EXTENDED REPORT

Treatment trials in ankylosing spondylitis: current and future considerations

D van der Heijde, J Braun, D McGonagle, J Siegel

Emerging treatment options in ankylosing spondylitis (AS) are giving new hope to patients with this chronic and potentially disabling disease. Clinical development of new treatments requires that rigorous and well controlled trials be conducted to demonstrate safety and efficacy. A number of classification systems have been developed in recent years as a result of enhanced understanding of the pathogenesis of AS. Although new outcome measures have been developed and a consensus has been reached on the use of assessment instruments in clinical trials, there is still need for improvement and implementation.

The ASsessments in Ankylosing Spondylitis (ASAS) Working Group has addressed some of these dilemmas by establishing a core set of domains for the evaluation of AS and by selecting specific assessment methods for each domain. They have also published improvement criteria for assessing short term improvement with symptom modifying antirheumatic drugs and are presently in the process of developing response criteria for disease controlling antirheumatic treatment. Various experts are also currently examining discrepancies and inadequacies of classification systems for AS. Imaging studies, magnetic resonance imaging, in particular, may provide better classification criteria in the near future.

In addition to consensus on outcome assessment and classification of AS, lessons learnt from clinical trials in rheumatoid arthritis (RA) may serve as a template for AS. Guidance provided by the United States Food and Drug Administration (FDA) for clinical trials in RA may be of particular use. The FDA has defined the claims that sponsors can receive for RA products and the clinical trial data that would be expected to be submitted to support such claims.

See end of article for authors' affiliations
....

Correspondence to:
Dr D van der Heijde, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, PO Box 5800, 6202 AZ Maastricht, The Netherlands;
dhe@sint.azm.nl
....

The emergence of new treatment options in ankylosing spondylitis (AS), resulting from recent advances in immunology, biotechnology, and pharmaceutical science, has accelerated the need for universal standards to assess systematically treatment response and indications in AS. Clinical trials to evaluate the safety and efficacy of these new treatments should be conducted using instruments agreed upon for measuring short and long term outcome and classification systems for differentiating patient populations based on disease activity, disease pattern, and previous treatment. Standardisation of these aspects will facilitate the collection of conclusive, reproducible, and comparable study results, which in turn might impact future treatment strategies. Although consensus has not yet been established on many aspects of disease outcome, classification, treatment indications, and clinical trial end points for medical treatments in AS, much progress has been made over the past 50 years as the aetiology and pathophysiology of AS have been further elucidated. Recent efforts of various individual people and groups towards devising comprehensive uniform outcome assessments and classification systems and the impact of these efforts on the design and implementation of future treatment trials are reviewed herein.

OUTCOME ASSESSMENT

Numerous assessment methods are currently available in AS, including laboratory measurements, metrology, radiographs, and questionnaires. In their review of published work, Bakker *et al*[1] and van der Heijde *et al*[2] documented more than 100 methods. Although the abundance of approaches suggests significant progress, continued work is needed to improve existing methods and to fill gaps within the spectrum of relevant outcomes.[3] In addition, consensus on the selection of

methods to apply in clinical trials and on guidelines for their use is needed.

In 1995 an international working group was formed to address these needs. The group, known as the ASsessments in Ankylosing Spondylitis (ASAS) Working Group, established a core set of domains for the evaluation of AS and selected specific assessment methods (instruments) for each domain.[2 4] They subsequently published improvement criteria for assessing short term improvement with symptom modifying antirheumatic drugs (SM-ARDs) using outcome data from placebo controlled trials of non-steroidal anti-inflammatory drugs (NSAIDs).[5] Presently, work is in progress to develop response criteria for disease controlling antirheumatic treatment (DC-ART) based on results of studies with infliximab and etanercept. The ASAS core set of domains, selection of instruments, improvement criteria for assessing short term improvement with SM-ARDs, and development of response criteria for DC-ART are reviewed in the subsequent sections.

....

Abbreviations: AS, ankylosing spondylitis; ASAS, ASsessments in Ankylosing Spondylitis; BASDAI, Bath AS Disease Activity Index; BASFI, Bath AS Functional Index; BASMI, Bath AS Metrology Index; BASRI, Bath AS Radiology Index; CRP, C reactive protein; DC-ART, disease controlling antirheumatic treatment; DFI, Dougados Functional Index; ESR, erythrocyte sedimentation rate; FDA, Food and Drug Administration; MISS, MR imaging in seronegative SpA; MRI, magnetic resonance imaging; NSAID, non-steroidal anti-inflammatory drug; OMERACT, Outcome Measure in Rheumatoid Arthritis Clinical Trials group; RA, rheumatoid arthritis; SASSS, Stoke AS Spinal Score; SF-36, Short Form-36; SM-ARD, symptom modifying antirheumatic drug; SpA, spondyloarthropathy; VAS, visual analogue scale

Table 1 Core set for studies on DC-ART, SM-ARDs/physical therapy, and clinical record keeping[2]

DC-ART	SM-ARDs/physical therapy	Clinical record keeping
Physical function	Physical function	Physical function
Pain	Pain	Pain
Spinal mobility	Spinal mobility	Spinal mobility
Spinal stiffness	Spinal stiffness	Spinal stiffness
Patient global assessment	Patient global assessment	Patient global assessment
Peripheral joints/entheses		Peripheral joints/entheses
Acute phase reactants		Acute phase reactants
Spine radiograph		
Hip radiograph		
Fatigue		

DC-ART, disease controlling antirheumatic treatment; SM-ARDs, symptom modifying antirheumatic drugs.

