From inflammatory back pain to ankylosing spondylitis

ISBN 10: 90-5278-563-5
ISBN 13: 978-90-5278-563-9

Layout: Tiny Wouters
Production: Datawyse | Universitaire Pers Maastricht

Scientific studies in this thesis were supported by an educational grant from the Dutch arthritis association.
Printing of this thesis was financially supported by Abbott, Amgen, Astra Zeneca, Sanofi Aventis, Schering Plough and Wyeth.

From inflammatory back pain to ankylosing spondylitis

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 6 oktober 2006 om 12.00 uur

door

Elisabeth Louise Johanna Heuft-Dorenbosch

Geboren op 27 april 1966 te Deurne

Promotor

Prof. dr. D.M.F.M. van der Heijde
Prof. dr. Sj. van der Linden

Co-promotor

Dr. R.B.M. Landewé

Beoordelingscommissie

Prof. dr. J.A. Knottnerus (voorzitter)
Prof. dr. R.A. de Bie
Prof. dr. J. Sieper, University Hospital Benjamin Franklin, Berlin, Germany
Prof. dr. P.P. Tak, Academisch Medisch Centrum, Amsterdam
Prof. dr. J.T. Wilmink

Voor mijn ouders
Richard, Anne en Koen

Contents

Chapter 1

Introduction

Clinical features of SpA

Clinical features of spondylarthropathy

All research described in this thesis focuses on a group of chronic, systemic rheumatic disorders called spondylarthropathy (SpA) - or spondyloarthritis. SpA forms a challenging condition to any rheumatologist working in daily medical practice or in research. One of the most prominent features is chronic axial inflammation (spine, hips and shoulders). SpA can have a varying presentation and course. It often affects people in young adulthood, may result in a substantial burden of disease and can be extremely difficult to diagnose in the early phase. During the last decades important developments have taken place with regard to insights in pathogenesis, disease course and outcome, as well as in treatment. Axial inflammation leads to inflammatory back pain, which differs clinically from back pain caused by non-inflammatory problems. Inflammatory back pain is typically chronic (>3 months duration), insidious in onset, accompanied by morning stiffness, ameliorates with exercise, deteriorates with rest and starts before the age of 40. Besides axial involvement SpA can present with peripheral arthritis and extra articular symptoms such as enthesitis, uveitis, gastrointestinal inflammation, skin lesions and cardiac involvement. SpA forms a group of related disorders, sharing a common genetic background and often a familial clustering, but with variable expression in individual patients. Included in this group of conditions are psoriatic arthritis (PsA), SpA related to inflammatory bowel disease, reactive arthritis (ReA), a group called undifferentiated spondylarthropathies (USpA) and ankylosing spondylitis (AS), the prototype of SpA. The spectrum of disease can range from mild, with patients not seeking medical attention, to severe, with a disabling course and substantial loss of quality of life, functional impairment and inability to work. The cause of SpA is multifactorial, including genetic, (endogenous) factors as well as exogenous factors. The most well-known genetic factor is the histocompatibility antigen HLA-B27, which is most prominent in AS, as 90% of patients with AS carries this gene. An exogenous factor in SpA is evident in the case of reactive arthritis, where a bacterial (e.g. gastrointestinal) infection triggers arthritis. Despite substantial efforts a bacterial infection has never been identified as a cause in the other subsets of SpA. Pro-inflammatory cytokines with tumor necrosis factor alfa (TNF-α) as a pivotal cytokine are recognized to play a central role in the inflammatory process. Therapies aiming at blocking this cytokine are extremely effective in ameliorating signs and symptoms of AS (and other forms of SpA).

History and classification in SpA and AS

The first clinical description of AS was by Connor in 1691. In the late 1800s Marie, Strumpel and von Bechterew described patients with AS. Previously used synonyms for AS include Marie-Strumpel's disease, Bechterew's disease, pelvospondylitis ossificans and rheumatoid spondylitis. For a long time AS was considered rheumatoid arthritis affecting the spine and not a distinct clinical entity. The first classification criteria for AS were developed in 1961 (Rome criteria), revised in 1966 (New York criteria) and again revised in 1984 (modified New York criteria). In a publication of Moll and Wright in 1984 the concept of spondylarthropathies as a whole group of related inflammatory disorders distinct from rheumatoid arthritis was introduced. Since then both in research and daily practice important developments focusing on SpA have taken place.
In SpA two classification criteria sets exist that are applied by researchers with the aim to compare homogeneous groups of patients in clinical studies: the Amor criteria and the European spondylarthropathy study group (ESSG) criteria, published in 1990 and 1991 respectively. Diagnostic criteria for SpA (criteria used by physicians to make a diagnosis in an individual patient) are still lacking, but in 2004 the first diagnostic algorithm for axial SpA was published. This algorithm was built on data available from the literature, but still has to prove its value when applied in clinical practice.

Treatment and assessment in SpA and AS

In AS basic treatment consists of education, physical therapy, exercises and medication. Until recent years, medical treatment consisted largely of non-steroidal anti inflammatory drugs (NSAIDs) with varying results on complaints in individual patients. The discovery of tumor necrosis factor alpha (TNF-α) as main pro-inflammatory cytokine in AS has lead to development of a potent therapy, the blocking of TNF-α, with excellent clinical results. This therapy was first introduced in rheumatoid arthritis but proved to be also very effective in AS and other forms of SpA, such as psoriatic arthritis. In parallel to clinical development of TNF-blocking drugs, the interest in outcome assessment in SpA has increased. Assessment in general has three main aims; to classify, to prognosticate and to measure change over time. With respect to assessment in rheumatology the "outcome measures in rheumatology clinical trials" (OMERACT) filter was developed. This "filter" consists of three aspects: truth, discrimination and feasibility that each includes a number of relevant questions that should be addressed with every measurement; First aspect, truth: is the measure truthful, does it measure what is intended to measure? Is the result

unbiased and relevant? Second aspect, discrimination: does the measure discriminate between patients and/or disease states with severe and less severe disease. Do we obtain the same results in unchanged condition by the same or different observers? Does the instrument pick up change over time? Third aspect, feasibility: can the measure be applied easily in light of time constraints, money and interpretability.

The assessment in ankylosing spondylitis (ASAS) working group, an international working group of international clinical experts from over 40 countries, clinical epidemiologists, representatives of the pharmaceutical industry, and representatives of patient leagues, which has started in Amsterdam 1995, has picked up these principles of measurement with regard to AS. ASAS aimed at developing international standardized endpoints for use in clinical trials on AS, as well as in clinical practice. The ASAS working group has defined three core sets of overlapping domains to evaluate disease controlling anti rheumatic therapy (DCART), symptom modifying anti rheumatic drugs and physical therapy (SMARD/physical therapy) and to uniform clinical record keeping. Each domain comprises different instruments available for assessment applicable to the specified domain. Assessments proposed by ASAS were later endorsed by OMERACT.

Table 1.1 Specific instruments for each domain in the core set for DCART, SMARD, physical therapy and clinical record keeping.

Domain	Instrument
Physical Function	BASFI or Dougados FI
Pain	VAS/NRS, last week, in spine, at night, due to AS and VAS/NRS, last week, in spine, due to AS
Spinal mobility	Chest expansion, and modified Schober, and occiput-to-wall distance, and lateral spinal flexion or BASMI
Patients global assessment	VAS/NRS last week
Morning stiffness	Duration of morning stiffness, In spine, last week
Fatigue	VAS/NRS last week
Peripheral joints and entheses	Number of swollen joint count Validated enthesitis index
Acute phase reactants	ESR
Spine radiographs	Anterioposterior + lateral lumbar and lateral cervical spinal and X-ray pelvis (to visualize sacroiliac joint and hips
Hip radiographs	As above

BASFI=Bath ankylosing spondylitis functional index; BASMI=Bath ankylosing spondylitis metrology index; CRP=C-reactive protein; DCART=disease-controlling anti-rheumatic therapy; ESR=erythrocyte sedimentation rate; NRS=numerical rating scale; SMARD=symptom modifying anti-rheumatic drug; VAS=visual analogue scale.

Developments

As mentioned before, making a diagnosis of SpA in the early phase of the disease is still a significant challenge to clinicians. Historically a long delay existed between the start of symptoms and a diagnosis, mainly with respect to AS. This had several reasons. First not all patients are seeking medical attention for their complaints, and second, recognition of inflammatory back pain, the hallmark of AS, by health care providers, is difficult. Third, until recently, a diagnosis in AS has relied upon the appearance of radiological changes in the sacroiliac joints. A mean time lag of 4-9 years between start of symptoms and appearance of radiological changes was not exceptional in various cohorts. These radiological changes appear because of structural changes in the SI joints as the result of the inflammatory process; the inflammatory process itself is not visualized on conventional X-rays. With the availability of magnetic resonance imaging (MRI), an imaging technique based on the use of electromagnetic fields that can directly visualize inflammation, an important step forward is made. With MRI, the inflammatory process in AS can be made visible in an early phase of the disease, and thus MRI can help in making an early diagnosis. The recently developed first diagnostic algorithm for axial SpA acknowledges the importance of inflammation assessed by MRI. With aid of this algorithm a diagnosis of axial SpA can be made on the basis of the combination of certain clinical, genetic and laboratory data, even before radiological changes have appeared. However the value of this algorithm has to be assessed in daily practice.

Aims of this Thesis

This thesis has two main topics. The first part of the thesis is on outcome assessment in established AS. Three widely used instruments are evaluated. For this propose, data of the OASIS cohort, an international, longitudinal, observational study on outcome in AS were used. The second part focuses on a newly formed cohort of patients with inflammatory back pain of short duration. Chapter 2 describes a simplification of the Mander enthesis index (MEI), that was developed in 1984 to assess enthesitis in patients with AS. Enthesitis was selected by the ASAS working group as a core set domain in clinical record keeping but no validated instrument was chosen. In the MEI a total number of 66 entheses is investigated by local pressure and intensity of pain is graded on a 0-3 scale. This instrument, however, is neither widely used in daily practice nor in clinical trials because it is time consuming, interpretation bias is possible and applying pressure on all sites in active enthesitis may be unacceptable to the patient. Therefore the ASAS working group suggested to perform more

research in this field before selecting an official instrument. In chapter 2 is described how in a process of data reduction a more feasible enthesitis index, the Maastricht ankylosing spondylitis enthesis index (MASES), was constructed with aid of the MEI, reducing the original 66 included entheses to 13.
In AS reduced spinal mobility with progressive thoracic kyphosis is a characteristic feature. In Chapter 3 different measurements for thoracic spine extension in AS are compared; the occiput-to-wall-distance (OWD) and the tragus-to-wall-distance (TWD).These widely applied clinical measures assess the same construct, namely the extent of thoracic kyphosis in AS, and a thorough comparison of validity was never performed.
Measuring disease activity is difficult in AS; erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) are frequently normal in patients with spinal AS. For assessing disease activity the Bath ankylosing spondylitis disease activity index (BASDAI), a patient based questionnaire, is a frequently used instrument. In Chapter 4 we investigated whether the BASDAI, differed between patients with and without peripheral arthritis/enthesitis.
Until now, a diagnose of AS depends on the radiological presence of sacroiliitis on plain radiographs. In chapter 5 the performance of detecting sacroiliitis on plain radiographs was assessed between radiologists and rheumatologists and sensitivity and specificity were assessed.
Inflammatory back-pain is a prominent feature in AS but frequently also in other forms of SpA. In chapter 6 we describe how a group 68 of patients with inflammatory back pain (IBP) of short duration as a common and main symptom (the early spondylo arthritis clinic –ESpAC cohort) can be classified when applying different sets of classification criteria for SpA, and which clinical and imaging features are of discernible importance. Subsequently, we challenged the diagnostic value of a recently published diagnostic algorithm to detect (pre-radiological) axial SpA, the Berlin algorithm, in this cohort.
Last decade the use of MRI has obtained interest in the diagnosis of axial inflammation, however especially in early AS its value has to be determined. In chapter 7 the prevalence of inflammation and structural changes of the SI joints on MRI and agreement between two independent readers were assessed.
In chapter 8 the contribution of MRI and conventional radiography (CR) of SI joints of patients in the ESpAC cohort in making an early diagnosis of SpA is described.
Finally, chapter 9 and 10 give a summary and general discussion in English and Dutch.

Chapter 2

Assessment of enthesitis in ankylosing spondylitis

L Heuft-Dorenbosch, A Spoorenberg, A van Tubergen, R Landewé,
H van der Tempel, H Mielants, M Dougados, D van der Heijde

Ann Rheum Dis 2004;62:127-132

Abstract

Objective
The aim of this study was firstly to assess the validity of the enthesis index published by Mander (Mander enthesis index=MEI) and secondly to investigate whether it s possible to define a new enthesis index that is less time consuming to perform with at least similar or better properties.

Methods
Data from the OASIS cohort, an international, longitudinal, observational study on outcome in ankylosing spondylitis (AS), were used. In this study, measures of disease activity, among which the Bath ankylosing spondylitis disease activity index (BASDAI) and the MEI, were assessed regularly in 217 patients. With regard to the MEI, for each measurement period independently, a process of data-reduction was performed to identify the entheses most frequently reported as painful by the patients. A more concise enthesis index was constructed with aid of the thus found entheses. Using correlations with measures of disease activity the validity of several entheses indices was tested.

Results
Reduction of the number of entheses from 66 to 13 and omitting grading of the intensity of pain resulted in an index that we called “Maastricht ankylosing spondylitis enthesitis score” (MASES). The MASES (range 0-13) has much greater feasibility compared to the MEI (range 0-90), however, up to 21% of patients with a score >0 on the MEI were not identified by a score on the MASES >0. Only 2.1% of the patients with an original enthesis score >0 had an original score on the MEI>3 (range 0-90) and it can be questioned whether a low sore on the MEI index represents clinically important enthesitis. Spearman correlation coefficient between the MASES score and the MEI was 0.90 and between the MASES and the BASDAI was 0.53 compared to a correlation of 0.59 between the MEI and the BASDAI.

Conclusions
MASES seems to be a good alternative to the MEI with much better feasibility.

Introduction

In 1995, the 'assessment in ankylosing spondylitis' (ASAS) working group was formed in order to select a core set for outcome assessment in ankylosing spondylitis (AS).[1] This international working group consists of clinical experts, clinical epidemiologists, representatives of the pharmaceutical industry, and representatives of patient leagues who share their expertise in the field of AS. In 1998, a definite core set was selected, to be used in various clinical settings, which consisted of different domains with specific instruments for each domain.[2] For example, the domains selected to evaluate disease-controlling treatment are function, pain, spinal mobility, patient's global, stiffness, peripheral joints and entheses, acute phase reactants, spine and hip radiology and fatigue. For all domains except entheses and fatigue, a specific instrument was selected.

Enthesitis is a primary clinical feature in ankylosing spondylitis (AS). In 1987, Mander *et al*. published an instrument to investigate enthesitis in AS (the Mander enthesis index (MEI)).[3] A total of 66 entheses are investigated by local pressure and intensity of pain is graded on a 0-3 scale (0=no pain, 1=mild tenderness, 2=moderate tenderness, 3=wince or withdraw). This instrument, however, is neither widely used in daily practice nor in clinical trials. The members of the ASAS working group questioned the feasibility of MEI. Applying the MEI is time consuming; there may be a potential discrepancy between the reaction of the patient and its interpretation by the doctor; and applying pressure on all sites in active enthesitis may be unacceptable to the patient. Therefore the ASAS working group suggested to perform more research in this field before selecting an official instrument.

To assess the appropriateness of an instrument to measure outcome, all aspects of validity need to be examined. To standardize the nomenclature of validity, the "outcome measures in rheumatoid arthritis clinical trials" (OMERACT) filter has been proposed.[4] The three domains of the OMERACT filter are truth (validity), discrimination (reproducibility and responsiveness) and feasibility.

To be able to relate the level of enthesitis with the overall level of disease activity, an instrument to assess disease activity should be selected. However, in AS, currently no "gold standard" exists for measuring disease activity. Objective measures such as C-reactive protein (CRP) and Westergren erythrocyte sedimentation rate (ESR) correlate poorly with clinical disease activity.[5] A self administered questionnaire has therefore been developed that better reflects clinical disease activity in AS: the Bath ankylosing spondylitis disease activity index (BASDAI). This instrument has been shown to be valid, reproducible and responsive to change.[6-8] Also widely used to measure

disease activity is a 10 cm visual analogue scale (VAS) to be completed by the patient and by the doctor separately.
This study aimed, firstly, at assessing the validity of the MEI and secondly, at investigating whether it is possible to modify the MEI into a less time consuming index with at least similar validity.

Patients and Methods

Patients

For this study we used data from the OASIS cohort; an international, longitudinal, observational study on outcome in AS with follow-up visits according to a fixed protocol. Data from this cohort have been previously reported.[5] Consecutive outpatients with an established diagnosis of AS according to the modified New York criteria were included in 1996. We here will report data of patients from the University Hospital Maastricht and the Maasland Hospital Sittard, the Netherlands and the Universital Hospital Gent, Belgium, all secondary and tertiary referral centers. Data from baseline, one-year follow-up and two-year follow-up will be presented. At these visits patients completed a number of questionnaires and underwent a clinical examination including the MEI.

Mander enthesis index

To assess the MEI, the investigator applies pressure over 66 different entheses accessible to palpation. The patients' response to firm palpation over these entheses is noted (0 = no pain, 1 = mild tenderness, 2 = moderate tenderness, 3 = wince or withdraw). The following sites are included in the index: the nuchal crests, the manubriosternal joint, the costochondral joints, the greater tuberosity and the medial and lateral epicondyles of the humerus, the iliac crests, the anterior superior iliac spines, the greater trochanter of the femur, the medial and lateral condyles of the femur, the insertion of the achilles tendons and plantar fascia to the calcaneus, the cervical, thoracic and lumbar spinous processes, the ischial tuberosties and the posterior superior iliac spines. Figure 2.1 represents the original figure published by Mander *et al*, showing the included entheses in their enthesis index. Some of the sites are scored individually while others are scored as a group, with the highest scoring site being recorded for the group as a whole. The sites that were grouped were: the nuchal crests, the costochondral joints, the cervical, thoracic and lumbar spinous processes. After grouping, a maximum score of 90 can be achieved.

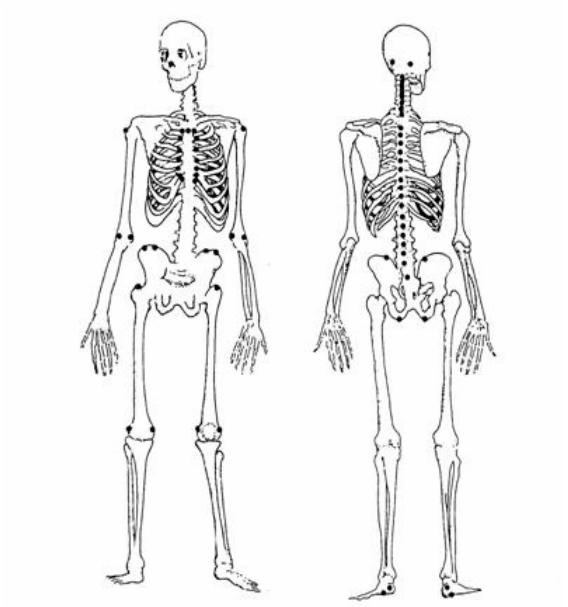

Figure 2.1 Entheses examined with patient lying (a) supine, (b) prone.

Measures of disease activity

The BASDAI consists of six questions on fatigue, pain of the spine and hips, pain or swelling of the peripheral joints, localized tenderness as a proxy for enthesitis, and severity and duration of morning stiffness. The questions are answered on a 10 cm VAS, anchored with the labels 'none' and 'very severe' at either end of the first five questions, and with '0 hours' and 'two hours' at either end of the question on duration of morning stiffness. The mean of the two scores regarding morning stiffness counts as one variable. The final score is defined by calculating the mean of the five items. Scores range from 0 (best) to 10 cm (worst). When values were missing, at most one out of six questions from the BASDAI was substituted by the patient's mean. In case of a missing value of question five or six, both dealing with morning stiffness, the remaining question counted as the mean of questions five and six.

A single item 10 cm VAS, concerning the degree of disease activity, and to be completed by both the patient and the doctor independently was anchored 'no disease activity' at 0 cm and 'very severe activity' at 10 cm.

The ESR was assessed using the Westergren method (mm/1st hour; normal range for men 0-7, for women 0-12) and CRP by the turbidimetric method (mg/l; normal range 2-9). The lowest detection limit for CRP was 2 mg/l and patients with undetectable levels were assigned 0.

Statistical analysis

Data at baseline, one- and two-years' follow-up were analyzed independently. Initially, the doctor recorded all 66 different entheses included in the MEI without grouping. After grouping, an enthesis score according to Mander was calculated for every patient. As a next stage, for each enthesis the original gradation of pain on the MEI was recoded dichotomously into 'no pain' or 'painful'. An original score of 0 was regarded as 'no pain'; original scores ranging from 1 to 3 were regarded as 'painful'. Only data from patients with a score on the MEI of >0 were used for developing modifications of the MEI.

In the following analyses a process of data-reduction was performed. Baseline, one-year follow-up and two-year follow-up were analyzed independently. Frequency tables were used to determine which specific enthesis was scored painful most frequently at the specific time. This enthesis was noted and all patients reporting this enthesis as painful were not taken into consideration in the subsequent step. In the remaining patients, a similar analysis with use of frequency tables was performed to determine which enthesis was scored most frequently as painful. Similarly, all patients reporting this particular enthesis as painful were not included in subsequent steps. This process was repeated until up to 80% (arbitrarily chosen) of all patients with a score on the MEI>0 were included. If an equal amount of patients reported two different entheses as most frequently painful at the same analysis, we selected the point that was anatomically the most easy to localize. The entheses detected in this way at baseline, one year and two years were taken together. If an enthesis was situated on the right or left side of the body, the contralateral enthesis was also included. Because we wanted an index as concise as possible, we decided, additionally, to decrease the number of entheses by omitting entheses more difficult to localize and entheses neighbouring those already selected.

As a last step the MEI, and the two modified MEIs were calculated with and without gradation of the intensity of pain.

Braun et al. recently used an enthesis index in a study on infliximab in AS, composed of 12 entheses which are reported to be commonly affected in the inflammatory process in AS (major entheses index).[9-11] This "major entheses index" includes the iliac crests, the great trochanters of the femur, the medial and lateral condyles of the femur, and the proximal insertion of the achilles tendon and insertion of the plantar fascia to the calcaneus. We also assessed how this "major entheses index " performed in our study group.

Spearman's correlation was used to determine the relationship between the original MEI and its proposed modifications with disease activity as measured by the BASDAI, the "enthesis question" of the BASDAI, patient and doctor VAS for disease activity, ESR and CRP. To assess a possible floor effect, the number of patients with a score on the original MEI of >0 were compared with the number of patients with a score on the modified enthesis indexes of >0.

Results

Table 2.1 presents the patient characteristics and scores on the ASAS core set measures. Of the patients, 112/162 (69%) were male, mean duration of complaints was 21.6 years, established disease duration was 11.4 years. Assessment of HLA-B27 was available in 123 patients; of these patients 85% was HLA-B27 positive. Table 2.2 presents the most frequently scored entheses at baseline, at one-year follow-up and at two-year follow-up. At baseline, the MEI was available for 149 patients, of which 115 reported a score >0. By pressure on five different entheses, 87 (76.%) of these 115 patients were detected.

Table 2.1 Baseline characteristics and scores on ASAS core set measures of the study patients (Mean (SD)).

	Total study group (n=162)
Male/ female	112 / 50
Age [years]	44.9 (12.3)
Duration of complaints [years]	21.6 (12.0)
Time since diagnosis [years]	11.4 (9.2)
HLA-B27 [present / absent / no data]	105 / 18 / 39
History of IBD [present / absent]	17 / 145
History of uveitis [present / absent]	62 / 100
History of psoriasis [present / absent]	10 / 152
BASFI	3.7 (2.4)
VAS pain of the spine [cm]	3.5 (2.3)
Night pain [4 point Likert]	1.2 (0.8)
Chest expansion [cm]	4.2 (1.9)
10 cm Schober [cm]	2.6 (1.4)
Occiput to wall distance [cm][a]	2.7 (0.0 - 7.5)
VAS patient global [cm]	3.4 (2.7)
Arthritis as measured by swollen joints [present / absent]	41 / 121
ESR [mm / hr][a]	9 (4 - 17)
CRP [mg / l][a]	7 (6-16)
VAS physician on disease activity [cm];	1.8 (1.8)
VAS physician on disease activity[a][cm]	1.0 (0.5 - 2.6)
VAS patient on disease activity [cm];	3.5 (2.6)
VAS patient on disease activity[a] [cm]	3.0 (1.1-5.2)
Duration of morning stiffness [min]	37 (29)

[a] Median (interquartile range); ASAS=assessments in ankylosing spondylitis; HLA=human leukocyte antigen; IBD=inflammatory bowel disease; BASFI=Bath ankylosing spondylitis functional index; VAS=visual analogue scale; ESR=erythrocyte sedimentation rate; CRP=C-reactive protein.

Table 2.2 Most frequently scored enthesis points at baseline, at one-year follow-up, and at two-year follow-up in order of detection during the data reduction process.

Baseline	One-year follow-up	Two-year follow-up
Proximal insertion of achilles tendon left[a]	Posterior superior iliac spine left[ab]	5th lumbar spinous processab
Iliac crest left[b]	7th costochondral joint right[ab]	7th costochondral joint rightab
5th lumbar spinous process[ab]	1st costochondral joint right[ab]	Posterior superior iliac spine left ab
7th costochondral joint left[ab]	Proximal insertion of achilles tendon left[a]	4th costochondral joint lefta
1st thoracic spinous process[a]	4th lumbar spinous process[a]	Anterior superior iliac spine rightb
	9th thoracal spinous process[a]	
76% of patients with MEI >0 detected	79% of patients with MEI >0 detected	83% of patients with MEI >0 detected

MEI=Mander enthesis index; [a] Included in reduced Mander enthesis index; [b] Included in Maastricht ankylosing spondylitis enthesis score.

At one-year follow-up, the MEI was available in 151 patients of whom 103 reported a MEI>0. By pressure on six different entheses out of 66, 81 (79%) of these 103 patients were identified. At two-year follow-up, the MEI was available in 129 patients of whom 86 patients reported a MEI>0; five entheses identified 71 (83%) of the 86 patients. Of the in total of 16 entheses selected at the three measurement periods, five appeared twice, which makes a total of 11 selected different entheses. If an enthesis was localized on the left or right side of the body, the enthesis on the contralateral side was also included, which increased the number of entheses to 18. We named this new index the 'reduced Mander enthesis index'.

To develop an enthesis index which was as concise as possible by further reducing the number of entheses, we excluded the following entheses: L4 because L5 was included; costochondral joint 4, because costochondral joints 1 and 7 were included and Th1 and Th9 because these are not easy to localize. Moreover L4, Th1, costochondral 4 and Th9 were only found in the last steps of the data reduction process and their inclusion increased the percentage of detected patients by only 4% at baseline, 6% at one-year follow-up and 8% at two-year follow-up. The total number of entheses was now reduced to 13. We named this group of 13 entheses the "concise enthesis index". Table 2.3 lists the entheses selected in the reduced MEI and the concise enthesis index.

Table 2.4a presents the number and percentage of patients reporting a score more than 0 on the MEI, the reduced MEI, the concise enthesis index and the major entheses index. At baseline, 24 patients (21%) with a score of >0 on the original MEI were not detected by a score >0 on the concise enthesis index. At one and two-years' follow-up these numbers were 20 patients (19%) and nine (10%) patients, respectively. The major entheses index did not detect 32, 42 and 35 patients at baseline, one and two-years' follow-up with an score >0 on the MEI (Table 2.5b). Table 2.4b presents the mean, median, p25-p75 and range of the 4 different enthesis indices.

Table 2.3 Entheses selected in reduced Mander enthesis index and concise enthesis index, listed, from head to toe.

Reduced Mander enthesis index	Concise enthesis index
1st costochondral joint left / right	1st costochondral joint left / right
4th costochondral joint left / right	
7th costochondral joint left / right	7th costochondral joint left / right
Posterior superior iliac spine left / right	Posterior superior iliac spine left / right
Anterior superior iliac spine left / right	Anterior superior iliac spine left / right
Iliac crest left / right	Iliac crest left / right
1st thoracal spinous process	
4th lumbar spinous process	
5th lumbar spinous process	5th lumbar spinous process
9th thoracal spinous process	
Proximal insertion of achilles tendon left / right	Proximal insertion of achilles tendon left / right

Table 2.4a Number of patients at baseline, one-year follow-up and two-year follow-up with a score on the Mander enthesis index, the reduced Mander enthesis index and the concise enthesis index of >0. Results are shown as number(%) of patients.

	Baseline n=149	One-year follow-up n=151	Two-year follow-up n=129
Mander enthesis index >0	115 (77)	103 (68)	86 (67)
Reduced Mander enthesis index >0	98 (66)	92 (61)	81 (63)
Concise enthesis index >0	91 (61)	83 (55)	77 (60)
Major enthesis index >0	83 (56)	61 (40)	51 (40)

Table 2.4b Values for the Mander enthesis index, reduced Mander enthesis index without gradation, concise enthesis index and major enthesis index without gradation at baseline, one-and two-years follow-up.

Enthesis index	Mean (SD)	Median (p25-p75)	Range
Mander enthesis index with gradation (range 0-90)			
baseline	8 (11)	4 (1-10)	0 - 56
1 year	6 (9)	2 (0-8)	0 - 47
2 years	7 (11)	2 (0-11)	0 - 56
Reduced Mander enthesis index without gradation (range 0-18)			
baseline	4 (5)	2 (0-7)	0 - 18
1 year	3 (5)	2 (0-7)	0 - 18
2 years	4 (5)	2 (0-6)	0 - 18
Concise enthesis index without gradation(range 0-13)			
baseline	3 (4)	2 (0-5)	0 - 13
1 year	2 (3)	1 (0-4)	0 - 13
2 years	3 (4)	1 (0-5)	0 - 13
Major enthesis index without gradation (range 0-12)			
baseline	2 (3)	1 (0-4)	0 - 12
1 year	2 (3)	0 (0-2)	0 - 12
2 years	2 (3)	0 (0-3)	0 - 12

Table 2.5a The number of patients missed by the concise index (score=0), but with a score of >0 on the Mander enthesis index.

Score on Mander enthesis index	Baseline[a]	One-year follow-up[a]	Two-year follow-up[a]
1	12	5	1
2	4	5	3
3	5	5	4
4	1	2	1
5	1	1	0
6	1	1	0
7	0	0	0
8	0	1	0
Total	24	20	9

[a] number of patients.

Table 2.5b The number of patients missed by the major enthesis index (score=0), but with a score of >0 on the Mander enthesis index.

Score on Mander enthesis index	Baseline[a]	One-year follow-up[a]	Two-year follow-up[a]
1	11	14	8
2	9	9	11
3	4	9	7
4	3	7	6
5	1	2	0
6	2	1	1
7	1	0	1
8	0	0	1
9	1	0	0
Total	32	42	35

[a] number of patients.

To assess a possible floor-effect of the MEI, Table 2.5 presents the patients with a score of 0 on the concise enthesis index with a score >0 on the MEI at baseline, one-year and two-years follow-up. The total of 53 patients missed with the concise enthesis index predominantly appeared to have low scores on the MEI. 44 out of these 53 patients have a score on the MEI≤3 (range 0-90).
Table 2.6 presents the correlation between the MEI (with and without gradation), the reduced Mander enthesis index (with and without gradation) the concise enthesis index and the major entheses index (with and without gradation) with BASDAI, the enthesis question of the BASDAI, ESR, CRP and the patients and the doctors' VAS concerning the degree of disease activity. The enthesis question of the BASDAI did not show a higher correlation coefficient with MEI (with and without gradation), the reduced MEI (with and without gradation) and the concise enthesis index. Furthermore, it can be seen that the correlation with measurements of disease activity does not change significantly when the three different entheses indices are applied. On applying the concise enthesis index without gradation compared with the MEI (which is

graded), the correlation is only slightly decreased. We propose to call the concise enthesis index without gradation the 'Maastricht ankylosing spondylitis entheses score' (MASES). Figure 2.2 shows the entheses of the MASES.

