Raising the bar for classification and outcome assessment for clinical studies in axial spondyloarthritis

Boel, A.

RAISING THE BAR FOR CLASSIFICATION AND OUTCOME ASSESSMENT FOR CLINICAL STUDIES IN AXIAL SPONDYLOARTHRITIS

Anne Boel

Raising the bar for classification and outcome assessment for clinical studies in axial spondyloarthritis

Anne Boel

ISBN: 978-94-6419-554-5

Cover Photo: Adobe Stock
Lay-out: Ilse Modder, www.ilsemodder.nl
Printing: Gildeprint, www.gildeprint.nl

The printing of this thesis was financially supported by UCB Pharma which is gratefully acknowledged.

Raising the bar for classification and outcome assessment for clinical studies in axial spondyloarthritis

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl
volgens besluit van het college voor promoties
te verdedigen op
dinsdag 18 oktober 2022
klokke 15.00 uur

door

Anne Boel
geboren te Terheijden
in 1992

Promotor
prof. dr. D.M.F.M. van der Heijde

Co-promotores
dr. F.A. van Gaalen
dr. M.V. Navarro-Compán — Hospital La Paz, Madrid

Leden promotiecommissie:
Prof. dr. T.W.J. Huizinga
Dr. M.S. Ramiro
Prof. Dr. M. Boers — Amsterdam UMC, locatie VUmc
Prof. Dr. A.M. van Tubergen — Maastricht UMC
Prof. Dr. G. Kloppenburg

A room without books is like a body without a soul

- Marcus Tallius Cicero-

CONTENT

CHAPTER 1

GENERAL INTRODUCTION

AXIAL SPONDYLOARTHRITIS

Axial spondyloarthritis (axSpA) is a chronic inflammatory rheumatic disease which usually starts in the second or third decade of life and is characterised by chronic back pain (present for at least 3 months) and spinal stiffness[1-3]. Contrary to other rheumatic diseases that are characterised by bone loss, axSpA is characterized by new bone formation, resulting in bone fusion and sclerosis of the sacroiliac joints and spine. The fusion of the sacroiliac joints and/or spinal vertebra contribute to limitations in mobility and physical function, further affecting many activities of daily living[4,5]. As axSpA usually occurs at a relatively young age, patients have to adjust to their disease for most of their lives[6]. Alongside pain and stiffness, many patients experience fatigue and sleep problems, all of which have a major impact on quality of life and their ability to partake in day-to-day activities, such as the ability to remain employed, conduct domestic work and participate in leisure activities[4,5]. Consequently, quality of life in patients with axSpA is reduced compared to the general population[5,7].

In addition to the characteristic spinal complaints, there are several other clinical features that are common among patients with axSpA, the so called spondyloarthritis (SpA) features[1,8]. Information on the following three SpA features can be collected by taking history of the patient: inflammatory back pain (IBP), a good response to non-steroidal anti-inflammatory drugs (NSAIDs) and a positive family history of SpA. IBP is considered if at least four of the following five parameters are present: 1) age at onset before the age of 40; 2) insidious onset; 3) improvement with exercise; 4) no improvement with rest; and 5) night pain with improvement upon getting up[9]. A good response to NSAIDs is reflected by a significant reduction in, or complete absence of back pain in the 24-48 hours after taking a full dose of NSAIDs[8]. Lastly, a positive family history of SpA is present in case of a family history of axSpA, psoriasis, reactive arthritis, uveitis, or inflammatory bowel disease (IBD) in a first-degree (i.e. parents, siblings and children) or second-degree relative (i.e. grandparents, aunt, uncle, niece, and nephew)[10]. Clinical examination can provide insight in peripheral manifestations, including (peripheral) arthritis, enthesitis and dactylitis. Peripheral arthritis can be present in any of the peripheral joints, but there is a preference for asymmetrical involvement of joints of the lower limbs such as the knee. Enthesitis is inflammation at the site of the insertion of the tendon, ligaments, or capsule into bone; the most common enthesitis is heel enthesitis[2,8]. Arthritis and enthesitis are the most common peripheral manifestations and found in ~30–50% of patients axSpA[2,11,12]. Dactylitis is swelling of an entire digit -finger or toe- and much less prevalent than arthritis or enthesitis (6-8% prevalence)[2,11,12]. Additional features are the so-called extra-musculoskeletal manifestations, which include psoriasis, IBD and uveitis. Uveitis is the most frequent extra-musculoskeletal manifestation and occurs in approximately 20-30% of patients with axSpA[12-14]. Uveitis is inflammation of the uveal tract (the middle

layer of the eye) and typically presents as uveitis anterior in axSpA, which is often of short duration, acute in onset, occurs unilaterally, and frequently alternates between eyes[1,2,13]. IBD includes both Crohn's disease and ulcerative colitis, both have a chronic character and are characterised by inflammation of the digestive tract. Prevalence of IBD among patients with axSpA ranges from ~5-10%[12,14,15]. Psoriasis is characterised by red, dry, thick, and raised patches on the skin, which are often covered with a silvery-white coating called scale, and they tend to itch. Prevalence is estimated to be ~10-20% in patients with axSpA[12,14,15]. For the three peripheral and three extra-musculoskeletal SpA features it is assessed whether they are currently present or were present in the past and if the diagnosis was confirmed by a physician[8]. Laboratory tests are used to evaluate whether acute phase reactants (i.e. C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR)) are elevated -which occurs in approximately 50–60% of patients with r-axSpA and 30–40% of patients with nr-axSpA[16]-, and to establish presence/absence of Human Leukocyte Antigen B27 (HLA-B27). Prolonged high levels of disease activity due to inflammation can results in irreversible structural damage to the sacroiliac joints and spine of patients with axSpA[16-20].

Finally, as axSpA affects the sacroiliac joints in most patients, imaging of sacroiliac joints has a pivotal role in diagnosis and classification of axSpA. Radiographs and Magnetic Resonance Imaging (MRI) are the most commonly used imaging techniques in clinical practice[8]. However, there are limitations to the use of radiographs of the sacroiliac joints in patients with early disease, because structural changes generally take years to occur[19]. Therefore, if radiographs are normal or ambiguous, MRI of the sacroiliac joints can provide valuable information, as it allows for the identification of active inflammation (i.e. presence of bone marrow oedema in subchondral bone), as well as the presence of post-inflammatory structural changes (i.e. erosions, sclerosis, and fatty lesions). Both imaging modalities are used to assess sacroiliitis, but use a different definition. Sacroiliitis on radiographs is defined as bilateral grade 2–4 or unilateral grade 3–4, according to the modified New York criteria[21], which represents irreversible structural damage to the sacroiliac joints, whereas sacroiliitis on MRI as a SpA feature is defined as active inflammatory lesions of the sacroiliac joints with definite bone marrow oedema/osteitis suggestive of sacroiliitis associated with SpA[8]. For making a diagnosis also the structural abnormalities on MRI are important. All SpA features are very useful in diagnosis of axSpA as well as classification of patients for clinical trials. Furthermore, these features can provide important information regarding disease prognosis[14].

There are two major subtypes of axSpA[1]: 1) radiographic axSpA (r-axSpA, also known as Ankylosing Spondylitis (AS)), characterised by substantial structural damage to the sacroiliac joints visible on radiographs; and 2) non-radiographic axSpA (nr-axSpA), characterised by clinical symptoms of axSpA in absence of definite sacroiliitis visible on radiographs. Nr-axSpA is often considered an early stage of the disease, which implies patients can progress from nr-axSpA to r-axSpA dependent on risk factors, such as male sex, HLA-B27 positivity, high inflammatory activity (i.e. elevated CRP or inflammation visible on MRI), and smoking status[3,17-19,22-25]. However, progression from nr-axSpA to r-axSpA occurs in approximately 5-20% of patients in a time-period of 2-5 years[17,19,26-29], whereas a proportion of patients may never develop radiographic sacroiliitis and thus never progress to r-axSpA[23,28], emphasising nr-axSpA is more than an early stage of disease, it is also an disease-expression[22,30,31]. Nevertheless, complaints and disease activity of the patients with nr-axSpA have been reported to be equally severe and limiting as those from patients with r-axSpA[32,33].