ASAS CORE SET OF DOMAINS

Treatment studies in AS have often employed inconsistent and excessive numbers of assessment methods, some of which are not validated. This can create multiple dilemmas. Establishing a uniform minimum core set of variables for inclusion in all research projects may help prevent these dilemmas by ensuring that:

- Chance occurrences of statistically significant differences between groups are minimised
- Investigators do not introduce bias by selectively publishing only favourable variables
- Comparisons can be made between studies
- Meta-analyses can be performed.[6]

The ASAS Working Group, composed of clinicians, researchers, industry representatives with expertise in AS, and patients with AS from more than 20 countries around the world, used a combination of expert consensus and statistical approaches to develop a core set of domains. The group defined core sets for three different settings:

- The evaluation of DC-ART
- The evaluation of SM-ARDs and physical function
- Use in clinical record keeping.

Table 1 lists the core set of domains for each of these settings. It should be noted that a number of other settings were initially considered based on all potential points of interest in AS. The three settings chosen reflect the most pertinent aspects while eliminating redundancy.[2]

The ASAS core sets for assessing outcome in AS were endorsed by the Outcome Measure in Rheumatoid Arthritis Clinical Trials group (OMERACT) and by the International League Against Rheumatism (ILAR) in 1998 and have been known since as the ASAS/OMERACT/ILAR core sets.[7] Validation for aspects of truth (Is the measure truthful? Does it measure what is intended? Is the result unbiased?) and discrimination (Does the measure discriminate between situations of interest?), set forth by the OMERACT filter, are continuing for several instruments.[8] However, results of the ASAS Working Group's combined expert consensus and statistical approach have shown considerable similarity with the results of the purely statistical approach of Calin *et al*[9] to the selection of a core set and to recommendations given by Bellamy[10] in his book on clinical metrology of musculoskeletal diseases. These similarities suggest the acceptability of the ASAS core set, although further validation is required.

ASAS SELECTION OF INSTRUMENTS

After determining the core set of domains for the previously mentioned settings, the ASAS Working Group proceeded to select specific instruments for each domain. One hundred and five instruments identified in the literature were evaluated for feasibility and relevance. Thirty five instruments were deemed not feasible or not relevant, or both, and were eliminated. The remaining 70 were ranked and discussed by members of the group to determine the final selection of instruments presented in table 2.[4]

Although the group selected only single instruments for each domain, combined indices such as the Bath AS Disease Activity Index (BASDAI), and the Bath AS Metrology Index (BASMI) need to be evaluated further. Overall, these combined indices scored high in percentage of feasibility and relevance. However, in some cases, single components within these indices scored low.[4] Moreover, a combined score similar

Table 2 Specific instruments for each domain in core sets for DC-ART, SM-ARDs/physical therapy, and clinical record keeping[4]

Domain	Instrument
Function	BASFI or Dougados Functional Index
Pain	VAS, past week, spine, at night, due to AS *and* VAS past week, spine, due to AS
Spinal mobility	Chest expansion *and* modified Schober *and* occiput-to-wall distance
Patient global	VAS, past week
Stiffness	Duration of morning stiffness, spine, past week
Peripheral joints and entheses	Number of swollen joints (44 joint count); currently no preferred instruments available for entheses
Acute phase reactants	ESR
Radiograph spine	AP + lateral lumbar and lateral cervical spine *and* X-pelvis (SI and hips)
Radiograph hips	See spine
Fatigue	Currently no preferred instruments

AP, anteroposterior; AS, ankylosing spondylitis; BASFI, Bath Ankylosing Spondylitis Functional Index; ESR, erythrocyte sedimentation rate; SI, sacroiliac; VAS, visual analogue scale.

Table 3 The 10 BASFI items (the patients must indicate their level of ability for each of the 10 activities on a 10 cm VAS) and the 20 items of the DFI (all begin with "Can you . . .") on a three point scale

BASFI (Questionnaire item No)	DFI (Questionnaire item No)
Items with similar content	
Putting on your socks or tights without help or aids (e.g., sock aid) (1)	Put on your shoes (1)
Bending forward from the waist to pick up a pen from the floor without an aid (2)	Bend over to pick up an object (11)
Getting up out of an armless dining room chair without using your hands or any other help (4)	Get up from a chair (9)
Getting up off the floor from lying on your back without help (5)	Turn in bed (14)
	Sleep on your back (16)
Standing unsupported for 10 minutes without discomfort (6)	Remain standing for 10 minutes (5)
Climbing 12–15 steps without using a handrail or walking aid (one foot for each step) (7)	Climb one flight of stairs (6)
Doing the full day's activities, whether it be at home or at work (10)	Do your job or housework (18)
Items with different content	
Reaching up to a high shelf without help or aids (e.g., helping hand) (3)	Pull on a pullover (3)
Looking over your shoulder without turning your body (8)	Get into a bathtub (4)
Doing physically demanding activities (e.g., physiotherapy, exercises, gardening, or sports) (9)	Run (7)
	Sit down (8)
	Get into a car (10)
	Crouch (12)
	Lie down (13)
	Sleep on your stomach (17)
	Cough or sneeze (19)
	Breathe deeply (20)

BASFI, Bath Ankylosing Spondylitis Functional Index; DFI, Dougados Functional Index; VAS, visual analogue scale.

to the disease activity score in rheumatoid arthritis (RA) is under development. The inclusion of different domains included in the ASAS core set, which are not included in other combined indices, might be an advantage. A description and rationale for the ASAS instruments within each domain are detailed in the following sections and ASAS instruments are compared with other widely used instruments.

Physical function

The Bath AS Functional Index (BASFI) and Dougados Functional Index (DFI) were selected for evaluation of function. The BASFI, published in 1994, includes 10 items on ability to perform and cope with activities of daily living (table 3). Each activity in this questionnaire is scored on a 10 cm visual analogue scale (VAS). The mean of the 10 scales yields the total score.[11] The DFI was first published in 1988 and consists of 20 Likert response items assessing the ability to perform distinct daily activities (table 3). A point score is assigned to each of the three possible answers (yes, with no difficulty=0; yes, but with difficulty=1; and no=2). The total score is calculated as a sum of the 20 item scores (range, 0–40) (table 3).[12] Although not published as an official modification of the DFI, a five point Likert scale has since replaced the original three point scale. The questions are scored as either 0, 0.5, 1, 1.5, or 2, so that the total range is the same but allows for greater detection of change.

