetoond dat metingen van schade op de MRI gevoeliger zijn voor verandering dan metingen op de 'conventionele' röntgenfoto's van de SI-gewrichten. Deze bevinding kan worden gebruikt voor het plannen van toekomstige studies die de progressie van structurele schade op het niveau van de sacroiliacale gewrichten evalueren, inclusief het testen van interventies die gericht zijn op het remmen van deze progressie.

Slotopmerkingen

In dit proefschrift hebben we innovatieve oplossingen toegepast op enkele van de meest uitdagende vragen op het gebied van SpA. Hierdoor hebben we een goed inzicht gekregen in wat SpA werkelijk is, wat zich in de toekomst kan vertalen in een betere herkenning van de ziekte in de klinische praktijk en in het onderzoek. Aangezien SpA een langzaam voortschrijdende ziekte is, zijn er enkele jaren nodig om betekenisvolle veranderingen te zien in de beeldvorming van de wervelkolom en de sacroiliacale gewrichten, wat methodologische uitdagingen met zich

meebrengt. We hebben aangetoond dat een doordachte analytische aanpak nuttig is bij het bepalen van de verandering in de loop van de tijd van deze uitkomsten en van de factoren die de verandering beïnvloeden. Er worden inspanningen geleverd om de uitkomstmeting in axiale SpA verder te verbeteren, inclusief de ontwikkeling van nieuwe beeldvormingstechnieken, die kunnen profiteren van onze voorgestelde oplossingen voor het scoren op lange termijn van de beeldvorming.

LIST OF PUBLICATIONS

1. Machado P, Bogas M, Ribeiro A, Costa J, Neto A, **Sepriano A**, Raposo A, Cravo AR, Vilar A, Furtado C, Ambrósio C, Miguel C, Vaz C, Catita C, Nour D, Araújo D, Vieira-Sousa E, Teixeira F, Brandão F, Canhão H, Cordeiro I, Gonçalves I, Ferreira J, Fonseca JE, da Silva JA, Romeu J, Ferreira J, Costa L, Maurício L, Cunha-Miranda L, Parente M, Coutinho M, Cruz M, Oliveira M, Salvador MJ, Santos MJ, Pinto P, Valente P, Abreu P, Roque R, Ramiro S, Capela S, Las V, Barcelos A. 2011 Portuguese recommendations for the use of biological therapies in patients with psoriatic arthritis. Acta Reumatol Port. 2012 Jan-Mar;37(1):26-39

2. Rodrigues AM, Caetano-Lopes J, Vale AC, Aleixo I, Pena AS, Faustino A, **Sepriano A**, Polido-Pereira J, Vieira-Sousa E, Lucas R, Romeu JC, Monteiro J, Vaz MF, Fonseca JE, Canhão H. Smoking is a predictor of worse trabecular mechanical performance in hip fragility fracture patients. J Bone Miner Metab. 2012 Nov;30(6):692-9

3. Rodrigues AM, Caetano-Lopes J, Vale AC, Vidal B, Lopes A, Aleixo I, Polido-Pereira J, **Sepriano A**, Perpétuo IP, Monteiro J, Vaz MF, Fonseca JE, Canhão H. Low osteocalcin/collagen type I bone gene expression ratio is associated with hip fragility fractures. Bone. 2012 Dec;51(6):981-9

4. Araújo F, Silva I, **Sepriano A**, Nero P, Branco JC. Off-label use of rituximab in systemic rheumatic diseases: case series and review. Acta Reumatol Port. 2013 Oct-Dec;38(4):290-4.