Table 2.6 Spearman correlation between Mander enthesis index, reduced Mander index, concise enthesis index with and without gradation versus measures of disease activity.

	Mander enthesis index	Mander enthesis index without gradation	Reduced Mander enthesis index	Reduced Mander enthesis index without gradation	Concise enthesis index without gradation	Concise enthesis index without gradation (MASES)	Major enthesis index without gradation
BASDAI[a]	0.59	0.58	0.57	0.56	0.54	0.53	0.49
Enthesis question BASDAI[a]	0.57	0.56	0.55	0.54	0.53	0.51	0.41
VAS physician[a]	0.30	0.27	0.29	0.27	0.24	0.22	0.21
VAS patient[a]	0.47	0.45	0.48	0.47	0.45	0.44	0.37
CRP[b]	0.07	0.06	0.06	0.05	0.05	0.04	0.02
ESR[b]	0.06	0.07	0.07	0.07	0.05	0.04	0.07

BASDAI=Bath ankylosing spondylitis disease activity index; VAS=visual analogue scale; CRP=C-reactive protein; ESR=erythrocyte sedimentation rate; MASES=Maastricht ankylosing spondylitis entheses score. [a] Correlation coefficients significant at the $p<0.01$ level; [b] Correlation coefficients not significant.

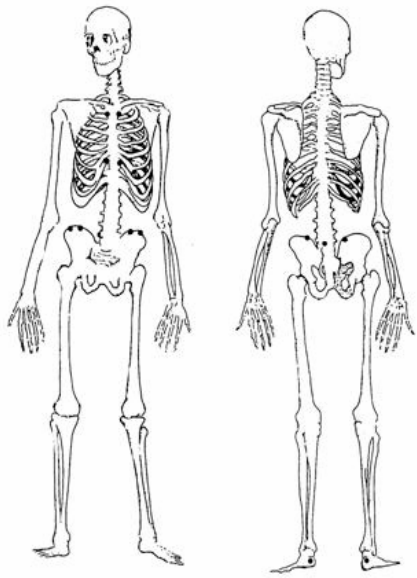

Figure 2.2. The entheses of the MASES.

Discussion

Some remarks should be made about applying the OMERACT filter to the enthesis index by Mander *et al* . As far as we know, no reports exist in which the MEI is correlated with histological proof or radiological signs (e.g. magnetic resonance imaging (MRI) or ultrasound) of enthesitis. Unfortunately we were not able to assess this aspect in our study. However, there are studies investigating the correlation between clinical signs of enthesitis (swelling or/and pain) and imaging methods as MRI or ultrasonography. MRI may show swelling of the enthesis and the peritendinous soft tissue, distension of adjacent bursae by fluid collection and oedema of the bone near the insertion. Ultrasonography may show the following signs of enthesitis; thickening of the tendon insertion, intratendinous focal changes, calcific deposits in the insertions, and periosteal changes. Recently, Marzo Ortega *et al.* performed a descriptive longitudinal study in which the efficacy of etanercept in the treatment of resistant spondylarthropathy was assessed in 10 patients.[12] In their study clinical enthesitis was investigated by clinical assessment of 78 entheses and by a patient VAS. In the nine patients with clinical enthesitis, clinical enthesitis resolved completely in seven patients and improved in two patients. In these nine patients, there were 44 MRI detectable entheseal lesions of which 86% resolved completely or improved. The concomitant amelioration of both MRI findings and clinical findings suggest that a clinical measure for enthesitis does indeed depict enthesitis; however, the study was not designed to assess the validity of a clinical measure to assess entheses. Lethinen *et al*. studied 31 consecutive patients with spondylarthropathy for the presence of enthesiopathy in the lower extremities both clinically and with ultrasonography[13]. Sonography detected inflammatory lesions in 44 entheses of 20 patients, clinical examination detected 56 symptomatic entheses in 20 patients, suspected of enthesitis. In 21 entheses, both examinations were positive.

In terms of truth, a measure to assess entheses should distinguish enthesitis from other causes of (joint) pain as arthritis. Mander *et al*. developed their enthesis index in a study in which only AS patients without peripheral arthritis were included.[3] In our patients peripheral arthritis, as measured as presence of at least one swollen joint, was present in 25% patients. In the process of developing the MASES, the only included enthesis neighbouring a peripheral joint, was the proximal insertion of the Achilles tendon. In this respect, less confounding with pain due to peripheral arthritis can occur compared to the MEI.

An enthesis index is meant to measure severity of enthesitis. Severity encompasses both intensity and extent of enthesitis. It is not known whether the patients' response correlates with the degree of enthesitis and as stated in the original article by Mander "it is not known whether the severity or simply the

presence of tenderness over the entheses is important".[3] Interpretation of the degree of patients' response reflected in grades can increase both inter-and intraobserver bias. In joint counts it has been proven that gradation did not improve the performance of the scores and is therefore not recommended.[14] Without the gradation, the MEI is easier to perform and in our study grading did not improve the correlation with measures of disease activity.

The MEI is not very practical to apply clinically because of its extensiveness. This study shows that a reduction of the entheses-index to 13 entheses instead of 66 still provides reasonable assessment of the entheses.

On 53 visits, patients who originally had a score on the MEI>0 are missed with the MASES (Table 2.5). Most of these patients had a low score on the MEI. It might be questioned whether a low sore on the MEI represents clinically important enthesitis. However, this indicates that there might be some floor effect in the MASES. In our view non- detection of a few patients is more than balanced by the enormous gain in feasibility. In many trials, including two recently published large trials on non-steroidal anti inflammatory drugs (NSAIDs),[15,16] no measure for entheses is included, illustrating the lack of acceptability of the MEI. This in contrast with the importance placed on entheses by the ASAS group and the opinion that enthesitis can be seen as the cornerstone of the site of inflammation in AS.[9,17,18]

The fourth question of the BASDAI can be considered to be an enthesiopathy VAS. It might be questioned why this patient VAS is not used in evaluating enthesitis rather than an instrument used by the doctor. However, it is usual for the doctor to differentiate between entheseal pain and articular, muscular or other causes of pain; this can be more difficult for patients. Therefore a combination of patient and doctor assessment of enthesitis would be preferable.

The index should not only discriminate between active and no active disease and between groups of patients but also within individual patients. Only a small study to investigate the discriminate capacity of the MEI is reported in the original paper where the MEI is compared in AS patients with and without NSAID treatment. So far we have not tested the discriminative capacity of the full MEI, the reduced Mander index, and the MASES. We contacted many authors of published and unpublished trials in AS. None of these had included the MEI included in their studies. Therefore we make a plea that the MEI is included in studies so that the discriminative power of the various enthesitis scores can be tested. New trials on tumor necrosis factor blockade treatment could provide this information, especially if MRI of the spine is carried out concomitantly to demonstrate sites of inflammation. While awaiting validation of the MASES we recommend its use as it seems to be a good alternative.

References

1. van der Heijde D, Bellamy N, Calin A, Dougados M, Khan MA, van der Linden S. Preliminary core sets for endpoints in ankylosing spondylitis. Assessments in Ankylosing Spondylitis Working Group. J Rheumatol 1997;24:2225-9.
2. van der Heijde D, van der Linden S, Bellamy N, Calin A, Dougados M, Khan MA. Which domains should be included in a core set for endpoints in ankylosing spondylitis? Introduction to the ankylosing spondylitis module of OMERACT IV. J Rheumatol 1999;26:945-7.
3. Mander M, Simpson JM, McLellan A, Walker D, Goodacre JA, Dick WC. Studies with an enthesis index as a method of clinical assessment in ankylosing spondylitis. Ann Rheum Dis 1987;46:197-202.
4. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology [editorial]. J Rheumatol 1998;25:198-9.
5. Spoorenberg A, van der Heijde D, de Klerk E, Dougados M, de Vlam K, Mielants H, van der Tempel H, et al. Relative value of erythrocyte sedimentation rate and C-reactive protein in assessment of disease activity in ankylosing spondylitis. J Rheumatol 1999;26:980-4.
6. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.
7. Jones SD, Calin A, Steiner A. An update on the Bath Ankylosing Spondylitis Disease Activity and Functional Indices (BASDAI, BASFI): excellent Cronbach's alpha scores. J Rheumatol 1996;23:407.
8. Calin A, Nakache JP, Gueguen A, Zeidler H, Mielants H, Dougados M. Defining disease activity in ankylosing spondylitis: is a combination of variables (Bath Ankylosing Spondylitis Disease Activity Index) an appropriate instrument? Rheumatology Oxford 1999;38:878-82.
9. Ball J. Enthesopathy of rheumatoid and ankylosing spondylitis. Ann Rheum Dis 1971;30: 213-23.
10. Gerster JC. Plantar fasciitis and Achilles tendinitis among 150 cases of seronegative spondarthritis. Rheumatol Rehabil 1980;19:218-22.
11. Braun J, Brandt J, Listing J, Zink A, Alten R, Golder W, Gromnica Ihle E, Kellner H, Krause A, Schneider M, Sorensen H, Zeidler H, Thriene W, Sieper J. Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial. Lancet 2002;359: 1187-93.
12. Marzo Ortega H, McGonagle D, O'Connor P, Emery P. Efficacy of etanercept in the treatment of the entheseal pathology in resistant spondylarthropathy: a clinical and magnetic resonance imaging study. Arthritis Rheum 2001;44:2112-7.
13. Lehtinen A, Taavitsainen M, Leirisalo Repo M. Sonographic analysis of enthesopathy in the lower extremities of patients with spondylarthropathy. Clin Exp Rheumatol 1994;12:143-8.
14. Prevoo ML, van Riel PL, van 't Hof MA, van Rijswijk MH, van Leeuwen MA, Kuper HH, et al. Validity and reliability of joint indices. A longitudinal study in patients with recent onset rheumatoid arthritis. Br J Rheumatol 1993;32:589-94.
15. Dougados M, Gueguen A, Nakache JP, Velicitat P, Veys EM, Zeidler H, et al. Ankylosing spondylitis: what is the optimum duration of a clinical study? A one year versus a 6 weeks non-steroidal anti-inflammatory drug trial. Rheumatology (Oxford) 1999;38:235-44.
16. Dougados M, Behier JM, Jolchine I, Calin A, van der Heijde D, Olivieri I, et al. Efficacy of celecoxib, a cyclooxygenase 2-specific inhibitor, in the treatment of ankylosing spondylitis: a six-week controlled study with comparison against placebo and against a conventional nonsteroidal antiinflammatory drug. Arthritis Rheum 2001;44:180-5.
17. McGonagle D, Gibbon W, Emery P. Classification of inflammatory arthritis by enthesitis. Lancet 1998;352:1137-40.
18. McGonagle D, Khan MA, Marzo Ortega H, O'Connor P, Gibbon W, Emery P. Enthesitis in spondyloarthropathy. Curr Opin Rheumatol 1999;11:244-50.

Chapter 3

Measurement of spinal mobility in ankylosing spondylitis: comparison of occiput-to-wall and tragus-to-wall distance

L Heuft-Dorenbosch, D Vosse, R Landewé, A Spoorenberg, M Dougados, H Mielants, H van der Tempel, Sj van der Linden, D van der Heijde

J Reumtol 2004;9:1779-1784

Abstract

Objective
To investigate if the tragus-to-wall distance (TWD) is more reliable compared to the occiput-to-wall distance (OWD) as a measurement for thoracic spine extension in ankylosing spondylitis (AS).

Methods
Data from the OASIS cohort, an international, longitudinal, observational study on outcome in AS, were used. Measurements of OWD and TWD were performed at baseline and at 6,12,18 and 24 months. Paired data of T_x and T_{x+6} months were used to perform test-retest measurements (intra-class correlations, limits of agreement and interperiod correlation matrix). Bland and Altman plots were constructed to investigate the agreement between both observations, assuming that there was no true change between 0 and 6 months. To investigate whether a change in disease activity would have influenced the results, limits of agreement were calculated in a subgroup of patients with a stable Bath ankylosing spondylitis disease activity index (BASDAI; defined as a maximum BASDAI change of ±1) between T_0 and T_6 and compared with the results of the whole group .
Limits of agreement were also calculated for kyphosed patients only.

Results
The test-retest intraclass correlations were between 0.94-0.96 for OWD and between 0.93-0.95 for TWD. The direct measurement-remeasurement correlation calculated by extrapolation of the interperiod correlation regression line was 0.92 for OWD and 0.90 for TWD. OWD and TWD showed comparable reliability on the entire value of scores. The lower 95% limit of agreement was between –3.4 cm and –2.5 cm for OWD and between –3.4 cm and –3.1 cm for TWD. The upper limit of agreement was between 3.1 cm and 4.2 cm for OWD and between 2.9 cm and 3.9 cm for TWD. In all patients as well as in kyphosed patients only, limits of agreement were comparable between OWD and TWD. The patterns of the scatterplots according to Band and Altman were similar for OWD and TWD. Measurement error was more pronounced in kyphosed patients compared to patients with a normal thoracic extension. However, over the entire range of kyphosis, measurement error was similar.

Conclusions
OWD and TWD are equally reliable in assessing thoracic spine extension. Although the TWD is in general easier to perform in AS patients compared to OWD, we recommend the OWD measurement over TWD: in OWD measurement a value of zero easily distinguishes patients with normal thoracic spine extension from kyphosed patients.

Introduction

Reduced spinal mobility and changes in posture are characteristic features of ankylosing spondylitis (AS). The typical patient with AS has a reduced lumbar lordosis and an increased thoracic kyphosis, and as a result the head is somewhat bent forward. In the course of the disease process these postural changes tend to progress and may become irreversible due to structural changes of the spine. To quantify thoracic kyphosis, the distance between occiput and wall (occiput-to-wall distance, OWD) is assessed when the patient is standing erect with stretched knees and the back against the wall. OWD is a measure for thoracic spine extension[1] and was selected as a core set instrument by the assessment in ankylosing spondylitis (ASAS) working group. This is an international working group of clinical experts, clinical epidemiologists, representatives of the pharmaceutical industry, and representatives of patient associations. This core set was selected in 1998, and consists of different domains with specific instruments per domain, and is to be used in various settings in clinical studies in AS. To assess the domain of spinal mobility, the ASAS working group selected three instruments as core set instruments.[2] In addition to the OWD, these instruments are chest expansion and modified Schober test.

In clinimetrics, important features of an instrument are (construct) validity, reliability and responsiveness.[3] Reliability (synonyms: reproducibility, repeatability) is an expression of the extent to which similar results are obtained on repeated applications of the same assessment technique, assuming no true interval change in the phenomenon under study.[1]

An other way to assess thoracic spine extension is by measuring the tragus-to-wall distance (TWD). Although not included in the core set, this is part of the Bath ankylosing spondylitis metrology index (BASMI) which is also widely used.[4,5]

There are some arguments that favor TWD over OWD measurement. Together with thoracic spine extension, involuntary flexion and extension of the head can occur. Flexion and extension of the head take place in the atlanto-occipital joint and the cervical spinal joints and can interfere with the measurement of both OWD and TWD. We hypothesized that compared to the measurement of TWD, the measurement of OWD would be more influenced by concomitant flexion or extension of the head. The explanation for this lies in the fact that, compared to the occiput, the tragus lies closer to the sagittal axis of the flexion or extension movements (Figure 3.1). On the other hand, TWD might be influenced by unintended rotation of the head. Second, the OWD is more difficult to measure since the patient's hair frequently obscures the view of the occiput. Third, with the patient standing erect against the wall, if the OWD is small, the observer's view of the ruler is difficult and prone to parallax problems. Because of these

arguments we wanted to test the hypothesis that the TWD instrument would be more reliable than the OWD as a measurement for thoracic spine extension.

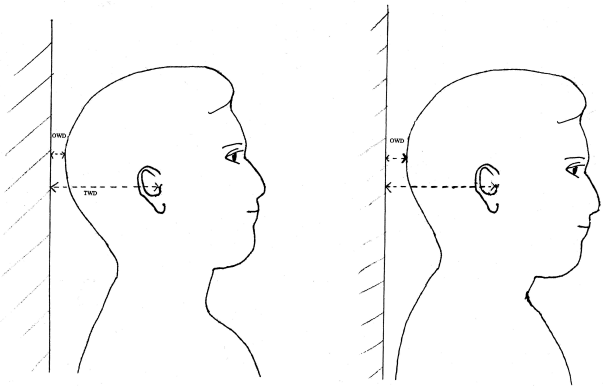

Figure 3.1 TWD in both positions of the head is the same, whereas the OWD is not, due to flexion of the head.

Patients and Methods

Patients

In our study we used the data from the OASIS cohort, an international longitudinal observational study on outcome in AS with follow-up visits according to a fixed protocol. Data from this cohort have been reported.[6] Participating centers are the University Hospital, Maastricht, the Netherlands, the Maasland Hospital, Sittard, the Netherlands; Hospital Cochin, Paris, France; Universital Hospital Gent, Belgium; which are all secondary and tertiary referral centers. Consecutive outpatients with an established diagnosis of AS according to the modified New York criteria were enrolled in 1996 and followed thereafter.

For this study we used data from visits at baseline (T_0) and at 6 (T_6), 12 (T_{12}), 18 (T_{18}) and 24 months (T_{24}). On each study visit, all patients completed a number of questionnaires and underwent a clinical examination. The same investigator performed all clinical examinations per country. In addition to the core set variables propagated by the ASAS working group, several additional measurements were done, including the TWD.

Occiput-to-wall distance

The patient stands with heels and buttocks touching the wall behind and with the knees straight. The patient is asked how far back he/she can get the head

still keeping the chin in the normal position. In the straight position, the distance between the posterior convexity of the occiput and the wall is measured to the nearest 0.1 centimeter using a rigid ruler.[1] The better of two attempts is recorded.

Tragus-to-wall distance

The patient is positioned as in measurement of the OWD. The distance between the tragus and the wall is measured to the nearest 0.1 centimeter using a rigid ruler.[1] The better of two attempts is recorded.
In our study, measurement of TWD immediately followed measurement of OWD without repositioning the patient. However, we are comparing assessments with an interval of six months, and obviously there was repositioning between the first and second assessment of the OWD and the TWD. Therefore, measurement variability due to repositioning (including flexion and rotation of the head) is included in the overall measurement error.

Statistical analysis

We first assessed the mean, range, and standard deviation of both OWD and TWD. Because of the character of the disease, with slow disease progression, we assumed on a group level that no relevant changes in spinal mobility occurred during a time interval of six months, and thus changes in this time interval would be due to measurement error rather than to real changes in spinal mobility. As we did not perform a true test-retest measurement to assess the measurement error of both the OWD and TWD, we used the paired data of T_x and $T_{x+6months}$. To determine if it was appropriate to do so, we used several methods. First, random-effect single-measure intraclass correlations (ICC; type 2.1) were calculated for T_x and $T_{x+6months.}$ Second, an interperiod correlation matrix was constructed. By this method the intercorrelation of the measurement periods are plotted against the intervening time intervals. These correlations can be linearly related to time, which may be represented by a well-fitting regression line. Extrapolation of this line through the y-axis gives the direct measurement-remeasurement correlation, which may be interpreted as a quality measure of the measurement.[7] This was done twice, first for all patients, then only for patients not able to reach normal thoracic extension. Third, limits of agreement as defined by Bland and Altman (explained below) were calculated in all patients with a stable Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) score (defined as a maximum BASDAI change of ±1) between T_0 and $T_{6.}$ The results of this analysis were compared with the results of the whole group.
To visualize the agreement between two observations and to check if observed differences are similar along the total range of scores, we plotted for all patients

the difference between the score at T_0 and at T_6 of the OWD and the TWD against each patient's mean of the two scores. This graphic representation by means of a scatterplot of two measurements is called Bland-Altman plot. In this method the difference between two observations (on the y-axis) is plotted against the mean of the same two observations (on the x-axis). Discrepancies between the observations are thus visualized as well as any possible relationship between the various parts of the scale and the corresponding measurement error. The advantage compared to a simple plot of the results of one method against the other is that in this latter method the data points will usually be clustered and between- method differences are difficult to assess. Thereafter we calculated the limits of agreement for OWD and TWD based on the 95% limits of agreement method by Bland and Altman using the formula: $\bar{d} \pm 1.96 * sd_{diff}$, where $\bar{d}$ represents the mean difference between the two observations and sd_{diff} the standard deviation of the difference.[8] Similar plots were made for the measurements comparing T_6 and T_{12}, T_{12} and T_{18}, T_{18} and T_{24}. In contrast with the OWD, which is zero in all patients with normal thoracic spine extension, the minimum TWD depends on the size of the head and the position of the ear. In patients able to touch the wall with their occiput, the possibility for measurement error due to patient variance seems to be smaller compared to patients not able to touch the wall, as the wall gives stability to the head. Thus a high proportion of patients with an OWD=0 may result in smaller limits of agreement. In daily practice the assessment of both OWD and TWD will largely be used in kyphosed patients. Thus we also calculated the limits of agreement for kyphosed patients exclusively. To be able to compare TWD with OWD in the same patients, we selected all patients based on an OWD>0 for calculations of limits of agreement for both OWD and TWD.
To assess whether the OWD and TWD instruments behave differently in different ranges of the scale, limits of agreement were calculated for the entire range of values, as well as for every quartile.

Results

At baseline, 217 patients were included. Characteristics of the patients are presented as mean with standard deviation (SD), or as median with interquartile range if appropriate (Table 3.1). The test-retest ICC was between 0.94 and 0.96 for OWD and between 0.93 and 0.95 for TWD. The direct measurement-remeasurement correlation calculated by extrapolation of the regression line was 0.92 for OWD and 0.90 for TWD. With calculations only for kyphosed patients, these figures were 0.88 for OWD and 0.92 for TWD.

Table 3.1 Baseline characteristics and scores on ASAS core set measures (Mean (SD)).

	Total study population (n=217)
Male / female	150 / 67
Age [years]	43.1 (12.7)
Duration of complaints [years]	19.6 (11.8)
Time since diagnosis [years]	10.8 (8.9)
HLA B27 [present / absent / no data]	155 / 32 / 30
History of IBD [present / absent / no data]	17 / 145 / 55
History of uveitis [present / absent / no data]	81 / 133 / 3
History of psoriasis [present / absent / no data]	10 / 152 / 55
BASFI [0-10]	3.4 (2.6)
VAS pain of the spine (0-10) [cm]	3.5 (2.4)
Night pain [4 point likert][a] (IQR)	1.0 (1.0-2.0)
Chest expansion [cm]	4.7 (2.2)
10 cm Schober [cm]	2.8 (1.4)
Tragus-to-wall distance [cm][a] (IQR)[range]	12.5 (11.0-16.0)[8.2-34.4]
Occiput-to-wall distance [cm][a] (IQR)[range]	1.6 (0.0-6.0)[0.0-26.1]
VAS patient global (0-10) [cm]	3.5 (2.8)
Peripheral arthritis [present / absent][b]	57 / 160
ESR [mm/hr][a] (interquartile range)	10 (5-19)
CRP [mg/l][a] (interquartile range)[a]	7 (6-19)
VAS physician on disease activity[a] [0-10] (interquartile range)	1.4 (0.5-3.3)
VAS patient on disease activity (0-10) [cm]	3.8 (2.8)
Duration of morning stiffness [min]	36 (30)

[a] Median (interquartile range); IBD=inflammatory bowel disease; BASFI=Bath ankylosing spondylitis functional index; VAS=visual analogue scale; IQR=inter quartile range; [b] defined as ≥1 swollen joint on physical examination.

Figure 3.2 shows this regression line for all patients. Scatterplots according to Band and Altman are shown for OWD and TWD for all patients for the comparison of T_0 and T_6 in Figures 3.3 and 3.4 respectively. The patterns of the plots were roughly similar for OWD and TWD. There is no influence of the magnitude of the OWD or TWD on the measurement error, i.e., the measurement error is similar along the entire scale. The summary results for the average scores and differences as well as the 95% limits of agreement comparing baseline and $T_{6,}$ T_6 and T_{12}, T_{12} and T_{18} and T_{18} and T_{24} are shown in Table 3.2. Results were comparable between all paired observations. Comparing OWD and TWD at various time intervals, there is no consistent difference between the means and the SD in either the upper or the lower limit of agreement. As well, the limits of agreement for both OWD and TWD in all patients did not differ importantly. The lower 95% limit of agreement was between –3.4 cm and –2.5 cm for OWD and between –3.4 cm and -3.1 cm for TWD. The upper limit of agreement was between 3.1 cm and 4.2 cm for OWD and between 2.9 cm and 3.9 cm for TWD (Table3.3).

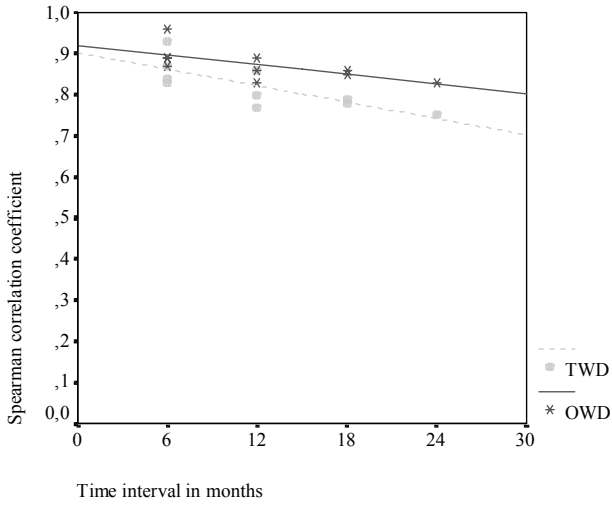

Figure 3.2 Graph of the spearman correlation coefficient against each interperiod with regression line for OWD and TWD.

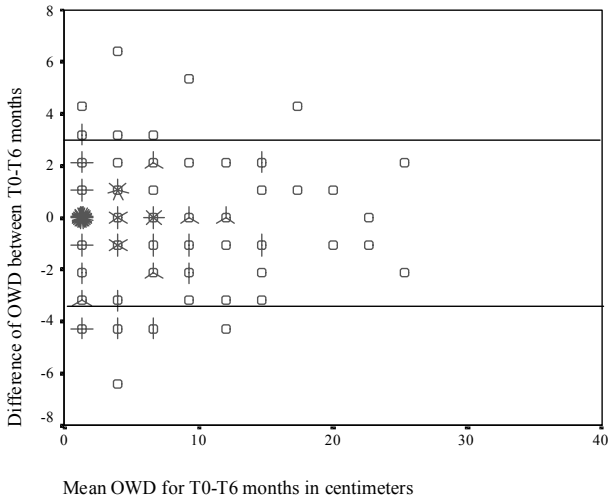

Figure 3.3 Graph of the difference scores against the mean score of OWD in cm in all patients, The upper line is indicating the upper limit (+2 SD) and the lower line (-2 SD) the lower limit of the 95% of agreement interval. □ is representing one patient; each additional spike of the sunflower is representing an additional patient. (For example the data point the most left in the lower part of the figure below the line of –2SD is representing five patients).

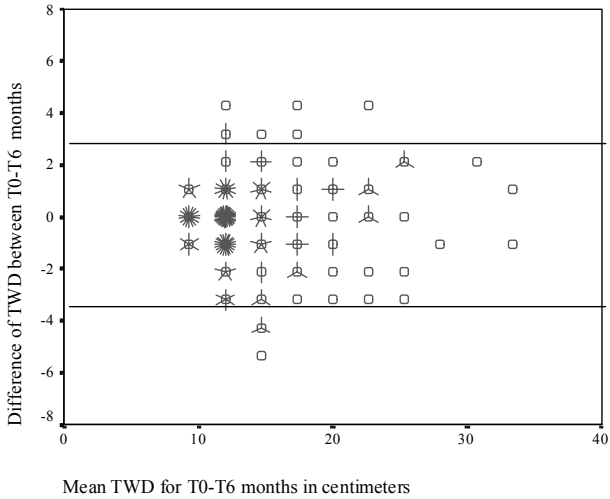

Figure 3.4 Graph of the difference scores against the mean score for TWD in cm in all patients, each spike of the sunflower representing one case.

Table 3.2 Summary statistics of the average scores of occiput-to-wall distance (OWD) and tragus-to-wall distance (TWD) with a six months interval, summary statistics of the difference scores of OWD and TWD with a six months interval and 95% limits of agreement of scores of OWD and TWD with a six months interval of 217 patients.

Paired data	Average of the 2 observations in cm[a]		Difference of paired observations in cm[b]	95% limits of agreement of paired observations in cm
	Mean (SD)	Median (min; max)	Mean (SD)[c]	
OWD T_0-T_6	3.8 (5.5)	1.2 (0.0 ; 24.9)	-0.2 (1.7)	-3.4 ; 3.1
TWD T_0-T_6	14.2 (4.7)	12.2 (8.2 ; 34.0)	-0.2 (1.6)	-3.4 ; 2.9
OWD T_6-T_{12}	3.8 (5.6)	1.4 (0.0 ; 25.9)	0.6 (1.8)	-2.9 ; 4.2
TWD T_6-T_{12}	14.0 (4.8)	12.3 (7.5 ; 34.0)	0.3 (1.8)	-3.3 ; 3.9
OWD T_{12}-T_{18}	4.1 (5.7)	2.0 (0.0 ; 28.2)	0.2 (1.7)	-3.1 ; 3.5
TWD T_{12}-T_{18}	14.3 (5.1)	12.6 (6.5 ; 35.0)	0.4 (1.8)	-3.1; 3.9
OWD T_{18}-T_{24}	4.4 (5.9)	2.5 (0.0 ; 29.5)	0.4 (1.5)	-2.5 ; 3.3
TWD T_{18}-T_{24}	14.6 (5.4)	12.9 (6.4 ; 39.0)	0.3 (1.7)	-3.1 ; 3.7

[a] Average score of the two observations = (observation 1 + observation 2) divided by 2; [b] Difference score between the two observations = observation 1 minus observation 2; [c] SD of the difference scores (i.e. the $SD_{difference}$) which is an estimate of random measurement error used to calculate the 95% limits of agreement.

Table 3.3 95% limits of agreement for the whole group of patients and 95% limits of agreement for the group of patients with OWD>0.

Paired data	95% limits of agreement of paired observations in all patients in cm n=217	95% limits of agreement of paired observations in cm in kyphosed patients n=115
OWD T_0-T_6	-3.4; 3.1	-4.8; 4.2
TWD T_0-T_6	-3.4; 2.9	-4.2; 3.6
OWD T_6-T_{12}	-2.9; 4.2	-3.7; 5.3
TWD T_6-T_{12}	-3.3; 3.9	-3.3; 4.5
OWD T_{12}-T_{18}	-3.1; 3.5	-3.9; 4.3
TWD T_{12}-T_{18}	-3.1; 3.9	-3.5; 4.3
OWD T_{18}-T_{24}	-2.5; 3.3	-4.3; 3.1
TWD T_{18}-T_{24}	-3.1; 3.7	-3.5; 4.7

Analyzing the 95% limits of agreement of OWD and TWD in all patients compared to kyphosed patients shows that the 95% limits of agreement are comparable for OWD and TWD (Table 3.4). However, the 95% limits of agreement in kyphosed patients is increased compared to all patients: for OWD –4.8 and 4.2 cm and for TWD –4.2 and 3.6 cm.

Table 3.4 Ranges of OWD and TWD of patients with OWD>0 at T_0 or T_6 are presented, as well as 95% limits of agreement between T_0 and T_6 for the whole group of patients with OWD>0 divided in quartiles.