Both self administered questionnaires have been shown to be valid and reliable measures of functional capacity in AS and are widely used. However, there are important differences between the two instruments, as well as limitations to their performance in specific situations. These are summarised in table 4.[11–13] Presently, the BASFI is the most widely used instrument to assess physical function in AS.

Pain

For the assessment of pain, two 10 cm VASs were selected: one for pain of the spine at night due to AS on average past week, and the other for pain of the spine (without time restraints)

due to AS on average past week. The BASDAI, which has been used in a number of studies, contains three VAS items relating to pain and discomfort over the past week. The BASDAI items assess three locations of pain: overall pain in neck, back, or hip; overall level of pain/swelling in joints other than the neck, back, or hip; and overall discomfort from any areas tender to touch or pressure (table 5).[14] The VAS on the peripheral joints combines pain and swelling; the last question is assumed to assess tenderness of the entheses. The swelling of the joints and the involvement of the entheses are assessed in separate domains in the ASAS core set. Although not included in the ASAS core set, there is also a pain measurement that is assessed by the doctor. This is, in fact, a combined assessment of tenderness on palpation in combination with limitation caused by pain and spasm. This gives a global grading from 0 to 4 for three areas in the spine: cervical, thoracic, and lumbar plus sacroiliac joints. Some think it is advantageous to have a doctor's assessment of pain in addition to various measures assessed by the patient.

Spinal mobility

The clinical relevance of spinal mobility in the assessment of AS depends upon two main factors: the stage of the disease and the method by which spinal mobility is measured. Thus, responses to questions about the relevance of assessing spinal mobility by ASAS members ranged widely, depending on how the questions were posed with respect to disease stage and method of measurement. Nevertheless, the majority of members agreed that over long periods of follow up, assessment of spinal mobility provides a sensitive measure of structural damage and disease activity. Chest expansion, modified Schober test, and occiput-to-wall distance were selected to represent the domain of spinal mobility. Whereas numerous modifications of the Schober test exist, the version recommended by the ASAS Working Group is performed as follows:

- With the patient standing erect, make a mark on the back at the midpoint on an imaginary line joining the posterior superior iliac spines.

Table 4 Differences and limitations of the BASFI and DFI based on literature review by Ruof and Stucki[13]

BASFI	DFI	Limitations
10 Items	20 Items	DFI requires more time to complete
Score based on 10 cm VAS	Score based on 3 point scale	DFI 3 point scale limits expression of change compared with BASFI (sensitivity has been enhanced for DFI with a modified 5 point scale)
7 Items similar to DFI, 3 additional items improve content validity and the spectrum of item difficulty	7 Items similar to BASFI, some items are redundant and some deal with symptoms rather than function	Construct and content validity of DFI inferior to those of the BASFI
Can discriminate the effects of a 3 week intensive physical therapy treatment period	Cannot discriminate the effects of a 3 week intensive physical therapy treatment period	Distribution of DFI scores shows tendency toward normal scores, which may inhibit further capture of improvement in patients with mild disability in physical therapy trials
		BASFI and DFI do not provide rules for handling missing data

BASFI, Bath Ankylosing Spondylitis Functional Index; DFI, Dougados Functional Index; VAS, visual analogue scale.

Table 5 BASDAI questionnaire items

Questions

(1) VAS overall level of fatigue/tiredness past week
(2) VAS overall level of AS neck, back, or hip pain past week
(3) VAS overall level of pain/swelling in joints other then neck, back, or hips past week
(4) VAS overall discomfort from any areas tender to touch or pressure past week
(5) VAS overall level of morning stiffness from time of awakening past week
(6) Duration of morning stiffness from time of awakening (up to 120 minutes)

BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; VAS, visual analogue scale.

- Make another mark 10 cm above the first.
- Ask the patient to bend forward maximally, keeping the knees fully extended.
- With the spine in fullest flexion, measure the distance between the two marks.
- The normal distance is greater than 15 cm owing to stretching of the skin overlying the mobile lumbar spine.[4]

The BASMI is a combined measure to assess spinal mobility and hip function, and consists of the following items:

- Tragus to wall
- Lumbar flexion
- Cervical rotation
- Lumbar side flexion
- Intermalleolar distance.[15]

Patient global
The patient global assessment, measured by a VAS on average for the past week, was selected. This differs from the Bath AS Global, which is a global assessment for one week and includes a VAS over six months.

Spinal stiffness
The selection for spinal stiffness was duration of morning stiffness of the spine past week. Two components of the BASDAI also assess spinal stiffness: overall level of morning stiffness from time of awakening past week and duration of morning stiffness from time of awakening past week (table 5).

The average of these two questions represents the stiffness component of the BASDAI. In the analyses to develop the ASAS improvement criteria, this combined measure for spinal stiffness performed better than information on the duration of stiffness alone.

Peripheral joints and entheses
The ASAS Working Group selected a joint count for evaluation of peripheral joints (without grading or weighting of the joints). The 44 joints included in the count are:

- Right and left sternoclavicular joints
- Acromioclavicular joints
- Shoulder joints
- Elbows
- Wrists
- Knees
- Ankles
- 10 Metacarpophalangeal joints
- 10 Proximal interphalangeal joints of the hands
- 10 Metatarsophalangeal joints.