5. Fonseca JE, Gonçalves J, Araújo F, Cordeiro I, Teixeira F, Canhão H, da Silva JA, Garcês S, Miranda LC, Ramiro S, Roxo A, Pimentel-Santos FM, Tavares V, Neto A, **Sepriano A**, Malcata A, Faustino A, Silva C, Ambrósio C, Duarte C, Miguel C, Barcelos F, Santos H, Cunha I, Ramos JC, Gomes JA, Pimentão JB, Costa L, Maurício L, Silva M, Bernardes M, Bogas M, Coelho PC, Monteiro P, Aguiar R, André R, Leitão R, Pimenta S, Meirinhos T, Fernandes S, Las V, Castelão W. The Portuguese Society of Rheumatology position paper on the use of biosimilars. Acta Reumatol Port. 2014 Jan-Mar;39(1):60-71

6. **Sepriano AR**, Pereira Pombo de Araújo FC, Mourão de Almeida Lourenço Dias PM, da Cunha Branco J. Normocalcemic Primary Hyperparathyroidism presenting as Calcinosis Cutis. J Clin Rheumatol. 2014 Sep;20(6):330-1

7. Araújo F, Silva I, **Sepriano A**, Reizinho C, Marques L, Nero P, Branco JC. Cervical spine involvement as initial manifestation of Rheumatoid Arthritis: a case report. Acta Reumatol Port. 2015 Jan-Mar;40(1):64-7

8. **Sepriano A**, Roman-Blas JA, Little RD, Pimentel-Santos F, Arribas JM, Largo R, Branco JC, Herrero-Beaumont G. DXA in the Assessment of Subchondral Bone Mineral Density in Knee Osteoarthritis – a Semi-Standardized Protocol after Systematic Review. Semin Arthritis Rheum. 2015 Dec;45(3):275-83

9. Rodrigues AM, Gouveia N, da Costa LP, Eusébio M, Ramiro S, Machado P, Mourão AF, Silva I, Laires P, **Sepriano A**, Araújo F, Coelho PS, Gonçalves S, Zhao A, Fonseca JE, de Almeida JM, Tavares V, da Silva JA, Barros H, Cerol J, Mendes J, Carmona L, Canhão H, Branco JC. EpiReumaPt-

the study of rheumatic and musculoskeletal diseases in Portugal: a detailed view of the methodology. Acta Reumatol Port. 2015 Apr-Jun;40(2):110-24

10. Branco JC, Rodrigues AM, Gouveia N, Eusébio M, Ramiro S, Machado PM, da Costa LP, Mourão AF, Silva I, Laires P, **Sepriano A**, Araújo F, Gonçalves S, Coelho PS, Tavares V, Cerol J, Mendes JM, Carmona L, Canhão H, EpiReumaPt study group. Prevalence and Physical and Mental Health Patterns of Rheumatic and Musculoskeletal Diseases in Portugal: Results from Epireumapt, A National Health Survey. RMD Open 2016;2:e000166

11. **Sepriano A**, Landewé R, van der Heijde D, Sieper J, Akkoc N, Brandt J, Braun J, Collantes-Estevez E, Dougados M, Fitzgerald O, Huang F, Gu J, Kirazli Y, Maksymowych WP, Marzo-Ortega H, Olivieri I, Ozgocmen S, Roussou E, Scarpato S, Sørensen IJ, Valle-Oñate R, Van den Bosch F, van der Horst-Bruinsma I, Weber U, Wei J, Rudwaleit M, on behalf of ASAS. 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):1034-42

12. **Sepriano A**, Rudwaleit M, Sieper J, van den Berg R, Landewé R, van der Heijde D. Five-year follow-up of radiographic sacroiliitis: progression as well as improvement? Ann Rheum Dis. 2016 Jun;75(6):1262-3

13. **Sepriano A**, Ramiro S, van der Heijde D, Ávila-Ribeiro P, Fonseca R, Borges J, Teixeira L, Carvalho P, Cerqueira M, Neves J, Meirinhos T, Barcelos A, Sequeira G, Salvador MJ, Canas da Silva JA, Santos H, Bernardes M, Vieira-Sousa E, Canhão H, Branco JC, Pimentel-Santos F, Landewé R. Effect of comedication with conventional synthetic DMARDs on TNF inhibitors-retention in patients with spondyloarthritis: A prospective cohort. Arthritis Rheumatol. 2016 Nov;68(11):2671-2679