		T_0-T_6 (n=115)				
		Whole range	1st quartile (n=29)	2nd quartile (n=29)	3rd quartile (n=29)	4th quartile (n=28)
OWD T_0-T_6	Range	0.4 - 24.9	0.4 - 2.6	2.7 - 5.3	5.4 – 9.8	10.0-24.9
	95% Limits of agreement T_0-T_6	-4.8 ; 4.2	-5.4 ; 4.3	-5.3 ; 4.5	-4.4 ; 3.8	-4.2 ; 3.7
TWD T_0-T_6	Range	10.0 - 34.0	10.0 - 15.5	11.8 - 17.1	13.1 - 20.6	16.6 - 34.0
	95% Limits of agreement T_0-T_6	-4.2 ; 3.6	-4.8 ; 3.0	-4.9 ; 4.1	-3.3 ; 3.7	-4.2 ; 4.0

The mean change in BASDAI between T_0 and T_6 on a group level was 0.09 (SD 1.76). In 100 patients the BASDAI was stable between T_0 and T_6. In this group of patients the limits of agreement for OWD were –3.2 and 3.2 centimeter (cm) and for TWD -3.3 and 3.3 cm. Examining only kyphosed patients with a stable BASDAI (n=56) the limits of agreement were – 4.4 and 4.4 cm for OWD and – 4.5 and 4.5 for TWD. These limits of agreement do not

differ from the limits of agreement found in analysis of the whole group. It thus seems justified in this situation to use measurements with a time interval of six months for test retest purposes. Furthermore, these data suggest there was no systematic error.
Of the 217 patients, 115 (53%) were not able to touch the wall with their occiput at both T_0 and T_6. These patients were classified as kyphosed. Table 3.4 shows the limits of agreement comparing baseline and T_6 for OWD and TWD, calculated for these patients. The results are presented for the whole range of OWD and TWD and divided into quartiles. No consistent pattern over the quartiles was detected for all time intervals. We concluded that the limits of agreement are comparable over the whole scale and thus a more severe kyphosis does not evoke more measurement error. For comparison of OWD and TWD it is appropriate to use the limits of agreement not divided in quartiles of the scale.

Discussion

In measurement of both OWD and TWD it is required that the chin is “in the neutral position”. We hypothesized that involuntary flexion or (hyper)extension of the head, which takes place in the atlanto-occipital joint and the cervical spinal joints and can occur together with thoracic spine extension, would influence the OWD more than the TWD because, compared to the occiput, the tragus lies closer to the sagittal axis about which the atlanto-occipital joint flexion and extension occur. In other words patient variance would influence the OWD more compared to TWD and thus the OWD would be more prone to measurement error. On the other hand, variation in rotation of the head could influence the TWD more compared to the OWD. However, analyzing our data, we found similar ICC and similar limits of agreement for OWD and TWD, which did not confirm a differential aspect in measurement error between the two instruments.
We did not perform a formal test-retest method in assessing the measurement error of the OWD and TWD with a short time interval. However, in our view it is appropriate to use the data with a 6-month interval in a disease with very slow progression. The various analyses we performed do underscore this assumption.
The OWD and TWD instruments did not behave differently across different ranges of the measurement. Auléley *et al*.[9] reported in an international study on 120 patients with AS a 95% limit of agreement between –2.8 and 2.9 cm for OWD, which is little lower than in our group. In their investigation TWD was not studied. However, we do not know the percentage of patients with a normal OWD in that study; if this was higher this may explain the small difference, as

we showed that measurement error is increased in kyphosed patients. In calculation of the BASMI[4,5], an intuitively created index based on the literature, the TWD and not the OWD was used. As stated, this measurement was among 20 separate measurements historically used in the assessment of AS. Pile *et al*.[10] found an excellent inter-observer reliability represented by an inter-observer coefficient of reliability of 0.97 for TWD. No interobserver coefficient of reliability was given for OWD. Viitanen *et al*. reported an interobserver ICC of 0.94 for both OWD and TWD and an intraobserver ICC of 0.98 for OWD and of 0.90 for TWD.[11] But these investigators did not present 95% levels of agreement and a high ICC does not rule out an unacceptable level of measurement error.[12]

Bellamy, *et al*. reported the median of the minimally clinically important difference estimates (MCIDE: as defined by an expert panel with use of Delphi rounds) for TWD as 3.0 cm and for OWD as 1.25 cm.[13] In both our study and the study of Auléley *et al*.[9], the 95% limits of agreement, are beyond the reported MCIDE from this expert panel, which suggests that the MCIDE cannot reliably be distinguished from measurement error.

Our conclusion is that we found no difference in reliability between OWD and TWD. Although we feel the TWD is in general easier to perform compared to OWD, we recommend the OWD measurement over TWD. Our main argument is that in OWD measurement a value of zero easily distinguishes patients with normal thoracic spine extension from kyphosed patients. This argument is especially valid in the research setting, where a value of 0 for OWD clearly means no kyphosis and a value for example, of 12 for TWD could mean no kyphosis or beginning kyphosis. We think this outweighs the minor disadvantages of the feasibility of measurement of OWD.

References

1. Bellamy N. Muscoloskeletal Clinical Metrology. Dordrecht / Boston/ London: Kluwer academic publisher; 1993.
2. van der Heijde D, Calin A, Dougados M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. Assessments in Ankylosing Spondylitis. J Rheumatol 1999;26:951-4.
3. Bellamy N. Clinimetric Concepts in Outcome Assessment: The Omeract Filter. Journal of Rheumatology 1999;26:948-50.
4. Jenkinson TR, Mallorie PA, Whitelock HC, Kennedy LG, Garrett SL, Calin A. Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. J Rheumatol 1994;21:1694-8.
5. Jones SD, Porter J, Garrett SL, Kennedy LG, Whitelock H, Calin A. A new scoring system for the Bath Ankylosing Spondylitis Metrology Index (BASMI). J Rheumatol 1995;22:1609.
6. Spoorenberg A, van der Heijde D, de Klerk E, Dougados M, de Vlam K, Mielants H, et al. A comparative study of the usefulness of the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index in the assessment of ankylosing spondylitis. J Rheumatol 1999;26:961-5.
7. van der Heijde DM, van 't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Ann Rheum Dis 1990;49:916-20.
8. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
9. Auléley GR, Benbouazza K, Spoorenberg A, Collantes E, Hajjaj-Hassouni N, van der Heijde D, et al. Evaluation of the smallest detectable difference in outcome or process variables in ankylosing spondylitis. Arthritis Rheum 2002;47:582-7.
10. Pile KD, Laurent MR, Salmond CE, Best MJ, Pyle EA, Moloney RO. Clinical assessment of ankylosing spondylitis: a study of observer variation in spinal measurements. Br J Rheumatol 1991;30:29-34.
11. Viitanen JV, Heikkila S, Kokko ML, Kautiainen H. Clinical assessment of spinal mobility measurements in ankylosing spondylitis: a compact set for follow-up and trials? Clin Rheumatol 2000;19:131-7.
12. Atkinson G, Nevill A. Statistical methods for assessing measurement error(reliability) in variables relevant to sports medicine. Sports Med 1998;26:217-38.
13. Bellamy N, Buchanan WW, Esdaile JM, Fam AG, Kean WF, Thompson JM, et al. Ankylosing spondylitis antirheumatic drug trials. III. Setting the delta for clinical trials of antirheumatic drugs--results of a consensus development (Delphi) exercise. J Rheumatol 1991;18:1716-22.

Chapter 4

The influence of peripheral arthritis on disease activity in ankylosing spondylitis patients as measured with the Bath ankylosing spondylitis disease activity index

L Heuft-Dorenbosch, A van Tubergen, A Spoorenberg, R Landewé,
M Dougados, H Mielants, H van der Tempel, D van der Heijde

Arthritis & Rheumatism 2004;15:154-159

Abstract

Objective
To assess the differences in disease activity as measured by the Bath ankylosing spondylitis disease activity index (BASDAI) between patients with and without peripheral arthritis/enthesitis. To investigate whether scores on the BASDAI change by omitting the 2 questions on peripheral disease.

Methods
Disease activity was assessed on a 10-cm visual analog scale and by BASDAI. Alternative BASDAIs were constructed by omitting the peripheral joints question and/or the enthesitis question. Correlations between the alternative BASDAIs and other measures of disease activity were calculated. Generalized estimating equations (GEE) were used to assess whether having peripheral arthritis influenced BASDAI and alternative BASDAI scores, and to assess whether peripheral arthritis influenced the score of the individual questions of the BASDAI.

Results
At baseline, the BASDAI was calculated in 214 patients. In patients with peripheral arthritis (n=56), the mean (SD) BASDAI score was 4.4 (2.3) as compared with 3.1 (1.9) (P<0.0001) in the patients without peripheral arthritis (n=158). The relationship between arthritis and the BASDAI score appeared to be truly longitudinal (GEE regression coefficient β=0.64; 95% confidence interval 0.28–1.00). Peripheral arthritis was significantly longitudinally associated with all separate item scores of the BASDAI. Omitting the peripheral joints and/or enthesitis question from the BASDAI questionnaire only partially explained the difference in BASDAI score between the 2 groups.

Conclusion
Disease activity measured by the BASDAI is higher in patients with concomitant peripheral disease compared with patients with disease restricted to the axial skeleton. The increased BASDAI score in patients with peripheral arthritis is partially explained by increased overall disease activity as well as by a disproportionate contribution of the peripheral joints question to the overall score.

Introduction

Ankylosing spondylitis (AS) is a chronic rheumatic condition characterized by a wide spectrum of symptoms.[1] These symptoms can be limited to the axial skeleton (spinal disease) or can be more extensive, whereby AS can present with peripheral arthritis, enthesitis, dactylitis, and uveitis (extraspinal disease). For measuring disease activity in AS, no gold standard currently exists. Objective measures such as C-reactive protein (CRP) and Westergren erythrocyte sedimentation rate (ESR) poorly correlate with clinical disease activity.[2] The assessment in ankylosing spondylitis (ASAS) working group selected in 1998 a core set of single instruments to assess disease activity and disease severity in AS.[3] Combined instruments are not included in the core set. The progress reports stated that the value of combined instruments, such as the Bath ankylosing spondylitis metrology index (BASMI)[4,5] and Bath ankylosing spondylitis disease activity index (BASDAI)[6], should be investigated because the combined instruments might not have the same validity in all types of disease, e.g., in spinal or extraspinal disease.

The BASDAI[6] is a widely used subjective measure to assess disease activity that proved to be valid, reproducible, and responsive to change.[7,8] The BASDAI consists of 6 questions, of which one deals with complaints in peripheral joints (question 3) and one with enthesitis (question 4). The peripheral joints question is worded, “How would you describe the overall level of pain/swelling in joints other than neck, back or hips you have had”. The enthesitis question is worded, “How would you describe the overall level of discomfort you have had from any areas tender to touch or pressure.” Both questions relate to the past week.

We hypothesized that patients with peripheral arthritis would show a higher BASDAI score compared with patients with AS restricted to the axial skeleton, and that this higher score might be attributed to a higher score on the BASDAI questions about peripheral joints and entheses. The first aim of the present study was to investigate whether a difference exists in degree of disease activity as measured by the BASDAI between patients with and without peripheral arthritis or enthesitis; in other words, to assess whether arthritis is a determinant of disease activity. The second goal was to investigate whether this difference would change by omitting the questions on arthritis and/or enthesitis. Finally, the correlation was assessed between the BASDAI, including and excluding the questions on arthritis and enthesitis, and other measurements of disease activity in patients both with and without peripheral arthritis or enthesitis.

Patients and Methods

Patients

In our study, we used data from the outcome in ankylosing spondylitis international study (OASIS) cohort, an international, longitudinal, observational study on outcome of AS with follow-up visits according to a fixed protocol. Data from this cohort have been reported previously.[2,9] Consecutive outpatients with an established diagnosis of AS according to the modified New York criteria[10] were included in 1996. We used data from visits at baseline (T_0), 6 months (T_6), 12 months (T_{12}), 18 months (T_{18}), and 24 months (T_{24}). All patients completed a number of questionnaires and underwent a clinical examination.

Measures of disease activity.

The core set of variables selected by the ASAS working group[3] were all assessed.

The BASDAI consists of six questions: the first on fatigue; the second on pain of the spine and hips; the third on pain or swelling of the peripheral joints; the fourth on enthesitis; and the last two on severity and duration of morning stiffness, respectively.[6] The questions are answered on a 10-cm visual analog scale (VAS). The mean of the two scores regarding morning stiffness counts as one variable. The final score is defined by calculating the mean of the five items. Scores range from 0 (best) to 10 (worst). In case of missing values for questions 1–4, the BASDAI was not calculated and reported as missing. In case of one missing value of question 5 or 6, the remaining question counted as the mean of questions 5 and 6. We calculated alternative BASDAIs in three different ways. First, by omitting question 3, the peripheral joints question (BASDAI-without-joints); second by omitting question 4, the enthesitis question (BASDAI-without-entheses); and third by omitting both the peripheral joints and the enthesitis question (BASDAI-without-joints-entheses). The mean of the remaining items was calculated to obtain the value of these alternative BASDAIs.

Disease activity was also assessed by a single-item 10-cm VAS concerning the degree of disease activity during last week (patient) and on the day of physical examination (physician).

Function was measured by the Bath ankylosing spondylitis functional index (BASFI).[11] During clinical examination, the physician performed a painful (53 joints) and swollen (44 joints) joint count. Peripheral arthritis was defined as the presence of at least one swollen joint. Peripheral arthritis data on follow-up visits at T_6, T_{18}, and T_{24} were available only for patients in The Netherlands (n=137). Pain was recorded in two ways: pain of the spine due to AS last week, on average, using a 10-cm VAS; and night pain last night using a 4-point Likert

scale. Patients' global well-being was measured on a 10-cm VAS. ESR was assessed using the Westergren method (normal range male 0–7 mm/hour, female 0–12 mm/hour) and CRP by the turbidimetric method (normal range 2-9 mg/liter). The lowest detection limit for CRP was 2 mg/l and patients with undetectable levels were assigned 0.

Statistical analysis

Each study visit was analyzed independently. Independent *t*-tests were used to compare patients with (A+ group) and without (A- group) peripheral arthritis with regard to scores on the BASDAI and alternative BASDAIs, and on the separate questions of the BASDAI.

Correlation between the original and alternative BASDAIs and disease activity by patients was calculated using Pearson's correlation coefficient. Correlation between the original and alternative BASDAIs and disease activity by physician, ESR, and CRP were calculated using Spearman's correlation coefficient because of skewness in data distribution.

In general, longitudinal data sets are characterized by observations with high variability between patients and rather low variability within patients (i.e., the BASDAI score at T_0 is highly correlated with the BASDAI score at T_{12}). Because of the high within-patient correlation, longitudinal relationships cannot be analyzed with ordinary regression methods. Generalized estimating equations (GEE) is a regression technique for studying intervariable relationships in longitudinal studies; this technique takes into account time, as well as time-independent and time dependent covariates.[12] The advantage of using GEE over other methods is that GEE uses all available longitudinal data, allows unequal numbers of repeated measurements and unequal time intervals, and does not require multivariate normality of the outcome variable. GEE does require an a priori "working" correlation structure to adjust for the within-subject correlation operating in repeated-measurement designs. A correlation structure must be chosen on the basis of the actual data set. Here we chose an exchangeable correlation structure.

Autoregressive GEE was used to assess the effect of having peripheral arthritis on the total score of the BASDAI and alternative BASDAIs, as well as to assess the effect of having peripheral arthritis on the individual question scores of the BASDAI. In autoregressive analysis, each value of the dependent variable at T_x is adjusted for the value of the dependent variable at T_{x-1}. The effect of having peripheral arthritis on the different scores is given as a regression coefficient (β). Betas are standardized by dividing the beta by the standard error (β/SE), which allows comparison of different scales and a different standard error.

Results

At baseline, 217 patients were included. Table 4.1 presents characteristics of the patients. The median score, interquartile range, and range of the individual questions of the BASDAI at T_0 are presented in Figure 4.1. All individual questions of the BASDAI scored significantly lower in the A- group as compared with the A+ group. This effect was observed at all time points. Particularly, the individual BASDAI question scores on peripheral joints, entheses, and stiffness were sensitive to having peripheral arthritis.

Table 4.2 Baseline characteristics and scores on ASAS core set measures (Mean (SD)).

	Total study population (n=217)
Male/ female	150 / 67
Age [years]	43.1 (12.7)
Duration of complaints [years]	19.6 (11.8)
Time since diagnosis [years]	10.8 (8.9)
HLA-B27 [present / absent / no data]	142 / 24 / 51
History of IBD [present / absent / no data]	17 / 145 / 55
History of uveitis [present / absent / no data]	81 / 133 / 3
History of psoriasis [present / absent / no data]	10 / 152 / 55
BASFI [0-10]	3.4 (2.6)
VAS pain of the spine [cm] (0-10)	3.5 (2.4)
Night pain [4 point likert][a] (0-4)	1.0 (1.0-2.0)
Chest expansion [cm]	4.7 (2.2)
10 cm Schober [cm]	2.8 (1.4)
Occiput to wall distance [cm][a]	1.0 (0.0-6.0)
VAS patient global [cm] (0-10)	3.5 (2.8)
Peripheral arthritis [present / absent][b]	57 / 160
ESR [mm/hr][a]	10 (5-19)
CRP [mg /l][a]	7 (6-19)
VAS physician on disease activity [cm](0-10)	1.4 (0.5-3.3)
VAS patient on disease activity [cm](0-10)	3.8 (2.8)
Duration of morning stiffness [min]	36 (30)

[a] Median (interquartile range); ASAS=assessment in ankylosing spondylitis' (ASAS) working group; IBD=inflammatory bowel disease; BASFI=Bath ankylosing spondylitis functional index; ESR=erythrocyte sedimentation rate; CRP=C-reactive protein VAS=visual analogue scale; [b] defined as >1 swollen joint by physical examination.

Table 4.2 shows the scores on the BASDAI, the alternative BASDAIs, and the patients' VAS concerning the degree of disease activity; it also presents the median with interquartile ranges on ESR, CRP, and physicians' single-item VAS concerning the degree of disease activity at baseline. All scores were significantly higher in the A+ group compared with the A- group. By omitting the questions on joints and entheses, the scores in the A- group increased slightly (BASDAI-without-joints or BASDAI-without-entheses) to moderately (BASDAI-

without-joints-entheses) and only slightly in the A+ group if both questions were omitted. Consequently, the difference in scores between the A+ group and the A- group decreased somewhat, particularly if both questions were omitted.

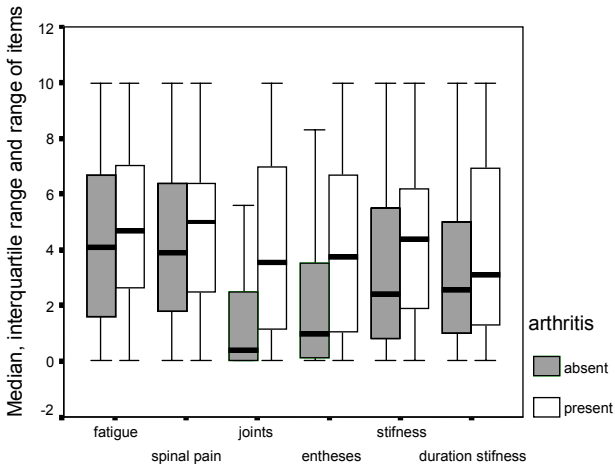

Figure 4.1 Items of BASDAI on T_0 in patients with and without peripheral arthritis.

Table 4.2 Scores on the BASDAI and 'alternative BASDAIs', ESR, CRP, patients' and physicians' disease activity in patients with (A+ group) and without (A-group) peripheral arthritis (Mean (SD)) at baseline.

	A+ group n=56	A- group n=158	Difference between A+ and A- group	p-value
BASDAI [cm]	4.4 (2.3)	3.1 (1.9)	1.3	<0.0001
BASDAI-without-joints [cm]	4.5 (2.4)	3.4 (2.1)	1.1	0.004
BASDAI-without-entheses [cm]	4.5 (2.2)	3.3 (2.0)	1.2	<0.0001
BASDAI-without-joints-entheses [cm]	4.7 (2.4)	3.9 (2.3)	0.8	0.023
ESR [mm/h][a]	14.0 (7-26)	8.0 (4-16)	6	0.006
CRP [mg/l][a]	11.0 (6-30)	7.0 (6-15)	4	0.024
Patient VAS on disease activity [cm]	4.5 (3.0)	3.5 (2.7)	1.0	0.04
Physician VAS on disease activity [cm][a]	2.7 (1.3-4.9)	1.0 (0.5-2.5)	1.7	<0.0001

[a] Median (interquartile range); BASDAI=Bath ankylosing spondylitis disease activity index; ESR=erythrocyte sedimentation rate; CRP=C-reactive protein; VAS=visual analogue scale. All scores (except ESR and CRP) range from 0-10.

Table 4.3 presents the correlation between the various BASDAIs and ESR, CRP, and the patients and the physicians' single-item VAS concerning the degree of disease activity at T_0. The correlations between ESR and CRP and the various BASDAIs were low in both groups. The correlations between the

patients' disease activity and all BASDAIs were consistently higher than those between physician disease activity and the BASDAIs. By leaving out the peripheral joint question, the enthesitis question, or both the peripheral joint and the enthesitis questions, there was no gain in correlation with the patient disease activity in the group without peripheral disease; however, there was substantial loss in the group with peripheral disease.

Table 4.3 Correlations between patients' and physicians' disease activity, ESR and CRP and BASDAI and alternatives on T_0.

	VAS physician on disease activity	VAS patient on disease activity	ESR	CRP	BASFI
A- group					
BASDAI	0.41	0.71	0.14	0.11	0.63
BASDAI-without-joints	0.41	0.73	0.14	0.14	0.60
BASDAI-without-entheses	0.38	0.70	0.13	0.09	0.60
BASDAI-without-joints-entheses	0.37	0.70	0.12	0.15	0.57
A+ group					
BASDAI	0.42	0.73	0.28	0.26	0.68
BASDAI-without-joints	0.35	0.68	0.16	0.17	0.66
BASDAI-without-entheses	0.36	0.69	0.24	0.21	0.68
BASDAI-without-joints-entheses	0.29	0.60	0.11	0.11	0.64

BASFI = Bath ankylosing spondylitis functional index; for additional abbreviations, see Table 4.2. $p<0.01$.

Longitudinal data analysis by GEE allowed us to investigate time trends in scoring by individual patients (Table 4.4). From this table it can be seen that the average patient without arthritis at T_0 will score 0.64 of 10 (6.4% of the scale) points higher at T_6 months if he or she has arthritis at that time point. It can also been seen that excluding the joint and/or enthesitis question(s) from the original BASDAI questionnaire did not significantly change this longitudinal relationship between peripheral arthritis and scoring: the regression coefficient was 0.66 if the BASDAI was calculated without the joint question, 0.66 without the enthesitis question, and 0.68 without both questions. This analysis suggests that peripheral arthritis in patients with AS reflects higher disease activity in general, as measured by the BASDAI questionnaire. The relationship between BASFI and true peripheral arthritis was not statistically significant, suggesting that function is less influenced by having arthritis.

Table 4.5 shows the longitudinal relationship between true peripheral arthritis and the 6 item scores from the BASDAI separately. The β in this table gives the amount of change of the calculated variable when assessment of arthritis changes from "no arthritis" to "arthritis" and thus the model can be interpreted as follows: A patient without arthritis will have an average increase on the fatigue question of the BASDAI of 0.75 if he has arthritis and an average

increase of 0.71 on the spine and hip pain question. As can be seen, presence of arthritis increased the score on all individual questions. The effect of having arthritis was the lowest on the question on spinal pain and the largest on the peripheral joints question.

Table 4.4 Longitudinal relationship between having peripheral arthritis and score on the BASDAI and its alternatives, and the BASFI in patients with AS: first order autoregressive analysis using generalized estimating equations (GEE).

	β	95 %CI	p-value	standardized β
BASDAI[a]				
arthritis (yes/no)	0.64	0.28 to 1.00	<0.0001	1.8
BASDAI-without-joints[a]				
arthritis (yes/no)	0.66	0.30 to 1.02	<0.0001	1.7
BASDAI-without-entheses[a]				
arthritis (yes/no)	0.66	0.28 to 1.04	0.001	1.7
BASDAI-without-joints-entheses[a]				
arthritis (yes/no)	0.68	0.28 to 1.07	0.001	1.7
BASFI[a]				
arthritis (yes/no)	0.15	-0.11 to 0.41	0.23	0.6

[a] Dependent variable in the GEE. BASDAI and its derivates as well as BASFI range from 0 to 10. In all analyses the values of the dependent variable at the previous time point is included as autoregressive variable; the coefficient for the autoregressive variable is not presented in the table. For abbreviations see table 4.1.

Table 4.5 Longitudinal relationship between having peripheral arthritis and score on the individual questions of BASDAI in patients with ankylosing spondylitis: first order autoregressive analysis using generalized estimating equations (GEE).

	β	95% CI	p-value	standardized β
Fatigue question BASDAI (question 1)[a]				
arthritis (yes/no)	0.75	0.50-1.0	0.003	3.0
Spinal pain and hips question BASDAI (question 2)[a]				
arthritis (yes/no)	0.71	0.43-0.99	0.011	2.5
Peripheral joint question BASDAI (question 3)[a]				
arthritis (yes/no)	1.39	1.08-1.70	<0.001	4.5
Entheses question BASDAI (question 4)[a]				
arthritis (yes/no)	0.87	0.61-1.13	0.001	3.3
Intensity Morning stiffness BASDAI (question 5)[a]				
arthritis (yes/no)	0.85	0.63-1.07	<0.001	3.9
Duration Morning stiffness BASDAI (question 6)[a]				
arthritis (yes/no)	0.63	0.43-0.83	0.002	3.2

[a] Dependent variable in the GEE. Individual BASDAI questions range from 0 to 10. In all analyses the values of the dependent variable at the previous time point is included as autoregressive variable; the coefficient for the autoregressive variable is not presented in the table.

Discussion

The BASDAI includes the components fatigue, spinal and hip pain, peripheral joint pain and swelling, enthesitis, and morning stiffness (both severity and duration). The designers of the BASDAI considered these as important and common symptoms in AS.[6] In a six-week, double-blind, placebo-controlled nonsteroidal anti-inflammatory drug study, the BASDAI proved to be a valid and appropriate composite index to define disease activity in AS.[7] In this study however, only patients with axial disease were accepted and those with active peripheral involvement were excluded.

Because of the individual weight of each question of the BASDAI, we hypothesized that patients with AS limited to the axial skeleton would have lower scores on the BASDAI compared with patients with concomitant peripheral disease. We further hypothesized that this would be due to a lower scoring on both the peripheral joint question and the enthesitis question. In our study, patients with peripheral arthritis indeed reported higher disease activity compared with patients without peripheral arthritis as measured by the BASDAI. Although this difference is most obvious in the peripheral joints question and the enthesitis question, it is also present in questions on fatigue, duration and intensity of morning stiffness, and spinal pain. Cautious observation of the data shows that the higher BASDAI in patients with peripheral arthritis might be explained by two separate phenomena. 1) Patients with peripheral arthritis have higher overall disease activity, which is reflected by higher scores in the nonarthritis items. 2) Having peripheral arthritis - and a higher score on the peripheral arthritis question - contributes disproportionately to the BASDAI score, which is reflected by the highest standardized regression coefficient for the peripheral joints question.

As a consequence of these separate effects, the difference in BASDAI between patients with and those without arthritis only partially disappears by omitting the joints and enthesitis questions from the BASDAI questionnaire.

Similarly, different aspects of peripheral joint disease in AS have been reported in the past. In 1958, Wilkinson and Bywaters concluded that the main factor influencing prognosis in AS was the presence or absence of extraspinal joint involvement.[13] Carette et al observed in 1983 in their study in 150 war veterans with AS that early peripheral joint disease predicts disease severity.[14] Claudepierre et al. reported in 1998 that disease activity in spondylarthropathy was linked to peripheral joint disease, disease duration, and dietary habits.[15] However, Claudepierre et al. evaluated a French version of the BASDAI in which a high score on spinal pain did not correlate with pain in peripheral joints.[16] In our study, we found a significant correlation coefficient between question 2 (spinal and hip pain) and question 3 (joint pain) of 0.36 at T_0 and ranging from 0.61 to 0.67 at T_6, T_{12}, T_{18}, and T_{24}. Differences in study population

could be an explanation of the difference in correlation found between the mentioned questions of the BASDAI.

Our most important conclusion is that disease activity as measured with the BASDAI is higher in patients with peripheral joint disease compared with patients with disease limited to the axial skeleton, and thus that peripheral arthritis is an expression of higher disease activity. Modification of the BASDAI by omitting the question(s) on peripheral joint pain and/or enthesitis results in neither disappearance of this difference nor in a better correlation between BASDAI and other measures of disease activity.

References

1. van der Linden S, van der Heijde D. Ankylosing spondylitis: clinical features. Rheum Dis Clin North Am 1998;24:663–76.
2. Spoorenberg A, van der Heijde D, de Klerk E, Dougados M, de Vlam K, Mielants H, et al. Relative value of erythrocyte sedimentation rate and C-reactive protein in assessment of disease activity in ankylosing spondylitis. J Rheumatol 1999;26:980–4.
3. van der Heijde D, van der Linden S, Bellamy N, Calin A, Dougados M, Khan MA. Which domains should be included in a core set for endpoints in ankylosing spondylitis: introduction to the ankylosing spondylitis module of OMERACT IV. J Rheumatol 1999;26:945–7.
4. Jenkinson TR, Mallorie PA, Whitelock HC, Kennedy LG, Garrett SL, Calin A. Defining spinal mobility in ankylosing spondylitis (AS): The Bath AS Metrology Index. J Rheumatol 1994;21:1694–8.
5. Jones SD, Porter J, Garrett SL, Kennedy LG, Whitelock H, Calin A. A new scoring system for the Bath Ankylosing Spondylitis Metrology Index (BASMI). J Rheumatol 1995;22:1609.
6. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286–91.
7. Calin A, Nakache JP, Gueguen A, Zeidler H, Mielants H, Dougados M. Defining disease activity in ankylosing spondylitis: is a combination of variables (Bath Ankylosing Spondylitis Disease Activity Index) an appropriate instrument? Rheumatology (Oxford) 1999;38:878–82.
8. Jones SD, Calin A, Steiner A. An update on the Bath Ankylosing Spondylitis Disease Activity and Functional Indices (BASDAI, BASFI): excellent Cronbach's alpha scores. J Rheumatol 1996;23:407.
9. Spoorenberg A, van der Heijde D, de Klerk E, Dougados M, de Vlam K, Mielants H, et al. A comparative study of the usefulness of the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index in the assessment of ankylosing spondylitis. J Rheumatol 1999;26:961–5.
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;27:361–8.
11. Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281–5.
12. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30.
13. Wilkinson MBE. Clinical features and course of ankylosing spondylitis as seen in a follow up of 222 hospital referred cases. Ann Rheum Dis 1958;17:209–28.
14. Carette S, Graham D, Little H, Rubenstein J, Rosen P. The natural disease course of ankylosing spondylitis. Arthritis Rheum 1983;26:186–90.
15. Claudepierre P, Sibilia J, Roudot Thoraval F, Flipo RM, Wendling D, Goupille P, et al. Factors linked to disease activity in a French cohort of patients with spondyloarthropathy. J Rheumatol 1998;25:1927–31.
16. Claudepierre P, Sibilia J, Goupille P, Flipo RM, Wendling D, Eulry F, et al. Evaluation of a French version of the Bath Ankylosing Spondylitis Disease Activity Index in patients with spondyloarthropathy. J Rheumatol 1997;24:1954–8.

Chapter 5

Radiographic assessment of sacroiliitis by radiologists and rheumatologists: does training improve quality?

A van Tubergen, L Heuft-Dorenbosch, G Schulpen, R Landewé, R Wijers, D van der Heijde, J van Engelshoven, Sj van der Linden

Ann Rheum Dis 2003;62:519-525

Abstract

Objective
To assess the performance of radiologists and rheumatologists in detecting sacroiliitis.

Methods
100 rheumatologists and 23 radiologists participated. One set of films was used for each assessment, another for training, and the third for confidence judgment. Films of HLA-B27+ patients with AS were used to assess sensitivity. For specificity films of healthy HLA-B27- relatives were included. Plain sacroiliac (SI) films with simultaneously taken computed tomographic scans (CTs) were used for confidence judgment. Three months after reading the training set, sensitivity and specificity assessments were repeated. Next, participants attended a workshop. They also rated 26 SI radiographs and 26 CTs for their trust in each judgment. Three months later final assessments were done.