Only one index for entheses, the Mander enthesis index, was found in the literature.[16] However, because members of the ASAS Working Group deemed this measure not feasible owing to the extensiveness of the instrument, no selection for entheses was made. One of the BASDAI questions deals with pain from entheses. Several modifications of the Mander index have been applied in recent trials. However, these modifications have not been validated. Currently, a simplified index to assess entheses is being developed and validated.

Acute phase reactants
Two recent evaluations of the validity of the erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP), by Ruof and Stucki[17] and Spoorenberg et al,[18] respectively, concluded that acute phase reactants do not comprehensively represent the disease process in AS and that their worth in AS clinical trials is limited. In contrast with RA, the ESR and CRP values are lower in AS and generally do not vary as much with respect to the severity of the disease. However, patients with AS with peripheral joint involvement and/or inflammatory bowel disease tend to have higher levels than patients with only axial involvement. Neither the ESR nor the CRP level had higher validity than the other. The advantages of lower costs, ease of

performance, standardised testing, and promptness of results led to the selection of the ESR by the ASAS Working Group to represent the acute phase reactant domain.[4 17 18] Interestingly, more recent evaluations suggest that acute phase reactants may reflect responsiveness in the evaluation of DC-ART. Thus, the perception of the usefulness of acute phase reactants in the assessment of AS might change in the near future with the availability of new information on the effect of biological treatments in AS.

Radiographs of spine and hips

The spine and sacroiliac joints are predominantly affected in AS. However, large axial joints, such as the hips and shoulders, and peripheral joints can also be involved. Plain radiographs of the cervical spine, lumbar spine, and pelvis (hip is included with the pelvis) were selected, although no evaluation for the thoracic spine and no scoring method have been recommended. In contrast with RA, for which radiographic change is an important end point and several validated scoring methods are available, radiographic change in AS is not a well established measure of outcome, and the choice of scoring method varies according to the type of study performed (for example, study of the natural history, prognosis, or effectiveness of treatment in AS).

The situation is complicated further by uncertainty about whether various radiographic abnormalities in AS represent the inflammatory disease process or whether they indicate healing. For example, fluffy periostitis or erosions are thought to represent inflammation, whereas bridging syndesmophytes are due to spinal ossification and are clinically asymptomatic. These are often present in late disease, and may be regarded as due to a non-inflammatory process.[19] Further elucidation of the relationship between inflammation and new bone formation in AS may provide answers. In the meantime, these and other issues need to be addressed before a scoring method can be chosen. Additional data obtained from other radiographic modalities (for example, computed tomography), ultrasonography, magnetic resonance imaging (MRI), scintigraphy, and dual energy x ray absorptiometry), some of which are discussed below, are becoming available, which may give more insight into the underlying pathophysiological processes in AS.[20]

Two scoring methods have been fully described in the literature to date: the Stoke AS Spinal Score (SASSS) published by Taylor et al in 1991[21] and the scoring method published by Kennedy et al in 1995,[22] which was subsequently modified in 1999 and named the Bath AS Radiology Index BASRI.[23] Both the SASSS (table 6) and the BASRI (table 7) demonstrated feasibility and good reproducibility.[20–26] In an unselected cohort of patients with AS, very little progression could be assessed by either method at a one and two year interval.[26] More data are needed, especially from patients with a high likelihood of progression.

Fatigue

No specific instrument was selected for the assessment of fatigue because the ASAS members judged none of the four measures identified in the literature to be relevant. However, recent research has demonstrated the validity of the BASDAI VAS item on overall level of fatigue, with good reliability and sensitivity to change in the assessment of overall fatigue (table 5).[27] If information is needed on several aspects of fatigue, the multifactorial fatigue index is a good alternative.[28] The Short Form (SF)-36 method has recently been used to assess fatigue and other elements of physical function in patients with AS.[29] Studies on the performance of other instruments to assess fatigue are in progress.

Table 6 SASSS scoring method

Description	Score
Sacroiliac joints	
Normal	0
Blurring of joint margin	1
1 + periarticular sclerosis or partial bony bridging or pseudowidening	2
2 + erosions or bony bridging	3
Complete ankylosis	4
Lumbar spine	
Erosion, sclerosis, or squaring	1
Syndesmophytes	2
Total bony bridging at each site	3
Composite score lumbar spine	72 (maximum)

SASSS, Stoke Ankylosing Spondylitis Spinal Score.

ASAS DEFINITION OF IMPROVEMENT WITH SYMPTOM MODIFYING AND DISEASE CONTROLLING ANTIRHEUMATIC DRUGS
Symptom modifying antirheumatic drugs

Therapeutic studies in AS to date have not shown any retardation of structural damage. Thus, it is not possible to say whether drugs can be considered truly disease modifying. In recognition of this fact, investigators in AS have used terms such as symptom modifying antirheumatic drugs (SM-ARDs) and disease controlling antirheumatic treatment (DC-ART). Standard criteria for defining improvement in SM-ARD evaluation were developed based on groundwork laid down by the ASAS Working Group for assessing outcome in AS and on outcome data from five randomised controlled trials with NSAIDs. Measures representing at least four of the five AS core domains (physical function, pain, spinal mobility, spinal stiffness/inflammation, and the patient global assessment) were included in each of the five trials.[12 30–33] Drawing from these data, relevant levels of change/improvement were defined within each of the five AS domains. A conceptual list of possible ways to define improvement in terms of some or all of the five core domains was then prepared. Clinical judgment, previous work in osteoarthritis and RA, and results of published clinical trials were applied to the development of the list, which consisted of 20 single item, multiple domain, and index definitions.[5]

The candidate definitions of improvement were tested and validated using χ^2 tests. The sensitivity and specificity of each definition for the identification of actively treated patients was determined using a random two thirds of the clinical trial data. From these, a range of good performing candidates for response criteria was selected based on high χ^2 test values and placebo response rates of 25% or less, and validated using the remaining one third of the data. Further validation was performed by examining the overlap between response rates and partial remission rates for each definition of response.