14. Vieira Sousa E, **Sepriano A**, Pimentel Santos F. Investigator initiated trials in Portuguese rheumatology. Acta Reumatol Port. 2016 Apr-Jun;41(2):101-3

15. Regel A, **Sepriano A**, Baraliakos X, van der Heijde D, Braun J, Landewé R, Van den Bosch F, Falzon L, Ramiro S. Efficacy and safety of non-pharmacological and non-biological pharmacological treatment: a systematic literature review informing the 2016 update of the ASAS/EULAR recommendations for the management of axial spondyloarthritis. RMD Open 2017;3:e000397

16. **Sepriano A**, Regel A, van der Heijde D, Braun J, Baraliakos X, Landewé R, Van den Bosch F, Falzon L, Ramiro S. Efficacy and safety of biological and targeted-synthetic DMARDs: a systematic literature review informing the 2016 update of the ASAS/EULAR recommendations for the management of axial spondyloarthritis. RMD Open 2017;3:e000396

17. van der Heijde D, Ramiro S, Landewé R, Baraliakos X, Van den Bosch F, **Sepriano A**, Regel A, Ciurea A, Dagfinrud H, Dougados M, van Gaalen F, Géher P, van der Horst-Bruinsma I, Inman RD, Jongkees M, Kiltz U, Kvien TK, Machado PM, Marzo-Ortega H, Moltó A, Navarro-Compán V, Ozgocmen S, Pimentel-Santos FM, Reveille J, Rudwaleit M, Sieper J, Sampaio-Barros P, Wiek D, Braun J. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis. 2017 Jun;76(6):978-991

18. **Sepriano A**, Rubio R, Ramiro S, Landewé R, van der Heijde D. 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

19. Ramiro S, **Sepriano A**, Chatzidionysiou K, Nam JL, Smolen JS, van der Heijde D, Dougados M, van Vollenhoven R, Bijlsma JW, Burmester GR, Scholte-Voshaar M, Falzon L, Landewé RBM. Safety of synthetic and biological DMARDs: a systematic literature review informing the 2016 update of the EULAR recommendations for management of rheumatoid arthritis. Ann Rheum Dis. 2017 Jun;76(6):1101-1136

20. Duarte C, Sousa-Neves J, Águeda A, Ribeiro P, Daniel A, Eugénio G, Serra S, Araújo F, Barcelos A, Filipe B, Bernardes M, Canhão H, Cerqueira M, Capela S, Cordeiro A, Costa F, Costa L, Cruz M, Cunha Miranda L, Duarte C, Falcão S, Faria D, Figueira R, Freitas JP, Gonçalves MJ, Madruga Dias J, Melo Gomes J, Mourão AF, Neto A, Oliveira Ramos F, Pimenta S, Pinto P, Polido Pereira J, Ponte C, Ramos J, Rodrigues A, Santos H, Santos MJ, **Sepriano A**, Silva C, Tavares Costa J, Teixeira F, Teixeira V, Valente P, Vieira-Sousa E, Barros R, Abreu P, Fonseca JE. Portuguese Recommendations for the use of biological therapies in patients with rheumatoid arthritis- 2016 update. Acta Reumatol Port. 2017 Jun 1;42(2)(Apr-Jun):112-126

21. Duarte AC, Santos-Faria D, Gonçalves MJ, **Sepriano A**, Mourão AF, Duarte C, Neves JS, Águeda AF, Ribeiro PÁ, Daniel A, Neto A, Cordeiro A, Rodrigues A, Barcelos A, Silva C, Ponte C, Vieira-Sousa E, Teixeira F, Oliveira-Ramos F, Araújo F, Barcelos F, Canhão H, Santos H, Ramos J, Pereira J, Costa J, Melo Gomes JA, Cunha-Miranda L, Costa L, Cerqueira M, Cruz M, Santos MJ, Bernardes M, Oliveira P, Abreu P, Figueira R, Barros R, Falcão S, Pinto P, Pimenta S, Capela S, Teixeira V, Fonseca JE. Portuguese recommendations for the use of methotrexate in rheumatic diseases - 2016 update. Acta Reumatol Port. 2017 Apr-Jun;42(2):127-140