Results
Sensitivity (84.3%/79.8%) and specificity (70.6%/74.7%) for radiologists and rheumatologists were comparable. Rheumatologists showed 6.3% decrease in sensitivity after self education ($p=0.001$), but 3.0% better specificity ($p=0.008$). The decrease in sensitivity was reversed after the work-shop. Difference in sensitivity three months after the workshop and baseline was only 0.5%. Sensitivity <50% occurred in 13% of participants. Only a few participants showed changes of >5% in both sensitivity and specificity. Intra-observer agreement for sacroiliitis grades 1 or 2 ranged from 65% to 100%. Sensitivity for CT (86%) was higher than for plain films (72%) ($p<0.001$) with the same specificity (84%). Confidence ratings for correctly diagnosing presence (7.7) or absence (8.3) of sacroiliitis were somewhat higher than incorrectly diagnosing the presence (6.6) or absence (7.4) of sacroiliitis ($p<0.001$).

Conclusion
Radiologists and rheumatologists show modest sensitivity and specificity for sacroiliitis And sizeable intraobserver variation. Overall, neither individual training nor workshops improved performance.

Introduction

Ankylosing spondylitis (AS) is a chronic inflammatory disease, in which predominantly the axial skeleton is affected. Sacroiliitis is considered as the hallmark of the disease. Involvement of the sacroiliac (SI) joints is usually established by plain radiographs, or - to a lesser degree - by computed tomography (CT) or magnetic resonance imaging (MRI). Rheumatologists in daily practice mostly order plain radiographs for diagnosing sacroiliitis, and they often read these films themselves. Radiologists read and report radiographs of SI joints exams requested by rheumatologists, general practitioners, or orthopaedic surgeons. Reading radiographs of the SI joints is considered difficult and the diagnosis of sacroiliitis is often missed or incorrectly established. Inter-and intraobserver variations have been reported to be large.[1-3] Kappa statistics to express intraobserver variation in these studies ranged from 0.07 to 1.0[2,3] and inter-observer variation from 0.19 to 0.79.[1-3]
We questioned whether the performance of radiologists and rheumatologists in reading of SI radiographs differs, and whether this performance could be improved by offering training sessions. We provided individual self education with a training set of radiographs and as a second step organized workshops to read radiographs of the SI joints. Our study aimed at defining variations in sensitivity and specificity of detecting sacroiliitis for both radiologists and rheumatologists, and at investigating to what degree sensitivity and specificity would persistently change after training sessions. We also aimed at quantifying the intraobserver variability in reading these radiographs. Furthermore, we wanted to estimate the self confidence of radiologists and rheumatologists in determining the presence or absence of sacroiliitis, for both plain radiographs and CT scan of the SI joints.

Patients and Methods

Participants

All 154 Dutch rheumatologists, except for the six involved in the study, were invited to participate in this project. In total 117 agreed, but for a variety of reasons 17 did not start. At baseline, 100 (85% of the 117 consenting rheumatologists) participated, after self-education 86 (74%), and after the workshop 75 (64%) (Figure 5.1). In addition, a random sample of 30 consenting radiologists (out of the total number of 687 radiologists in the Netherlands) was taken from a list of members of the Dutch Radiology Society. The percentage of their daily work which consisted of viewing skeletal radiographs was as follows: for three <10%, 15 radiologists 10-30%, ten radiologist 30-50%, and

this was unknown for two radiologists. At baseline 23 (77% of the random sample of radiologists) participated, after self-education 22 (73%), and after the workshop eight (27%) (Figure 5.1).

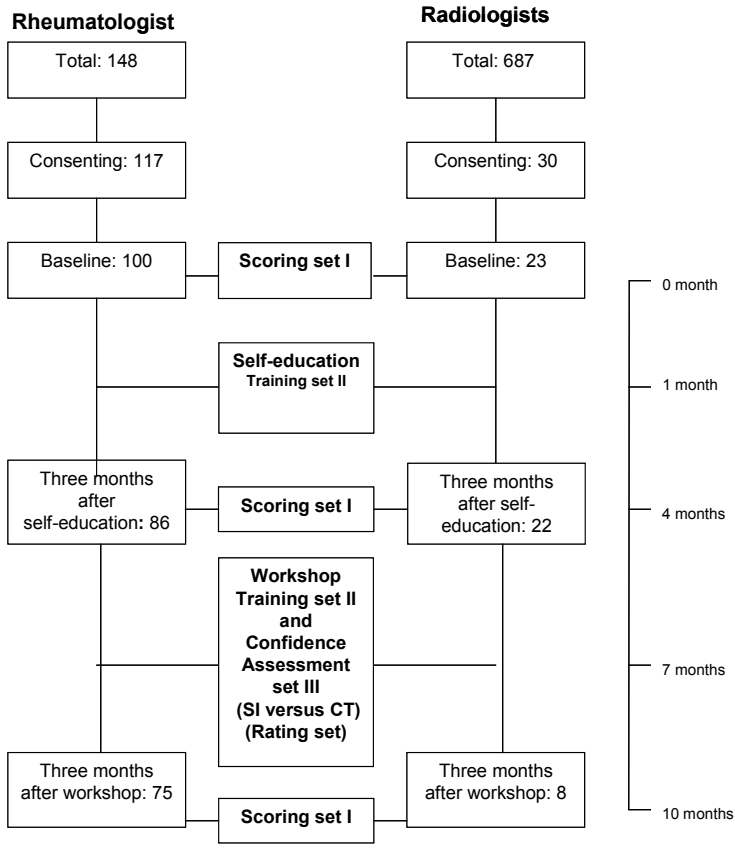

Figure 5.1 Flow chart of participants and examinations.

Radiographs of the sacroiliac joints

For this study, three different sets of radiographs of the SI joints were composed: a *scoring* set for the assessments, a *training* set for individual training, and a *confidence assessment* set to estimate the observer's perceived certainty in diagnosing sacroiliitis by plain radiograph or CT scan of the SI joints. All radiographs were derived from a large Swiss family survey among 275 HLA-B27 positive patients with AS and 511 first degree relatives, who all completed questionnaires and underwent physical examination, HLA-typing,

and radiographic studies of the SI joints.[4] The radiographs were projected from an anteroposterior view, limited to the SI joints, and not including the hip joints. All radiographs had been scored twice "blindly" by each of four experts (two rheumatologists, one epidemiologist, and one radiologist). The mean score of the eight readings of each SI joint was rounded to the next whole figure. Only SI films of HLA-B27 positive patients with AS showing definite sacroiliitis (as defined by the expert panel) and fulfilling the modified New York criteria were included in this study to assess observer's sensitivity. For the assessment of specificity we used the films of HLA-B27 negative first degree relatives of the HLA-B27 positive patients with AS. These relatives had no signs or symptoms suggestive of AS. In addition, a subset of radiographs with simultaneously taken CTs was selected. In the same way, the four experts had judged these CTs separately from the conventional plain SI films.

The New York scoring method for the SI joints was followed: 0=no abnormalities; 1=suspicious changes (no specific abnormalities); 2=minimal sacroiliitis (loss of definition at the edge of the SI joints, there is some sclerosis and perhaps minimal erosions, there may be some joint space narrowing); 3=moderate sacroiliitis (definite sclerosis on both sides of the joint, blurring and indistinct margins, and erosive changes with loss of joint space); 4=complete fusion or ankylosis of the SI joint (without some residual sclerosis).[5] According to the modified New York criteria at least grade 2 bilaterally or grade 3 or 4 unilaterally is necessary for the diagnosis of AS.[6] The gradings for the left and right SI joint were recoded into one final grading representing sacroiliitis according to the New York criteria.

The *assessment* or *scoring* set comprised 50 radiographs, of which 10 appeared twice (reversed-that is, the left joint was now marked as the right joint and visa versa): 12 (+3 repeats) radiographs with a final New York grading of "no sacroiliitis", 12 (+3 repeats) with a grading of "dubious abnormalities", 12 (+3 repeats) with bilateral definite sacroiliitis grade 2, and four films (+ 1 repeated) with a grading 3 or 4. For each radiograph, only the age and sex of the subject was provided. The mean age of the 16 patients with AS represented in the scoring set was 44.1 years; the mean disease duration was 14.9 years at the time the SI radiographs for this study were taken. The *training* set also comprised 50 radiographs (10 for each grading 0 to 4). For each radiograph, information on the grading was provided. The *confidence assessment* set was composed of 17 plain SI radiographs and 17 corresponding CTs with sacroiliitis and nine SI/CT pairs without sacroiliitis. For these films no information on grading, age, gender or clinical findings was provided. All radiographs in the *scoring* set and in the *confidence assessment* set appeared in a completely random order.

Assessments

In total three assessments took place with the scoring set (Figure 5.1). At each of these three occasions all participants individually graded each SI joint of the 50 radiographs according to the New York criteria. First, a baseline score of sensitivity and specificity was established with the scoring set. One month later, each participant received individually the training set in order to practice reading radiographs individually. Three months after this training by self-education, the participants again received the scoring set to assess the presence of persisting effects of this training procedure.
Another three months later, the participants attended one of several workshops organized throughout the country (with a maximum of 20 participants per workshop) in which the full spectrum of normal and abnormal sacroiliac joints, and the grading of these joints according to the New York criteria, were shown and discussed with the participants. In this workshop the same training set as for individual self education was used. Finally, in the same workshop, the participants also judged another 26 radiographs of the confidence assessment set for the presence or absence of sacroiliitis (according to the modified New York criteria), and rated their self confidence about their "yes/no" judgment from 1 (maximally uncertain) to 10 (absolutely sure). The same exercise was done for the 26 CTs from the same patients. Half of the participants first assessed the plain SI joints, and the other half started with reading the 26 CT films. When all scoring data were completed, all the radiographs and CTs were discussed referring to the judgments by the expert panel. Three months after this workshop, the final assessment took place with the *scoring* set as before.
All radiograph sets were sent by post to the participants, who were requested to read the radiographs within the next two weeks. If necessary, a reminder was sent. After the last measurement, all participants received feedback on their personal scores together with the aggregated (centiles) results of all participants.

Statistical analysis

After completion of each *scoring* set, the final grading for each radiograph was dichotomized into presence or absence of sacroiliitis. Grade 2 or more bilaterally, and grade 3 or 4 unilaterally was taken as the presence of sacroiliitis according to the New York criteria. The results from each participant were compared with the "gold standard" as defined by the expert panel. Sensitivity and specificity for each participant were calculated using 2x2 tables for each assessment period. The results of sensitivity and specificity at baseline were compared with the results after self education and after the workshop by paired *t*- tests.

The mean intraobserver agreement (concordance rate) was calculated by means of 2x2 tables for both radiologists and rheumatologists for each of the ten films that were presented in duplicate (although in reversed right-left order). The Kappa statistic was not applied. Repeated films were included only once in the analysis of sensitivity and specificity.
To calculate sensitivity and specificity the judgments on the presence or absence of sacroiliitis for the 26 SI radiographs and CTs were compared with the standard defined by the experts. Differences in sensitivity and specificity between SI radiographs and CTs were calculated with paired *t* tests. The ratings that the participants had provided for their self confidence were subdivided into two groups: films *correctly* (according to the expert panel) diagnosed for the presence or absence of sacroiliitis and films *incorrectly* diagnosed for the presence or absence of sacroiliitis. This was done for plain radiographs and CTs separately. The differences between ratings on correctly versus incorrectly diagnosed sacroiliitis, and the differences in ratings of radiographs versus CTs, were analyzed by independent *t*- tests.

Results

Sensitivity and specificity

Table 5.1 shows the mean sensitivity and specificity scores for both rheumatologists and radiologists at baseline before any training, after individual self-education, and after the workshop. In general, the radiologists showed somewhat higher sensitivity scores and lower specificity scores than the rheumatologists, but the differences were not statistically significant at any time. The scores for the radiologists did not significantly change after self education or the workshop, possibly partly reflecting the smaller numbers involved, especially at the last assessment. In contrast, the rheumatologists showed statistically significant decrease in sensitivity after individual self-education, but at the same time a statistically significant improvement in specificity. The decrease in sensitivity was reversed after the workshop, because at that time a statistically significant increase compared with the results after self education was observed. The differences between sensitivity three months after the workshop and sensitivity at baseline were not statistically significant.
Because the differences in sensitivity and specificity between rheumatologists and radiologists were not statistically significant at each of the three assessments, further analyses were performed with pooled data from both groups.

Table 5.1 Sensitivity and specificity of radiographs of the sacroiliac joints by rheumatologists and radiologists in comparison with a gold standard.

	Rheumatologists			Radiologists		
	Mean (SD)	Median	Range	Mean (SD)	Median	Range
Sensitivity at baseline	80 (18)	81	31-100	84 (22)	88	25-100
Sensitivity after self-education	74 (19)[a]	75	25-100	80 (19)	78	44-100
Sensitivity after workshop	79 (17)[b]	81	25-100	83 (17)	84	50-100
Specificity at baseline	75 (15)	75	38-100	71 (16)	71	46-100
Specificity after self-education	78 (14)[c]	79	29-100	70 (18)	73	38-96
Specificity after workshop	76 (13)	79	38-96	80 (16)	85	50-96

Numbers provided represent percentages (SD). [a] Statistically significant decrease compared with baseline scores (p=0.001); [b] Statistically significant increase compared with scores after self-education (p=0.002); [c] Statistically significant improvement compared with baseline scores (p=0.008) (all paired t-test). Differences between rheumatologists and radiologists were not statistically significant at any time point.

Table 5.2a presents the distribution in sensitivity and specificity of all participants at each assessment period. Clearly, a relatively large group of participants had difficulties in diagnosing sacroiliitis (for example, sensitivity of ≤50% at baseline for 13% of the participants). Participants with a low sensitivity or a low specificity score appeared to show high specificity or sensitivity scores respectively (Table 5.2b). Similarly, high sensitivity and high specificity scores were associated with low specificity and low sensitivity scores respectively.

Table 5.2a The distribution of sensitivity and specificity of both rheumatologists and radiologists at baseline, after self-education, and after the workshop are presented.

	Distribution						
	≤50%	50-60%	60-70%	70-80%	80-90%	90-100%	Total
Sensitivity							
At baseline	13	2	11	11	23	40	100
After self-education	17	8	16	13	20	26	100
After workshop	7	6	13	19	20	35	100
Specificity							
At baseline	8	12	16	25	21	17	100
After self-education	7	8	12	27	29	17	100
After workshop	6	4	13	36	26	16	100

The values are percentages of participants.

Change in sensitivity and specificity at individual level

Minor changes in sensitivity or specificity at the *group* level do not preclude larger changes for *individual* participants. Therefore, to find out what kind of changes in sensitivity and specificity occurred after self education or after the

workshop for *individual* participants, we recoded the change scores of each participant dichotomously by considering up to ±5% change in sensitivity or specificity as “no important change” and all other changes above or below this cut off point as a “relevant change”. Table 5.3 shows the results of the profiles of the participants three months after self education and three months after the workshop. Relevant increases in both sensitivity and specificity occurred only in a minority of the participants. Most of the participants showed an increase in either sensitivity or specificity without relevant change in specificity or sensitivity, respectively.

Table 5.2b The corresponding scores of either sensitivity or specificity is represented when respectively specificity or sensitivity is low or high.

	If sensitivity ≤50, than specificity:	If specificity ≤50, than sensitivity:	If sensitivity ≥90, than specificity:	If specificity ≥90, than sensitivity:
Sensitivity				
At baseline		99 (2)		59 (17)
After self-education		99 (2)		50 (13)
After workshop		99 (3)		61 (17)
Specificity				
At baseline	90 (6)		61 (11)	
After self-education	91 (5)		58 (13)	
After workshop	88 (7)		63 (10)	

Mean percentages (standard deviation).

Table 5.3 Profiles of both rheumatologists and radiologists of change after the interventions.

	After self-education compared with baseline scores (n=108)	After workshop compared with baseline scores (n=84)	After workshop compared with scores after self-education (n=84)
Sensitivity and specificity remained equal	13	16	13
Sensitivity and specificity improved	1	4	4
Sensitivity improved, specificity decreased *or* remained equal	21	27	39
Specificity improved, sensitivity decreased *or* remained equal	31	26	21
Both sensitivity and specificity decreased *or* either decreased and the other remained equal	34	27	23
Total	*100*	*100*	*100*

The values are percentages of participants. “Equal” was defined as a change of up to 5%, “improved” as an increase of more than +5%, and “decreased” as a change of more than -5%.

The participants with an improvement in sensitivity after self-education compared to baseline of >5%, showed a mean(SD) sensitivity at baseline of

63 (20) (n=23), and those with a change of >5% after the workshop a mean baseline sensitivity of 64(16) (n=23). Similarly, an improvement in specificity of >5% after self-education and after the workshop was associated with lower specificity scores at baseline of 60(11) (n=33) and 63(10) (n=22), respectively, than the mean specificity scores of the whole group as shown before (Table 5.1). Furthermore, the participants with a decrease of >5% in sensitivity after self education and the workshop compared with baseline, showed higher baseline scores in sensitivity (84(14) (n=57) and 86(15) (n=33), respectively), and those with a decrease of >5% in specificity showed higher baseline scores in specificity (77 (16) (n=28) and 84 (14) (n=23), respectively).

Intra-observer variation

Table 5.4 presents the intra-observer variation for both rheumatologists and radiologists for each of the 10 radiographs – together representing the spectrum of sacroiliitis – that were used for this assessment. No major differences in agreement between each of the repeated radiographs were found between rheumatologists and radiologists. The most extreme grades 0 and 4 showed the highest agreement. Sacroiliitis grades 1 and 2 showed substantial variation represented by lower agreement rates.

Table 5.4 Intra-observer agreement for both rheumatologists and radiologists for each of 10 duplicated radiographs of sacroiliac joints showing different degrees of sacroiliitis as defined by an expert panel. Dubious abnormalities show lower concordance.

	Rheumatologists			Radiologists		
Grading (sacroiliitis)	Baseline	After self-education	After workshop	Baseline	After self-education	After workshop
0 (no)	99	100	100	100	100	100
0 (no)	96	99	97	100	100	100
0 (no)	98	95	99	83	91	100
1 (dubious)	82	83	83	87	91	88
1 (dubious)	83	76	70	78	82	88
1 (dubious)	65	66	70	78	82	63
2 (yes)	70	76	76	83	52	75
2 (yes)	74	76	70	83	82	63
2 (yes)	83	80	87	100	86	88
4 (yes)	99	94	99	96	100	100

The values represent mean percentages of agreement.

Plain radiograph versus computed tomography

During the workshop, each participant had to read independently in random order 26 radiographs and 26 CTs of the same patients for the presence or

absence of sacroiliitis according to the New York criteria. In addition, for every judgment the participants also had to provide a rating on a 1-10 scale for their perceived self confidence for each of these diagnostic decisions. Table 5.5 presents the results of sensitivity and specificity, and the confidence ratings of the participants for both plain SI radiographs and CTs. The sensitivity score for CTs was significantly higher compared with plain radiographs, whereas no difference in specificity was found.

Table 5.5 Sensitivity, specificity, and mean confidence ratings of participants in defining presence or absence of sacroiliitis on radiographs versus computed tomography (CT) of the sacroiliac joints.

	Radiograph	CT
Sensitivity	72.0 (15.1)[a]	85.6 (14.1)[a]
Specificity	84.5 (13.8)	83.5 (17.0)
Mean rating, presence of sacroiliitis correctly diagnosed	7.7 (1.0)	7.8 (1.1)
Mean rating, presence of sacroiliitis incorrectly diagnosed	6.6 (1.2)	6.7 (1.3)
Mean rating, absence of sacroiliitis correctly diagnosed	8.3 (1.0)	8.0 (1.1)
Mean rating, absence of sacroiliitis incorrectly diagnosed	7.4 (1.4)	7.0 (1.6)

Means (standard deviation). Sensitivity and specificity range from 0 to 100, the range of the ratings is from 1 to 10. [a] $p<0.001$ with paired *t*-test. The differences in ratings between incorrectly and correctly diagnosed presence or absence of sacroiliitis were all statistically significantly ($p<0.001$). The differences in ratings for radiographs and CTs were not statistically significant.

Discussion

The presence of sacroiliitis is mandatory for the diagnosis of AS. The SI joints are unilaterally or bilaterally affected with mild to severe inflammation, which may eventually lead to partial or complete ankylosis.[7] The recognition of sacroiliitis is, however, often considered as difficult and requires experience. In this study the performance of rheumatologists and radiologists in detecting sacroiliitis has been evaluated. Three features of this nationwide study - in which more than 50% of all Dutch rheumatologists and a small sample (4.4% at baseline; 1.2% at completion) of radiologists participated - are striking.

Firstly, the diagnosis of radiographic sacroiliitis by radiologist and rheumatologists was comparable. Secondly, sensitivity and specificity scores were relatively moderate: 15-25% of the radiographs were incorrectly classified as if sacroiliitis was present (false positives), and 20-30% of the radiographs were incorrectly classified as if sacroiliitis was absent (false negatives) (Table 5.1). A high sensitivity score was associated with a low specificity score (Table 5.2b), and an increase in sensitivity was often accompanied with decreased specificity and *vice versa* (Table 5.3). Thirdly, improvement in both sensitivity and specificity that will persist for at least three months after a training session

appeared to be difficult to achieve. It is worrying that both sensitivity and specificity decreased in a large group of participants. Thus, the individual training sessions and the workshops as provided cannot be regarded as effective in promoting the performance of "blindly" diagnosing the presence or absence of radiographic sacroiliitis. Although individual compliance was not assessed, non-compliance overall cannot explain the apparent lack of effects of the workshop which was attended by the majority (75%) of the participants.

It is difficult to explain these observations. It seems that an improvement or a decrease in sensitivity (or specificity) after a training session was associated with correspondingly lower or higher sensitivity (or specificity) scores at baseline as compared with the mean score of the group at baseline (Table 5.2b). This effect may be attributed to regression to the mean or a floor-ceiling effect. Possibly, after training sessions, the attitude towards interpreting radiographs might have changed. Participants with initially low sensitivity scores might now have considered every spot or blurring at SI joints as aberrant, thereby improving the sensitivity score, but at the cost of specificity. Conversely, participants with initially low specificity scores might now have considered every spot or blurring at the SI joint more cautiously, at the cost of sensitivity. However, the participants were not informed about their sensitivity and specificity scores during the study period. Clearly, even after training sessions it remains difficult to distinguish between the normal and abnormal. Possibly, the same intervention should not have been offered to every participant. It might have been better to have assessed sensitivity and specificity first and then provide different targeted intervention to those participants with low sensitivity (and high specificity) and those with high sensitivity (and low specificity). This might be an area for future research.

The relative roles of plain radiographs, CT, and MRI in the radiographic diagnosis of sacroiliitis remain a matter of debate. The high sensitivity of CT and MRI is well known. Several studies have reported that CT and MRI are better than to plain radiographs in detecting early sacroiliitis.[8-13] However, because of to cost and other limitations to resources it is not always possible to use these techniques in the routine diagnosis of sacroiliitis.[14,15] Therefore, plain SI radiographs remain mostly the initial diagnostic tool. CT or MRI may be particularly helpful as an additional diagnostic aid in the early stages of sacroiliitis (when plain radiographs may be negative) if there is a high probability of sacroiliitis, or conventional radiographs are inconclusive.

Owing to the difficulties in interpreting plain radiographs of SI joints, large inter- and intraobserver variations have been reported.[1-3] In our study, intra-observer variability was expressed as the percentage of agreement for each of 10 radiographs that appeared twice in the *scoring* set. Clearly, concordance is highest if SI joints are definitely normal (grade 0) or definitely abnormal (grade 4). The use of kappa statistics would not have been useful in this situation

because of the high levels of expected agreement. The amount of intraobserver variation was comparable for both rheumatologists and radiologists. Most variation was seen in grades 1 and 2. However, the diagnostic -and possibly therapeutic- consequences of such seemingly small differences in grading of SI joints are most important. Patients with grade 2 sacroiliitis bilaterally will usually be diagnosed and treated as having AS, whereas people with grade 1 will normally not be considered as having an inflammatory rheumatic disease. Especially in these cases, CT or MRI may be helpful.[16] It should be noted, however, that there are clear differences in properties – and therefore also in appropriateness of their application - among plain SI films, CT and MRI. Plain radiographs provide an image where all sections are added to each other, while CT and MRI give information in slices. Further, plain films and CT can assess mainly bone and bone destruction, whereas MRI can assess cartilage and inflammation in the acute stage. It should also be realized that AS might sometimes occur in the absence of radiographic sacroiliitis.[17]

Another aim of our study was to assess the degree of confidence of rheumatologists and radiologists in determining the presence or absence of sacroiliitis on plain radiographs and CTs of the same patients. The ratings for the *correctly* diagnosed presence or absence of sacroiliitis were on average higher than the *incorrectly* diagnosed presence or absence of sacroiliitis ($p<0.001$). Although the participants felt less certain about their judgments of those radiographs and CTs which they misdiagnosed than those which they correctly diagnosed, the ratings for the incorrectly diagnosed radiographs remained, somewhat surprisingly, relatively high (mean 6.6 versus 7.4 on a 0-10 scale) (Table 5.5.). Clearly, the use of CTs compared with the use of radiographs did not increase the self confidence of the participants. However, many rheumatologists felt they did not have sufficient experience in reading CTs of SI joints and, therefore, these results might improve after training. On the other hand, the number of radiologists who participated in this part of the study is too small to generalize the findings.

The prevalence of definite sacroiliitis in the *scoring* set was 40%. This high à priori likelihood was unknown to the participants. It is unrealistically high in daily practice of radiologists, but on the other hand, diagnostic gain is at its highest level if the pretest probability is about 50%. Therefore, for rheumatologists this prevalence would indicate proficiency in making use of diagnostic tools. If a large number of normal (grade 0) SI films had been included in the *scoring* set this would have inflated the specificity artificially without clearly predictable effects on the sensitivity of diagnosing radiographic sacroiliitis.

Finally, except for data on age and sex of the patient in the *scoring* set, no clinical findings from the patient's history or physical examination were provided. Therefore, only radiographs were presented to radiologists and

rheumatologists in order to evaluate their performance in detecting sacroiliitis. This radiological diagnosis is an indispensable condition for the diagnosis of AS. In daily practice, however, rheumatologists mostly take into consideration the clinical information of the patient before they come to a final judgment. Rheumatologists may decide to re-evaluate the patient at a later time, or to refer the patient for additional CT or MRI. Recently a study has assessed the real performances of (Dutch) rheumatologists in daily practice, visited by patients incognito.[18] In particular, a female patient mimicking symptoms suggestive of AS and referred by her general practitioner visited a total of 25 rheumatologists. She brought with her a radiograph from another hospital clearly showing bilateral sacroiliitis. After history taking and physical examination, in which nearly all rheumatologists performed spinal mobility tests, more than 50% of the rheumatologists proposed additional radiographic imaging.[18] Evidently, a large group of rheumatologists felt uncertain about interpreting radiographs. Unfortunately, this study does not seem to contribute towards increasing their performance. It should again be emphasised that our study assessed sensitivity, specificity, and observer variation in reading films of SI joints, but did not take into consideration the effects on these characteristics of any clinical information. Such clinical data might already be known before reading the films or may be provided to the observer afterwards. The final effect of such additional information on the precision of establishing sacroiliitis as an indispensable condition for the diagnosis of AS is not yet known.

In conclusion, longlasting improvements in the performance of diagnosing sacroiliitis seems difficult to achieve, at least through self education using a training set of SI films or through uniform workshops. No statistically significant differences in sensitivity, specificity, and intraobserver variation of reading radiographs of SI joints were found between the radiologists and rheumathologists. Currently, CT of SI joints as compared with plain SI radiographs does not improve self confidence in diagnosing sacroiliitis.

References

1. Hollingsworth PN, Cheah PS, Dawkins RL, et al. Observer variation in grading sacroiliac radiographs in HLA-B27 positive individuals. J Rheumatol 1983;10:247-54.
2. Yazici H, Turunc M, Ozdogan H, et al. Observer variation in grading sacroiliac radiographs might be a cause of 'sacroiliitis' reported in certain disease states. Ann Rheum Dis 1987;46:139-45.
3. Taylor HG, Wardle T, Beswick EJ, Dawes PT. The relationship of clinical and laboratory measurements to radiological change in ankylosing spondylitis. Br J Rheumatol 1991;30: 330-5.
4. van der Linden S, Khan MA, Rentsch HU, et al. Chest pain without radiographic sacroiliitis in relatives of patients with ankylosing spondylitis. Journal of Rheumatology 1988;15:836-839.
5. Dale K. Radiographic gradings of sacroiliitis in Bechterew's syndrome and allied disorders. Scand J Rheumatol 1979;32 (Suppl):92-7.
6. 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;27: 361-8.
7. Braun J, Bollow M, Sieper J. Radiologic diagnosis and pathology of the spondylo-arthropathies. Rheum Dis Clin North Am 1998;24:697-735.
8. Ryan LM, Carrera GF, Lightfoot RW, Hoffman RG, Kozin F. The radiographic diagnosis of sacroiliitis. A comparison of different views with computed tomograms of the sacroiliac joint. Arthritis Rheum 1983;26:760-3.
9. Docherty P, Mitchell MJ, MacMillan L, et al. Magnetic resonance imaging in the detection of sacroiliitis. J Rheumatol 1992;19:393-401.
10. Battafarano DF, West SG, Rak KM, Fortenbery EJ, Chantelois AE. Comparison of bone scan, computed tomography, and magnetic resonance imaging in the diagnosis of active sacroiliitis. Semin Arthritis Rheum 1993;23:161-76.
11. Braun J, Bollow M, Eggens U, et al. Use of dynamic magnetic resonance imaging with fast imaging in the detection of early and advanced sacroiliitis in spondylarthropathy patients. Arthritis Rheum 1994;37:1039-45.
12. Blum U, Buitrago Tellez C, Mundinger A, et al. Magnetic resonance imaging (MRI) for detection of active sacroiliitis--a prospective study comparing conventional radiography, scintigraphy, and contrast enhanced MRI. J Rheumatol 1996;23:2107-15.
13. Yu W, Feng F, Dion E, et al. Comparison of radiography, computed tomography and magnetic resonance imaging in the detection of sacroiliitis accompanying ankylosing spondylitis. Skeletal Radiol 1998;27:311-20.
14. Murphey MD, Wetzel LH, Bramble JM, et al. Sacroiliitis: MR imaging findings. Radiology 1991;180:239-44.
15. Fenton P. Magnetic resonance imaging of the sacroiliac joints: worth the cost? J Rheumatol 1996;23:2020-1.
16. Braun J, Sieper J, Bollow M. Imaging of sacroiliitis. Clin Rheumatol 2000;19:51-7.
17. Khan MA, van der Linden SM, Kushner I, Valkenburg HA, Cats A. Spondylitic disease without radiologic evidence of sacroiliitis in relatives of HLA-B27 positive ankylosing spondylitis patients. Arthritis Rheum 1985;28:40-3.
18. Gorter S, van der Linden S, Brauer J, et al. Rheumatologists' performance in daily practice. Arthritis Rheum 2001;45:16-27.

Chapter 6

Performance of various criteria sets in patients with inflammatory back pain of short duration; the Maastricht early spondyloarthritis clinic

L Heuft-Dorenbosch, R Landewé, R Weijers, H Houben, Sj van der Linden, P Jacobs, D van der Heijde

Ann Rheum Dis doi:10.113/ard.2006.053918

Abstract

Objective
To describe how patients presenting with inflammatory back pain (IBP) of short duration can be classified by different sets of classification criteria for spondyloarthritis (SpA) and ankylosing spondylitis (AS), and which clinical and imaging features are of discernible importance.

Methods
68 patients with IBP of two-year duration at most were included in the early spondyloarthritis cohort (ESpAC). Detailed history, clinical examination and imaging of sacroiliac joints by plain radiography and magnetic resonance imaging (MRI) were obtained. The Berlin criteria set for SpA that has a prominent place for MRI and HLA-B27 was used to quantify the relative contribution of MRI in classifying SpA.