The resulting preliminary definition of short term improvement in AS incorporates four outcome domains (all scored on a scale of 0–100):

- Physical function (BASFI score)
- Pain (VAS pain score)
- Patient global assessment (VAS global assessment score)
- Inflammation (mean of the two morning stiffness related BASDAI VAS scores (first choice) or by morning stiffness duration with a maximum of 120 minutes (second choice)).

Table 7 BASRI scoring method

Grade		Description
Sacroiliac joints (NY Criteria)		
0	Normal	No change
I	Suspicious	Suspicious change
II	Minimal	Small localised areas with erosion or sclerosis without alteration in joint width
III	Moderate	Moderate or advanced sacroilitis with erosions, evidence of sclerosis, widening, narrowing, or partial ankylosing
IV	Severe	Severe abnormality with total ankylosis
Hip		
0	Normal	No change
I*	Suspicious	(Possible) focal joint space narrowing
II*	Minimal	Definite narrowing leaving a circumferential joint space >2 mm
III	Moderate	Narrowing with circumferential joint space ≤2 mm or bone-on-bone apposition of <2 cm
IV	Severe	Bone deformity or bone-on-bone apposition >2 cm
AP and lateral lumbar and lateral cervical spine		
0	Normal	No change
I	Suspicious	Suspicious but no definite change
II	Minimal	Any number of erosions, squaring, sclerosis ± syndesmophytes on ≤2 vertebrae
III	Moderate	Syndesmophytes on ≥3 vertebrae ± fusion involving 2 vertebrae
IV	Severe	Fusion involving ≥3 vertebrae
BASRI spine		Maximum=12
BASRI total		Maximum=16

*Increase grades I and II by one grade each if two of the following bony changes are present: erosions, osteophytes, protrusio.
AP, anteroposterior; BASRI, Bath Ankylosing Spondylitis Radiographic Index.
Note: BASRI spine is the sum of sacroiliac joints (combined as one score 0–4), AP and lateral lumbar spine combined as one score 0–4, and lateral cervical spine 0–4; BASRI total is the sum of BASRI spine and BASRI hips (as one score 0–4).

Improvement of AS is defined as improvement of ≥20%, and net improvement of ≥10 units, on a scale of 0 to 100 in each of three domains with no worsening of ≥20% and no worsening of ≥10 units in the fourth, whereas the reverse defines worsening of the disease. A definition of partial remission was also provided that requires the patient to have a low level of disease activity (that is, <20 units on a scale of 0–100 in each of the four domains).[5]

Arguments in favour of a single domain for defining improvement can be made because single domain definitions for improvement performed well in tests, and improvement in each of the four outcome domains tended to be consistent. However, a multiple domain definition is preferable owing to greater content validity and reliability.[5] As a follow up and for further validation, a Delphi exercise was held among the ASAS members to investigate the clinical relevance of the criteria for short term improvement. Those patients who fulfilled the ASAS improvement criteria had a clinically relevant improvement by the judgment of the members. However, there was also a group of patients who did not fulfil the criteria but showed a clinically relevant improvement according to the ASAS members, implying that the criteria were strict.[34]

Disease controlling antirheumatic treatments
A similar approach for defining improvement in DC-ART evaluation is currently being applied based on outcome data from trials with infliximab and etanercept. Owing to the difference in selection of domains for DC-ART it can be assumed that the domains included in the improvement criteria will also be different. More data will become available in the near future.

CLASSIFICATION
The classification of AS has gradually evolved as knowledge of the clinical features, natural history, genetic predisposition, and pathophysiology of the disease have unfolded. Disease classification is vital because it governs not only the development and implementation of clinical trials, as is the subject of this paper, but also guides many aspects of disease management and international communications.

Currently accepted classification systems consist of those that provide criteria for classifying the spondyloarthropathies (SpAs),[35 36] of which AS is a subtype, and those that provide criteria for classifying AS as a distinct disease entity.[37] Because patients with AS often have highly variable clinical presentations and outcomes, the current classification systems often lack sensitivity and specificity for certain populations and situations. For instance, considerable debate is currently taking place about the diagnostic and staging criteria for AS and about the overlapping relationship of these criteria with classification criteria and with outcome assessment in AS. This debate and a new proposal for disease staging in AS are discussed in "Staging of patients with ankylosing spondylitis: a preliminary proposal" within this supplement (p iii19).

The results of a questionnaire sent to 30 international experts from countries in Europe and North and Central America surveying the experts' opinions on nomenclature, disease classification, and study design for future trials illustrate the widespread discrepancies in the terminology and criteria for diagnosis and classification of AS used by rheumatologists (see "Building consensus on nomenclature and disease classification for ankylosing spondylitis" within this supplement (p iii61)). The discrepancies underline the need for further discussion and resolution. Experts are currently searching for solutions in histological, immunological, and imaging studies of AS.

MRI for diagnosis and as outcome
Imaging studies with MRI may also provide helpful and sensitive classification criteria and clues as to where to look for key pathogenic mechanisms in AS. In diagnosis, MRI has been used to visualise acute sacroiliitis, spondylitis, and spondylodiscitis in patients with SpA in recent studies.[38–40] In contrast

with radiographic imaging, which may take several years to detect the effects of inflammation, bony changes, and ankylosis, MRI detects acute inflammation of the enthesis, bone, and synovium in addition to bony changes and ankylosis. This is best detected when applying contrast with gadolinium-DPTA or fat suppression techniques.