22. Dougados M, **Sepriano A**, Moltó A, van Lunteren M, Ramiro S, de Hooge M, van den Berg R, Navarro Compán V, Demattei C, Landewé R, van der Heijde D. 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

23. Araújo FC, **Sepriano A**, Teixeira F, Jesus D, Rocha TM, Martins P, Tenazinha C, Cordeiro A, Mourão AF, Silva C, Vaz C, Duarte C, Ponte C, Dos Santos FP, Canhão H, Santos H, Pimentão JB, da Silva JC, Pereira J, da Silva JAP, Miranda LC, Oliveira M, Saavedra MJ, Gonçalves P, Falcão S, Capela S, Fonseca JE. The Portuguese Society of Rheumatology position paper on the use of biosimilars - 2017 update. Acta Reumatol Port. 2017 Jul-Sep;42(3):219-228

24. Cruz EB, Ramiro S, Machado P, Sousa S, Aguiar R, **Sepriano A**, Manica SR7, Kiltz U, Branco JC, Pimentel-Santos FM. Translation and cross-cultural adaptation of the ASAS Health Index and ASAS Environmental Factors Item Set into European Portuguese Language. Acta Reumatol Port. 2017 Jul-Sep;42(3):256-262

25. Rodríguez-Carrio J, Putrik P, **Sepriano A**, Moltó A, Nikiphorou E, Gossec L, Kvien TK, Ramiro S. Improving the peer review skills of young rheumatologists and researchers in rheumatology: the EMEUNET Peer Review Mentoring Program. RMD Open 2018;4:e000619. doi: 10.1136/rmdopen-2017-000619

26. Duftner C, Dejaco C, **Sepriano A**, Falzon L, Schmidt WA, Ramiro S. Imaging in diagnosis, outcome prediction and monitoring of large vessel vasculitis: A systematic literature review and meta-Analysis informing the EULAR recommendations. RMD Open. 2018 Feb 2;4(1):e000612. doi: 10.1136/rmdopen-2017-000612

27. Rodrigues Manica S, **Sepriano A**, Ramiro S, Pimentel Santos F, Putrik P, Nikiphorou E, Norton S, Moltó A, Dougados M, van der Heijde D, Landewé RBM, van den Bosch FE, Boonen A. Work participation in spondyloarthritis across countries: analysis from the ASAS-COMOSPA study. Ann Rheum Dis. 2018 Sep;77(9):1303-1310

28. Ramiro S, Claudepierre P, **Sepriano A**, van Lunteren M, Moltó A, Feydy A, Antonietta d'Agostino M, Loeuille D, Dougados M, Reijnierse M, van der Heijde D. 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

29. van Lunteren M, **Sepriano A**, Landewé R, Sieper J, Rudwaleit M, van der Heijde D, van Gaalen F. Do ethnicity, degree of family relationship, and the spondyloarthritis subtype in affected relatives influence the association between a positive family history for spondyloarthritis and HLA-B27 carriership? Results from the worldwide ASAS cohort. Arthritis Res Ther. 2018 Aug 3;20(1):166

30. Stamm TA, Machold KP, Aletaha D, Alasti F, Lipsky P, Pisetsky D, Landewé R, van der Heijde D, **Sepriano A**, Aringer M, Boumpas D, Burmester G, Cutolo M, Ebner W, Graninger W, Huizinga T, Schett G, Schulze-Koops H, Tak PP, Martin-Mola E, Breedveld F, Smolen J. Induction of sustained remission in early inflammatory arthritis with the combination of infliximab plus methotrexate: the DINORA trial. Arthritis Res Ther. 2018 Aug 9;20(1):174