Results
Fourteen of the 68 patients had AS according to the modified New York criteria, 57 patients fulfilled the European spondylarthropathy study group (ESSG) criteria for SpA, 48 patients fulfilled the Amor criteria for SpA (43 patients fulfilled both criteria sets) and 44 patients fulfilled the Berlin criteria for SpA. Only 4 patients did not fulfill any criteria set; 36 patients fulfilled ESSG, Amor and Berlin criteria. The 14 patients with AS fulfilled all three SpA criteria sets.

Conclusion
Among our selected cohort of patients with early IBP the prevalence of SpA according to three different criteria sets is high. The ESSG criteria were most sensitive, followed by the Amor criteria and the Berlin criteria. The modified New York criteria for AS appeared to be most specific. In this cohort, the contribution of MRI and HLA-B27 to purely clinical criteria in making a diagnosis of axial SpA was rather limited.

Introduction

Diseases belonging to the group of spondylarthropathy or spondyloarthritis (SpA) share clinical and genetic characteristics, which distinguish them from rheumatoid arthritis.[1] The most prominent clinical feature of SpA is chronic inflammatory back pain (IBP). Other common features are arthritis, enthesitis, uveitis, and inflammatory bowel disease. The common genetic feature is the human leukocyte antigen B27 (HLA-B27). Included in the spectrum of SpA are ankylosing spondylitis(AS), psoriatic arthritis (PsA), SpA related to inflammatory bowel disease, reactive arthritis (ReA) and undifferentiated spondylarthropathy (USpA). An early diagnosis of SpA may move forward the initiation of effective treatment, and thus diminish the burden of illness as well as to avoid costs.

Current criteria for classification of SpA are the European spondylarthropathy study group criteria (ESSG)[2] and the Amor criteria[3], and for the classification of AS the modified New York criteria.[4] Especially in the case of early AS, criteria sets fall short, because the classification of AS depends on the presence of radiological sacroiliitis, which frequently appears late in the course of the disease. Thus, a long delay may exist between start of symptoms and establishing a diagnosis.[5] In a recent paper Rudwaleit *et al.* underscored the need for new classification criteria for axial SpA.[6] Axial SpA is a new concept that recognizes the involvement of the spine. Rudwaleit *et al.* from Berlin has recently proposed a diagnostic algorithm for axial SpA, to be used in the individual patient. A somewhat other approach proposed by the Berlin group, that served as a template for their diagnostic algorithm, was based on the calculation of the likelihood ratio (LR) product of currently available diagnostic tests for SpA.[6] Magnetic resonance imaging (MRI) and HLA-B27 testing have a prominent place in both the diagnostic algorithm and the LR product method.

In the present paper we describe how a group of patients presenting with IBP as a common and main symptom (the early spondyloarthritis clinic –ESpAC cohort) can be classified according to different sets of classification criteria for SpA and AS, and which features were of discernible importance. Subsequently, we challenged the diagnostic value of the Berlin algorithm by changing the contribution of MRI and HLA-B27.

Patients and Methods

Patients

Inclusion/exclusion criteria of the cohort

Patients with IBP present for two years at most were eligible. IBP was defined according to Calin.[7] IBP is considered present if four of the five following characteristics are present: age at onset of back pain <40 years, insidious onset, duration of back pain >three months, association with morning stiffness, improvement of back pain with exercise. Chronic back pain that awakens a patient at night is also suggestive of IBP, and is included in the Amor criteria set. Therefore we included also patients fulfilling only three of the Calin criteria plus night pain.

All practicing rheumatologists from the region of Limburg in both the Netherlands and Belgium and orthopedic surgeons of the university hospital Maastricht were asked to refer patients whom they considered eligible. All patients with a known history of psoriasis, who had visited the dermatology outpatient clinic between 2000 and 2002 received a questionnaire inquiring about the presence of IBP. Patients with a known diagnosis of IBD received a questionnaire when visiting the outpatient clinic, and the ophthalmologist handed out this questionnaire to patients with a diagnosis of acute anterior uveitis. The members of the regional branch of the AS society were informed about our investigation by means of a leaflet sent to them with their AS journal, and were encouraged to refer family members with IBP of short duration.

Assessments

Clinical evaluation

All patients completed the Bath ankylosing spondylitis activity index (BASDAI)[8] and the ankylosing spondylitis quality of life questionnaire (ASQol)[9] the Bath ankylosing spondylitis functional index (BASFI),[10] a numeral rating scale for night pain (range 0-3) and visual analogue scales (VAS) for spinal pain last week, global assessment for well being last week, and fatigue last week (range from 0 to 10). Duration of morning stiffness last week (minutes) was recorded, and a 44 joint swollen joint count was performed. Chest expansion (cm), modified Schober (cm), occiput-to-wall distance (cm) and lateral spinal flexion (cm) were assessed.

Erythrocyte sedimentation rate (ESR) (Westergren's method)(normal range: $\leq$7 for males; $\leq$12 for females) and C-reactive protein (turbidimetric method) (normal range: 2-9 mg/l) were measured. Further, HLA-B27 typing was performed.

Radiological evaluation

Conventional pelvic radiographs in the anterior-posterior view were performed, and SI-joints were scored by two independent observers without knowledge of clinical characteristics according to the modified New York criteria (from zero: normal to 4: complete fusion). SI joints with a score of 0 or 1 were considered normal; SI joints with a score of 2 or higher were considered abnormal (radiographic sacroiliitis (SI-itis)). In case of disagreement between readers, the independent judgment of a third reader was conclusive. Patients with bilateral grade 2 or more, or unilateral grade 3 or 4 were classified as fulfilling the radiological criterion of the modified New York criteria.
A MRI of the sacroiliac joints was performed using a 1,5 Tesla Philips Gyro scan ACS-NT. The following sequences were used: T1-weighted Spin Echo (SE), Short Tau Inversion Recovery (STIR), T2-weighted fast SE, dynamic T1-gradient sequence, and T1-weighted SE with fat suppression after the administration of contrast medium (Gadolinium diethylenetriaminepentate (Gd), 0.1 mmol/kg body weight). MRIs were independently scored by two readers (LH and RW) who were blind for the patient's identity, and for clinical, laboratory and radiological data. Both inflammation and structural changes (erosions, sclerosis and ankylosis) were scored. The MRI was considered normal or abnormal (for inflammation or structural changes) if both observers agreed. In case of disagreement between both readers, the independent judgment of a third reader was decisive.[11]

Classification of patients

Patients were classified according to the modified New York criteria for AS, according to the ESSG and Amor classification criteria for SpA, and according to the Berlin criteria for axial SpA. A Berlin classification was made both by the LR product method and by following the Berlin diagnostic algorithm (two different methods). Table 6.1 shows the different criteria sets.

Table 6.1 Sets of classification criteria for spondyloarthritis (SpA) and ankylosing spondylitis (AS).

Amor classification criteria for SpA (score)	ESSG criteria for SpA	Berlin criteria for SpA (likelihood ratio)	New York criteria for AS
Clinical findings: Lumbar or dorsal pain at night or morning stiffness of lumbar or dorsal spine: 1 Asymmetrical oligoarthritis: 2 Buttock pain: 1 If alternating buttock pain: 2 Sausage like toe or digit: 2 Heel pain or other well defined enthesopathy: 2 Iritis: 1 Non-gonococcal urethritis or cervicitis ≤1 month before onset of symptoms: 1 Acute diarrhoea ≤1 month before symptoms: 1 Psoriasis, balanitis or inflammatory bowel disease (ulcerative colitis or Crohn's disease): 2 Radiological finding: Sacroiliitis, bilateral grade 2 or unilateral grade 3 or 4: 3 Genetic background: Presence of HLA-B27 and/or family history of AS, reactive arthritis, acute anterior uveitis, psoriasis or IBD: 2 Response to treatment: Clear-cut improvement within 48 hours of non-steroidal anti-inflammatory drug: 2 A sum score ≥6 is required for a positive classification	Inflammatory spinal pain ór synovitis (asymmetric, predominantly of the lower extremities) and at least one of the following: Family history: first or second degree relative with AS, spondylitis, psoriasis, acute iritis, reactive arthritis or IBD Past or present psoriasis diagnosed by a physician Past or present ulcerative colitis or Crohn's disease) diagnosed by a physician and confirmed by radiography or endoscopy Past or present pain alternating between the two buttocks Past or present spontaneous pain or tenderness on examination of the site of insertion of the Achilles tendon or plantar fascia (enthesis) Episode of diarrhoea occurring ≤1 month before onset of symptoms Non-gonococcal urethritis or cervicitis ≤1 month before onset of symptoms Bilateral grade 2-4 sacroiliitis or unilateral grade 3 or 4 sacroiliitis (where grade 0 is normal, grade 1 possible, 2 minimal, 3 moderate	Inflammatory spinal pain (3.1) Alternating buttock pain (4.0) Heel pain (enthesiopathy) (3.4) Peripheral arthritis (4.0) Dactylitis (4.5) Anterior uveitis (7.3) Psoriasis (2.5) Inflammatory bowel disease (4.0) Positive family history of AS, reactive arthritis, IBD, psoriasis, or acute anterior uveitis (6.4) Response to NSAIDs (5.1) Raised acute phase reactants (2.5) HLA-B27 (9.0) MRI (9.0) SpA is considered present if the product of likelihood ratios ≥200	1. Low back pain for at least 3 months duration that is improved by exercise and not relieved by rest 2. Limited lumbar spinal motion in saggital and frontal (anterior and posterior planes) 3. Chest expansion decreased relative to normal values for sex and age Bilateral sacroiliitis grade 2-4 or unilateral sacroiliitis grade 3 or 4

AS=ankylosing spondylitis; HLA-B27=human leucocyte antigen B27; IBD=inflammatory bowel disease; NSAID=non-steroidal anti-inflammatory drug.

Alternatively patients could be classified as axial SpA according the Berlin diagnostic algorithm if they had IBP, AND: a) if ≥3 features of SpA are present, patients are classified as axial SpA; b) if ≤2 features of SpA are present, HLA-B27 typing is performed. Patients with one or two features of SpA are classified as axial SpA if HLA-B27 is present. Patients with zero SpA features in whom HLA-B27 is present should undergo MRI of the SI-joints, and are classified as axial SpA only if MRI shows inflammation of the SI-joints. Patients with zero, one or two features of SpA in whom HLA-B27 is absent are not classified as axial SpA. The clinical SpA features enthesitis, positive family history, uveitis, alternating buttock pain, peripheral arthritis, dactylitis, psoriasis, Crohn's disease and positive response to NSAIDs.
A classification according to the Berlin LR product method requires the multiplication of the LRs of those SpA features that are present. In case of a LR product ≥200 a classification of axial SpA is made, since a LR product ≥200 results in a positive predictive value of approximately 90% in a population of patients with chronic low back pain and an assumed disease prevalence of axial SpA of 5% (pre-test probability). For example, a patient with chronic IBP (LR+: 3.1), acute anterior uveitis (LR+: 7.3) and HLA-B27 (LR+: 9.0) has a LR product of 3.1 × 7.3 × 9.0 = 203.7 and can therefore be classified as axial SpA.

Results

Most patients were referred by rheumatologists. Other patients were referred by dermatologists, gastroenterologists, ophthalmologists and orthopedic surgeons, or via the local branch of the AS society (detailed information in Figure 6.1). The most important reason for not including referred patients was that they either denied back pain or reported back pain with a non-inflammatory character. Patients from the rheumatology, orthopedic- and ophthalmology outpatient clinics did not receive a questionnaire but were referred immediately in case of IBP. Of the 94 patients that were considered eligible 68(72%) consented in participation. Of these 68 patients only one patient did not return questionnaires but participated in the clinimetric-and radiological part of the study.

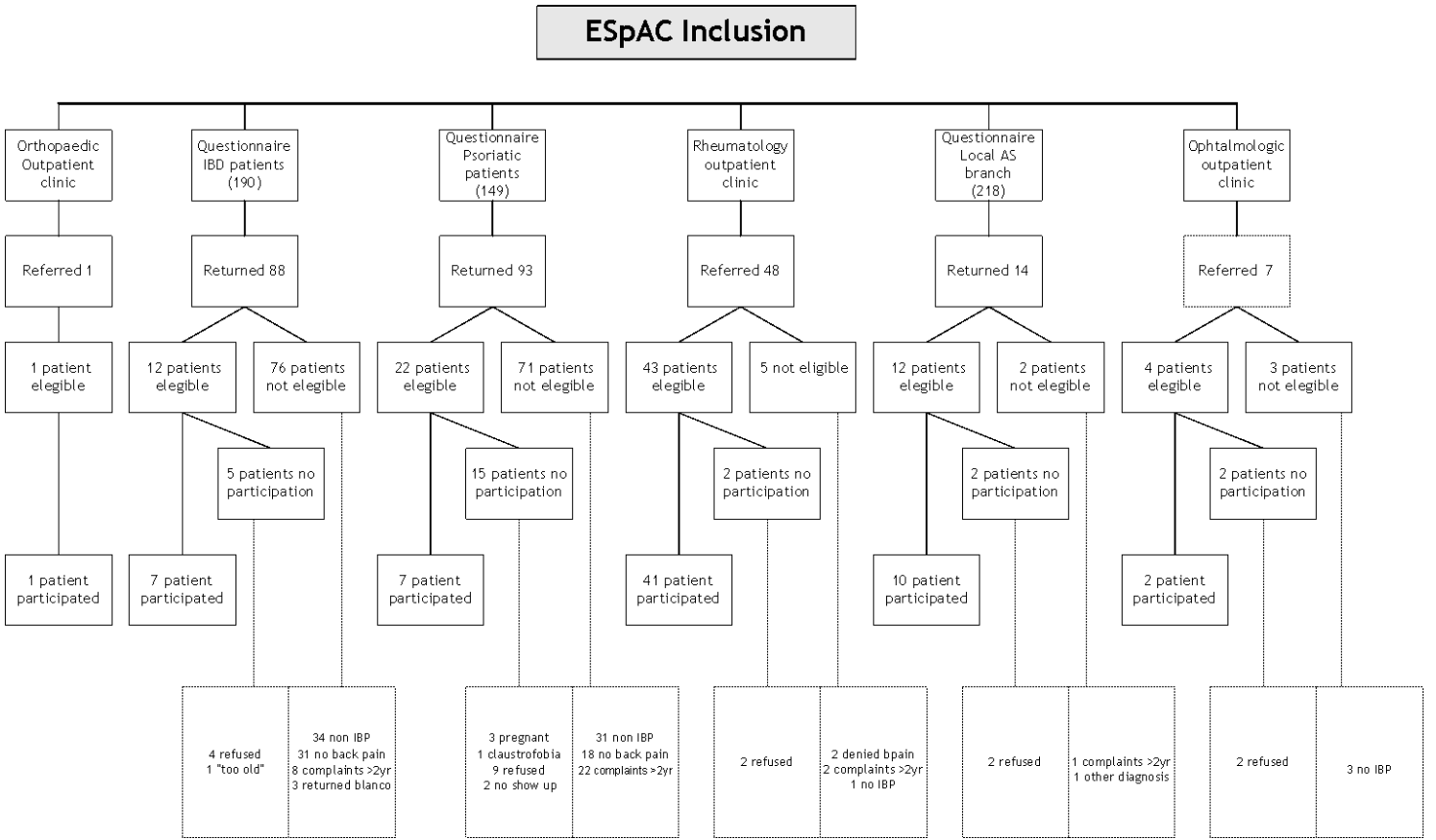

Figure 6.1 Route of inclusion of the 68 patients in the ESpAC cohort. In dotted boxes reason of non participation.

Table 6.2a shows main demographic data and data about disease history. The median duration of complaints was 18 months (interquartile range:12-24 months). Of all patients, 97% responded confirmatory to at least four of the five questions about IBP (Calin-criteria). The only two patients mentioning only three of the five questions, reported night-pain disturbing their sleep. Of all patients 46% were HLA-B27 positive, 15% had IBD confirmed by a gastro-enterologist, 15% reported uveitis, 24% psoriasis, confirmed by a dermatologist or present at physical examination, and 54% had a positive family history for AS (first- or second-degree relative). Non-steroidal anti-inflammatory drugs (NSAIDs) were used by 29 patients, and disease modifying drugs by ten patients(five patients IBD-related and five other patients for arthritis). Of these latter, three patients used salazopyrine (one of them in combination with methotrexate), one patient used glucocorticoids, and one patient used methotrexate in combination with IBD-related drugs.

Table 6.2a Baseline characteristics of 68 patients included in the early spondyloarthropathy clinic (ESpAC) cohort. Demographic data and history.

		ESpAC study population (n=68)
Male		38
HLA-B27+		46
IBD[a]		15
Uveitis[a]		15
Psoriasis[a]		24
Family history of AS		37
Dactylitis[a]		10
Heel pain[a]		6
Peripheral arthritis[a]		28
IBP improving with NSAIDs[a]		59
Buttock pain[a]		71
Current NSAID-use		43
Current DMARD-use		9
Current glucocorticoid use		1.5
Current use of IBD-related drugs		6
Criteria for inflammatory low back pain:	3 criteria present	3
	4 criteria present	41
	5 criteria present	56

Figures in the table are %, [a] history- or presence of this symptom. HLA-B27=human leukocyte antigen B27; IBD=inflammatory bowel disease; ESR=erythrocyte sedimentation rate; CRP=C-reactive protein; NSAID=non steroidal anti-inflammatory drug; DMARD=disease modifying anti-rheumatic drug

Table 6.2b shows the clinimetric and laboratory data. A high proportion of patients had a reduced Schober test, chest expansion or cervical rotation. The mean BASDAI was moderately high. Forty percent of patients had raised acute phase reactants (ESR and/or CRP). The percentage of IBD patients in the

group of patients with normal acute phase reactants (18%) was comparable to the percentage of IBD patients in the group with raised acute phase reactants (11%).

Table 6.2b Baseline characteristics of 68 patients included in the early spondyloarthropathy clinic (ESpAC) cohort. Clinimetry and acute phase reactants.

	ESpAC study population (n=68)
VAS pain of the spine last week [cm] (0-10)	4.7 (2.6)
Night pain [4-point Likert-scale] (median (IQR))	1 (1; 2)
Global well being last week (on a 10cm VAS)	3.8 (3.8)
Duration of morning stiffness last week (in minutes) (median (IQR))	23 (10; 60)
Severity of fatigue last week (on a 10cm VAS)	4.5 (2.6)
44-swollen joint count	0.19 (0.6)
Finger to Floor distance (in cm) (median (IQR))	0.0 (0-12.5)
Chest expansion (in cm) [% of patients <4cm]	4.9 (2.0) [43]
Lateral spinal flexion (in cm) [% of patients <20cm]	15.2 (4.8) [90]
Modified Schober (in cm) [% of patients <5cm]	15.0 (1.1) [50]
Occiput-to-wall distance (in cm) [% of patients >0cm]	0.4 (1.7) [7]
Cervical rotation (in º) [% of patients <70º]	73 (14) [32]
Bath ankylosing spondylitis disease activity index (BASDAI)	3.6 (2.1)
Bath ankylosing spondylitis functional index (BASFI)	2.6 (2.1)
Bath ankylosing spondylitis metrology index (BASMI)	1.1 (1.2)
Maastricht ankylosing spondylitis enthesopathy index (MASES)	2.8(2.4)
Ankylosing spondylitis quality of life (ASQoL)	6.4 (4.8)
Erythrocyte sedimentation rate (ESR) (in mm) (median (IQR))	6 (4; 16)
C-reactive protein (CRP) (in mg/l) (median (IQR))	7 (3; 9)
Patients with elevated acute phase reactants (in %)	41

Figures are means (standard deviations) unless otherwise indicated. VAS=visual analogue scale; IQR=interquartile range.

Figure 6.2 shows the number of patients fulfilling three different criteria sets for SpA. A total of 36 patients (53%) fulfilled all three criteria sets, 13 patients (19%) fulfilled two criteria sets and 15 patients (22%) fulfilled only one criteria set. Only four of the 68 patients (6%) did not fulfill any criteria set for AS or SpA. The highest classification rate was found with the ESSG criteria (84%), followed by the Amor criteria (71%) and by the Berlin criteria (65%). The Amor criteria and the Berlin criteria have a high level of conceptual similarity, but differ with respect to the contribution of MRI. We therefore investigated the contribution of MRI in explaining mismatches between Amor and berlin classification. Of the five patients that were negative for Amor and positive for Berlin, only one patient had a positive MRI. This patient would have lost the Berlin classification if this MRI would have been negative. All patients that were positive for Amor and negative for Berlin had a negative MRI. A positive MRI would have changed the classification for seven of the nine patients. In

summary, MRI was of distinctive contribution in eight of the 14 patients with a mismatch and irrelevant or redundant in six of them.

Of the 36 patients fulfilling all three criteria sets for SpA 21 (58%) showed inflammation on MRI and 23 (64%) were HLA-B27 positive. Of the 13 patients fulfilling two criteria-sets for SpA no patient showed inflammation on MRI and five patients (39%) were HLA-B27 positive. Of the 15 patients fulfilling one criteria-set for SpA, one (7%) showed inflammation on MRI and three (20%) were HLA-B27 positive. A detailed description of all MRI findings in this cohort has been reported elsewhere.[12] None of the four patients who did not fulfil any criteria set showed inflammation on MRI, and none of these patients were HLA-B27 positive. But two of these patients had grade 2 radiographic sacroiliitis in one SI-joint only. Of note, one additional patient would fulfil the Amor classification criteria if the radiographic sacroiliitis criterion could also be fulfilled by a positive MRI scan (chronic or inflammatory lesions were considered positive)

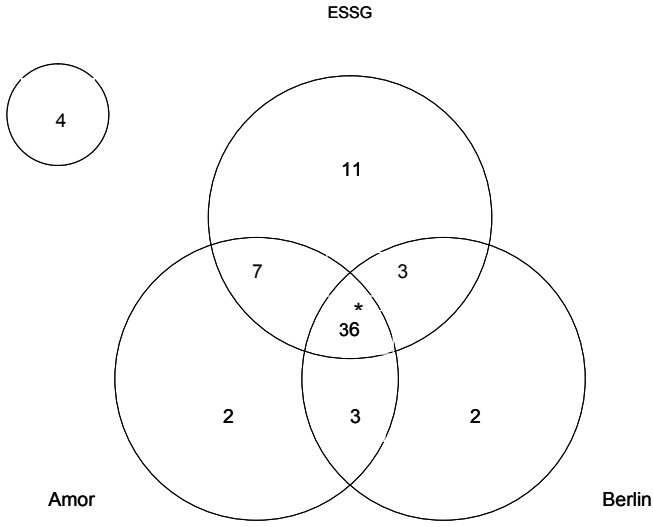

Figure 6.2 The number of patients fulfilling three different criteria sets for SpA; ESSG, Amor and Berlin criteria set. * 14 of these patients fulfill also the modified New York criteria.

In summary, 22 patients (32%) showed inflammation on MRI, and 31 patients (46%) were HLA-B27 positive. Of the 14 patients meeting the modified New York criteria for AS 12 patients (86%) were HLA-B27 positive, and 12 (86%) showed inflammation on MRI including both HLA-B27 negative patients who

met the criteria for AS. One of these HLA-B27 negative patients had a history of psoriasis. Both AS patients without inflammation on MRI were HLA-B27 positive.

We have analyzed whether HLA-B27 positive patients differed from HLA-B27 negative patients, and whether patients with inflammation on MRI differed from patients without inflammation on MRI, with regard to demographic and clinical characteristics (Table 6.3). This comparison revealed a statistically significant male predominance in patients that are HLA-B27 positive (P=0.01), and/or have MRI inflammation present (P=0.004). HLA-B27 positive patients (P=0.002) and/or patients with MRI-inflammation present (P<0.001) reported significantly more often improvement on NSAIDs as compared to patients without these characteristics. In addition, HLA-B27 was significantly over-represented in patients with MRI inflammation present (P<0.001). All other demographic and clinical characteristics were equally distributed among HLA-B27 positive- and HLA-B27 negative patients, and among those with- and without MRI-inflammation.

Table 6.3 A comparison between HLA-B27 positive and negative patients and between patients with- and without inflammation on magnetic resonance imaging of the sacroiliac joints in the early spondyloarthritis clinic (ESpAC) cohort.

	HLA-B27		Inflammation on MRI	
	present (n=31)	absent (n=37)	present (n=22)	absent (n=46)
Male	52[a]	24	64	24
Inflammation on MRI	52	16		
HLA-B27 positive			73	33
History or presence of IBD	7	23	5	20
History or presence of acute anterior uveitis	19	11	14	15
History or presence of psoriasis	13	32	18	26
Family history of ankylosing spondylitis	48	27	32	39
Raised acute phase reactants	45	38	46	44
Improvement on NSAIDs (%)	81	41	91	44
Mean BASDAI (SD)	3.6 (1.8)	3.7 (4.4)	3.8 (2.1)	3.6 (2.2)
Mean BASFI (SD)	2.5 (1.8)	2.7 (2.3)	2.7 (2.0)	2.6 (2.2)
Checklist IBP:				
<5 criteria present	30	57	36	48
5 criteria present	70	43	64	52
Presence of extra-spinal SpA features:				
0 SpA features	39	24	36	28
1 or 2 SpA features	55	67	55	65
≥3 SpA features	6	8	9	7

[a] data are percentages unless otherwise indicated. BASDAI=Bath ankylosing spondylitis disease activity index; BASFI=Bath ankylosing spondylitis functional index. Extra-spinal SpA features: asymmetric arthritis, Achilles tendinitis, dactylitis, inflammatory bowel disease, acute anterior uveitis, psoriasis.

These differentiating results were somewhat more obvious if the analysis was applied in patients that were HLA-B27 positive Or had MRI inflammation present.

The performance of the Berlin algorithm is shown in Figure 6.3; Fourteen patients directly classified as AS, an additional 24 patients fulfilled the Berlin criteria for axial SpA on clinical grounds (presence of ≥3 SpA features), and an other ten patients with one to two SpA features were HLA-B27 positive and thus fulfilled the Berlin classification criteria; so a total of 48 patients were classified as axial SpA.

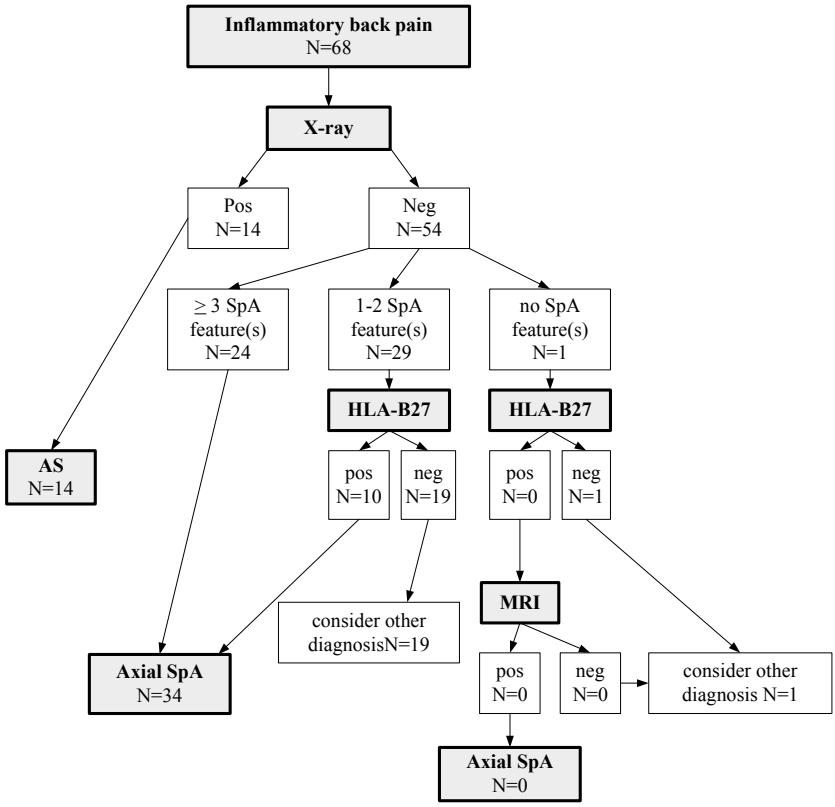

Figure 6.3 Results of application of the Berlin algorithm.

In our cohort only one patient had not one single SpA feature. This patient was HLA-B27 negative, and thus should - according to the Berlin algorithm - not further be investigated by MRI. If we modify the algorithm to apply MRI first

instead of HLA-B27 (Figure 6.4), only five patients were selected on the basis of one or two clinical SpA criteria in combination with a positive MRI. Table 6.4 shows the diagnostic performance of the Berlin and the modified Berlin algorithm, both with- and without the patients classifying as AS at inclusion. The diagnostic performance of the modification of the Berlin algorithm was better as compared to the original algorithm, when taking fulfillment of both the ESSG criteria and the Amor criteria as an arbitrary gold standard. According to this gold standard, 11 patients were classified as false positive for axial SpA in the original algorithm (six with ≥ three, and five with one or two SpA features who were HLA-B27 negative) and six patients were classified as false negative (all with one to two SpA features who had inflammation on MRI). If MRI was applied first (modified algorithm), seven patients were classified as false positive(six with ≥ three SpA features and one with one to two SpA feature who had inflammation on MRI) and seven patients as false negative(all had 1-2 SpA features and had no inflammation); The modified algorithm outperforms the original algorithm, but still resulted in misclassifications.

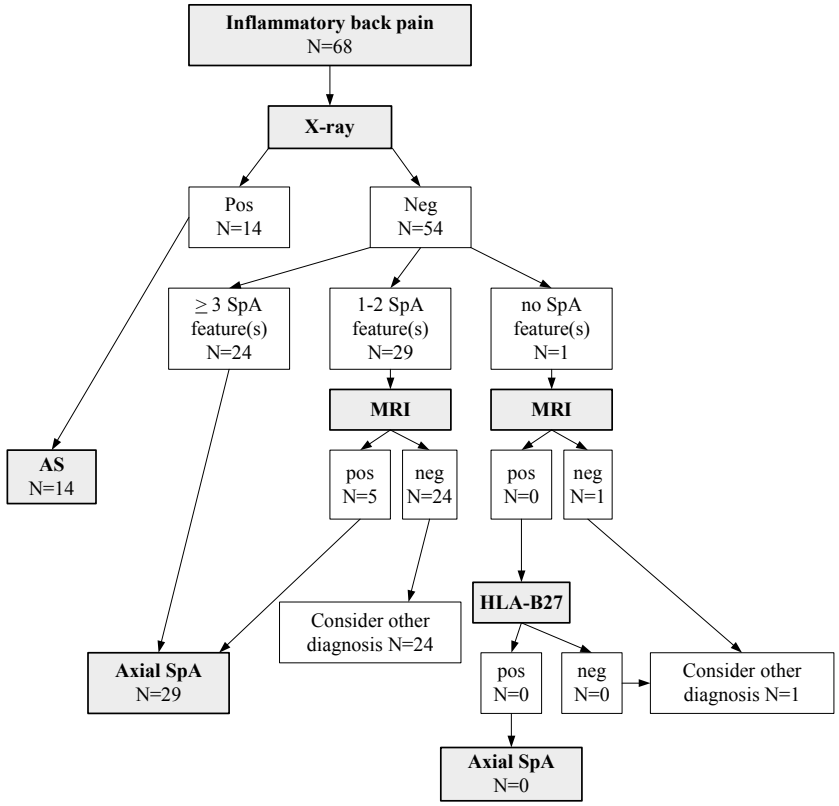

Figure 6.4 Results of application of the modified Berlin algorithm (MRI first).

Table 6.4 Test performance of modifications of the Berlin algorithm in the early spondyloarthritis clinic (ESpAC) cohort with "fulfillment of Amor- and ESSG criteria for spondyloarthritis" as gold standard.