In relation to pathogenic concepts, MRI studies have shown that the earliest lesion in the sacroiliac joint in AS is subchondral osteitis,[41] and that human leucocyte antigen (HLA)-B27 determines the severity of osteitis. MRI studies have also shown that osteitis or enthesitis is common at all diseased sites, including synovial joints, suggesting common unifying mechanisms for disease at disparate sites. Importantly, the reported MRI abnormalities have been validated as being representative of acute inflammatory processes within the bone.[42]

MRI has recently been used to investigate the impact of anti-tumour necrosis factor α treatment on axial, peripheral, and entheseal disease in patients with AS. In a study by Stone *et al*, improvement in axial MRI was seen in seven of eight patients after the first two infusions with infliximab.[43] In an open label trial of etanercept in SpA,[44] scoring was performed on paired MRI scans of entheseal lesions, osteitis in the sacroiliac joints, lumbar and cervical spine, and peripheral joints at baseline and week 24. Thirty eight of 44 lesions (86%) either resolved completely or improved after treatment, and no new lesions developed.

Grade II radiographic evidence of sacroiliitis, a hallmark of the SpAs, is currently a criterion for the diagnosis of AS based on the modified New York criteria.[45] However, the difficulty of detecting early inflammation may be resolved with the use of MRI, and therefore improve current classification by providing earlier diagnosis and more accurate staging of AS. Dynamic MRI with gadolinium-DTPA enhancement has been shown to measure effectively the inflammatory process and its consequences. Bollow *et al* compared this method with inflammation quantified by cellular analysis of immunostained sacroiliac biopsy specimens, and found good correlation, showing that T cells and macrophages are frequent cells of early and active sacroiliitis in the SpAs.[46]

Realising the need for early diagnosis and intervention in SpA and for better methods of assessing bone abnormalities, an international multicentre collaborative of researchers has initiated the MR imaging in seronegative SpA (MISS) initiative to develop a scoring system for spinal disease and to study the use of MRI as an outcome measure.

As alluded to previously, radiographic scoring methods for AS and other SpAs are wrought with methodological issues, such as which abnormalities to score, which sites to include in the scoring method, which radiographic views to use, what order to score, and how to handle interobserver and/or intraobserver variation.[20 47] Many of these MRI related issues will be examined by the MISS initiative. However, the overall advantages of MRI, including detection of early inflammation, better visualisation of lesions, especially cartilage and enthesitis, make it a useful assessment tool.[7] Furthermore, the MRI changes in the bone are often quite marked, which may facilitate scoring. For example, κ values >0.8 for scoring plantar fasciitis have been reported.[48]

DESIGN AND CONDUCT OF FUTURE TREATMENT TRIALS

Recent promising results in preliminary AS trials with infliximab and etanercept have raised hopes that symptom modifying, disease controlling, and disability reducing indications can be obtained for these treatments. Future trials must be designed to determine whether such claims can be made. Trial duration, patient population, efficacy end points, assessment methods, and data analyses are among the many issues that will need to be considered. Although the Food and Drug Administration (FDA) has not, as yet, endorsed any specific outcome measures for trials in AS, the contributions of the ASAS Working Group and others towards developing comprehensive standardised outcome assessments and classification systems will help guide the design and implementation of future trials. Conversely, new measures and classification systems may be validated by the clinical data obtained in future trials.

The United States FDA, in collaboration with members of the academic community, pharmaceutical industry, and the public, has, from time to time, issued guidance documents to define the claims that sponsors can receive for their products and the clinical trial data that would be expected to be submitted to support such claims. Although a guidance document has not yet been developed for AS products by the FDA or by any other government or private entity within or outside the US, a guidance document on the clinical development of programmes for drugs, devices, and biological products for the treatment of RA was issued in 1999 by the US Department of Health and Human Services, FDA, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), and Center for Devices and Radiological Health (CDRH).[49] A number of issues were addressed in the guidance document which may have application for AS. However, it must be emphasised that RA and AS are distinct disease entities for which direct associations and comparisons cannot always be made.

FDA GUIDANCE FOR CLINICAL TRIALS IN RA

The FDA guidance document for treatment trials in RA identifies six claims that can be considered for therapeutic agents based on a wide range of potentially achievable outcomes:

- Reduction in signs and symptoms of RA
- Major clinical response
- Complete clinical response
- Remission
- Prevention of disability
- Prevention of structural damage.

Although relief of signs and symptoms has been the central therapeutic effect of therapeutic drugs marketed for RA since approximately 1997, the addition of claims for major clinical response, complete clinical response, and remission, helps to distinguish products further by providing a means for demonstrating patient benefit of greater magnitude than is needed for a claim of symptomatic relief. The claim for prevention of structural damage does not in itself define a patient benefit and therefore should be combined with one of the other claims. The claim for prevention of disability reflects the potential for long term benefits in the course of the disease.

Table 8 lists the duration and end points for the six claims suggested in the guidance document. Trials of at least six months' to two years' duration are recommended. These relatively lengthy trials are desirable for a number of reasons:

- RA is a disease of long duration
- Interventions that provide only short term benefit are less valuable to patients than those that provide long term benefit
- Products with the potential to elicit antibody formation should be assessed for durability because antibodies may develop to block effectiveness.

A trial of three months' duration for the claim of reduction in signs and symptoms, however, is acceptable for products belonging to an already established pharmacological class (that is, NSAIDs). Regardless of the trial duration, methods that evaluate response over time are preferable to methods that incorporate only the baseline value and the final observation.