31. Ramiro S, van der Heijde D, **Sepriano A**, van Lunteren M, Moltó A, Feydy A, d'Agostino MA, Loeuille D, Dougados M, Reijnierse M, Claudepierre P. Spinal radiographic progression in early axial Spondyloarthritis: Five-year results from the DESIR cohort. Arthritis Care Res (Hoboken). 2019 Dec;71(12):1678-1684

32. **Sepriano A**, Ramiro S, Landewé R, Dougados M, van der Heijde D, Rudwaleit M. 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.

33. Boone NW, **Sepriano A**, van der Kuy PH, Janknegt R, Peeters R5, Landewé RBM. Cotreatment with methotrexate in routine care patients with rheumatoid arthritis receiving biological treatment yields better outcomes over time. RMD Open. 2019 Jan 22;5(1):e000836

34. Gomes JL, **Sepriano A**, Eusébio M, Serra S, Fonseca JE, Saavedra MJ, Cunha-Miranda L, Silva C5, Bernardes M, Rosa-Gonçalves D, Costa J, Castelão W, Branco JC, Santos MJ. Predictors and causes of first-line biologic agent discontinuation in rheumatoid arthritis: data from Reuma.pt. Acta Reumatol Port. 2019 Jan-Mar;44(1):57-64

35. **Sepriano A**, Ramiro S, Araújo FC, Machado PM, Rodrigues AM, Gouveia N, Eusébio M, Canhão H, Branco J. Performance of referral strategies for spondyloarthritis: a population-based nationwide study. Rheumatology (Oxford). 2019 Jun 1;58(6):1086-1094

36. **Sepriano A**, Ramiro S, Landewé R, Dougados M, van der Heijde D. Percentage of progressors in imaging: can we ignore regressors? RMD Open. RMD Open. 2019 Feb 26;5(1):e000848.

37. Bergstra SA, **Sepriano A**, Ramiro S, Landewé R. Three handy tips and a practical guide to improve your propensity score models. RMD Open. 2019 May 1;5(1):e000953

38. Maksymowych WP, FitzGerald O, Østergaard M, Homik J, van der Heijde D, Lambert RG, Elkayam O, Ramiro S, Thorne JC, Larché MJ, Ferraccioli G, Backhaus M, Burmester GR, Boire G, Combe B, Schaeverbeke T, Saraux A, Dougados M, Rossini M, Govoni M, Sinigaglia L, Cantagrel A, Barnabe C, Bingham CO 3rd, Tak PP, van Schaardenburg D, Berner Hammer H, Paschke J, Dadashova R, Hutchings E, **Sepriano A**, Landewé R. Outcomes and Findings of the International Rheumatoid Arthritis (RA) BIODAM Cohort for Validation of Soluble Biomarkers in RA. J Rheumatol. 2019 Sep 1. pii: jrheum.190302

39. **Sepriano A**, Ramiro S, FitzGerald O, Østergaard M, Homik J, van der Heijde D, Elkayam O, Thorne JC, Larché MJ, Ferraccioli G, Backhaus M, Burmester GR, Boire G, Combe B, Schaeverbeke T, Saraux A, Dougados M, Rossini M, Govoni M, Sinigaglia L, Cantagrel A, Barnabe C, Bingham CO 3rd, Tak PP, van Schaardenburg D, Berner Hammer H, Paschke J, Dadashova R, Hutchings E, Landewé R, Maksymowych WP. Adherence to Treat-to-target Management in Rheumatoid Arthritis and Associated Factors: Data from the International RA BIODAM Cohort. J Rheumatol. 2019 Sep 15. pii: jrheum.190303

40. van Lunteren M, van der Heijde D, **Sepriano A**, Berg IJ, Dougados M, Gossec L, Jacobsson L, Ramonda R, Rudwaleit M, Sieper J, Landewé R, van Gaalen FA. 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