	Order of application:	
	HLA-B27 first	MRI first
All patients (n=68)		
Sensitivity	77%	84%
Specificity	56%	72%
NPV	74%	72%
PPV	69%	84%
LR positive test	1.8	3.0
LR negative test	0.41	0.22
Only patients NOT fulfilling the modified New York criteria for AS (n=54)		
Sensitivity	68%	76%
Specificity	70%	72%
NPV	56%	72%
PPV	79%	76%
LR positive test	2.3	2.7
LR negative test	0.46	0.33

HLA-B27=human leucocyte antigen B27; MRI=magnetic resonance imaging; AS=ankylosing spondylitis; NPV=negative predictive value; PPV=positive predictive value; LR=likelihood ratio; LR positive test=sensitivity / (1-specificity); LR negative test=(1-sensitivity) / specificity.

Discussion

In this cohort of patients with IBP of short duration as the main inclusion criterion, a remarkably high proportion classified for SpA according to three different criteria sets. Half of patients fulfilled all three criteria sets, and an additional fifth fulfilled at least two criteria sets. Explanations may be that the majority of patients was referred by rheumatologists and that we screened for IBP in patients with a SpA-related disease (psoriasis, uveitis, IBD) and in a population of relatives with AS or SpA. Undoubtedly, the prevalence of SpA in our cohort will not reflect the prevalence of SpA in individuals with IBP from the general population.

Already14 patients fulfilled the modified New York criteria for AS. Feltkeller *et al.* reported a long delay between start of symptoms and a diagnosis of AS.[5] Active case finding based on IBP in combination with SpA feature(s) may importantly reduce the time interval between the start of symptoms and a diagnosis of AS in at least a part of patients with AS.

The proportion of women in this SpA cohort seems high (62%) in light of the male predominance in established AS cohorts, but is in line with sex rates in other studies with patients with IBP of short duration.[13-15] One explanation may be that, unlike AS, SpA is disease that more often occurs in women than in men. Another explanation is that these women in fact do not have "true SpA" or

SpA-related disease, and that the specificity of SpA criteria is too low. The observation that this group of women is often HLA-B27 negative, without MRI inflammation present, adds to the validity of this latter explanation. As a consequence, we found a lower rate of HLA-B27 positive patients (46%), as compared to AS. Bollow *et al*.[14] found 69% of their patients with SpA positive for HLA-B27, Puhakka *et al.*[15] found a prevalence of 63%, and Hanly *et al*.[13] reported 50%. The data described above are consistent with the hypothesis that HLA-B27 positive male patients with axial SpA may evolve into AS. This hypothesis can be tested in the follow up of this cohort.

The ESpAC cohort is an appropriate cohort to investigate the diagnostic performance of the Berlin diagnostic algorithm for axial SpA. We had to define an external criterion, since a gold standard for a diagnosis of AS in an early, pre-radiographic phase is lacking. We assumed the fulfillment of both the ESSG- and the Amor-criteria as a sufficiently robust substitute for a gold standard. Interestingly, and despite an important overlap, there were five patients who fulfilled the Amor criteria that did not fulfill the ESSG criteria. The major discrepancy between both criteria sets includes the contribution of asymmetrical oligoarthritis to a classification according to Amor. Because our aim was to test the performance of the Berlin criteria for axial SpA (not peripheral SpA), we felt that we should exclude those patients that only fulfilled the Amor (but not the ESSG) classification criteria because of arthritis. Such an approach may be criticized. The distinction between axial and peripheral SpA is rather artificial, and there may be a important overlap. Both the ESSG and Amor variables for SpA were developed as classification criteria and we have used them in a diagnostic setting, which may lead to underperformance.[16] And testing one criteria set against a combination of other criteria sets with an overlap in items has a danger of circularity. With our artificial construct as a gold standard, the performance of the Berlin algorithm in our study was moderate. The yield improved after conversion of the order of HLA-B27 and MRI. The advantage of MRI, unlike HLA-B27 testing, is that MRI directly visualizes the inflammatory process. Limitations are that interpretation of MRI abnormalities is observer-dependent, that MRI does not yield a clear positive or negative result, and that it is more expensive.

A noteworthy finding in this study is that it clearly demonstrates the relation between HLA-B27 and SI-joint inflammation (Table 6.3). Of the patients with sacroiliac inflammation, 73% was positive for HLA-B27, compared to only 33% of the patients without inflammation. Conversely, in the HLA-B27 positive patients, half of patients showed sacroiliac inflammation on MRI, compared to one sixth of HLA-B27 negative patients. There was a male predominance in the group with inflammation, and more patients with SI inflammation on MRI responded to NSAIDs. The “inflammatory group” resembling a group of established AS patients, may ultimately develop definite AS. A limitation of our

study is that we did not analyze spinal inflammation, Recent data showed a high proportion of patients with inflammation of the spine in the thoracic segment.[17] We also did not test the usefulness of our recent observation that the assessment of structural changes on conventional radiography (CR) in combination with inflammation on MRI is most sensitive for detecting abnormalities in SI joints in patients with recent onset IBP.[11] It is our intention to prospectively investigate which patients in our cohort will develop AS over time and which Berlin algorithm is most useful to predict such a course.

References

1. Moll JM, Johnson G, Wright V. Psoriatic arthritis: a unique family. Rheumatol Rehabil. 1974;13(3):154-7.
2. 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;34:1218-27.
3. Amor B, Dougados M, Mijiyawa M. [Criteria of the classification of spondylarthropathies]. Rev Rhum Mal Osteoartic. 1990;57(2):85-9.
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;27(4):361-8.
5. Feldtkeller E, Bruckel J, Khan MA. Scientific contributions of ankylosing spondylitis patient advocacy groups. Curr Opin Rheumatol. 2000;12(4):239-47.
6. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new criteria? Arthritis Rheum. 2005;52(4):1000-8.
7. Calin A, Porta J, Fries JF, Schurman DJ. Clinical history as a screening test for ankylosing spondylitis. Jama. 1977;237(24):2613-4
8. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol. 1994;21:2286-91.
9. Doward LC, Spoorenberg A, Cook SA, et al. Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis. Ann Rheum Dis. 2003;62(1):20-6.
10. 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;21:2281-5.
11. Heuft-Dorenbosch L, Landewe R, Weijers R, et al. Combining information obtained from MRI and conventional radiographs in order to detect sacroiliitis in patients with recent-onset inflammatory back pain. Ann Rheum Dis. 2006;65(6):804-8.
12. Heuft-Dorenbosch L, Weijers R, Landewe R, van der Linden S, van der Heijde D. 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. 2005;8(1):R11.
13. Hanly JG, Mitchell MJ, Barnes DC, MacMillan L. Early recognition of sacroiliitis by magnetic resonance imaging and single photon emission computed tomography. J Rheumatol. 1994;21(11):2088-95.
14. Bollow M, Braun J, Hamm B, et al. Early sacroiliitis in patients with spondyloarthropathy: evaluation with dynamic gadolinium-enhanced MR imaging. Radiology. 1995;194(2):529-36.
15. Puhakka KB, Jurik AG, Schiottz-Christensen B, et al. Magnetic resonance imaging of sacroiliitis in early seronegative spondylarthropathy. Abnormalities correlated to clinical and laboratory findings. Rheumatology (Oxford). 2004;43(2):234-7.
16. Inman RD. Clinical stratification in the spondylarthropathies: E. pluribus unum or ex uno plures? Arthritis Rheum. 2001;45(6):475-7.
17. Baraliakos X, Hermann KG, Landewe R, et al. Assessment of acute spinal inflammation in patients with ankylosing spondylitis by magnetic resonance imaging: a comparison between contrast enhanced T1 and short tau inversion recovery (STIR) sequences. Ann Rheum Dis. 2005;64(8):1141-4.

Chapter 7

Magnetic resonance imaging changes of sacroiliac joints in patients with recent-onset inflammatory back pain: inter-reader reliability and prevalence of abnormalities

L Heuft-Dorenbosch, R Weijers, R Landewé, Sj van der Linden, D van der Heijde

Arthritis Res Ther 2006;8:R11.

Abstract

To study the inter-reader reliability of detecting abnormalities of sacroiliac (SI) joints in patients with recent onset inflammatory back pain (IPB) by magnetic resonance imaging (MRI), and to study the prevalence of inflammation and structural changes at various sites of the SI joints.

Sixty-eight patients with IBP (at least 4 of the 5 following criteria: symptom onset before age 40, insidious onset, morning stiffness, duration >three months, improvement with exercise- or three out of five of these plus night pain) were included (38% male; mean age 34.9 year (standard deviation 10.3); 46% HLA-B27 positive; mean symptom duration 18 months) with symptom duration <2 years. A MRI scan of the SI joints was made in the coronal plane with the following sequences: T1-weighted spin echo (SE), short-tau inversion recovery (STIR), T2-weighted-fast spin echo with fat saturation, and T1-spin echo with fat saturation after the administration of gadolinium (Gd). Both SI-joints were scored for inflammation (separately for subchondral bone and bone marrow, joint space, joint capsule, ligaments) as well as for structural changes (erosions, sclerosis and ankylosis), by two observers independently. Agreement between the two readers was analyzed by concordance and discordance rates and by kappa statistics.

Inflammation was present in 32 SI joints of 22 patients; most frequently located in bone marrow and/or subchondral bone (29 joints in 21 patients). Readers agreed on the presence of inflammation in 85% of the cases in the right SI joint and in 78% of the cases in the left SI joint. Structural changes on MRI were present in 11 patients. Ten of these 11 patients showed also signs of inflammation.

Agreement on the presence or absence of inflammation and structural changes of SI joints by MRI was acceptable, and was sufficiently high to be useful in ascertaining inflammatory and structural changes due to sacroiliitis. About one-third of patients with recent onset inflammatory back pain show inflammation, and about one-sixth show structural changes in at least one SI joint.

Introduction

Ankylosing spondylitis (AS) is a chronic rheumatic condition, characterized by inflammation of the axial skeleton, particularly the sacroiliac (SI) joints. Patients fulfill classification criteria for AS if characteristic radiological changes of the SI joint are present, together with defined clinical symptoms and findings.[1] AS belongs to the group of seronegative spondyloarthritides (SpA). The European spondylarthropathy study group has developed classification criteria for SpA.[2] Sacroiliitis is a characteristic feature of AS, and is frequently found in patients with SpA, although it is not obligatory. Patients with sacroiliitis experience chronic low back pain with an inflammatory pattern that often begins in young adulthood. Because chronic low back pain is common in the population, sacroiliitis is often not considered as a cause of back-pain. Besides, early sacroiliitis is often not visible on conventional radiographs, or is difficult to interpret, which may lead to a long delay in establishing a diagnosis. Frequently, a mean duration of more than eight years between the start of symptoms and the diagnosis of AS is reported.[3,4] Such a delay is increasingly unwarranted because of the availability of effective treatment. Magnetic resonance imaging (MRI) is an imaging modality that may shorten the delay between the start of symptoms and a classifying diagnosis of AS or SpA since MRI can detect inflammation early.[5-7] Algorithms for diagnostic purposes have recently been proposed in which MRI of the SI joints was attributed a prominent place.[8] In order to judge whether MRI is helpful in making an early diagnosis, the psychometric properties of assessing inflammation and structural changes by MRI should appropriately be tested in patients with very early disease, and not only in those with advanced AS. The aim of this study was to evaluate whether MRI could reliably assess inflammation and structural damage of SI joints in patients with short-term inflammatory back pain.

Patients and Methods

Patients

Patients with inflammatory low back pain present for two years at most, and without a confirmed rheumatologic diagnosis, were eligible for this study. Inflammatory back pain was defined according to the Calin criteria.[9] Inflammatory back pain by these criteria is defined if at least four of the five following characteristics are present: insidious onset; onset before the age of 40 years; persistence for at least three months; association with morning stiffness; and improvement with exercise. Patients also could be included if three out of five of these criteria were present plus night pain. Preferably, but

this is not obligatory, patients should have at least one feature of SpA according to the European spondylarthropathy study group criteria: presence of a family member with AS; presence or history of psoriasis, inflammatory bowel disease (IBD) or uveitis.
The study was approved by the institutional review board and all patients gave written informed consent.

MRI

A MRI examination of the SI joints was performed using a 1,5 Tesla Philips Gyro scan ACS-NT (Philips, Best, the Netherlands). Patients were scanned in a supine position using a Synergy-spine coil as the surface coil. We chose a coronal oblique scan plane parallel to the length of the sacrum and two slabs: one transversal slab was positioned cranially to the region of interest to diminish flow-artefacts, and one was positioned frontally through the bowel and anterior abdominal wall, to diminish motion artefacts of breathing and bowel movements. The following sequences were used: T1-weighted spin echo (SE), short tau inversion recovery (STIR), T2-weighted fast SE with fat saturation and T1-weighted SE with fat suppression after the intravenous administration of contrast medium (gadolinium diethylenetriaminepentate (Gd), 0.1 mmol/kg body weight).
Different relevant MRI findings with regard to sacroiliitis were identified from the literature; a differentiation was made between inflammatory changes and structural changes and different localization of these changes. Pathological changes of interest were defined as inflammation and structural changes including erosions, sclerosis and ankylosis. Regions of interest were the subchondral region, the bone marrow, the joint capsule, the joint space, and the retro-auricular ligaments.
Firstly in different sessions, MRI scans were reviewed and scored together by two observers (LHD and RW) and discrepancies in scoring were extensively discussed. After these training sessions, inter-reader reliability was assessed for a small subset of MRI scans. As the reliability appeared sufficiently high, each MRI was thereafter independently scored by these two observers, who were blind for the patient identity and for clinical, laboratory and radiological data. Findings were graded 0 (absent), 1 (minimal), 2 (moderate) and 3 (extensive). Inflammation was scored per SI-joint in the subchondral region (the region adjacent to the cortical lamella, extending 0.5 cm into the bone marrow cavity), the bone marrow, the joint capsule (the transition of the joint space to para-articular soft tissue), the joint space (defined as the space between the cortical lamellae), and the retro-auricular ligaments. Inflammation was defined as a low signal intensity on T1, with enhancement after gadolineum administration, and/or high signal intensity on STIR and/or T2 fast SE.

Inflammation in ligaments was defined as areas of low signal intensity running trough high signal intensity tissue on T1, which reflects interosseous ligaments crossing juxta-articular fatty tissue.
Structural changes were scored per SI joint, and included erosions (an irregularly delineated joint space on T1), sclerosis (low signal intensity on T1, STIR and T2 fast SE, without enhancement after gadolineum administration) and ankylosis (the disappearance of the joint space in all sequences).
Inflammation and sclerosis were scored on the iliac and sacral side of both SI-joints separately. Erosions and ankylosis were scored for the entire left and right SI joint. Active inflammation was defined as inflammation in at least one of the joint regions (subchondral bone, bone marrow, ligaments, joint capsule, joint space), and presence of structural damage as erosions, sclerosis and/or ankylosis per SI joint.

Analysis

Agreement between both MRI readers with respect to inflammation (per site) and chronic changes (sclerosis, erosions and ankylosis) was analyzed by cross-tabulation, by concordance- and discordance rates and by kappa statistics (unweighted Cohen's kappa).

Results

Patients

Of the 70 patients that were selected for the study, two patients were excluded (one because of claustrophobia and one because of withdrawal of consent); therefore, complete data for 68 patients were available for analysis. The characteristics of the patients are presented in Table 7.1. One-half of the patients were HLA-B27 positive. One-third of the patients reported a history of either psoriasis, IBD or uveitis. Two additional patients reported a history of both uveitis and IBD, and one patient reported psoriasis and IBD. Of the 25 patients (37%) with a family history of AS, four patients had a medical history of uveitis, and two patients had a history of IBD. Fifteen patients (22%) did not have any of the additional SpA features. Of these 15 patients, seven were HLA-B27 positive.

Agreement on MRI findings

Figure 7.1 shows the frequency and localization of inflammation (Figure 7.1a to Figure 7.1d) and chronic changes (sclerosis (Figure 7.1e), erosions (Figure 7.1f) and ankylosis (Figure 7.1g)) per SI joint per observer. Inflammation of the

subchondral region and the bone marrow was the most frequently observed finding. It can be seen that both readers use all grades, but pathological findings were mostly scored as grade 2, representing moderate involvement. For the further analyses all positive findings are grouped together irrespective of the grade applied.

Table 7.1 Baseline characteristics of 68 patients with chronic inflammatory low back pain.

Characteristics		All patients (n=68)
Sex (% male)		38
Mean age (SD) [years]		34.9 (10.3)
Symptom duration [months] (Median) [IQR]		18.0 [12.0-24.0]
Criteria for inflammatory low back pain:	3 criteria present [%]	56
	4 criteria present [%]	41
	5 criteria present [%]	3
Night pain present [%]		96[a]
HLA-B27 present [%]		46
History of inflammatory bowel disease present [%]		15
History of uveitis present [%]		15
History of psoriasis present [%]		24
Family history of ankylosing spondylitis present [%]		37

[a] 45 of the 47 patients in whom night pain was explored reported confirmatory.

Table 7.2 presents data on the inter-observer agreement with respect to the inflammatory and structural findings. Readers agreed on the presence of inflammation at any site in 85% of the cases in the right SI joint, and in 78% of the cases in the left SI joint. Kappa values reflecting the agreement for detecting inflammation were reasonable (right SI joint: 0.68; left SI joint: 0.51). Table 7.2 also provides insight into the prevalence of the various findings in this population with early inflammatory back pain. Based on the concordance and discordance rates, agreement is very similar for all assessed sites. However, due to the low prevalence of inflammation at several of the locations and structural changes overall, Cohen's kappa values are influenced negatively. The lowest kappa and also the lowest concordance rate is found for inflammation of the joint capsule of the left SI joint, which was present in only three patients.

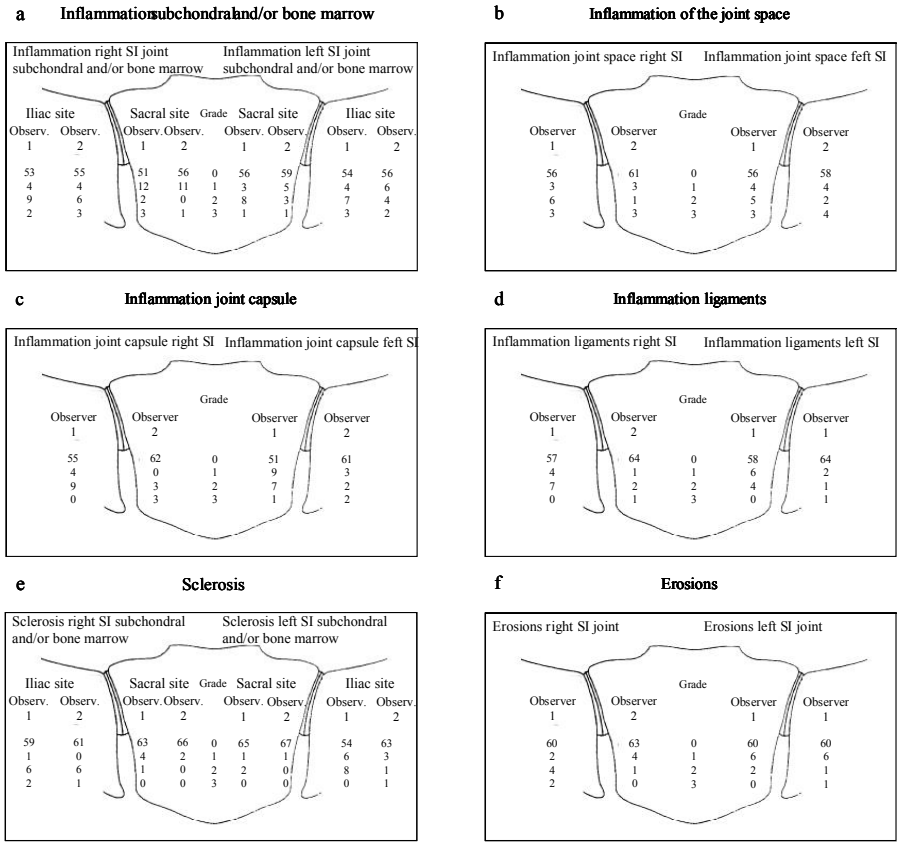

Figure 7.1 Schematic presentation of the number of pathological findings by observer one and two. All abnormalities are graded for extensiveness as 0 (absent), 1 (minimal), 2 (moderate) and 3 (extensive) in the right and left SI joint in 68 patients. For inflammation in subchondral bone and bone marrow, and for sclerosis the data are presented for each site of the joint. For the other abnormalities the grading for the entire joint per observer is presented.

Table 7.3. summarizes the concordant findings of both readers with respect to inflammation and structural changes on a patient level. Twenty-two out of 68 patients showed any sign of inflammation; ten patients in both SI joints and 12 unilaterally. Only one patient had inflammation excluding the bone marrow and subchondral bone. Ten of these 22 patients with inflammation in one or both joints also had structural changes. Only one patient had structural changes without inflammation. Similarly to inflammation, about half of the patients with structural changes showed these bilaterally (six out of 11).

Table 7.2 Agreement between two observers with respect to MRI characteristics per site per SI-joint in 68 patients with inflammatory low back pain.

MRI characteristic per site	SI-joint	Present (number)	Absent (number)	Concordance rate	Discordance rate	Cohen's kappa
Inflammation:						
At any site	Right	18	40	0.85	0.15	0.68
	Left	14	39	0.78	0.22	0.51
Bone marrow +/- subchondral bone	Right	17	43	0.88	0.12	0.73
	Left	12	47	0.87	0.13	0.65
Joint capsule	Right	5	54	0.87	0.13	0.46
	Left	3	47	0.74	0.26	0.12
Ligaments	Right	4	57	0.90	0.10	0.49
	Left	3	57	0.88	0.12	0.38
Joint space	Right	6	55	0.90	0.10	0.58
	Left	8	54	0.91	0.09	0.68
Chronic changes:	Right	6	49	0.81	0.19	0.37
	Left	11	49	0.88	0.12	0.66

Inflammation was scored for each SI joint in joint space (the space between the cortical lamellae), subchondral bone (the region adjacent to the cortical lamella, extending 0.5 cm into the bone marrow cavity), bone marrow, ligaments (defined as areas of low signal intensity running trough high signal intensity tissue on T1) and joint capsule (the transition of the joint space to para-articular soft tissue), and was defined as a low signal intensity on T1, with enhancement after Gd-administration, and/or high signal intensity on STIR and/or T2 fast SE. Chronic changes were scored for each SI joint in joint space, subchondral bone, and bone marrow, and included erosions, sclerosis and/or ankylosis. Each SI joint was labeled as showing inflammation and/or structural (chronic) changes if these respective features were present at least once in at least one site.

Table 7.3 Number of patients with abnormalities (inflammation, chronic changes or both) based on concordant observations by both readers

	Number of patients with involvement of			
	Only the left SI-joint	Only the right SI-joint	Both SI-joints	One or two SI joints
Inflammation	4	8	10	22
Structural changes (ankylosis, sclerosis, erosions)	5	0	6	11
Inflammation as well as structural changes	5	3	2	10
Inflammation: Bone marrow or subchondral bone	4	9	8	21
Joint capsule	0	2	4	6
Ligaments	0	1	3	4
Joint space	4	2	4	10

Inflammation was scored for each SI joint in joint space (the space between the cortical lamellae), subchondral bone (the region adjacent to the cortical lamella, extending 0.5 cm into the bone marrow cavity), bone marrow, ligaments (defined as areas of low signal intensity running trough high signal intensity tissue on T1) and joint capsule (the transition of the joint space to para-articular soft tissue), and was defined as a low signal intensity on T1, with enhancement after Gd-administration, and/or high signal intensity on STIR and/or T2 fast SE. Structural changes were scored for each SI joint in joint space, subchondral bone, and bone marrow, and included erosions, sclerosis and/or ankylosis. Each SI joint was labeled as showing inflammation and/or structural changes if these respective features were present at least once in at least one site.

Discussion

One of the important aims of this study was to establish whether inflammation and structural changes on MRI could reliably be assessed. In order to allow a detailed judgment, and to trace redundancies, we decided to score inflammation and structural changes per site and per type of lesion. It can be concluded that the agreement between both readers about the presence or absence of pathological findings on MRI was reasonable, especially for inflammation at sites were it was most prevalent. With agreement levels mostly around 85% for the presence of inflammation overall and at different locations, it seems sufficiently high to justify a conclusion of inflammation made by one observer in clinical practice. Expectedly with reference to the population under study, the prevalence of chronic changes on MRI was low. Because of this low prevalence of structural changes, reliability of scoring these changes is more difficult to assess. This similarly applies to the assessment of inflammation in the joint space, capsule and ligaments. Notwithstanding this limitation, the overall agreement for the different sites of the joint was comparable, with a possible exception for inflammation in the joint capsule. An explanation may be that the delineation of the joint capsule is poorly defined, which may give rise to misinterpretations.

Another important finding in this study was that it is probably sufficient to look for bone marrow edema and/or subchondral inflammation. The contribution of other sites of the joint to make a diagnosis of inflammation was only marginal. We found only one patient in whom inflammation was restricted to joint capsule and ligaments.

A few studies reported agreement with respect to lesions found on MRI examinations of the SI-joints, but none of the studies was performed in patients with recent onset inflammatory back pain. Bigot *et al*. proposed 11 criteria referring to both the synovial and the fibrous part of the sacroiliac joint that point to sacroiliitis and showed a good intra- observer and inter-observer reliability (a kappa value of 0.89 for detecting bone marrow edema).[10] However, this was a study in 22 SpA patients with established disease, in which radiological sacroiliitis grade 2 according to the New York criteria was present in 80% of the SI-joints. Puhakka *et al*. have proposed a scoring system for MRI abnormalities of the SI-joints, which distinguished inflammatory activity as well as joint damage.[11] Inter-observer reliability in this study, which included 41 patients with SpA of whom 20 patients had grade 2 sacroiliitis or more of at least one SI-joint on radiography, was importantly lower (a kappa of 0.47 for bone marrow enhancement, and 0.67 for joint space enhancement) as compared with our study. Finally, Docherty *et al*. found a kappa value of 0.63 for inter-observer agreement with respect to inflammation on MRI in a study of 20 patients with established or suspected sacroiliitis on radiographs, but

contrast administration was not performed.[12] Our results are largely in accordance with the published literature, although comparability is limited due to the differences in study population, the prevalence of the abnormalities largely influencing kappas.

In this cohort of inflammatory back pain of less than two years duration, inflammation in the SI joints on MRI could be detected in about one third of the patients (22/68). Moreover, one out of sixth of patients already showed signs of structural changes on the MRI (11/68). Although the number of patients with structural changes is low, this finding indicates that MRI might be a useful tool in the assessment of patients with early inflammatory back pain. It is the amount of variation in the outcome of interest (inflammation and/or structural changes) rather than the number of patients under study that is important in judging whether the sample size is sufficient to test reliability. As long as all kinds of abnormalities are covered, it is possible to test reliability even in situations like this, with only 11 patients showing abnormalities. Undoubtedly, however, the likelihood of covering all kinds of abnormalities will increase by increasing patient number. The real value of MRI will therefore be ascertained in future, by following the patients longitudinally and obtaining more data, which will occur in this cohort. The development and choice of an appropriate scoring system for sacroiliitis on MRI to be used in clinical studies and trials will be subject of interest in an ongoing ASAS-OMERACT working group.[13]

Conclusion

MRI can reliably detect inflammation and structural changes in SI joints in patients with early inflammatory back pain. Assessing bone marrow and/or subchondral bone enhancement suffices to detect inflammation. Inflammation in joint space, joint capsule and ligaments does hardly contribute to this detection, because it is associated with inflammation in bone marrow and/or subchondral bone. About one-third of patients with recent onset inflammatory back pain show inflammation, and about one-sixth shows structural changes in at least one SI joint, indicating that MRI might be a useful tool to diagnose sacroiliitis in patients with inflammatory back pain.

References

1. van der Linden SM, Valkenburg HA, de Jongh BM, Cats A. The risk of developing ankylosing spondylitis in HLA-B27 positive individuals. A comparison of relatives of spondylitis patients with the general population. *Arthritis Rheum* 1984,27:241-9.
2. Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, Cats A, Dijkmans B, Olivieri I, Pasero G, Veys E, Zeidler H, The European Spondylarthropathy Study Group. The European spondylarthropathy study Group preliminary criteria for the classification of spondylarthropathy. *Arthritis Rheum* 1991,34:1218-27.
3. Feldtkeller E, Bruckel J, Khan MA. Scientific contributions of ankylosing spondylitis patient advocacy groups. *Curr Opin Rheumatol* 2000,12:239-47.
4. Spoorenberg A, de Vlam K, van der Heijde D, de Klerk E, Dougados M, Mielants H, van der Tempel, Boers M, van der Linden Sj. Radiological scoring methods in ankylosing spondylitis: reliability and sensitivity to change over one year. *J Rheumatol* 1999,26:997-1002.
5. Braun J, Bollow M, Eggens U, Konig H, Distler A, Sieper J. Use of dynamic magnetic resonance imaging with fast imaging in the detection of early and advanced sacroiliitis in spondylarthropathy patients. *Arthritis Rheum* 1994,37:1039-45.
6. Bollow M, Loreck D, Banzer D, Brandt H, Zerbes K, Kourik W, Mellorowicz H, Backhaus M, Schmidt W, Bohl-Buhler M, Hauer R, Eggens U, Braun J. Diagnostic Imaging of inflammation in the axial skeleton. *Zeitschrift fur Rheumatologie* 1999,58:61-70.
7. Braun J, van der Heijde D. Imaging and scoring in ankylosing spondylitis. *Best Pract Res Clin Rheumatol* 2002,16:573-604.
8. Rudwaleit M, van der Heijde D, Khan MA, Braun J, Sieper J. How to diagnose axial spondyloarthritis early. *Ann Rheum Dis* 2004,63:535-43.
9. Calin A, Porta J, Fries JF, Schurman DJ. Clinical history as a screening test for ankylosing spondylitis. *JAMA* 1977,237:2613-4.
10. Bigot J, Loeuille D, Chary-Valckenaere I, Pourel J, Cao MM, Blum A. Determination of the best diagnostic criteria of sacroiliitis with MRI. *J Radiol* 1999,80:1649-57.
11. Puhakka KB, Jurik AG, Egund N, Schiottz-Christensen B, Stengaard-Pedersen K, van Overeem Hansen G, Christiansen J. Imaging of sacroiliitis in early seronegative spondylarthropathy. Assessment of abnormalities by MR in comparison with radiography and CT. *Acta Radiol* 2003,44:218-29.
12. Docherty P, Mitchell MJ, MacMillan L, Mosher D, Barnes DC, Hanly JG. Magnetic resonance imaging in the detection of sacroiliitis. *J Rheumatol* 1992,19:393-401.
13. Landewé R, Hermann K-G, van der Heijde D, Baraliakos X, Jurik A-G, Lambert R, Ostergaard M, Rudwaleit M, Salonen D, Braun J, and the ASAS/OMERACT MRI in AS working group. Scroing sacroiliac joints by magnetic resonance imaging. A multiple-reader reliability experiment. *J Rheumatol* 2005;32(10):2050-5.

Chapter 8

Combining information obtained from MRI and conventional radiographs in order to detect sacroiliitis in patients with recent-onset inflammatory back pain

L Heuft-Dorenbosch, R Landewé, R Weijers, A Wanders, H Houben, Sj van der Linden, D van der Heijde

Ann Rheum Dis 2006;65(6):804-8.

Abstract

Objective

To compare the contribution of changes on magnetic resonance imaging (MRI) and changes on conventional radiography (CR) in sacroiliac (SI) joints of patients with recent onset inflammatory back pain (IBP) in making an early diagnosis of seronegative spondyloarthritides (SpA).

Patients and Methods

68 patients with IBP (38% male; mean age 34.9 year (SD 10.3)) were included if symptom duration was <two years. Coronal MRI scans of the SI joints were scored for inflammation and structural changes and pelvic radiographs were scored according to the modified New York grading (mNY). Agreement between MRI and CR was analyzed by cross-tabulation per SI joint and per patient.