Table 8 FDA guidance on design and conduct for RA trials

Claim	Duration	End point(s)
Reduction in signs and symptoms of RA	≥6 Months or 3 months if the product belongs to an already well characterised pharmacological class (e.g., NSAIDs)	Composite end points, such as ACR 20 or Paulus criteria (1990), or well accepted signs/symptoms measures (i.e., joint pain, tenderness, and swelling, using a 66 or 28 joint count, and global assessments
Major clinical response	≥7 Months	ACR 70 for 6 consecutive months
Complete clinical response	≥1 Year	Remission by ACR criteria and radiographic arrest by Larsen or modified Sharp method for 6 consecutive months
Remission	≥1 Year	Remission by ACR criteria and radiographic arrest by Larsen or modified Sharp method for >6 consecutive months while not receiving any antirheumatic treatment
Prevention of disability	2–5 Years	HAQ or AIMS and SF-36 in addition to demonstration of improvement of signs and symptoms
Prevention of structural damage	1 Year	Slowing x ray progression using either Larsen or Modified Sharp, or other validated radiographic index in addition to one of the other claims

AIMS, Arthritis Impact Measurement Scales; ACR, American College of Rheumatology; FDA, Food and Drug Administration; HAQ, Health Assessment Questionnaire; NSAID, non-steroidal anti-inflammatory drug; RA, rheumatoid arthritis.

In addition to identifying potential claims for RA products and advising on the duration and efficacy end points for trials to support various claims, the guidance document provides detailed discussions of special considerations for phase I–IV trials, safety analyses, biological products, medical devices, and juvenile RA. For efficacy trials (phase III work), the guidance document considers global considerations, including patient selection, concomitant antirheumatic treatment and other concomitant treatments, stratification, blinding, and effects of dropouts and non-compliance; trial design considerations, including design of superiority, equivalence, and novel trials; and analytical issues, including handling dropouts, comparison with baseline outcome measures, and statistical considerations.

For biological products, characteristics and issues unique to biological agents as compared with traditional drugs are detailed, such as species specificity, toxicity response, and product homogeneity. In addition, the role of antibodies is also discussed. For instance, homogeneity of a biological agent often has a critical role in determining the activity and toxicity of a compound. Thus, biological agents should demonstrate consistency from lot to lot while under development and should be well characterised in order to be appropriately evaluated.

The comprehensive FDA guidance document for treatment trials in RA may serve as a model for AS by helping to identify key issues in the development of therapeutic drugs. Although some overlap exists between RA and AS, and a number of lessons can be learnt from experience in RA, issues unique to AS must also be identified and all issues must be addressed specifically for patients with AS. An FDA guidance document for AS is currently under development.

CONCLUSIONS
Until recently, treatment options for patients with AS have been limited. Despite the relatively high prevalence of the disease, no disease modifying agents have been available as they have been for RA. Promising preliminary results from studies with infliximab and etanercept have raised hopes that anti-tumour necrosis factor α agents may provide symptom control and disease modification for patients. However, universal standards for assessing treatment response and indications in AS are key to an evaluation of these agents. A number of challenges to standardisation have been identified and addressed. However, much work still needs to be done.

. . . .

Authors' affiliations
D van der Heijde, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, Maastricht, The Netherlands
J Braun, Rheumazentrum Ruhrgebiet, Herne, Germany, Department of Gastroenterology and Rheumatology, Free University, Berlin, Germany
D McGonagle, Department of Rheumatology, The University of Leeds, Leeds, UK
J Siegel, Center for Biologics Evaluation and Research, Food and Drug Administration, Rockville, Maryland, USA

REFERENCES
1 **Bakker C**, Boers M, van der Linden S. Measures to assess ankylosing spondylitis: taxonomy, review, and recommendations. J Rheumatol 1993;20:1724–30.
2 **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.
3 **van der Linden**, van der Heijde DM. Ankylosing spondylitis and other B27 related spondylarthropathies. Baillieres Clin Rheumatol 1995;9:355–73.
4 **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.
5 **Anderson JJ**, Baron G, van der Heijde D, Felson DT, Dougados M. Ankylosing spondylitis assessment group preliminary definition of short-term improvement in ankylosing spondylitis. Arthritis Rheum 2001;44:1876–86.

6 **van der Heijde DM**, van der Linden S. Measures of outcome in ankylosing spondylitis and other spondyloarthritides. Baillieres Clin Rheumatol 1998;12:683–93.

7 **OMERACT IV Conference**. Cancun, Mexico. J Rheumatal Suppl, 1998.

8 **Boers M**, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology [editorial]. J Rheumatol 1998;25:198–9.

9 **Calin A**, Nakache JP, Gueguen A, Zeidler H, Mielants H, Dougados M. Outcome variables in ankylosing spondylitis: evaluation of their relevance and discriminant capacity. J Rheumatol 1999;26:975–9.

10 **Bellamy N**. Musculoskeletal clinical metrology. Dordrecht: Kluwer, 1993:177–83.

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 **Dougados M**, Gueguen A, Nakache JP, Nguyen M, Mery C, Amor B. Evaluation of a functional index and articular index in ankylosing spondylitis. J Rheumatol 1988;15:302–7.

13 **Ruof J**, Stucki G. Comparison of the Dougados Functional Index and the Bath Ankylosing Spondylitis Functional Index. A literature review. J Rheumatol 1999;26:955–60.

14 **Garrett S**, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis disease activity index. J Rheumatol 1994;21:2286–91.

15 **Kennedy LG**, Jenkinson TR, Mallorie PA, Whitelock HC, Garrett SL, Calin A. Ankylosing spondylitis: the correlation between a new metrology score and radiology. Br J Rheumatol 1995;34:767–70.

16 **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.

17 **Ruof J**, Stucki G. Validity aspects of erythrocyte sedimentation rate and C-reactive protein in ankylosing spondylitis: a literature review. J Rheumatol 1999;26:966–70.

18 **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.

19 **Ball J**. The enthesopathy of ankylosing spondylitis. Br J Rheumatol 1983;22(suppl 2):25–8.

20 **van der Heijde D**, Spoonenberg A. Plain radiographs as an outcome measure in ankylosing spondylitis. J Rheumatol 1999;26:985–7.

21 **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.