INTERNATIONAL CHARACTERISATION OF AXIAL SPONDYLOARTHRITIS

Ever since it was recognised that axSpA is in fact a spectrum of disease rather than only AS, there was no longer one single feature (i.e. radiographic damage to the sacroiliac joints) that was present in all patients with axSpA. Therefore, axSpA is a good example of a disease that lacks pathognomonic symptoms and signs, and in particular, specific serological or immunological biomarkers[30]. Disease features are hardly ever identical among patients, hence a clinical diagnosis of axSpA requires careful consideration and exclusion of differential diagnoses as well as pattern recognition using clinical, laboratory, and imaging findings characteristic of axSpA by an experienced rheumatologist[22,31,34]. Furthermore, geographic disease prevalence and the clinical setting affect pre-test probability to make a diagnosis[35]. Thus, a complex multistep process using expert opinion is required to make a diagnosis that cannot be captured by counting features or ticking boxes[34-36]. This is why there are no diagnostic criteria for axSpA and it is currently unlikely that they will ever be developed[36].

Contrary, classification criteria are primarily intended to create well-defined, relatively homogeneous groups of patients for clinical research and validated classification criteria are critical to the interpretation of study findings and comparisons of results between studies[36]. Classification criteria do not capture the whole spectrum of manifestations of a disease, but should be highly specific in order to minimize false-positive errors (i.e. incorrectly labelled as having a disease)[31]. As rheumatic diseases are heterogeneous in nature, classification criteria would fail to identify some patients with axSpA.

This is due to the fact that classification criteria are aimed at a more homogeneous population and a narrower range of disease severity than that seen in routine clinical practice, thus classification criteria should not be used to diagnose patients, but solely to include patients in clinical studies[3,36].

Prior to the currently employed Assessment in SpondyloArthritis international Society (ASAS) classification criteria for axSpA, the most well-known and widely used classification criteria were the 1984 modified New York (mNY) criteria[21], which are used to classify patients with r-axSpA (figure 1).

Modified New York Criteria:

Clinical criteria (need at least one of the three):

1. Low back pain and stiffness >3 months which improves with exercise, but is not relieved by rest

2. Limitation of lumbar motion in sagittal and frontal planes

3. Limitation of chest expansion relative to normal values corrected for age and sex

PLUS

Radiographic criterion

Sacroiliitis grade 2 bilaterally, or sacroiliitis grade 3-4 unilaterally, or bilaterally

Figure 1 Modified New York criteria for ankylosing spondylitis by van der Linden et al. (1984)[21]

The most prominent feature in the mNY criteria is the definition for radiographic sacroiliitis, which is used as the working definition for sacroiliitis to this day. However, the mNY criteria do not allow identification of patients with axSpA early in the course of the disease when radiographic changes in the sacroiliac joints-which as described earlier can take years to manifest- are not yet present[1,2]. Furthermore, radiographic damage reflects the consequences of inflammation, rather than inflammation itself[37,38]. This is why MRI was included in the new set of classification criteria in 2009 by ASAS[37,39]. In the ASAS criteria, the radiographic criterion remains unchanged compared to the mNY criteria, but is complemented with the presence of sacroiliitis on MRI. Patients with sacroiliitis on either MRI or radiographs and at least one other SpA feature fulfil the so-called imaging arm (figure 2, left panel). Including sacroiliitis on MRI in addition to sacroiliitis on radiographs allows for classification of patients with early disease as well as established axSpA, and subsequently for the inclusion of these patients in clinical trials investigating the efficacy and safety of treatments[37,39].

Patients with r-axSpA may be classified using the mNY criteria, or the more recent ASAS criteria. Both the mNY and the ASAS axSpA classification criteria use an identical radiographic criterion (as shown in figures 1 and 2). However, the additionally required (clinical) features of the mNY and ASAS classification criteria differ. Patients without the inflammatory character of back pain fulfil the ASAS criteria if another SpA feature is present, but only fulfil the mNY criteria if spinal mobility is limited. In this thesis we look into the differences and similarities between the two criteria sets, and assess whether both classify the same patients with axSpA if radiographic sacroiliitis is present.

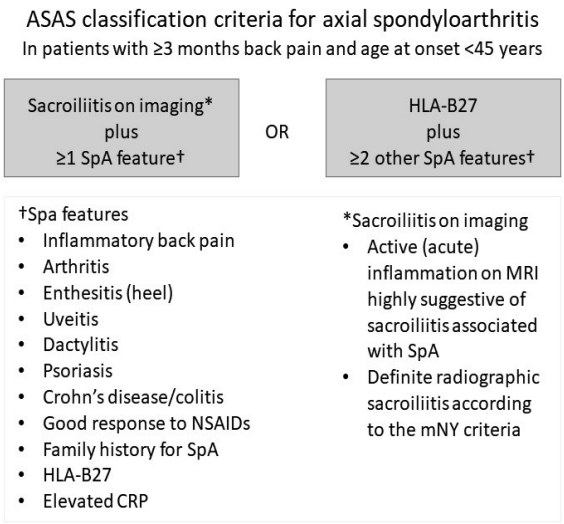


Figure 2 ASAS classification criteria for axial spondyloarthritis by Rudwaleit et al. (2009)[10]
ASAS, Assessment in SpondyloArthritis international Society; CRP, C-reactive protein; HLA-B27, Human Leukocyte Antigen B27; MRI, Magnetic Resonance Imaging; NSAIDs, non-steroidal anti-inflammatory drugs; SpA, spondyloarthritis

In case of absence of sacroiliitis on imaging, the presence of HLA-B27 is required to fulfil the ASAS classification criteria, which represents the so-called clinical arm. The clinical arm was originally intended for situations when imaging is not available (e.g. in large epidemiological studies), and patients are classified using the clinical arm if they are HLA-B27 positive and have at least 2 other SpA features (figure 2, right panel)[3]. The prominent role of HLA-B27 is understandable when one considers that the prevalence of axSpA, ranging between 0.3% and 1.4% is linked to the prevalence of HLA-B27 in a given population[40,41]. Both the prevalence of axSpA[40] and HLA-B27[42] vary considerably throughout the world. Furthermore, the association between HLA-B27 and axSpA varies between races and different subtypes of HLA-B27 are found in different parts of the world (e.g. HLA-B*27:05 and HLA-B*27:09 in Europe and HLA-B*27:04 and HLA-B*27:06 in Asia)[42]. The vast majority (>80%) of patients with r-axSpA is HLA-B27 positive, and this

percentage is only slightly lower in nr-axSpA[22,32,33,43]. HLA-B27 has been related to an earlier age at symptom onset[16,44,45] and better disease prognosis with appropriate treatment[46-48], yet also to an increased likelihood of developing radiographic damage[29,45,49].

There is a genetic link between HLA-B27 carriership and a positive family history of (ax) SpA, as HLA-B27 positive first-degree relatives of HLA-B27 positive patients with axSpA are more likely to develop axSpA than HLA-B27 positive individuals in the general population[50]. Furthermore, the risk of developing axSpA in HLA-B27 positive first-degree relatives of patients with axSpA is approximately one-in-five whereas the risk in HLA-B27 negative relatives is very low (≤1%)[51]. Hence, a positive family history of SpA can be useful in identifying patients with axSpA[52].