Results

A structural change was detected in 20 SI joints by MRI and in 37 SI joints by CR. Inflammation was detected in 36 SI joints by MRI, and 22 of these joints showed radiographic sacroiliitis. Fourteen patients fulfilled the mNY criteria based on CR. Based on structural changes on MRI eight patients would fulfill the mNY criteria, 14 (partly different patients as with CR) based on MRI for inflammation only, 16 based on MRI for inflammation and structural changes combined, 19 based on CR combined with MRI for inflammation, and 19 (same patients) based on CR combined with MRI for inflammation and structural changes.

Conclusions

CR can detect structural changes in SI joints with higher sensitivity than MRI. But inflammation on MRI can be found in a substantial proportion of patients with IBP with (yet) normal radiographs. Assessment of structural changes by CR followed by assessment of inflammation on MRI in patients with negative findings gives the highest returns in terms of detecting involvement of the SI joints by imaging in patients with recent onset IBP.

Introduction

Ankylosing spondylitis (AS) is the prototype disease in the group of spondyloarthritides (SpA). For the classification of patients as having AS according to the most widely used criteria, the modified New York criteria (mNY), radiographic sacroiliitis is obligatory.[1] A classification as SpA can also be made without sacroiliitis on radiographs, according to the Amor or European spondylarthropathy study group (ESSG) criteria.[2,3] It has been hypothesized that SpA with axial involvement not (yet) fulfilling the modified New York criteria may involve an earlier and less severe part of the spectrum of AS.[4]
Making an early diagnosis of SpA with axial involvement is challenging. One of the reasons is that sacroiliitis on radiographs is a rather late phenomenon and difficult to interpret reliably.[5] Magnetic resonance imaging (MRI) has been proposed as an imaging method to detect sacroiliitis earlier.[6] MRI can provide insight in both inflammation as well as in structural changes caused by inflammation, while radiographs show only structural changes. MRI may be particularly useful in making a diagnosis of SpA in patients presenting with inflammatory back pain (IBP). In the present study we compared the performance of MRI and conventional radiographs (CR) of sacroiliac (SI) joints in patients with recent onset IBP with a relatively high level of suspicion of SpA. We compared both imaging modalities with respect to structural changes, and we compared inflammation on MRI and structural changes on CR. Comparisons were made on the level of single joints and on the level of patients, by applying the mNY criteria by substituting CR information by MRI information in the radiographic criterion.

Patients and Methods

Patients

Patients with inflammatory low back pain present for two years at most, were eligible to this study. IBP was defined according to the Calin criteria which are positive if four out of the five following characteristics are present: insidious onset; onset before the age of 40 years; persistence for at least three months; association with morning stiffness; and improvement with exercise.[7] Patients could also be included if three out of five criteria were present plus night pain. Preferably, but not obligatory, patients should have at least one feature of SpA according to the ESSG criteria: presence of a family member with AS; presence or history of psoriasis, inflammatory bowel disease (IBD) or uveitis. The study was approved by the institutional review board and all patients gave written informed consent.

MRI

A MRI examination of the sacroiliac joints was performed using a 1,5 Tesla Philips Gyro scan ACS-NT (Philips, Best, the Netherlands). Patients were scanned in supine position using a Synergy-spine coil as surface coil. We chose a coronal oblique scan plane parallel to the length of the sacrum and two slabs: one transversal slab positioned cranially to the region of interest to diminish flow-artefacts, and one frontally through the bowel and anterior abdominal wall, to diminish motion artefacts of breathing and bowel movements. The following sequences were used: T1-weighted spin echo (SE), short tau inversion recovery (STIR), T2-weighted fast SE with fat sat and T1-weighted SE with fat suppression after the intravenous administration of contrast medium (Gadolinium diethylenetriaminepentate (Gd), 0.1 mmol/kg body weight).

Inflammation was scored for each SI joint in joint space, subchondral bone, bone marrow, ligaments and joint capsule. Inflammation was defined as a low signal intensity on T1, with enhancement after Gd-administration, and /or high signal intensity on STIR and/or T2 fast SE. Inflammation in ligaments was defined as areas of low signal intensity running trough high signal intensity tissue on T1, which reflects interosseous ligaments crossing juxta-articular fatty tissue. Structural changes (erosions, sclerosis, ankylosis) were scored in joint space, subchondral bone, and bone marrow. Each SI joint was labeled as showing inflammation or structural changes if these respective features were present in at least one of the investigated areas. Each set of MRIs was scored independently by two observers, who were blind for the patient identity and for clinical, laboratory and radiological data. All joints that showed a discrepancy between the readers for inflammation and/or structural damage were offered to a third reader. In total 21 discrepant joints were scored for the assessment of structural changes. The final score attributed to the joint was based on a two out of three majority score. A similar process was followed for discrepancies in inflammation. For this purpose 25 discrepant joints were offered to a third reader.

Conventional radiography

Antero-posterior conventional pelvic radiographs were scored independently by two observers, who were not involved in the MRI reading, without knowledge of clinical information, according to the mNY criteria (from zero (normal) to four (complete ankylosis)).[1] In case of a discrepancy between the readers a third reader (who was not involved in the MRI scoring) scored the SI joint. In total 42 discrepant SI joints were offered to the third reader. A final score for each SI joint was assigned on the basis of the majority score of the three observers.

Analysis

For CR, the scores were dichotomized. A SI joint with a majority score of zero or one was considered as normal; a SI joint with a majority score of two or higher was considered as having radiographic sacroiliitis. For fulfillment of the mNY criteria we also substituted CR information by MRI information in five different ways so that patients could fulfill the mNY criteria as follows; 1) according to the original method based on radiographs; 2) based on structural changes present on MRI in both SI joints; 3) based on inflammation present on MRI in both SI joints; 4) based on inflammation and/or structural changes present on MRI in both SI joints; 5) based on structural changes on CR combined with inflammation with or without structural changes on MRI in both SI joints. If based on radiographs patients with bilateral grade 2 sacroiliitis, or at least unilateral grade 3 sacroiliitis were classified as fulfilling the mNY criteria. If based on MRI, the mere presence of structural damage for definition two, the mere presence of inflammation for definition three, or one of both for definition four were considered sufficient, and severity or extent of the lesions was ignored. For definition 5 the grading for CR and presence of inflammation on MRI were combined.
Agreement between structural changes on MRI and sacroiliitis on CR as well as agreement between inflammation on MRI and sacroiliitis on CR was analyzed by cross tabulation. This was done both per SI joint and per classification according to the mNY criteria. Specificity, sensitivity and positive and negative predictive values were calculated with CR as gold standard.

Results

Patients

The characteristics of the 68 patients included in the study are presented in Table 8.1. Fifteen patients (22%) did not have any of the additional SpA features. Of these 15 patients, seven were HLA B27 positive. Fifty-seven patients fulfilled the ESSG criteria, 48 fulfilled the Amor criteria, and 43 fulfilled both sets of criteria.

Adjudication for structural and inflammatory changes on MRI

After the read of the two observers, there was agreement on the presence of structural changes in 17 joints, and on the absence of structural changes in 98 joints. Adjudication of 21 joints with a discrepancy for structural changes led to a positive assignment of erosions in three joints. One joint was scored positive in a patient in which already the other joint was scored positive, and the other

two joints were in one patient that was scored as having structural changes by one of the two readers in the original read in both joints.
There was agreement on the presence of inflammation in 32 joints, and on the absence of inflammation in 79 joints after the read of the two observers. Adjudication of 25 joints with a discrepancy for inflammatory changes led to an assignment of inflammation in four joints. These four joints all occurred in patients in which both readers already considered unilateral inflammatory changes present. Adjudication did not yield additional patients with inflammation.

Table 8.1 Baseline characteristics of 68 patients with chronic inflammatory low back pain.

Characteristics		All patients (n=68)
Sex (% male)		38
Mean age (SD) [years]		34.9 (10.3)
Symptom duration [months] (Median) [IQR]		18.0 [12.0-24.0]
Criteria for inflammatory low back pain:	3 criteria present [%]	56
	4 criteria present [%]	41
	5 criteria present [%]	3
Night pain present [%]		96[a]
HLA-B27 present [%]		46
History of inflammatory bowel disease present [%]		15
History of uveitis present [%]		15
History of psoriasis present [%]		24
Family history of ankylosing spondylitis present [%]		37
Fulfilling ESSG criteria		84%
Fulfilling Amor criteria		71%
Fulfilling both ESSG and Amor criteria		63%

[a] 45 of the 47 patients in whom night pain was explored reported confirmatory.

Adjudication for changes on CR

There was agreement on the presence of sacroiliitis in 29 SI joints, and on the absence of sacroiliitis in 65 SI joints. The adjudication of the 42 discrepant joints led to the assignment of eight joints as positive for sacroiliitis. In four patients adjudication resulted in fulfillment of the mNY criteria, as these patients showed already grade 2 abnormalities in the contra-lateral SI joint.

Abnormalities on MRI

In total, 20 SI joints in 12 patients showed structural changes on MRI (Tables 8.2a and 8.2b). Thirty-six joints in 22 patients showed signs of inflammation on MRI (Tables 8.3a and 8.3b). Twelve of these 22 patients also had structural

changes on MRI. None of the patients had structural changes on MRI without inflammation.

Table 8.2a Single SI joint analysis comparing structural changes observed on MRI with structural changes on conventional radiographs based on the modified New York criteria

Structural changes on MRI	Radiographs of SI joint[a]		
	abnormal	normal	total
Present	18	2	20
Absent	19	97	116
Total	37	99	136

Table 8.2b Per patient analysis comparing structural changes observed on MRI with structural changes on conventional radiographs based on the modified New York criteria.

Structural changes on MRI	Radiographs of SI joint[a]		
	sacroiliitis	normal	total
Present in both SI joints	8	0	8
Present in one SI joint	2	2	4
Absent	4	52	56
Total	14	54	68

[a] Patients fulfilled the modified New York criteria on MRI if structural changes were present and on CR if bilateral at least grade 2 or unilateral at least grade 3.

Table 8.3a Per joint analysis comparing inflammation observed on MRI with structural changes on conventional radiographs based on the modified New York criteria.

Inflammation on MRI	Radiographs of SI joint[a]		
	sacroiliitis	normal	total
Present	22	14	36
Absent	15	85	100
Total	37	99	136

Table 8.3b Per patient analysis comparing inflammation observed on MRI with structural changes on conventional radiographs based on the modified New York criteria.

Inflammation on MRI	Radiographs of SI joint[a]		
	sacroiliitis	normal	total
Present in both SI joints	9	5	14
Present in one SI joint	3	5	8
Absent	2	44	46
Total	14	54	68

[a] Radiographs of the SI joints were scored according to the modified New York criteria, which were met if bilateral grade 2 or unilateral grade 3 or four sacroiliitis was scored.

Abnormalities on CR

On CR 37 SI joints of 23 patients showed radiographic sacroiliitis; 28 SI joints grade 2, and nine SI joints grade 3. Fourteen patients fulfilled the radiographic criterion of the mNY criteria for AS: nine patients due to bilateral grade 2 sacroiliitis; four patients because of bilateral grade 3 sacroiliitis, and the remaining patient because of grade 2 and grade 3 sacroiliitis combined. The remaining nine patients showed unilateral grade 2 sacroiliitis.

MRI findings compared with radiographic findings

A comparison between structural changes on MRI and CR is presented in Tables 8.2a and 8.2b. In total 12 patients showed structural changes on MRI: eight in both joints and four in one joint. In 20 SI joints structural changes on MRI were detected, as compared to 37 SI joints with radiographic sacroiliitis (54%). In two SI joints that were normal on CR, structural changes were scored on MRI. If radiographic sacroiliitis graded according to the mNY criteria is considered the gold standard, the sensitivity of detecting chronic changes by MRI per SI joint is 49% (18/37) and the specificity is 98% (97/99). Corresponding positive predictive value and negative predictive value were 90% and 84% respectively.
Only eight of the 14 patients (57%) fulfilling the mNY criteria for AS would fulfil the radiographic criterion if this was based on the presence of structural changes on MRI (Table 8.2b), but nine of these 14 patients had signs of inflammation on MRI in both joints and three in one joint (Table 8.3). The two remaining patients showed only structural changes on MRI in one joint.
In 22 of the 37 SI joints with radiographic sacroiliitis (59%) inflammation was observed on MRI (Table 8.3a). Of the nine patients with unilateral radiographic sacroiliitis, only one patient had signs of inflammation on MRI (in both SI joints). Of the 36 SI joints with inflammation on MRI (17 left, 19 right) radiographic sacroiliitis was detected in 22 SI joints (61%; 11 left and 11 right).

If we consider either inflammation *or* structural changes on MRI as positive findings and compare this to radiographic abnormalities, there is only a small gain as compared to the information provided by inflammation alone (Tables 8.4a and 8.4b). Two more joints with MRI abnormalities could be identified, that appeared to be concordant with the findings on the radiographs. This resulted in two more patients fulfilling the mNY criteria. Based on the combined information of either inflammation or structural changes on MRI 11 of the 14 patients (79%) can be picked up that fulfill the mNY criteria according to CR and another five patients showing abnormalities on MRI while there were no abnormalities on CR. Based on MRI findings 16 patients would fulfill the mNY criteria. If we combine the information obtained by CR with that obtained by

MRI, 19 patients would fulfill the mNY criteria. In addition there are five patients showing abnormalities on MRI in a single joint in patients not fulfilling the mNY criteria on CR. CR in combination with inflammation on MRI, or CR in combination with inflammation and structural changes on MRI is equally informative.

Summarizing the above mentioned information, classification according to the mNY criteria would be justified for eight patients, if classification is based on MRI for structural changes only, 14 patients if classification is based on structural changes on CR, 14 (partly) different patients if classification is based on MRI for inflammation only, 16 patients if classification is based on MRI for inflammation and structural changes, 19 patients if classification is based on CR combined with MRI for inflammation, and 19 patients (the same patients as for CR combined with MRI inflammation only) if classification is based on CR combined with MRI for inflammation and structural damage. All patients defined as fulfillment of the radiographic mNY criteria according to the various definitions fulfill both the ESSG and the Amor criteria, except one patient with bilateral inflammation on MRI, but without structural changes on MRI or CR, who did not fulfill any of the SpA criteria. This patient, which presented with arthritis and was HLA-B27 positive, would fulfill the ESSG criteria and the Amor criteria if MRI inflammation would substitute structural changes on CR with a similar weight. In the six patients with structural changes on CR but not on MRI, other than radiographic features made these patients fulfill the ESSG and Amor criteria.

Table 8.4a Per joint analysis comparing inflammation or structural changes observed on MRI with structural changes on the conventional radiograph based on the modified New York criteria.

Inflammation and/or structural changes on MRI	Radiographs of SI joint[a]		
	sacroiliitis	normal	total
Present	24	14	38
Absent	13	85	98
Total	37	99	136

Table 8.4b Per patient analysis comparing inflammation or structural changes observed on MRI with structural changes on the conventional radiograph based on the modified New York criteria.

Inflammation and/or structural changes on MRI	Radiographs of SI joint[a]		
	sacroiliitis	normal	total
Present in both SI joints	11	5	16
Present in one SI joint	1	5	6
Absent	2	44	46
Total	14	54	68

[a] Radiographs of the SI joints were scored according to the modified New York criteria, which were met if bilateral grade 2 or unilateral grade 3 or four sacroiliitis was scored.

Discussion

The comparison of abnormalities of SI joints found on MRI and on CR in patients with recent onset IBP yielded important information. First, radiographs are more sensitive than MRI in the detection of structural changes. Second, the majority of the joints showing structural changes on MRI and/or CR also show inflammation on MRI, as did almost all patients with sacroiliitis on CR. Third, quite a number of patients show inflammation on MRI, but without signs of structural changes on either MRI of CR. Combining this information may lead to several conclusions.

To start with, the data add to the hypothesis that inflammation comes first, and structural changes are a subsequent feature. Depending on the lag time between inflammation and structural changes, a diagnosis of sacroiliitis could be made importantly earlier by using MRI inflammation as an early sign of disease. The real causality between MRI inflammation and radiographic sacroiliitis has to be proven in a longitudinal analysis, which will be possible with this cohort once follow up images have been made.

Another, rather unexpected conclusion is that CR is the preferred method for the assessment of structural changes in SI joints. We postulated CR as the gold standard for assessing structural changes. This is arguable for two reasons: CR may either overestimate or underestimate structural changes. In case of underestimation of structural changes by CR, MRI is performing worse as compared to what we already demonstrated. Overestimation of structural changes cannot be ruled out, but is considered unlikely as 12 of the 14 patients fulfilling the mNY criteria show either inflammation or structural changes in one or both SI joints on MRI (MRI confirms CR), and all these patients fulfill the ESSG and the Amor criteria for SpA. We selected CR as comparator as this is the most widely used method in clinical practice to make a diagnosis of sacroiliitis. It is know that computer tomography (CT) is a more sensitive method to detect structural changes but if true in this cohort, the difference with MRI would even be larger. Combining information on inflammation and structural changes from MRI seems the most logical way to use the information in clinical practice. By doing so, still only 11 of the 14 patients classified as fulfilling the mNY criteria according to CR can be classified according to MRI. If we would use information from MRI only (both inflammation and structural changes), another five patients would be classified according to the mNY criteria which do not fulfill the criteria on CR. Combining information on structural changes on CR with the information of inflammation on MRI classifies the highest number of patients: 14 based on structural changes on CR and five additional patients based on inflammation on MRI. Note that we used abnormalities in both SI joints on MRI as a requirement for substituting the mNY criteria. Another five patients would have classified if

unilateral MRI abnormalities had been sufficient. But in view of the literature, in which the positive predictive value of inflammation in a SI joint with respect to structural changes on the radiograph three years later was disappointingly low (60%)[8], we considered unilateral MRI inflammation insufficient. In our view this is the most appropriate way of using imaging modalities in patients with early IBP: first CR of the pelvis, followed by a MRI for the assessment of inflammation only, if patients are not fulfilling the mNY criteria. MRI for chronic changes does not seem to add much information to what is already provided by CR. By following the above-proposed flow diagram we combine the strengths of the two imaging methods.

It is well-known that assessment of sacroiliitis on CR has high inter-observer variation.[5] Therefore we decided to use a two out of three majority judgment. Assessment was first done by two experienced readers. In case of discrepancy, the joints were offered to a third independent reader. Two trained observers did assessment of the SI joints on MRI. Overall, there was good agreement on the presence on inflammation and structural changes.[9] A similar process was followed in case of discrepancy between the two readers as described for CR. So differences in scoring methodology or handling of data could not influence the likelihood of positive findings. Readers could also not be influenced by prior knowledge: the entire team reading CR was different from the team reading MRI.

Further validation of our results can be derived from follow-up of the patients, which are underway. Follow-up of the patients will be especially interesting for the ten patients that show inflammation on MRI (5 in both joints, 5 in one joint) but not (yet?) structural changes on CR. The data we presented are valid for patients with recent onset IBP with a high suspicion of SpA seen by a rheumatologist. Whether the results are also generalisable to patients with a lower likelihood of SpA is not known.

In conclusion, CR can detect structural changes in SI joints with higher sensitivity than MRI. But inflammation on MRI can be found in a substantial proportion of patients with IBP with (yet) normal radiographs. Applying only MRI for the assessment of both structural changes and inflammation would underestimate sacroiliitis. Assessment of structural changes by CR followed by assessment of inflammation on MRI in patients with negative findings yields the highest probability of detecting involvement of the SI joints in patients with recent onset IBP.

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;27: 361-8.
2. Amor B, Dougados M, Listrat V, Menkes CJ, Roux H, Benhamou C, et al. Are classification criteria for spondylarthropathy useful as diagnostic criteria? Rev Rhum Engl Ed 1995;62:10-5.
3. Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum 1991;34:1218-27.
4. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new criteria? Arthritis Rheum 2005;52:1000-8
5. van Tubergen A, Heuft-Dorenbosch L, Schulpen G, Landewé R, Weijers R, van der Heijde D, et al. Radiographic assessment of sacroiliitis by radiologists and rheumatologists: does training improve quality? Ann Rheum Dis 2003;62:519-25.
6. Rudwaleit M, van der Heijde D, Khan MA, Braun J, Sieper J. How to diagnose axial spondyloarthritis early. Ann Rheum Dis 2004;63:535-43.
7. Calin A, Porta J, Fries JF, Schurman DJ. Clinical history as a screening test for ankylosing spondylitis. JAMA 1977;237:2613-4
8. Oostveen J, Prevo R, den Boer J, van de Laar M. Early detection of sacroiliitis on magnetic resonance imaging and subsequent development of sacroiliitis on plain radiography. A prospective, longitudinal study. J Rheumatol 1999;26: 1953-8
9. Heuft-Dorenbosch L, Weijers R, Landewé R, van der Linden Sj, van der Heijde D. 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.

Chapter 9

Summary and discussion

Summary and discussion

The studies described in this thesis focus on outcome assessment in established ankylosing spondylitis (AS) (chapters 2-5) and in the second part on the clinical features in patients with early inflammatory back pain (IBP), of whom the majority has established spondyloarthritides (SpA) (chapters 6-8).
Chapter 2 focuses on the feasibility and in a part on the discriminatory capacity of an instrument for measurement of enthesitis in AS. Before the start of this study the measurement and quantification of enthesitis in AS was still a difficult and ill defined subject. The lack of consensus is reflected in different trials on AS, in which, if included at all, different and mainly self constructed enthesitis scores are reported. Our study shows that simplification of the most comprehensive and therefore time consuming enthesitis score, the Mander enthesis index (MEI), to a more feasible index named the Maastricht ankylosing spondylitis enthesis index (MASES) is an acceptable alternative. Reduction from the original 66 to 13 entheses sites captured almost all information. For a further assessment of the aspect truth of the instrument measuring enthesitis, a study investigating the correlation between a method directly visualizing inflammation (as for example MRI or ultra sound) and clinical assessment of entheses could be helpful.
In chapter 3 we applied the “outcome measures in rheumatology clinical trials” (OMERACT) filter in two measurements assessing thoracic spine extension. This is relevant as in AS patients progressive impairment in thoracic spine extension eventually adds to the characteristic bent forward posture of many AS patients. We concluded from our study that the reliability of both the occiput-to-wall distance (OWD) and the tragus-to-wall distance (TWD) measurement is similar. But the OWD is zero centimeter for all patients with a normal posture whereas the TWD is dependent on the size of the head. Therefore, the same result for TWD could mean hyperkyphosis in one patient and a normal posture in another patient. Therefore we recommend the use of OWD measurements for research purposes.
In chapter 4 we investigated how the Bath ankylosing spondylitis disease activity index (BASDAI), a six-item patient-oriented questionnaire measuring disease activity, differed between patients with- and without peripheral arthritis or enthesitis. We concluded that patients with peripheral arthritis reported a higher level of disease activity measured by means of BASDAI compared with patients without peripheral arthritis. Patients with peripheral arthritis not only scored higher on the question on peripheral arthritis but also had higher scores in the four items not dealing with peripheral arthritis or enthesitis. The difference in scoring of the BASDAI between the two groups did not disappear when omitting the two questions dealing specifically with peripheral arthritis and

enthesitis. Peripheral arthritis thus seems a sign of overall disease activity in AS.

In chapter 5 we investigated an aspect of truth in the reading of X-rays of the sacroiliac (SI) joints. We did this by assessing the comparative performance of radiologists and rheumatologists in reading SI joints on radiographs and on CT and by assessing the result of both individual and group training in the reading of radiographs. This study revealed that sensitivity and specificity scores were moderately high and comparable between the rheumatologists and radiologists. A relatively large group of participants had difficulties in diagnosing sacroiliitis, reflected by a low sensitivity for this group at baseline. Of the radiographs 15-25% were classified in a false positive manner and 20-30% in a false negative manner. The sensitivity for CT reading was significantly higher compared to radiographs (86 versus 72 %) and specificity comparable (84%). An unexpected and somewhat disappointing result was that individual training and group training in a workshop did not improve the results

In chapter 6 we present the results of the early spondyloarthritis cohort (ESpAC), a newly formed cohort of 68 patients with IBP of less than two years duration. These patients were classified according to the New York criteria for AS, and the European spondylarthropathy study group (ESSG) and Amor criteria for SpA. The patients were also classified according to the Berlin likelihood ratio product rule (Berlin LR rule). The Berlin LR rule predicts in an individual patient the likelihood of having axial SpA by assessing the presence of different SpA features in a patient and multiplying the known positive likelihood ratios (from literature) of these features. Of the patients a slight minority did not fulfill any criteria set for AS or SpA. The highest classification rate (84%) was found with ESSG criteria followed by Amor (71%) and by the Berlin LR rule (65%). An unexpectedly high proportion of one fifth of patients already met the modified New York criteria for AS and all these patients met all three criteria set for SpA and all but one showed inflammation on MRI of the SI joints. Half of patients fulfilling all three criteria sets for SpA showed inflammation on MRI. We also applied the recently published Berlin algorithm, a diagnostic algorithm for axial spondyloarthritis to test its performance. Subsequently, we challenged the diagnostic value by changing the place of MRI and HLA-B27 in the algorithm. The diagnostic performance of the Berlin algorithm with MRI instead of HLA-B27 first appeared to be better than that of the Berlin algorithm but was still not optimal.

In chapter 7 we studied the prevalence and inter-reader reliability of detecting abnormalities of SI joints on MRI. We did this by scoring SI joints for inflammation as well as for structural changes (erosions, sclerosis and ankylosis) by two observers independently. Inflammation was present in one fifth of SI joints mostly in the bone marrow and/or subchondral bone. Inflammation in other structures of the SI joint (joint space, joint capsule and

ligaments) was almost always associated with inflammation in bone marrow and/or subchondral bone. We concluded that assessing bone marrow and/or subchondral bone enhancement suffices to detect if inflammation in SI joints is present. Structural changes (erosions, sclerosis and ankylosis) were present in 12 patients. With conventional radiography as gold standard, the sensitivity of detecting structural changes on MRI was 49% and the specificity 98%. Agreement on the presence or absence of inflammation and structural changes of SI joints by MRI was acceptable, and sufficiently high to be useful in ascertaining inflammatory and structural changes due to sacroiliitis.
In chapter 8 the contribution of both MRI and conventional X-ray in making an early diagnosis of SpA were compared. Structural changes were noted in 37 SI joints on CR and in 20 SI joints by MRI. Inflammation was noted in 36 SI joints on MRI. Almost all SI joint showing structural changes on MRI or CR showed inflammation on MRI. 14 patients fulfilled the MNY criteria when based on CR, 8 patients when based on bilateral structural changes on MRI and 14 patients when based on the presence of bilateral inflammation on MRI. Combining structural changes by CR followed by assessment of inflammation on MRI in patients with negative findings gives the highest returns in terms of detecting involvement of the SI joints in patients with recent onset IBP. Thus defined 19 patients would be classified for fulfilling the radiographic criterion for AS in our cohort. We further concluded from our study that CR can detect structural changes in SI joints with higher sensitivity than MRI and that inflammation on MRI can be found in a substantial proportion of patients with IBP with (yet) normal radiographs.

Perspective

Rheumatology has faced tremendous changes during the last decade.
Not more than 30 years ago, many patients with an inflammatory rheumatic disease such as rheumatoid arthritis (RA) or AS were condemned to wheelchairs. RA and to a lesser extent AS were associated with work disability and an important increase in mortality. Therapy of inflammatory rheumatic diseases was aiming at relieving pain and preserving mobility, which was often pursued by physical therapy rather than by drugs. Rheumatologists followed their out-patients (which were in-patients during part of the year) writing “doing well” or “no complaints” in their records. They actually meant that there was no change in the condition of the patient that was already importantly disabled due to disease activity and its consequences. Since the ability of clinicians to pick up slight worsening during a half-year period in between visits is probably limited, it could happen that patients tremendously deteriorated over time without objective notice by their doctors. Another important drawback was that

effective drug treatment was hardly available, and disease modifying antirheumatic drugs (DMARDs) were known for their toxicity rather than for their efficacy, justifying a conservative application.

Rheumatology has now entered a new era by the development of biological drugs that were able to block or inhibit the activity of the pivotal pro-inflammatory cytokine Tumor Necrosis Factor- alpha (TNF-α). First introduced in RA during the last decade of the previous century, TNF-blocking drugs appeared to be able to almost completely suppress the signs and symptoms of inflammation to an unprecedented extent, without important side effects. Equally important, TNF-blocking drugs showed their efficacy in slowing down, or almost stopping radiographic progression in this disease, thus making an important impact on outcome in RA.

Based on ex vivo biopsy data from the sacroiliac joints of patients with AS, Braun and Sieper drew attention to the importance of TNF in the inflammatory process of AS, and hypothesized that TNF-blocking treatment could be effective in this disease. They proved right, and anno 2006 three currently available TNF-blocking drugs have been approved for application and reimbursement in AS.
Being asked for, most experts will rightfully claim the development of TNF-blocking drugs as the most important contribution to progress in rheumatology at the turn of the last century. But it is easy to forget that in order to be able to scientifically prove that TNF-blocking drugs indeed are effective drugs, appropriate assessment tools for application in clinical trials should be available. And it can easily be argued that the developments in the methodology of measurement in rheumatic diseases have been equally impressive as compared to drug development itself. In the field of spondyloarthritis in general, and AS in particular, it is the assessment in ankylosing spondylitis (ASAS) working group of international experts in the field of AS that has paved the way for such methodological developments. ASAS has determined a core set of pivotal domains in AS, and has validated the appropriate instruments to actually measure these domains. ASAS has also developed sets of response criteria to be used in clinical trials, as well as (partial) remission criteria, and criteria for disease modification. Such initiatives pertaining to measuring disease activity and severity, which have worldwide recognition and appreciation, tremendously empower the clinical trials designed by pharmaceutical industry to prove the efficacy of new drugs. An important consequence of methodological developments in the field of outcome research is also that research aiming at prognostication can be started, using ubiquitous outcome measures, so that studies can be compared and even pooled. Prognostication is an area of great interest for the next decade in view

of the costs of modern treatment and the increasing appreciation that the most effective (and expensive) drugs should be served for those patients that need them most (unfavorable outcome) and have the best chance on a favorable response. Anno 2006, this kind of prognostication is still in its infancy, despite a lot of endeavors in the past.
Another important field of interest which is actually inherent to better prognostication is an earlier diagnosis of AS. For many years an early diagnosis was not particularly relevant since efficacious drugs that potentially influence the course of the disease were lacking. Though the evidence that things have changed is still formally lacking, TNF-blocking drugs bear the potential to really influence the outcome of the disease in a positive manner. One could think of preserving functional ability, slowing of structural damage of the spine and keeping young patients at work. An important prerequisite will be that the disease AS can be recognized in its earliest stages, putatively before patients formally fulfill the classification criteria for AS, that require radiographic abnormalities of the SI joints, which makes them rather insensitive to early detection.

Magnetic resonance imaging may be important in filling in this gap since it detects sacroiliac- and spinal inflammation in a relatively early stage. Several groups are now working on diagnostic algorithms to improve the detection of patients with signs and symptoms that may evolve into AS. These endeavors may eventually lead to an earlier treatment start with efficacious therapies such as TNF-blocking drugs. Putatively, there is an analogy with RA, in which the window of opportunity hypothesis has evolved sufficiently well to justify an early and intensive treatment start in an attempt to modify the natural course of the disease. One could envision a situation in which a patient with inflammatory back pain of short duration and SI joint inflammation on MRI but still “clean” radiographs will be treated immediately with TNF-blocking drugs in an attempt to modify the natural evolution of a clinical picture best summarized as axial spondyloarthritis into classic AS. Such a patient does not fulfill current classification criteria for AS, and further research should shed light on the individual prognosis of this (kind of) patient(s) in order to find out whether TNF-blocking therapy could be justified in light of natural prognosis and probability of response. The ESpAC study described in this thesis forms the conceptual framework within which such research should be conducted. It is to be expected that further elaboration of the ESpAC study will give insight into prognosis, prognostic factors, value of MRI, biomarkers of diagnosis and progression, and the putative relation between early signs on MRI and late radiographic phenomena.
In follow up of the recent developments in the field of RA, where prognostication in the individual patient is nowadays one of the most pregnant

clinical questions, the field of AS has now reached a stage in which individual prediction of outcome will become of utmost importance. As outlined above, appropriate measurement instruments and insight into the earliest phases of the disease – where prediction of outcome is most important – are pivotal. The results described in this thesis may have contributed to this appreciation.