41. Boone NW, **Sepriano A**, van der Kuy PH, Janknegt R, Peeters R, Landewé RBM. Routine Assessment of Patient Index Data 3 (RAPID3) alone is insufficient to monitor disease activity in rheumatoid arthritis in clinical practice. RMD Open. 2019 Nov 28;5(2):e001050

42. Smolen JS, Landewé RBM, Bijlsma JWJ, Burmester GR, Dougados M, Kerschbaumer A, McInnes IB, **Sepriano A**, van Vollenhoven RF, de Wit M, Aletaha D, Aringer M, Askling J, Balsa A, Boers M, den Broeder AA, Buch MH, Buttgereit F, Caporali R, Cardiel MH, De Cock D, Codreanu C, Cutolo M, Edwards CJ, van Eijk-Hustings Y, Emery P, Finckh A, Gossec L, Gottenberg JE, Hetland ML, Huizinga TWJ, Koloumas M, Li Z, Mariette X, Müller-Ladner U, Mysler EF, da Silva JAP, Poór G, Pope JE, Rubbert-Roth A, Ruyssen-Witrand A, Saag KG, Strangfeld A, Takeuchi T, Voshaar M, Westhovens R, van der Heijde D. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Ann Rheum Dis. 2020 Jan 22. pii: annrheumdis-2019-216655

43. **Sepriano A**, Ramiro S, van der Heijde D, van Gaalen F, Hoonhout P, Moltó A, Saraux A, Ramonda R, Dougados M, Landewé R. What is axial spondyloarthritis? A latent class and transition analysis in the SPACE and DESIR cohorts. Ann Rheum Dis. 2020 Mar;79(3):324-331

44. Kerschbaumer A, **Sepriano A**, Smolen JS, van der Heijde D, Dougados M, van Vollenhoven R, McInnes IB, Bijlsma JWJ, Burmester GR, de Wit M, Falzon L, Landewé R. Efficacy of pharmacological treatment in rheumatoid arthritis: a systematic literature research informing the 2019 update of the EULAR recommendations for management of rheumatoid arthritis. Ann Rheum Dis. 2020 Feb 7. pii: annrheumdis-2019-216656

45. **Sepriano A**, Kerschbaumer A, Smolen JS, van der Heijde D, Dougados M, van Vollenhoven R, McInnes IB, Bijlsma JW, Burmester GR, de Wit M, Falzon L, Landewé R. Safety of synthetic and biological DMARDs: a systematic literature review informing the 2019 update of the EULAR recommendations for the management of rheumatoid arthritis. Ann Rheum Dis. 2020 Feb 7. pii: annrheumdis-2019-216653

46. Ramiro S, Landewé RB, van der Heijde D, **Sepriano A**, FitzGerald O, Østergaard M, Homik J, Elkayam O, Thorne JC, Larche M, Ferraciolli G, Backhaus M, Boire G, Combe B, Schaeverbeke T, Saraux A, Dougados M, Rossini M, Govoni M, Sinigaglia L, Cantagrel AG, Allaart CF, Barnabe C, Bingham CO, Tak PP, van Schaardenburg D, Hammer HB, Dadashova R, Hutchings E, Paschke J, Maksymowych WP. Is treat-to-target really working in rheumatoid arthritis? a longitudinal analysis of a cohort of patients treated in daily practice (RA BIODAM). Ann Rheum Dis. 2020 Apr;79(4):453-459

47. Marona J, **Sepriano A**, Rodrigues-Manica S, Pimentel-Santos F, Mourão AF, Gouveia N, Branco JC, Santos H, Vieira-Sousa E, Vinagre F, Tavares-Costa J, Rovisco J, Bernardes M, Madeira N, Cruz-Machado R, Roque R, Silva JL, Marques ML, Ferreira RM, Ramiro S. Eligibility criteria for biologic disease-modifying antirheumatic drugs in axial spondyloarthritis: going beyond BASDAI. RMD Open. 2020 Jan;6(1). pii: e001145