In recent years, the value of a positive family history has been questioned[3]. Research has shown that its diagnostic value is limited once HLA-B27 status is known, and the value of a positive family history is probably restricted to identifying chronic back pain patients that might be HLA-B27 positive[52]. Furthermore, studies investigating the association between HLA-B27 and all individual components in (ax)SpA cohorts suggest the association is driven by a positive family history of axSpA and possibly uveitis, but not by other forms of SpA[53,54]. This might leave one to wonder whether a positive family history is overvalued in the classification criteria, as HLA-B27 and a positive family history have an equal weight, and the definition might be too broad by including all five diseases. Nonetheless, it is important to keep in mind that a family history is easily accessible and can provide valuable information in identifying patients suspected of axSpA who first present to the general practitioner with chronic back pain complaints.

Another valuable and easily accessible piece of information in identifying chronic back pain patients suspected of axSpA is the age at onset of back pain complaints[55]. Multiple studies showed that the vast majority of patients with axSpA develop symptoms before the age of 45 years[44,56,57] (figure 3), thereby further emphasizing the importance of the age at symptom onset. These findings provided the basis to include the age criterion in definitions for inflammatory back pain[58,59] and also as an entry criterion in the ASAS criteria[37,39]. However, the majority of the data on which the age at onset criterion was based originates from Western Europe.

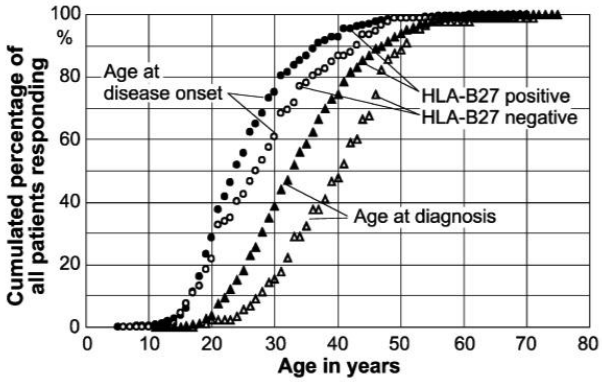


Figure 3 The cumulative distribution in age at disease onset stratified by HLA-B27 status by Feldtkeller et al. (2003)[44]

Given that both prevalence of axSpA[40] and its main genetic risk factor HLA-B27[42] vary considerably throughout the world, a similar distribution in age at onset to the patients in the Feldtkeller study[44] in other parts of the world is not a given. Yet, classification criteria should be applicable to all patients with axSpA worldwide to ensure consistency in the patients who get selected for participation in clinical trials. The same principle applies to the value of a positive family history. Here too, the definition was created using data limited to mostly Western European patients. Should the definition of a positive family history be revised in the future, a new definition must be applicable to patients all around the world. In this thesis we aim to provide an international perspective on the characterisation of patients with axSpA-specifically with regards to the age at symptom onset and positive family history of axSpA-, to investigate whether classification criteria are indeed applicable worldwide.

ASAS-OMERACT CORE SET FOR AXIAL SPONDYLOARTHRITIS

Similarly, it is equally important that all trials executed in various parts of the world assess the same outcomes and report them in a similar way, such that data from American studies can be compared to those initiated in Asia. This is where core outcome sets come in.

Core outcome sets describe the minimum set of mandatory outcomes that should be assessed and reported in all clinical studies of a specific health condition, population and setting[60,61]. A core outcome set consists of domains (*what to measure*) and instruments (*how to measure*). Through standardisation of measurements and reporting, the use of a core outcome set enables direct comparisons between clinical trials on the

effectiveness and safety of the investigated therapies and avoids selective reporting (i.e. only the favourable outcomes)[61]. Using a core outcome set for axSpA thereby reduces heterogeneity of outcomes between studies and the risk of reporting bias because it ensures that all trials contribute relevant and valuable information, which will ultimately result in better research[62]. In this light -taking into account the development of new outcome instruments-, regular review and update of existing core sets is important to ensure the included instruments are still relevant and important[62].

In 1997, ASAS developed a preliminary core outcome set for ankylosing spondylitis, followed by the selection of instruments for each domain[63,64]. The core set for AS was endorsed by OMERACT (Outcome Measures in Rheumatology) in 1999[65,66]. Figure 4 represents the domains that were selected as part of the original core set. The original core set was developed for three different scenarios, which are represented by the different ellipsoids in figure 4[65]:

1. Disease modifying antirheumatic drugs (DMARDs, here indicated as DC-ART where T stands for therapy), presented in the outer ellipsoid in dark grey: therapy that changes the course of disease, both by decreasing inflammatory manifestations, improving or preserving function and preventing or significantly decreasing progression of structural damage.
2. Symptom modifying antirheumatic drugs (SMARD), presented in the inner ellipsoid in white: therapy which improves the symptoms and clinical features of inflammatory manifestations in axSpA. Nonpharmacological interventions belong also to this scenario (e.g. physical therapy).
3. Clinical record keeping in daily practice (presented in the middle ellipsoid in light grey), to facilitate uniform clinical record keeping to strengthen research from clinical records and to monitor patient care in a standardized way.

The ASAS-OMERACT core set was well implemented since its introduction 20 years ago[67], but its development was limited to patients with r-axSpA, whereas it is now well-known that the axSpA spectrum includes both patients with r-axSpA and nr-axSpA[2,31,68]. Additionally, many new outcome instruments have been developed and validated for use in axSpA (such as the Ankylosing Spondylitis Disease Activity Score (ASDAS)[69], the ASAS Health Index[70], and validated enthesitis scores[71]), and with time it became known that it is important to include all stakeholders that will use the core set in its development too[62]. These advances combined with the improvements in the methodology surrounding the development of core sets made ASAS decide it was necessary to update the original ASAS-OMERACT core set for AS. The new core set needs to be applicable to the entire spectrum of axSpA and be developed according to the current recommended methodology.

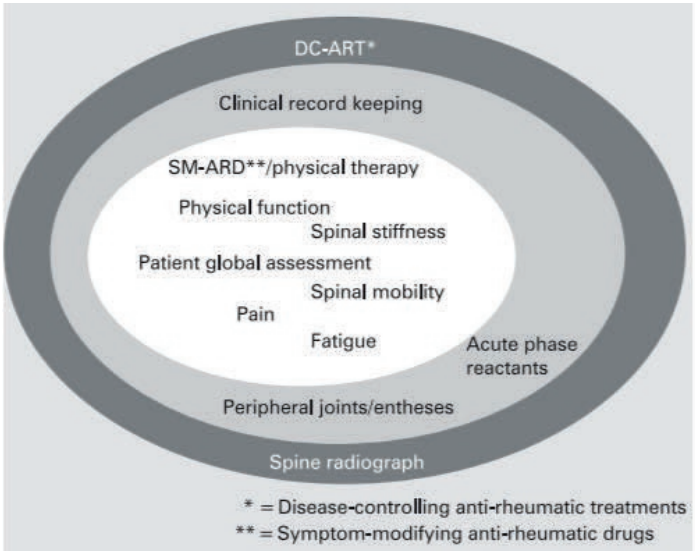


Figure 4 ASAS-OMERACT core domains for ankylosing spondylitis by van der Heijde et al. (1999)[64]
ASAS, Assessment in SpondyloArthritis international Society; OMERACT, Outcome Measures in Rheumatology; DC-ART, disease-controlling antirheumatic treatments; SM-ARD, symptom modifying antirheumatic drugs

The first step to be taken when developing or updating a core set is the selection of *what to measure*, which is defined in so-called domains that are combined to form the core domain set. Thereafter, it needs to be defined *how to measure* each of the chosen domains – through the selection of instruments or tools (core measurement set). The final end product will include both the selected domains and instruments, which will define the new or updated core outcome set.

To enable selection of *what to measure*, a complete overview of all potential domains is required to determine which domains should eventually be included in the core set. Herein, the domains included in the original core outcome set[65,66] and the literature (i.e. all domains reported in trials assessing (non)pharmacological therapies) provide a good starting point, as these represent domains currently measured in clinical research. Additionally, there is an important role for the stakeholders who will end up using the core outcome set to ensure no domains of importance are missed[62]. This includes direct input from patient representatives, which can be collected through qualitative studies and patient focus group interviews[72,73]. Once an extensive list of all potential domains is established, all stakeholders (e.g. patients, rheumatologists, physiotherapists, radiologists, researchers and representatives from pharmacological agencies and drug regulatory agencies) should be involved to ascertain which domains are relevant and should be considered for inclusion in the core outcome set, as they will become the end-users once the core outcome set is in place[74].