Samenvatting en discussie

Samenvatting en discussie

Algemeen

Het onderzoek dat in dit proefschrift beschreven wordt, heeft betrekking op patiënten met ankyloserende spondylitis (AS) (synoniem: de ziekte van Bechterew, spondylitis ankylopoëtica) en andere vormen van spondylartropathie (SpA). Spondylartropathie is een verzamelterm voor een groep van ziektebeelden. Hierbij horen naast AS als meest bekende voorbeeld, ook de ziektebeelden artritis psoriatica (gewrichtsontsteking bij de huidziekte psoriasis), reactieve artritis (gewrichtsontsteking die kan voorkomen na bepaalde infecties elders in het lichaam), gewrichtsontstekingen bij chronische darmontstekingen (ziekte van Crohn of colitis ulcerosa) en een restgroep van “ongedifferentieerde spondylartropathie”. Spondylartropathieën delen bepaalde kenmerken met elkaar; een belangrijk kenmerk is chronische ontsteking in gewrichten en in de wervelkolom. Het meest typerend zijn de ontsteking van de gewrichten tussen het heiligbeen en het darmbeen aan de achterzijde van het bekken, de sacroiliacaal (SI) gewrichten. Andere kenmerken zijn dat deze ziektebeelden vaker voorkomen bij patiënten die drager zijn van een bepaald erfelijke kenmerk, het HLA-B27 gen, dat ze in clusters binnen bepaalde families voorkomen en dat er ontstekingen kunnen optreden buiten het bewegingsapparaat. Voorbeelden hiervan zijn ontsteking van het regenboogvlies van het oog, de aorta (grote lichaamsslagader), de huid en darmen. Of en hoe ernstig bepaalde verschijnselen voorkomen, is per persoon verschillend. Er zijn patiënten met heel weinig klachten, weinig “ziektelast”, en er zijn patiënten met veel klachten en forse vermindering van ervaren levenskwaliteit en zelfs invaliditeit.

De rugpijn die veroorzaakt wordt door de ontsteking van bekkengewrichten en wervelkolom verschilt van rugpijn door bijvoorbeeld slijtage of overbelasting. Kenmerkend voor rugklachten door ontsteking zijn; sluipend begin van klachten, ontstaan van klachten voor de leeftijd van 40 jaar, ochtendstijfheid in de rug, verbetering van klachten door bewegen en verergering door rust, duur van klachten meer dan drie maanden. Wanneer er vier van deze kenmerken aanwezig zijn spreken we van “inflammatoire rugpijn”. De ontstekingen in de rug kunnen leiden tot schade aan het bot: structurele veranderingen. Wanneer de ontstekingen in de SI gewrichten na verloop van tijd schade hebben veroorzaakt die duidelijk zichtbaar is op een röntgenfoto, wordt er met zekerheid gesproken over AS. Voor die tijd kan er een lange fase van vier tot negen jaren bestaan met typische klachten, waarbij er een vermoeden bestaat op ontsteking van de SI gewrichten maar zonder bewijs hiervoor op de röntgenfoto’s. De laatste jaren zijn er belangrijke ontwikkelingen op het gebied van de (vroege) diagnostiek van AS, voornamelijk door het gebruik van

magnetic resonance imaging (MRI). Dit is een afbeeldingtechniek waarbij magnetische velden gebruikt worden en waarmee ontstekingen kunnen worden afgebeeld.

Behandeling

De behandeling van AS bestaat uit voorlichting, oefentherapie en medicijnen. Tot enkele jaren geleden bestonden de medicijnen voor AS alleen uit pijnstillers, meestal van de ontstekingsremmende soort (in het Engels: non steroidal anti inflammatory drugs; NSAIDS) zoals bijvoorbeeld ibuprofen of naproxen. Recent zijn er nieuwe medicijnen ontwikkeld die voor patiënten met een ernstige vorm van AS en ook andere vormen van SpA een enorme verbetering van klachten kunnen geven; een tot voor kort onhaalbaar resultaat. Deze medicijnen remmen het ontstekingseiwit tumor necrosis factor alfa (TNF-α) en worden periodiek via infuus of injectie toegediend.

Historie en wetenschappelijke ontwikkeling

Hoewel wij nu weten dat er al beschrijvingen van de ziekte van Bechterew bekend zijn uit 1691, heeft men lange tijd geen onderscheid gemaakt tussen reumatoïde artritis, een andere meer bekende chronische reumatische ontstekingsziekte maar met andere karakteristieken, en de groep van spondylartropathiën. Pas in 1984 werden spondylartropathiën als aparte groep van chronische reumatische ontstekingsziekten beschreven en werden kenmerken vastgelegd. In 1995 werd de “assessment in ankylosing spondylitis” (ASAS) werkgroep opgericht, met als doel het maken van internationale afspraken ten aanzien van onderzoek en dagelijkse medische zorg voor patiënten met AS. In deze werkgroep hebben meer dan 60 klinische experts, epidemiologen, vertegenwoordigers van patiëntenverenigingen en de farmaceutische industrie zitting. Deze werkgroep functioneert onder de paraplu van de outcome measurements in rheumatology clinical trials (OMERACT), een groot internationaal verband van onderzoekers dat zich richt op uitkomstmaten binnen het gehele hele veld van de reumatologie. Binnen OMERACT werd het OMERACT filter ontwikkeld, waarin de elementen “waarheid”, “onderscheidend vermogen” en “uitvoerbaarheid” getoetst worden door een set van relevante vragen die ten aanzien van ieder meetinstrument gesteld kunnen worden.
Aspect 1. *Waarheid*: Is het meetinstrument waarheidsgetrouw, meet het dat wat de bedoeling is? Is de meting onbevooroordeeld en relevant?
Aspect 2. *Onderscheidend vermogen*: Kan het meetinstrument een onderscheid maken tussen patiënten met- en zonder ziekte, of tussen patiënten met ernstige en minder ernstige ziekte? Krijg je eenzelfde

meetresultaat bij ongewijzigde omstandigheden? Krijg je hetzelfde resultaat wanneer gemeten wordt door een andere observator?
Aspect 3. *Toepasbaarheid*: Kan het meetinstrument gemakkelijk gebruikt worden uit oogpunt van tijdsinvestering, kosten en interpreteerbaarheid?
De ASAS werkgroep heeft voor onderzoek en medische zorg bij AS patiënten verschillende domeinen gedefinieerd waarover in klinisch onderzoek gerapporteerd moet worden. Deze domeinen zijn: pijn, fysiek functioneren, spinale mobiliteit (beweeglijkheid van de wervelkolom), ochtendstijfheid, moeheid, welbevinden, gewrichts- en peesontstekingen, ontstekingsreacties in het bloed, röntgenfoto's van de wervelkolom en röntgenfoto's van de heup. Voor elk domein zijn daarbij meetinstrumenten aangewezen die voor dit doel geschikt zijn.

Indeling proefschrift

De grote lijn van dit proefschrift is als volgt: In hoofdstuk 2-5 worden verschillende aspecten belicht van enkele meetinstrumenten zoals gebruikt bij patiënten met AS. In hoofdstuk 6, 7 en 8 worden kenmerken van een groep van patiënten met minder dan twee jaar bestaande inflammatoire rugklachten beschreven. Hier volgt van ieder hoofdstuk een beknopte samenvatting.

Hoofdstuk 2 gaat over de toepasbaarheid en het onderscheidend vermogen van een instrument om enthesitis (peesontstekingen) te meten. Enthesitis is een belangrijk kenmerk van SpA en geeft pijnklachten op de aanhechtingsplek van de pezen aan het bot. Vóór de start van deze studie was het meten en kwantificeren van enthesitis een moeizaam en slecht beschreven onderwerp. In verschillende gepubliceerde studies over AS worden vaak verschillende klinische meetinstrumenten gebruikt. Een bekende maar omslachtige meetmethode is de Mander enthesis index, die wordt gemeten door het uitoefenen van druk op 66 peesaanhechtingen. Hierbij wordt de pijnreactie van de patiënt gescoord op een schaal van 0 tot 3. Wij tonen aan dat het terugbrengen van de te onderzoeken punten van 66 tot 13 en het weglaten van de pijngradatie geen substantieel verlies aan informatie oplevert. Bovendien wordt hierdoor de scoringsmethode voor zowel onderzoeker als patiënt beter hanteerbaar. In deze studie hebben wij ons niet bezig gehouden met de vraag of er een relatie is tussen het klinisch onderzoek en zichtbare ontstekingen op echografie of MRI en zo ja, hoe sterk die is. Dit is een zinvolle onderzoeksvraag voor de toekomst.
In hoofdstuk 3 wordt het OMERACT filter toegepast op twee meetinstrumenten voor spinale mobiliteit, de "occiput-to-wall distance" (OWD) en de "tragus-to-wall distance" (TWD). Bij patiënten met AS ontstaat vaak in de loop der jaren een karakteristieke voorovergebogen houding die deels veroorzaakt wordt door

een langzaam toenemende beperking in het strekken van de thoracale wervelkolom. (borstgedeelte van de wervelkolom). De mate van voorover-buigen kan bepaald worden door het recht laten staan van de patiënt langs een muur en de afstand te meten van of het occiput (achterhoofd) of de tragus (het kraakbenig middenstukje van de voorzijde van het oor) tot de muur. Uit onze studie blijkt dat de betrouwbaarheid van beide metingen hetzelfde is. Echter, iemand zonder beperkingen in beweeglijkheid en vorm van de wervelkolom kan met zijn of haar achterhoofd tegen de muur staan, de OWD is dan dus altijd 0. De TWD is afhankelijk van de grootte van het hoofd en de positie van het oor. De TWD kan bij een rechtstaande patiënt dus 10 centimeter zijn maar ook 12. De TWD onderscheidt dus niet in één oogopslag patiënten met beperkingen van patiënten zonder beperkingen. Voor onderzoeksdoeleinden adviseren wij dan ook de OWD.

In hoofdstuk 4 onderzochten we of de BASDAI (Bath ankylosing spondylitis disease activity index) verschilde tussen patiënten met en zonder perifere gewrichtsontsteking of peesontstekingen. De BASDAI is een door de patiënt in te vullen lijst van zes vragen en meet ziekteactiviteit. Twee van de zes vragen hebben specifiek betrekking op gewrichts- en peesklachten, de overige vier vragen gaan over rugpijn, moeheid, duur- en intensiteit van ochtendstijfheid. Patiënten met gewrichtsontstekingen en/of peesontstekingen, bleken een hogere ziekteactiviteit te hebben vergeleken met patiënten zonder. Dit verschil verdween niet, wanneer de vragen over gewrichts- en peesontstekingen uit de vragenlijst werden weggelaten; gewrichts- en peesontstekingen lijken dus een uiting van algeheel toegenomen ziekteactiviteit.

Het beoordelen van SI gewrichten op bekkenfoto's is moeilijk. In hoofdstuk 5 onderzochten we hoe goed reumatologen en radiologen deze foto's kunnen beoordelen, hoe ze computertomografieën (CTs) van de SI gewrichten beoordelen en of het trainen van het beoordelen tot betere resultaten zou leiden. Deze onderzoeksvragen gaan over het aspect "waarheid" zoals opgenomen in het OMERACT filter. In deze studie bleek dat de sensitiviteit (de kans op de uitslag "afwijkend SI gewricht" bij een ziek persoon) en specificiteit (de kans op de uitslag "normaal SI gewricht" bij een gezond persoon) van beide groepen matig hoog en vergelijkbaar was. Een vrij grote groep van deelnemers had bij aanvang van het onderzoek moeite met het vaststellen van sacroiliitis. Van de röntgenfoto's werd 15 tot 25% fout positief beoordeeld (het ten onrechte diagnosticeren van afwijkingen) en 20 tot 30% fout negatief (het ten onrechte diagnosticeren van afwezigheid van afwijkingen). De sensitiviteit van het beoordelen van CT was hoger dan die van röntgenfoto's (86 versus 72%) en de specificiteit vergelijkbaar (84%). Een ietwat teleurstellende bevinding was, dat noch individuele training noch groepstraining de resultaten verbeterden.

In hoofdstuk 6 presenteren we de resultaten van het ESpAC cohort, een groep van 68 patiënten met minder dan twee jaar bestaande inflammatoire lage rugpijn. Omdat inflammatoire rugpijn een vaak voorkomende klacht is bij patiënten met enige vorm van SpA, hebben we gekeken of deze patiënten voldeden aan de verschillende criteria zoals die voor SpA bestaan, namelijk de gemodificeerde New York criteria voor AS en de European spondylarthropathy study group (ESSG) en Amor criteria voor SpA. Ook werden deze patiënten geclassificeerd volgens de Berlijn likelihood ratio product regel (Berlin LR rule). likelihood is een engelse statistische term, letterlijk vertaald aannemelijkheidverhouding, die de verhouding weergeeft tussen de kans op een bepaalde testuitslag bij zieken en de kans op dezelfde testuitslag bij gezonden. De Berlijn likelihood ratio product regel voorspelt voor een individuele patiënt de waarschijnlijkheid van de aanwezigheid van axiale spondylartropathie (spondylartropathie van de wervelkolom). Dit wordt gedaan door de bij een patiënt de aanwezige SpA kenmerken vast te stellen en de hierbij behorende likelihood ratio's zoals die vanuit de literatuur bekend zijn, met elkaar te vermenigvuldigen. Een patiënt kan hierdoor nog voordat de bewijzende afwijkingen op de röntgenfoto aanwezig zijn al voldoen aan de diagnostische criteria voor axiale SpA. De hoogste classificatie in het cohort werd gevonden voor de ESSG criteria (84%), en daarna voor de Amor (71%) en Berlin LR rule (65%). Een vijfde deel van de patiënten voldeed onverwacht al bij inclusie aan de gemodificeerde New York criteria voor AS, en deze 14 patiënten voldeden ook aan alle drie criteria sets voor SpA. In totaal voldeden 36 patiënten aan alledrie de criteria sets en bij deze patiënten konden we in de helft van de gevallen op de MRI ontstekingsactiviteit zien. In 2004 werd voor het eerst een diagnostisch algoritme voor axiale SpA gepubliceerd. Ook dit hebben we getest in onze patiëntenpopulatie. Het bleek dat nadat we in dit algoritme een kleine verandering toepasten, dit iets beter werd.
In hoofdstuk 7 bestuderen we het vóórkomen en de betrouwbaarheid van de beoordeling door twee onderzoekers van MRI afwijkingen in het ESpAC cohort. De SI gewrichten werden beoordeeld op aanwezigheid van ontstekingen, en structurele veranderingen (schade). Ontstekingen waren aanwezig in een vijfde deel van de SI gewrichten en betroffen meestal het beenmerg, of het bot gelegen direct onder het kraakbeen (subchondraal). Ontstekingen in bijvoorbeeld kapsel, gewrichtsspleet en ligamenten ging vrijwel altijd samen met ontstekingen in beenmerg of subchondraal bot. We concluderen dan ook, dat het voor het beoordelen van ontstekingsactiviteit voldoende is om beenmerg of subchondraal bot te beoordelen. Structurele schade werd gezien bij ongeveer een zesde van de patiënten. Wanneer we de gewone röntgenfoto als gouden standaard beschouwen is de sensitiviteit van MRI 49% en de specificiteit 98%. De overeenstemming tussen de twee beoordelaars over aan-

en afwezigheid van ontstekingen, was acceptabel en hoog genoeg voor een individuele beoordeling door één van de beoordelaars.
In hoofdstuk 8 wordt de bijdrage van zowel röntgenfoto's als MRI bepaald voor de vroeg-diagnostiek van AS. Hiertoe construeerden we 4 alternatieve manieren waarop iemand aan de criteria voor AS zou kunnen voldoen, naast de gangbare gemodificeerde New York criteria. Deze alternatieven waren:

1. Beide SI-gewrichten vertonen structurele schade op MRI.
2. Beide SI-gewrichten vertonen ontstekingsactiviteit op MRI.
3. Beide SI-gewrichten vertonen ontstekingsactiviteit en/of structurele schade op MRI
4. Beide gewrichten tonen een combinatie van structurele schade op röntgenfoto's met ontsteking en/of structurele schade op MRI.

Structurele schade werd gezien in 37 SI gewrichten op de röntgenfoto en in 20 SI gewrichten op MRI. Ontstekingsactiviteit werd gezien in 36 SI gewrichten op MRI. In vrijwel alle gewrichten met structurele afwijkingen werd ook ontstekingsactiviteit gezien. Veertien patiënten voldeden aan de gemodificeerde New York (mNY) criteria, gebaseerd op röntgenfoto's en 14 gedeeltelijk andere patiënten voldeden aan alternatief 2 en hadden op MRI ontstekingsactiviteit in beide SI gewrichten. Gebaseerd op structurele schade op MRI voldeden slechts acht patiënten. Negentien patiënten voldeden aan alternatief 4, een combinatie van een eerste beoordeling op de röntgenfoto's met bepaling van ontstekingsactiviteit op MRI bij die patiënten zonder afwijkingen op de röntgenfoto's. We stelden vast dat op gewone röntgenfoto's structurele afwijkingen met een hogere gevoeligheid kunnen worden waarnemen dan met MRI, en dat ontstekingsactiviteit bij een aanzienlijk aantal van de patiënten aangetroffen wordt.

Perspectief

De laatste tien jaar is de reumatologie ingrijpend veranderd. Nog geen 30 jaar geleden waren veel patiënten met ontstekingsziekten zoals reumatoïde artritis (RA) en AS uiteindelijk veroordeeld tot de rolstoel en zowel RA als AS ging gepaard met een verhoogde sterfte. Therapie was gericht op pijnverlichting en behoud van mobiliteit maar meestal werd méér bereikt door fysio-therapeutische behandeling dan door de voorgeschreven medicatie. Reumatologen noteerden over hun poliklinische patiënten in hun dossier steeds "geen klachten" of "gaat redelijk". Eigenlijk bedoelden ze, dat er geen verandering in de toestand van de al ernstig beperkte patiënt was en bovendien blijkt het signaleren van een langzame achteruitgang moeilijk. Zo kon het gebeuren dat patiënten in de loop van de jaren enorm verslechterden onder het toeziend oog van de dokter. Er waren weinig methodes om te meten, er was aanvankelijk nauwelijks effectieve behandeling, en de medicatie die het

verloop van ziekte kon beïnvloeden, de zogenaamde “disease modifying anti rheumatic drugs” (DMARDS) waren meer bekend vanwege hun schadelijke bijwerkingen, dan vanwege hun effectiviteit.
Door de ontwikkeling van de zogenaamde “biologicals” die de activiteit van het centrale ontstekingseiwit “Tumor Necrosis Factor Alpha” (TNF-α) remmen of blokkeren, is de reumatologie een geheel nieuwe weg ingeslagen. Anti-TNF behandeling werd in de negentiger jaren in eerste instantie geïntroduceerd bij de behandeling van RA. Bij RA is anti-TNF in staat het ontstekingsproces in belangrijke mate af te remmen, en de schade zoals die op röntgenfoto’s gezien wordt te vertragen of te stoppen. Gebaseerd op bevindingen in biopten van SI gewrichten bij AS, wezen Braun en Sieper op een belangrijke rol van TNF bij AS en werd de blokkade van TNF ook bij AS patiënten als mogelijke behandeling gesuggereerd. Ze kregen gelijk en in 2006 zijn drie TNF blokkerende preparaten toegelaten voor de behandeling van AS. Veel reumatologen zullen dan ook de ontwikkeling van anti-TNF beschouwen als een van de belangrijkste ontwikkelingen binnen de reumatologie. We moeten daarbij echter niet vergeten dat, om wetenschappelijk te kunnen bewijzen dat een behandeling daadwerkelijk effectief en veilig is, ook het gebruik van juiste meetmethoden essentieel is. De verdere ontwikkeling van de methodologie van meetmethodes binnen de reumatologie is daarom wellicht even belangrijk als het ontwikkelen van behandelmethodes zelf. Het werk van de ASAS werkgroep heeft deze methodologische ontwikkelingen enorm bevorderd en gestimuleerd. Door de werkgroep is zowel een kernset van onderzoeksgebieden als van de te gebruiken meetinstrumenten vastgesteld. Ook zijn responscriteria voor studies gedefinieerd, en voor de dagelijkse praktijk criteria voor (partiele) remissie (vermindering) van ziekte. Deze initiatieven worden wereldwijd erkend en gewaardeerd en hebben gezorgd voor verbetering van de kwaliteit van klinische farmacologische studies. De volgende belangrijke ontwikkeling zal onderzoek zijn dat zich richt op het voorspellen van het beloop en de uitkomst van ziekte (prognosticeren). Hierbij moeten dan algemeen beschikbare uitkomstmaten gebruikt worden, zodat studies goed vergeleken kunnen worden, en soms kunnen worden samengevoegd. Prognosticeren zal de komende tien jaar een belangrijk plaats innemen met het oog op de kosten van behandeling, maar ook met het oog op de afweging van de verwachte effectiviteit en mogelijke bijwerkingen voor een individuele patiënt. Op dit moment staat het prognostisch onderzoek nog in de kinderschoenen.

In het licht van de hierboven geschetste ontwikkelingen lijkt vroegdiagnostiek van AS dus een belangrijk thema te worden Op het moment dat er geen effectieve behandeling voor AS mogelijk was, en het beloop van de ziekte niet beïnvloed kon worden, was het belang van een vroege diagnose van AS minder duidelijk. Hoewel formeel het bewijs nog niet geleverd is dat blokkade

van TNF-α de uiteindelijke uitkomst van het ziekteproces bij AS beïnvloedt, bestaat de hoop dat het optreden van structurele schade in de wervelkolom vertraagd zal worden, het fysiek functioneren van patiënten beter bewaard blijft, en werkuitval minder vaak optreedt. In dat licht is het belangrijk dat AS in een vroeg stadium, voordat er structurele schade ontstaan is, herkend kan worden. Tot op heden is juist de aanwezigheid van structurele veranderingen op de röntgenfoto een voorwaarde waaraan voldaan moet worden voordat formeel de diagnose AS gesteld kan worden. Het zal duidelijk zijn dat deze voorwaarde vroegdiagnostiek nu nog in de weg staat. Het gebruik van MRI zou een mogelijkheid zijn om dit probleem te ondervangen, omdat MRI in een relatief vroeg stadium ontstekingen aan kan tonen. Verschillende onderzoeksgroepen werken nu aan het ontwikkelen van betrouwbare methodieken voor de vroegdiagnostiek van AS waarin MRI onderzoek is opgenomen. Dit zou kunnen leiden tot vroegere behandeling met bijvoorbeeld anti-TNF, waarbij er een analogie is met de behandeling van patiënten met RA. Bij RA patiënten is de theorie van de “window of opportunity” de belangrijkste reden om met een vroege en agressieve behandelstrategie het natuurlijke beloop van de ziekte in positieve zin te beïnvloeden. Men zou zich kunnen voorstellen dat een patiënt met inflammatoire rugpijn en ontstekingen van het SI gewricht op MRI ondanks een normale röntgenfoto onmiddellijk met anti-TNF behandeld zou worden in een poging de natuurlijke ontwikkeling naar AS tegen te gaan. Verder onderzoek in deze zal duidelijk moeten maken of een dergelijke aanpak gerechtvaardigd is in het licht van natuurlijk beloop, respons op behandeling, kosten en bijwerkingen. De ESpAC studie zoals beschreven in dit proefschrift zal in de toekomst een stuk informatie verschaffen over prognose, waarde van MRI, biologische markers voor diagnose en progressie en de relatie tussen vroege tekenen van inflammatie op MRI en latere radiologische afwijkingen. In analogie met de situatie bij RA is het onderzoeksveld rond AS nu in een stadium gekomen waarin de voorspelling van uitkomst en ziektebeloop bij een individuele patiënt een van de belangrijkste onderwerpen is geworden. De keuze van de juiste meetinstrumenten en inzicht in het vroege ziektebeloop zijn hiervoor noodzakelijk, hopelijk dragen de onderzoeksresultaten beschreven in dit proefschrift hieraan bij.

Dankwoord

Dankwoord

Ik ben een dankbaar mens. Dankbaar om het leven, dat mij zo rijk bedeelt. Op deze plek past het om u allen te noemen, die mijn wetenschappelijke bouwwerk van fundamenten, bouwstenen of verfraaiingen hebben voorzien. De kans om te promoveren kwam op het juiste moment op mijn pad. Zeer gewaardeerde professor van der Linden, u was diegene die mij als eerste deze kans bood, nadat ik een kleine proeve van bekwaamheid had afgelegd door de publicatie van een literatuurstudie over Yttrium synovectomieën. Bij u kreeg mijn “reumatologisch fundament” tijdens de opleiding tot reumatoloog voldoende sterkte voor de komende decennia. Dank voor de menselijke maat daarbij, ik waardeer u zeer. Professor van der Heijde, beste Désirée, met een ongelofelijke drive en inzet ben je wetenschapper. Dat je naast de vele vaak buitenlandse verplichtingen nog tijd hebt gevonden om mijn artikelen te lezen, met me te bespreken en te corrigeren is eigenlijk een mirakel. Dank voor je educatie, wetenschap zit niet in mijn genen, niet in mijn “nature”, maar door jou “nurture” heb ik mijn wetenschappelijk bouwwerk van voldoende stevigheid kunnen voorzien. Dank ook voor de momenten naast de wetenschap, waarin we even tijd hadden voor “het leven, het heelal en de rest”. Dr. Landewé, beste Robert, dank. Ook jou is de wetenschap op het lijf geschreven en jij hebt met veel geduld mijn te veel aan “Libelle” schrijfstijl voorzien van de juiste wetenschappelijke bewoordingen. Ik deed nooit tevergeefs een beroep op je. Van alle andere leden en (opleidings)assistenten van de werkgroep reumatologie in het azM kreeg ik tijdens mijn opleiding en onderzoeksperiode op het juiste moment concrete hulp of een steuntje in de rug; Professor Piet Geusens, Marijke van Santen, Annelies Boonen, Debby Vosse, Thea Schoonbrood, Simone Gorter, Hatice Demirel, allen dank. Mede onderzoekers Astrid Wanders, Astrid van Tubergen, Karin Bruynesteyn, Erik de Klerk, in jullie aanwezigheid was onderzoek doen zeker geen straf. Er waren in onze onderzoekskamers veel bespiegelingen van meer of minder relevantie, ik denk er met enige weemoed aan terug. René Weijers, radioloog, heel wat MRIs hebben we samen bekeken en heb jij in een latere fase voor mijn onderzoek gescoord. Je had het geweldig druk met je eigen promotie, hoe je het extra werk voor de ESpAC studie ingepland kreeg, blijft voor mij een raadsel. En hoe ik moet uitleggen dat het serieus kijken naar MRIs ook nog gezellig kan zijn, gelardeerd met anekdotes over het thuisfront, daar begin ik niet eens aan. Anneke Spoorenberg, eerste oppasmoeder van het OASIS cohort, inmiddels ver weg verhuisd, maar niet uit het hart. Dank dat je mijn paranimf wilt zijn. Frank Jungbauer, samen studeren, samen in een studentenhuis, getuigen bij elkaars trouwen, getuigen bij elkaars promoveren, Frank, ik zou bijna zeggen “in voor en tegenspoed” (maar dat past over het algemeen in een andere situatie) dank voor je vriendschap. Ja, wie moet ik eigenlijk niet bedanken? Alle

stenen van een huis hebben hun eigen plek en belang. Weet dat ik ieders bijdrage evenzeer gewaardeerd heb, en excuses als ik iemand onbedoeld niet noem. Gewaardeerde patiënten van zowel het OASIS als ESpAC onderzoek, u heeft belangeloos uw tijd gespendeerd, dank hiervoor. Zonder u was dit onderzoek niet mogelijk. Hopelijk dragen de resultaten uiteindelijk bij tot optimalisering van de medische zorg rond uw aandoening. Collegae reumatologen uit Nederlands en Belgisch Limburg, collegae vanuit de andere specialismen, dank voor het meedenken over mijn onderzoek, dank voor het verwijzen van patiënten. Yolanda Soons, zo lang al secretaresse van professor van der Linden ben je een spil op het reuma secretariaat van het azM en een luisterend oor voor velen. Ook voor mij was je altijd belangstellend, bij jou is het goed toeven. Onderzoeksassistentes Anita Legtenberg en Mieke Witte, zo verschillend qua stijl, zo hetzelfde wat betreft toewijding, met jullie was het prima samenwerken! Dank voor alle organisatie en improvisatie. Ook Janine, Els, Dyonne en Margriet bedankt voor jullie zorgen voor OASIS en ESpAC. Röntgen laboranten van de onderzoeks-MRI in het azM, dank voor jullie werk op de vrijdagmiddag, steeds in een prettige sfeer. Ook aan alle medewerkers van het secretariaat en de poli reumatologie in het azM, dank. Tiny Wouters, je bent inmiddels een echte “ouwe rot” in het drukproef klaar maken van proefschriften. Ook ik ben je zeer erkentelijk voor al het werk dat je in de laatste fase hebt verzet.
Inmiddels heb ik mijn plek gevonden als reumatoloog in de ziekenhuizen in Roermond en Weert. Maat Piet, medewerkers Corrie, Anita, Chrit en Marga, ook jullie steun is onontbeerlijk geweest en zal het nog lang blijven.
En dan is het ook gebruikelijk op deze plek je naasten te bedanken. Natuurlijk, het meeste dankbaar ben ik om mijn basis, geen fundament of bouwwerk in wetenschappelijk opzicht, maar de essentie van het bestaan. Jullie zijn de bloemen voor de ramen, maar door mij zo moeilijk met woorden te schilderen. Pappa en mamma, wat ben ik blij dat jullie er altijd waren, wat ben ik dankbaar dat jullie er zijn. Ma Heuft, broers en zus, schoonbroers en schoonzussen, een hechte familie is wat je iedereen toewenst, ik prijs mezelf gelukkig met jullie allemaal. Lieve vrienden, lieve familie ver of dichtbij, het juiste woord op het juiste moment lijkt zo weinig maar is zo essentieel. Dank voor alle momenten samen, dat geeft het leven kleur, dank dat ik erbij mag horen, dank voor jullie praktische hulp als het te hectisch werd. Lieve Miep, Tony en Fiona, jullie zorg voor onze kinderen was en is onbetaalbaar. Bedankt voor jullie mee-moederen. Richard, mijn liefste, steun, toeverlaat en sparringpartner, vader van onze kinderen, wie weet ons wonder uit te leggen? Lieve, lieve Anne en Koen, jullie hebben me blijmoedig leren koorddansen. Kom, we dansen verder in de zon…..

Curriculum vitae

Curriculum vitae

Liesbeth Heuft-Dorenbosch werd geboren op 27 april 1966 te Deurne in Noord Brabant als oudste van een gezin met vier kinderen. Haar gymnasium diploma behaalde zij in 1984 aan het Bisschoppelijk college Broekhin te Roermond. Van 1984 tot 1990 studeerde zij geneeskunde aan de Universiteit van Maastricht. In 1991 en 1992 was zij werkzaam als AGNIO interne geneeskunde, eerst in het academisch ziekenhuis Maastricht en later in het Catharina ziekenhuis Eindhoven waar zij in oktober 1992 begon met de opleiding tot internist. In 1996 verlegde zij haar professionele pad en in oktober van dat jaar werd begonnen met de opleiding tot reumatoloog in het academisch ziekenhuis Maastricht onder supervisie van Prof. dr. Sj. van der Linden. In 1999 werd zij geregistreerd als reumatoloog. Zij bleef werkzaam in het Academisch Ziekenhuis Maastricht waar zij in 2001 startte met het in dit proefschrift beschreven promotieonderzoek onder leiding van Prof. dr. D.M.F.M. van der Heijde. Daarnaast begon zij in 2001 parttime praktijk uit te oefenen in het Laurentius ziekenhuis te Roermond en het St. Jans Gasthuis te Weert in maatschap met de daar al werkzame reumatoloog Piet Jacobs. Sinds 1990 is zij getrouwd met Richard Heuft, samen hebben zij twee kinderen, Anne (1996) en Koen (1998).