48. Rodríguez-Carrio J, Putrik P, Gwinnutt J, **Sepriano A**, Alunno A, Ramiro S, Leipe J, Nikiphorou E. Mentoring for postdoctoral researchers in rheumatology: the Emerging EULAR Network (EMEUNET) post-doc mentoring programme. RMD Open 2020;6:e001139

49. Madari Q, **Sepriano A**, Ramiro S, Moltó A, Claudepierre P, Wendling D, Dougados M, van der Heijde D, van Gaalen FA. 5-year follow-up of spinal and sacroiliac MRI abnormalities in early axial spondyloarthritis: Data from the DESIR cohort. RMD Open 2020;6:e001093

50. **Sepriano A**, Ramiro S, van der Heijde D, Dougados M, Claudepierre P, Feydy A, Reijnierse M, Loeuille D, Landewé R. Integrated longitudinal analysis does not compromise precision and reduces bias in the study of imaging outcomes: A comparative 5-year analysis in the DESIR cohort. Semin Arthritis Rheum. 2020 Mar 3. pii: S0049-0172(20)30070-6

51. Stal R, van Gaalen F, **Sepriano A**, Braun J, Reijnierse M, van den Berg R, van der Heijde D, Baraliakos X. Facet joint ankylosis in r-axSpA: detection and 2-year progression on whole spine low-dose CT and comparison with syndesmophyte progression. Rheumatology (Oxford). 2020 May 17;keaa155.

52. **Sepriano A**, Ramiro S, van der Heijde D, Landewé RBM. Response to: 'Reactive arthritis, a missing link: comment on the recent article from Sepriano et al' by Zeidler and Hudson. Ann Rheum Dis. 2020 Apr 2. pii: annrheumdis-2020-217347

53. Nikiphorou E, Ramiro R, **Sepriano A**, Witrand A, Landewé RBM, van der Heijde D. Do smoking and socioeconomic factors influence imaging outcomes in axial spondyloarthritis? Five-year data from the DESIR cohort. Arthritis Rheumatol. 2020 Jun 20. doi: 10.1002/art.41408

54. **Sepriano A**, Ramiro S, Landewé R, Moltó A, Claudepierre P, Wendling D, Dougados M, van der Heijde D. Inflammation of the sacroiliac joints and spine on MRI predicts structural changes on MRI in axial spondyloarthritis: 5-year data from DESIR. Accepted for publication, Arthritis Care & Research

55. **Sepriano A**, Ramiro S, van der Heijde D, Dougados M, Claudepierre P, Feydy A, Reijnierse M, Loeuille D, Landewé R. Which imaging outcomes for axSpA are most sensitive to change? A 5-Year analysis of The DESIR Cohort. Accepted for publication, Arthritis Care & Research

56. van der Burg L, **Sepriano A**, Landewé RBM, Geuskens GA, Ottenheijm RPG, Dinant GJ, Boonen A. Comparative construct validity of three presenteeism instruments in workers with musculoskeletal complaints: a prospective cohort study. RMD Open. 2020 Aug;6(2):e001281

57. Rodrigues Manica S, **Sepriano A**, Pimentel Santos F, Gouveia N, Barcelos A, Branco J, Bernardes M, Miriam-Ferreira R, Vieira-Sousa E, Barreira S, Vinagre F, Roque R, Santos E, Madeira N, Rovisco J, Daniel A, Ramiro S. Effectiveness of switching between TNF inhibitors in patients with axial spondyloarthritis: is the reason to switch relevant? Arthritis Res Ther. 2020 Aug 21;22(1):195