For this purpose, the Delphi survey is a valuable tool in collecting opinions from a large group of participants, as it is easily accessible, guarantees anonymity, does not require travel -thereby enabling the inclusion of participants from different continents and time-zones- and does not require public speaking, which increases patient participation[75]. A common application of the Delphi survey is ranking a set of concepts in order of importance or decreasing a voluminous list to a more workable list by prioritizing concepts[76,77], making it the perfect tool to determine which domains are considered relevant by the stakeholders who will use the updated core outcome set. A Delphi survey consists of multiple rounds, which provides participants the opportunity to alter their responses in between rounds in light of the responses of peers. For this purpose, participants receive the aggregated information of peers as well as their own score after each round, which allows them to take the opinions of others into account when answering the questions for a second/third time[78]. The Delphi process ends when (the predefined level of) consensus is achieved, or when the prespecified number of rounds has been completed[79]. In this thesis we have employed the Delphi survey to gather opinions of patients and experts to define the most relevant disease domains to be included in the core set. Next, the results of this Delphi survey will be discussed amongst the ASAS members and shaped into a proposal for the core domain set for axSpA. Once consensus is reached a formal voting session will decide whether the proposed core domain set will be accepted.

Standardisation of methodology is as important as standardisation in measurements. However, there is little guidance on the methodology underlying a Delphi survey[76,79,80], which results in large variability in its execution. One of the aspects that lacks specification is how to invite participants to consecutive rounds of the Delphi survey, which can have an impact on the results and conclusions that are drawn from these results. There are two invitation approaches: 1) Invite only participants that have completed the previous round for the consecutive round; 2) Invite every participant for all consecutive rounds irrespective of whether they have responded or not. Scientific evidence to guide Delphi researchers on whether participants who miss a round can be included in a subsequent round is sparse. In this thesis we investigate whether a different invitation approach influences the final results of the Delphi survey.

After defining the core domain set, the next step is to determine the core measurement set (i.e. the selection of instruments that can be used to measure the domains). At least one instrument needs to be chosen for each selected domain. Herein, once again, there is an important role for previously published literature, as a thorough literature search can provide inside in all instruments currently assessed in clinical trials evaluating treatment effects in axSpA as well as ensure the most recently developed instruments are included too. Once all candidate instruments are identified, all psychometric properties of the

instruments should be collected, as these provide valuable information on the performance of the instruments. They include truth (domain match, face and content validity), feasibility, construct validity, and discrimination (test-retest reliability, responsiveness, clinical trial discrimination and thresholds of meaning)[81]. The truth aspect informs users whether the instrument measures what it is intended to measure and whether the scores are truthful. Feasibility relates to the ease of use, the burden related to completing the instrument for the respondent and/or administrator and the cost related to the use of the instrument. Discrimination describes whether the instrument is able to discriminate between situations of interest, this includes discrimination between treatment arms in a trial, as well as change over time as a result of treatment[81]. Herein it is important to determine whether the same result will be obtained if assessed twice in a situation where there is no change, and improvements/deteriorations reported in trials can thus be ascribed to the treatment rather than measurement error. Test-retest reliability assesses just that, and is therefore an important psychometric in choosing the instruments with the best fit for a given domain. In this thesis we will describe the test-retest reliability of the instruments used in the most recent randomised controlled trials in axSpA, to provide a basis for the selection of the most appropriate instruments for the axSpA core set.

For many outcomes -such as pain or health-related quality of life- rheumatologists and researchers rely on data provided by the patient, as there are no objective measures available for these outcomes. Hence, a large proportion of the outcome measures often used in the assessment of axSpA are so-called patient reported outcome measures.

PATIENT REPORTED OUTCOMES IN EARLY AXIAL SPONDYLOARTHRITIS

Patient reported outcomes are of great importance in axSpA, as most outcomes such as pain, physical functioning and quality of life are subjective measures that cannot be fully captured using objective outcome measures. Furthermore, previous studies have shown that physicians and patients have different perceptions of disease activity[82-85] and physical functioning[86], further emphasizing the importance of patient reported outcome measures. Building on this, patient reported outcomes played a crucial role in the recognition that the burden of disease is comparable between patients with r-axSpA and patients with nr-axSpA[32,33,87].

AxSpA can have a detrimental impact on health-related quality of life[30,88-91], which is why optimising long-term health-related quality of life and social participation has been defined as the main treatment goal in axSpA in the ASAS-EULAR (European Alliance of Associations for Rheumatology) recommendations[92,93]. Limitations of health-related quality of life in

patients with axSpA can be assessed using generic and disease-specific questionnaires. Generic instruments are less specific for a certain disease but allow for comparisons between diseases or with the general population[94]. The SF-36 (Medical Outcomes Study 36-Item Short-Form Health Survey)[95] is an example of a generic questionnaire that is often used in the field of axSpA. The SF-36 has 2 main components: sub-scores for physical health (physical component score) and mental health (mental component score). Standardized population scores are available for the SF-36, which facilitates comparisons between patients with axSpA and healthy individuals, as well as comparisons with other (rheumatic) diseases. Previous research has shown that health related-quality of life is already affected in patients with early axSpA and can be improved by reducing disease activity with effective treatment[96,97].

Additionally, as complaints start early in life in the majority of patients, axSpA is associated with significant risk of limiting work productivity over the patient's life course[30,88-91], which contributes to substantial societal costs of axSpA[22,96]. Unemployment rates and work disability rates are substantially increased compared to the general population, and switching to a less physically demanding job or early retirement are common among patients with axSpA compared to the general population[98,99]. Reduced ability to perform one's job adequately (presenteeism) and an increase in the hours missed from work due to disease (absenteeism) result in reduced work productivity[100]. Furthermore, patients with axSpA report that the disease influences their job choice, and that they require workplace adaptation, which adds to the personal and societal impact of the disease[101].

The majority of studies on work productivity loss focused on patients with r-axSpA and patients with a long disease duration. However, as participating in work has a large impact not only on societal cost, but also on an individual's social and psychological well-being[100] and axSpA usually starts in young adulthood-which tend to be the most productive years[6], it is equally-or even more- important to assess work outcomes in early disease.

At the time of initiation of the SpondyloArthritis Caught Early (SPACE) cohort[102], little was known on the long-term impact of the early phase of axSpA on quality of life and the accompanying socio-economic burden of this disease. Hence, one of the research aims was to study the burden of axSpA in patients in an early stage of the disease, and the implications of diagnosis. In order to do so, the SPACE cohort includes patients with recent onset chronic back pain, referred to the rheumatology outpatient clinic with a suspicion of axSpA. Since then, data from the SPACE cohort and other early (ax)SpA cohorts have shown great improvement in quality of life and work productivity following diagnosis, suggesting a beneficial effect of early diagnosis and subsequent treatment[96,103,104]. However, in absence of a comparator group these results are difficult to interpret and it is particularly difficult to attribute the observed improvement to axSpA treatment. In this thesis we will

provide additional insight in the burden of disease in terms of health-related quality of life and work productivity of patients with a diagnosis of axSpA in the first two years after diagnosis, by making a comparison with the patients who did not get a diagnosis of axSpA but were suspected of axSpA.