22 **Kennedy LG**, Jenkinson TR, Mallorie PA, Whitelock HC, Garrett SL, Calin A. Ankylosing spondylitis: the correlation between a new metrology score and radiology. Br J Rheumatol 1995;34:767–70.

23 **Calin A**, Mackay K, Brophy S. A new dimension to outcome: application of the Bath Ankylosing Spondylitis Radiology Index. J Rheumatol 1999;26:988–92.

24 **Dawes PT**. Stoke Ankylosing Spondylitis Spinal Score. J Rheumatol 1999;26:993–6.

25 **Averns HL**, Oxtoby J, Taylor HG, Jones PW, Dziedzic K, Dawes PT. Radiological outcome in ankylosing spondylitis: use of the Stoke Ankylosing Spondylitis Spine Score (SASSS). Br J Rheumatol 1996;35:373–6.

26 **Spoorenberg A**, de Vlam K, van der Heijde D, de Klerk E, Dougados M, Mielants H, et al. Radiological scoring methods in ankylosing spondylitis: reliability and sensitivity to change over one year. J Rheumatol 1999;26:997–1002.

27 **van Tubergen A**, Coenen J, Landewé R, Spoorenberg A, Chorus A, Boonen A, et al. Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis. Arthritis Rheum 2002;47:8–16.

28 **Smets EM**, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995;39:315–25.

29 **Braun J**, Brandt J, Listing J, Zink A, Alten R, Golder W, et al. Treatment of active ankylosing spondylitis with infliximab—a randomised controlled multicentre trial. Lancet 2002;359:1187–93.

30 **Dougados M**, Caporal R, Doury R, Thiese A, Pattin S, Laffez B, et al. A double blind crossover placebo controlled trial of ximoprofen in AS. J Rheumatol 1989;16:1167–9.

31 **Dougados M**, Nguyen M, Caporal R, Legeais J, Bouxin-Sauzet A, Pellegrini-Guegoault B, et al. Ximoprofen in ankylosing spondylitis: a double blind placebo controlled dose ranging study. Scand J Rheumatol 1994;23:243–8.

32 **Dougados M**, Gueguen A, Nakache J-P, Velicitat P, Veys EM, Zeidler H, et al. Ankylosing spondylitis: what is the optimum duration of a clinical study? One year versus a 6 weeks nonsteroidal anti-inflammatory drug trial. Rheumatology (Oxford) 1999;38:235–44.

33 **Dougados M**, Behier J-M, 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 anti-inflammatory drug. Arthritis Rheum 2001;44:180–5.

34 **van Tubergen A**, van der Heijde D, Anderson J, Landewé R, Dougados M, Braun J, et al and van der Linden S for the ASAS Working Group. Comparison of statistically derived ASAS improvement criteria for ankylosing spondylitis with clinically relevant improvement according to an expert panel. Ann Rheum Dis (in press).

35 **Dougados M**, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al. The European Spondyloarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum 1991;341:1218–26.

36 **Amor B**, Dougados M, Mijijawa M. Criteres de classification des spondyloarthorpathies. Rev Rhum 1990;57:85–9.

37 **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.

38 **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 spondyloarthropathy patients. Arthritis Rheum 1994;37:1039–45.

39 **Braun J**, Bollow M, Sieper J. Radiology and pathology of the spondyloarthropathies. Rheum Dis Clin North Am 1998;24:697–735.

40 **Bresnihan B**, Alvaro-Gracia JM, Cobby M, Doherty M, Domljan Z, Emery P, et al. Treatment of rheumatoid arthritis with recombinant human interleukin-1 receptor antagonist. Arthritis Rheum 1998;41:2196–204.

41 **Ahlstrom D**, Feltelius N, Nyman R, Hallgren R. Magnetic resonance imaging of sacroiliac joint inflammation. Arthritis Rheum 1990;33:1763–9.

42 **Bollow M**, Fischer T, Reisshauer H, Backhaus M, Sieper J, Hamm B, et al. Quantitative analyses of sacroiliac biopsies in spondyloarthropathies: T cells and macrophages predominate in early and active sacroiliitis—cellularity correlates with the degree of enhancement detected by magnetic resonance imaging. Annals Rheum Dis 2000;59:135–40.

43 **Stone M**, Salonen D, Lax M, Payne U, Lapp V, Inman R. Clinical and imaging correlates of response to treatment with infliximab in patients with ankylosing spondylitis. J Rheumatol 2001;28:1605–14.

44 **Marzo-Ortega H**, McGonagle D, O'Connor P, Emery P. Efficacy of etanercept in the treatment of entheseal pathology in resistant spondylarthropathy. Arthritis Rheum 2001;44:2112–17.

45 **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.

46 **Bollow M**, Fischer T, Reisshauer H, Backhaus M, Sieper J, Hamm B, et al. Quantitative analyses of sacroiliac biopsies in spondyloarthropathies: T cells and macrophages predominate in early and active sacroiliitis—cellularity correlates with the degree of enhancement detected by magnetic resonance imaging. Ann Rheum Dis 2000;59:135–40.

47 **van der Heijde D**, Boers M, Lassere M. Methodological issues in radiographic scoring methods in rheumatoid arthritis. J Rheumatol 1999;26:726–30.

48 **McGonagle D**, Marzo-Ortega H, O'Connor P, Gibbon W, Pease C, Reece R, et al. The role of biomechanical factors and HLA-B27 in magnetic resonance imaging-determined bone changes in plantar fascia enthesopathy. Arthritis Rheum 2002;46:489–93.

49 **United States Food and Drug Administration**. Clinical development programs for drugs, devices, and biological products for the treatment of rheumatoid arthritis (RA). Food and Drug Administration website. Available at: http://www.fda.gov/cder/guidance/1208fnl.htm. Accessed 1 November 2001.