CURRICULUM VITAE

Alexandre Sepriano was born on the 29th of September, 1985 in Vila Franca de Xira (30 km north of Lisbon), Portugal. He studied medicine at the University of Lisbon in Portugal and obtained his Integrated Master degree in Medicine in 2009. He did his general medicine rotation at Hospital Santa Maria in Lisbon between 2010 and 2011 and trained in Rheumatology at the Department of Rheumatology of Hospital Egas Moniz also in Lisbon from 2011 to 2016. During his training, Alexandre became a Fellow of the Harvard Medical School - Portugal Program's Clinical Scholars Research Training, a 2-year training program in epidemiology and biostatistics organized by the Harvard Medical School with the support from the Portuguese government, which he successfully completed. As part of his training, he also did a 9-month rotation in clinical research at the Department of Rheumatology of the Leiden University Medical Centre, in Leiden, The Netherlands. This rotation in Leiden was instrumental for his choice to pursue a clinical academic career and to undertake the research work that resulted in this thesis at the same institution. He obtained his specialist accreditation in Rheumatology in 2016. He has been awarded 5 grants and 6 scientific prizes. He is currently a Rheumatologist at the Hospital Egas Moniz in Lisbon, Portugal.

ACKNOWLEDGEMENTS

While writing the last words of this thesis, I cannot help to feel a major sense of accomplishment. But the journey that took me here was not made in solitude. I was very fortunate to count with the support from many people to whom I would like to address a few words of appreciation.

I would like to start by thanking all the patients, researchers and health professionals involved in the development and implementation of the DESIR, ASAS and SPACE cohorts.

A heartfelt word of gratitude to my dear promotors, Prof. dr. Désirée van der Heijde and Prof. dr. Robert Landewé, for their continuous support that extend way beyond science. Dear Désirée and dear Robert, you are so much more than promotors. You are true mentors in all possible ways. Over these years, you helped me not only to grow as a researcher, but also, and most importantly, as a person.

Dear co-promotor, Dr. Sofia Ramiro, dear Sofia, words cannot express my immense gratitude for all your support over these years. My friend, looking back to all we have achieved (look at this thesis) it is still not enough to pay tribute to all you did for me. So much work, often not visible, so many words, so much guidance that was key to help me be the person I am today.

A word of gratitude to my rheumatology training supervisor Prof. dr. Fernando Pimentel-Santos and to Prof. dr. Jaime Branco, head of the rheumatology department of *Hospital Egas Moniz* (HEM). Your support was key since day one. A special word to my colleagues and friends at the HEM rheumatology department. Also thank you Prof. dr. Helena Canhão, Prof. dr. João Eurico, Prof. dr. Maria José Santos, Prof. dr. José António Pereira da Silva for being a constant source of inspiration.

A decisive moment in my 'research journey' took place when I moved to Leiden to do a research fellowship at the LUMC. Prof. dr. Tom Huizinga. Dear Tom, thank you for making me feel at home and for the privilege of learning with you in your department. Thank you to all the LUMC PhD students for making me part of the group. A especial word to Sytske Anne Bergstra and Miranda van Lunteren for being my Dutch 'godmothers'.

I had the privilege to work with researchers from all over the world, both in the context of EMEUNET and in collaborative research projects. Prof. Dr. Maxime Dougados, dear Maxime thank you for your always thoughtful comments in data interpretation and manuscript writing. So often, you helped me to see from the perspective of the practicing rheumatologist while reporting complex data. Dear Prof. Dr. Martin Rudwaleit, dear Martin, your work inspired a large part of this thesis. Thank you for the many discussions around the concept of spondyloarthritis that were decisive for my current understanding of this complex disease.

Leonor, Carlos and Claudia my dearest aunt, uncle and sister, thank you for all your support over the years. I can say with confidence, I would not be where I am today without you.

Last, but definitely not the least, Cátia, my beloved wife, no words written here can give the proper credit to everything I have to thank you for. You have always supported my career as a researcher. I could not do this thesis without you. My dearest daughters Catarina and Beatriz, years from now when you read this thesis, I want you to know that you played a central role in it. You are my continuous source of inspiration and unconditional love which adds meaning to all that I do.