AIMS AND OUTLINE OF THIS THESIS

The first research aim of this thesis is to provide an international perspective on the characterisation of patients with axSpA, for which we investigated different aspects of the classification criteria. Herein, we started with assessing the similarities and differences between the modified New York criteria and ASAS axSpA criteria in the classification of patients with radiographic axSpA, which is described in **chapter 2**. For this purpose we included patients from eight cohorts from various countries and with various disease duration and severity: ASAS[37,39], Esperanza[105], GErman SPondyloarthritis Inception Cohort (GESPIC)[16], Outcome in Ankylosing Spondylitis International Study (OASIS)[83], Reuma.pt[106], Swiss Clinical Quality Management (SCQM)[107], SPondyloArthritis Caught Early cohort (SPACE)[102], and University of California San Francisco (UCSF) axSpA cohort[108]. Two cohorts (OASIS and UCSF axSpA) included patients with r-axSpA only, the other 6 cohorts included patients with r-axSpA as well as nr-axSpA.

The majority of the 8 cohorts included European patients, the ASAS cohort included American and Asian patients as well as European patients, and the UCSF axSpA cohort included only American patients. Esperanza, GESPIC and SPACE included patients with early disease, ASAS, Reuma.pt and SCQM patients with early or established disease and OASIS and UCSF axSpA patients with established disease only. Esperanza and SPACE required a maximum symptom duration of 2 years, GESPIC a maximum of 10 years, and the other cohorts did not employ a maximum symptom duration.

Another important aspect in the classification of patients with axSpA is age at onset, yet this criterion is based on mostly European data. Therefore, **Chapter 3** evaluates the age at onset of axial symptoms in a worldwide cohort of patients diagnosed with axSpA: the ASAS-PerSpA (ASAS peripheral symptoms in spondyloarthritis) cohort. In this international observational study with a cross-sectional design, 4465 consecutive patients with a diagnosis of axSpA, peripheral SpA or psoriatic arthritis (according to the treating rheumatologist) were included in 24 countries from 4 different geographical regions (Asia, Europe & North America, Latin America, and Middle East & North Africa).

Its international character, size and variety of symptom duration and disease severity enabled worldwide comparisons regarding age at symptom onset in the patients with axSpA. Using data from this same cohort, **chapter 4** provides insight in the geographical prevalence of a family history of a SpA-related disease and its relationship with HLA-B27.

The second research aim is to describe the process of the development of the core set for axSpA by updating the domains of the ASAS-OMERACT core set for ankylosing spondylitis, which is described in **chapters 5 to 8**. The first step in this process was to collect the opinions of patients with axSpA and experts in the field of spondyloarthritis regarding the importance of the domains. For this purpose, a 3-round Delphi survey was deployed, the results of which are presented in **chapter 5**. **Chapter 6** illustrates an additional unique aspect of the Delphi survey design: in a randomised experiment was assessed which invitation approach should be used when performing a Delphi survey. Using the results from the Delphi survey, the mandatory domains for the core set were formulated and endorsed by OMERACT, which is presented in **chapter 7**. Finally, **chapter 8** provides information on the test-retest reliability of measurement instruments used in axSpA, which is a vital step in the final selection of instruments that will become part of the core outcome set for axSpA.

The third and final research aim is to increase knowledge on work and activity outcomes and health-related quality of life over time in chronic back pain patients with a diagnosis of axSpA or a suspicion thereof. For this purpose, work and activity outcomes and health-related quality of life are assessed over time and a comparison is made between patients who get a definite diagnosis of axSpA after two years of protocolised follow-up and those who get diagnosed as no axSpA. This has been investigated in the SPACE cohort and can be found in **chapter 9**. The SPACE cohort is an ongoing international inception cohort[102]. Data was collected from Dutch, Italian, Norwegian and Swedish patients visiting the outpatient clinic with persistent back pain (>3 months and <2 years) with an onset before the age of 45, starting in 2009.

The final two chapters include a summary and general discussion of the findings of this thesis, in English in **chapter 10** and in Dutch for lay persons in **chapter 11**.

REFERENCES

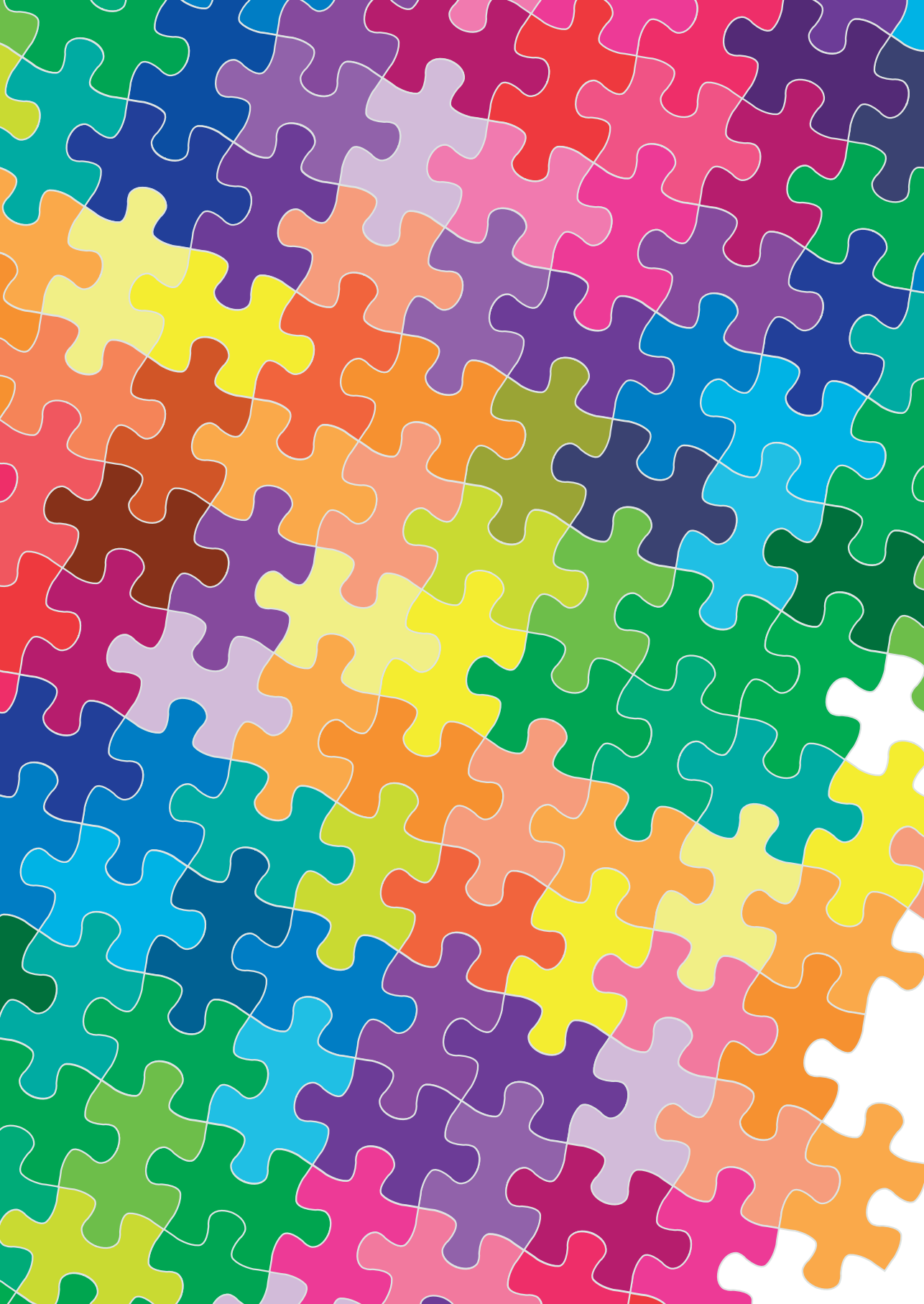
1. Navarro-Compán V, Sepriano A, El-Zorkany B, et al. Axial spondyloarthritis. Ann Rheum Dis 2021;80(12):1511-21.
2. Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet 2017;390(10089):73-84.
3. Poddubnyy D. Classification vs diagnostic criteria: the challenge of diagnosing axial spondyloarthritis. Rheumatology 2020;59(Suppl4):iv6-iv17.
4. Strand V, Singh JA. Patient Burden of Axial Spondyloarthritis. J Clin Rheumatol 2017;23(7):383-91.
5. Dagfinrud H, Mengshoel AM, Hagen KB, et al. Health status of patients with ankylosing spondylitis: a comparison with the general population. Annals of the Rheumatic Diseases 2004;63(12):1605-10.
6. Boonen A, van der Linden SM. The burden of ankylosing spondylitis. J Rheumatol Suppl 2006;78:4-11.
7. Kiltz U, van der Heijde D. Health-related quality of life in patients with rheumatoid arthritis and in patients with ankylosing spondylitis. Clin Exp Rheumatol 2009;27(4 Suppl 55):S108-11.
8. Sieper J, Rudwaleit M, Baraliakos X, et al. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. Ann Rheum Dis 2009;68(Suppl 2):ii1-ii44.
9. Sieper J, van der Heijde D, Landewé R, et al. New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the Assessment of SpondyloArthritis international Society (ASAS). Ann Rheum Dis 2009;68(6):784-88.
10. Rudwaleit M, van der Heijde D, Landewé R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009;68(6):777-83.
11. López-Medina C, Molto A, Sieper J, et al. Prevalence and distribution of peripheral musculoskeletal manifestations in spondyloarthritis including psoriatic arthritis: results of the worldwide, cross-sectional ASAS-PerSpA study. RMD Open 2021;7(1)
12. de Winter JJ, van Mens LJ, van der Heijde D, et al. Prevalence of peripheral and extra-articular disease in ankylosing spondylitis versus non-radiographic axial spondyloarthritis: a meta-analysis. Arthritis Res Ther 2016;18(1):196.
13. Muñoz-Fernández S, Martín-Mola E. Uveitis. Best Pract Res Clin Rheumatol 2006;20(3):487-505.
14. Stolwijk C, van Tubergen A, Castillo-Ortiz JD, et al. Prevalence of extra-articular manifestations in patients with ankylosing spondylitis: a systematic review and meta-analysis. Ann Rheum Dis 2015;74(1):65-73.
15. Essers I, Ramiro S, Stolwijk C, et al. Characteristics associated with the presence and development of extra-articular manifestations in ankylosing spondylitis: 12-year results from OASIS. Rheumatology 2015;54(4):633-40.
16. Rudwaleit M, Haibel H, Baraliakos X, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. Arthritis Rheum 2009;60(3):717-27.
17. Dougados M, Demattei C, van den Berg R, et al. Rate and Predisposing Factors for Sacroiliac Joint Radiographic Progression After a Two-Year Follow-up Period in Recent-Onset Spondyloarthritis. Arthritis Rheum 2016;68(8):1904-13.
18. Ramiro S, van der Heijde D, van Tubergen A, et al. Higher disease activity leads to more structural damage in the spine in ankylosing spondylitis: 12-year longitudinal data from the OASIS cohort. Ann Rheum Dis 2014;73(8):1455-61.
19. Poddubnyy D, Rudwaleit M, Haibel H, et al. Rates and predictors of radiographic sacroiliitis progression over 2 years in patients with axial spondyloarthritis. Ann Rheum Dis 2011;70(8):1369-74.
20. Hoppe B, Schwedler C, Haibel H, et al. Predictive value of C-reactive protein for radiographic spinal progression in axial spondyloarthritis in dependence on genetic determinants of fibrin clot formation and fibrinolysis. RMD Open 2021;7(2)
21. 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.
22. Carvalho PD, Machado PM. How to investigate: Early axial spondyloarthritis. Best Pract Res Clin Rheumatol 2019;33(4):101427.
23. Wang R, Gabriel SE, Ward MM. Progression of Nonradiographic Axial Spondyloarthritis to Ankylosing Spondylitis: A Population-Based Cohort Study. Arthritis Rheum 2016;68(6):1415-21.
24. Protopopov M, Proft F, Sepriano A, et al. Radiographic sacroiliitis progression in axial spondyloarthritis: central reading of 5year follow-up data from the Assessment of SpondyloArthritis international Society cohort. Rheumatology 2021;60(5):2478-80.
25. Sepriano A, Ramiro S, Landewé R, et al. Is active sacroiliitis on MRI associated with radiographic damage in axial spondyloarthritis? Real-life data from the ASAS and DESIR cohorts. Rheumatology 2019;58(5):798-802.
26. Costantino F, Zeboulon N, Said-Nahal R, et al.

Radiographic sacroiliitis develops predictably over time in a cohort of familial spondyloarthritis followed longitudinally. Rheumatology 2017;-56(5):811-17.
27. Sieper J, van der Heijde D. Review: Nonradiographic axial spondyloarthritis: new definition of an old disease? Arthritis Rheum 2013;65(3):543-51.
28. Deodhar A, Strand V, Kay J, et al. The term 'non-radiographic axial spondyloarthritis' is much more important to classify than to diagnose patients with axial spondyloarthritis. Ann Rheum Dis 2016;75(5):791-94.
29. Dougados M, Sepriano A, Molto A, et al. Sacroiliac radiographic progression in recent onset axial spondyloarthritis: the 5-year data of the DESIR cohort. Ann Rheum Dis 2017;76(11):1823-28.
30. Marzo-Ortega H. Axial spondyloarthritis: coming of age. Rheumatology 2020;59(Suppl4):iv1-iv5.
31. Robinson PC, van der Linden S, Khan MA, et al. Axial spondyloarthritis: concept, construct, classification and implications for therapy. Nat Rev Rheumatol 2021;17(2):109-18.
32. Kiltz U, Baraliakos X, Karakostas P, et al. Do patients with non-radiographic axial spondylarthritis differ from patients with ankylosing spondylitis? Arthritis Care Res (Hoboken) 2012;64(9):1415-22.
33. López-Medina C, Ramiro S, van der Heijde D, et al. Characteristics and burden of disease in patients with radiographic and non-radiographic axial Spondyloarthritis: a comparison by systematic literature review and meta-analysis. RMD Open 2019;5(2):e001108.
34. Ortolan A, Kiltz U, Doria A, et al. Do we believe in non-radiographic axial spondyloarthritis? A debate. Autoimmunity Reviews 2021;20(1):102703.
35. June RR, Aggarwal R. The use and abuse of diagnostic/classification criteria. Best Pract Res Clin Rheumatol 2014;28(6):921-34.
36. Aggarwal R, Ringold S, Khanna D, et al. Distinctions Between Diagnostic and Classification Criteria? Arthritis Care & Research 2015;67(7):891-97.
37. Rudwaleit M, van der Heijde D, Landewé R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009;68(6):777-83.
38. Bennett AN, McGonagle D, O'Connor P, et al. Severity of baseline magnetic resonance imaging-evident sacroiliitis and HLA-B27 status in early inflammatory back pain predict radiographically evident ankylosing spondylitis at eight years. Arthritis Rheum 2008;58(11):3413-8.
39. Rudwaleit M, Landewé R, van der Heijde D, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. Ann Rheum Dis 2009;68(6):770-6.
40. Dean LE, Jones GT, MacDonald AG, et al. Global prevalence of ankylosing spondylitis. Rheumatology 2014;53(4):650-7.
41. Stolwijk C, van Onna M, Boonen A, et al. Global Prevalence of Spondyloarthritis: A Systematic Review and Meta-Regression Analysis. Arthritis Care Res (Hoboken) 2016;68(9):1320-31.
42. Khan MA. Polymorphism of HLA-B27: 105 subtypes currently known. Curr Rheumatol Rep 2013;15(10):362.
43. Baraliakos X, Braun J. Non-radiographic axial spondyloarthritis and ankylosing spondylitis: what are the similarities and differences? RMD open 2015;1(Suppl 1):e000053-e53.
44. Feldtkeller E, Khan MA, van der Heijde D, et al. Age at disease onset and diagnosis delay in HLA-B27 negative vs. positive patients with ankylosing spondylitis. Rheumatol Int 2003;23(2):61-6.
45. Chung HY, Machado P, van der Heijde D, et al. HLA-B27 positive patients differ from HLA-B27 negative patients in clinical presentation and imaging: results from the DESIR cohort of patients with recent onset axial spondyloarthritis. Ann Rheum Dis 2011;70(11):1930-36.
46. Glintborg B, Sørensen IJ, Østergaard M, et al. Ankylosing Spondylitis versus Nonradiographic Axial Spondyloarthritis: Comparison of Tumor Necrosis Factor Inhibitor Effectiveness and Effect of HLA-B27 Status. An Observational Cohort Study from the Nationwide DANBIO Registry. J Rheumatol 2017;44(1):59-69.
47. Sieper J, Landewé R, Magrey M, et al. Predictors of remission in patients with non-radiographic axial spondyloarthritis receiving open-label adalimumab in the ABILITY-3 study. RMD Open 2019;5(1):e000917.
48. Yahya F, Gaffney K, Hamilton L, et al. Tumour necrosis factor inhibitor survival and predictors of response in axial spondyloarthritis-findings from a United Kingdom cohort. Rheumatology 2018;57(4):619-24.
49. Coates LC, Baraliakos X, Blanco FJ, et al. The phenotype of axial spondyloarthritis: is it dependent on HLA-B27 status? Arthritis Care Res (Hoboken) 2020
50. Fahed H, Mauro D, Ciccia F, et al. What Does Human Leukocyte Antigen B27 Have to Do with Spondyloarthritis? Rheumatic Disease Clinics of North America 2020;46(2):225-39.
51. Brown MA, Laval SH, Brophy S, et al. Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis. Ann Rheum Dis 2000;59(11):883-86.
52. van Lunteren M, van der Heijde D, Sepriano A, et al. Is a positive family history of spondyloarthritis

relevant for diagnosing axial spondyloarthritis once HLA-B27 status is known? Rheumatology 2019;58(9):1649-54.
53. van Lunteren M, Sepriano A, Landewé R, et al. Do ethnicity, degree of family relationship, and the spondyloarthritis subtype in affected relatives influence the association between a positive family history for spondyloarthritis and HLA-B27 carriership? Results from the worldwide ASAS cohort. Arthritis Res Ther 2018;20(1):166.
54. Ez-Zaitouni Z, Hilkens A, Gossec L, et al. Is the current ASAS expert definition of a positive family history useful in identifying axial spondyloarthritis? Results from the SPACE and DESIR cohorts. Arthritis Res Ther 2017;19(1):118.
55. Olivieri I, Salvarani C, Cantini F, et al. Ankylosing spondylitis and undifferentiated spondyloarthropathies: a clinical review and description of a disease subset with older age at onset. Curr Opin Rheumatol 2001;13(4):280-4.
56. van der Linden SM, Valkenburg HA, de Jongh BM, et al. 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(3):241-9.
57. Said-Nahal R, Miceli-Richard C, Berthelot JM, et al. The familial form of spondylarthropathy: a clinical study of 115 multiplex families. Groupe Français d'Etude Génétique des Spondylarthropathies. Arthritis Rheum 2000;43(6):1356-65.
58. Rudwaleit M, Metter A, Listing J, et al. Inflammatory back pain in ankylosing spondylitis: a reassessment of the clinical history for application as classification and diagnostic criteria. Arthritis Rheum 2006;54(2):569-78.
59. Calin A, Porta J, Fries JF, et al. Clinical history as a screening test for ankylosing spondylitis. Jama 1977;237(24):2613-4.
60. Boers M, Kirwan JR, Tugwell P. OMERACT Handbook, 2018.
61. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. Trials 2017;18(3):280.
62. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13:132.
63. van der Heijde D, Bellamy N, Calin A, et al. Preliminary core sets for endpoints in ankylosing spondylitis. Assessments in Ankylosing Spondylitis Working Group. J Rheumatol 1997;24(11):2225-9.
64. van der Heijde D, Calin A, Dougados M, et al. 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(4):951-4.
65. van der Heijde D, van der Linden S, Bellamy N, et al. 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(4):945-47.
66. van der Heijde D, van der Linden S, Dougados M, et al. Ankylosing spondylitis: plenary discussion and results of voting on selection of domains and some specific instruments. J Rheumatol 1999;26(4):1003-5.
67. Bautista-Molano W, Navarro-Compan V, Landewe R, et al. How Well Are The Assessment Of Spondyloarthritis International Society (ASAS)/ Outcome Measures In Rheumatology (OMERACT) Core Outcome Sets For Ankylosing Spondylitis Implemented In Randomized Clinical Trials? A Systematic Literature Review. Arthritis Rheum-Us 2013;65:S638-S39.
68. Sieper J, Braun J, Dougados M, et al. Axial spondyloarthritis. Nat Rev Dis Primers 2015;1:15013.
69. Lukas C, Landewé R, Sieper J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. Ann Rheum Dis 2009;68(1):18-24.
70. Kiltz U, van der Heijde D, Boonen A, et al. Development of a health index in patients with ankylosing spondylitis (ASAS HI): final result of a global initiative based on the ICF guided by ASAS. Ann Rheum Dis 2015;74(5):830-35.
71. Heuft-Dorenbosch L, Spoorenberg A, van Tubergen A, et al. Assessment of enthesitis in ankylosing spondylitis. Ann Rheum Dis 2003;62(2):127-32.
72. Boonen A, van Berkel M, Kirchberger I, et al. Aspects relevant for functioning in patients with ankylosing spondylitis according to the health professionals: a Delphi study with the ICF as reference. Rheumatology 2009;48(8):997-1002.
73. Boonen A, Braun J, van der Horst Bruinsma IE, et al. ASAS/WHO ICF Core Sets for ankylosing spondylitis (AS): how to classify the impact of AS on functioning and health. Ann Rheum Dis 2010;69(1):102-7.
74. Maxwell LJ, Beaton DE, Shea BJ, et al. Core Domain Set Selection According to OMERACT Filter 2.1: The OMERACT Methodology. J Rheumatol 2019;46(8):1014-20.
75. Khodyakov D, Grant S, Denger B, et al. Practical Considerations in Using Online Modified-Delphi Approaches to Engage Patients and Other Stakeholders in Clinical Practice Guideline Development. The patient 2020;13(1):11-21.
76. Humphrey-Murto S, de Wit M. The Delphi method—more research please. J Clin Epidemiol 2019;106:136-39.

77. Keeney S, McKenna H, Hasson F. The Delphi technique in nursing and health research: John Wiley & Sons 2010.
78. Walker A, Selfe JJBJoT, Rehabilitation. The Delphi method: a useful tool for the allied health researcher. 1996;3(12):677-81.
79. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. 2000;32(4):1008-15.
80. Sinha IP, Smyth RL, Williamson PR. Using the Delphi Technique to Determine Which Outcomes to Measure in Clinical Trials: Recommendations for the Future Based on a Systematic Review of Existing Studies. Plos Med 2011;8(1)
81. Beaton DE, Maxwell LJ, Shea BJ, et al. Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. J Rheumatol 2019;46(8):1028-35.
82. Desthieux C, Molto A, Granger B, et al. Patient-physician discordance in global assessment in early spondyloarthritis and its change over time: the DESIR cohort. Ann Rheum Dis 2016;75(9):1661-66.
83. Spoorenberg A, van Tubergen A, Landewé R, et al. Measuring disease activity in ankylosing spondylitis: patient and physician have different perspectives. Rheumatology 2005;44(6):789-95.
84. Lindström Egholm C, Krogh NS, Pincus T, et al. Discordance of Global Assessments by Patient and Physician Is Higher in Female than in Male Patients Regardless of the Physician's Sex: Data on Patients with Rheumatoid Arthritis, Axial Spondyloarthritis, and Psoriatic Arthritis from the DANBIO Registry. J Rheumatol 2015;42(10):1781-5.
85. Michelsen B, Ørnbjerg LM, Kvien TK, et al. Impact of discordance between patient's and evaluator's global assessment on treatment outcomes in 14 868 patients with spondyloarthritis. Rheumatology 2020;59(9):2455-61.
86. Berkanovic E, Hurwicz ML, Lachenbruch PA. Concordant and discrepant views of patients' physical functioning. Arthritis Care Res 1995;8(2):94-101.
87. López-Medina C, Molto A, Claudepierre P, et al. Clinical manifestations, disease activity and disease burden of radiographic versus non-radiographic axial spondyloarthritis over 5 years of follow-up in the DESIR cohort. Ann Rheum Dis 2020;79(2):209-16.
88. Boonen A, Sieper J, van der Heijde D, et al. The burden of non-radiographic axial spondyloarthritis. Semin Arthritis Rheum 2015;44(5):556-62.
89. Husky MM, Ferdous Farin F, Compagnone P, et al. Chronic back pain and its association with quality of life in a large French population survey. Health and quality of life outcomes 2018;16(1):195.
90. Sadosky AB, Taylor-Stokes G, Lobosco S, et al. Relationship between self-reported low-back pain severity and other patient-reported outcomes: results from an observational study. J Spinal Disord Tech 2013;26(1):8-14.
91. Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014;73(6):968-74.
92. Smolen JS, Schöls M, Braun J, et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. Ann Rheum Dis 2018;77(1):3-17.
93. van der Heijde D, Ramiro S, Landewé R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis 2017;76(6):978-91.
94. Kiltz U, Kiefer D, Boonen A. (Health-Related) Quality of Life as an Outcome in Studies of Axial Spondyloarthritis. Rheum Dis Clin North Am 2020;46(2):379-93.
95. Ware JEJ, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. Medical Care 1992;30(6):473-83.
96. van Lunteren M, Ez-Zaitouni Z, de Koning A, et al. In Early Axial Spondyloarthritis, Increasing Disease Activity Is Associated with Worsening of Health-related Quality of Life over Time. J Rheumatol 2018;45(6):779-84.
97. Fernández-Carballido C, Navarro-Compán V, Castillo-Gallego C, et al. Disease Activity As a Major Determinant of Quality of Life and Physical Function in Patients With Early Axial Spondyloarthritis. Arthritis Care Res (Hoboken) 2017;69(1):150-55.
98. Boonen A, Chorus A, Miedema H, et al. Withdrawal from labour force due to work disability in patients with ankylosing spondylitis. Ann Rheum Dis 2001;60(11):1033-39.
99. Cakar E, Taskaynatan MA, Dincer U, et al. Work disability in ankylosing spondylitis: differences among working and work-disabled patients. Clin Rheumatol 2009;28(11):1309-14.
100. Nikiphorou E, Ramiro S. Work Disability in Axial Spondyloarthritis. Curr Rheumatol Rep 2020;22(9):55.
101. Garrido-Cumbrera M, Poddubnyy D, Gossec L, et al. The European Map of Axial Spondyloarthritis: Capturing the Patient Perspective-an Analysis of 2846 Patients Across 13 Countries. Curr Rheumatol Rep 2019;21(5):19.
102. van den Berg R, de Hooge M, van Gaalen F, et al. Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and

performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. Rheumatology 2013;52(8):1492-9.
103. van Lunteren M, Ez-Zaitouni Z, Fongen C, et al. Disease activity decrease is associated with improvement in work productivity over 1 year in early axial spondyloarthritis (SPondyloArthritis Caught Early cohort). Rheumatology 2017;56(12):2222-28.
104. López-Medina C, Dougados M, Collantes-Estévez E, et al. Adherence to recommendations for the use of anti-tumour necrosis factor and its impact over 5 years of follow-up in axial spondyloarthritis. Rheumatology 2018;57(5):880-90.
105. Muñoz-Fernández S, Carmona L, Collantes E, et al. A model for the development and implementation of a national plan for the optimal management of early spondyloarthritis: the Esperanza Program. Ann Rheum Dis 2011;70(5):827-30.
106. Canhão H, Faustino A, Martins F, et al. Reuma.pt - the rheumatic diseases portuguese register. Acta Reumatol Port 2011;36(1):45-56.
107. Ciurea A, Scherer A, Exer P, et al. Tumor necrosis factor α inhibition in radiographic and nonradiographic axial spondyloarthritis: results from a large observational cohort. Arthritis Rheum 2013;65(12):3096-106.
108. Lee W, Reveille JD, Davis JC, et al. Are there gender differences in severity of ankylosing spondylitis? Results from the PSOAS cohort. Ann Rheum Dis 2007;66(5):633-38.



PART I

AXIAL SPONDYLOARTHRITIS DISEASE CHARACTERISTICS

CHAPTER 2

DO PATIENTS WITH AXIAL SPONDYLOARTHRITIS WITH RADIOGRAPHIC SACROILIITIS FULFIL BOTH THE MODIFIED NEW YORK CRITERIA AND THE ASAS AXIAL SPONDYLOARTHRITIS CRITERIA? RESULTS FROM EIGHT COHORTS.

Anne Boel, Anna Moltó, Désirée van der Heijde, Adrian Ciurea, Maxime Dougados, Lianne S. Gensler, Maria-José Santos, Eugenio de Miguel, Denis Poddubnyy, Martin Rudwaleit, Astrid van Tubergen, Floris van Gaalen, Sofia Ramiro

Ann Rheum Dis 2019;78(11):1545-1549.

ABSTRACT

Background

Patients with spondyloarthritis with radiographic sacroiliitis are traditionally classified according to the modified New York (mNY) criteria as ankylosing spondylitis (AS) and more recently according to the Assessment of SpondyloArthritis international Society (ASAS) criteria as radiographic axial spondyloarthritis (r-axSpA).

Objective

To investigate the agreement between the mNY criteria for AS and the ASAS criteria for r-axSpA and reasons for disagreement.

Methods

Patients with back pain ≥3 months, diagnosed as axSpA with radiographic sacroiliitis (mNY radiographic criterion) were selected from eight cohorts (ASAS, Esperanza, GESPIC, OASIS, Reuma.pt, SCQM, SPACE, UCSF). Subsequently, we calculated the percentage of patients who fulfilled the ASAS r-axSpA criteria within the group of patients who fulfilled the mNY criteria, and vice- versa in six cohorts with complete information.

Results

Of the 3882 patients fulfilling the mNY criteria, 93% also fulfilled the ASAS r-axSpA criteria. Inversely, of the 3434 patients fulfilling the ASAS r-axSpA criteria, 96% also fulfilled the mNY criteria. The main cause for discrepancy between the two criteria sets was the reported age at onset of back pain.

Conclusion

Almost all patients with axSpA with radiographic sacroiliitis fulfil both ASAS and mNY criteria, which supports the interchangeable use of the terms AS and r-axSpA.

INTRODUCTION

Traditionally, patients with axial spondyloarthritis (axSpA) with definite structural changes on conventional radiographs are classified according to the modified New York (mNY) criteria as ankylosing spondylitis (AS). However, they may also be classified according to the more recent Assessment of SpondyloArthritis international Society (ASAS) axSpA criteria as radiographic axSpA (r-axSpA).

Both the mNY and the ASAS axSpA classification criteria use the radiographic criterion as defined by the mNY criteria (ie, sacroiliitis of at least grade 2 bilaterally or at least grade 3 unilaterally). However, the additionally required (clinical) features of the classification criteria differ (table 1). Importantly, patients with age at onset of back pain ≥45 years cannot fulfil the ASAS criteria, but there is no age limit for the mNY criteria.[1,2] Patients without the inflammatory character of back pain fulfil the ASAS criteria if another SpA feature is present, but only fulfil the mNY criteria if there is limitation in spinal mobility. These differences in the clinical part of both criteria sets raise the question whether the two sets classify the same patients with axSpA with radiographic sacroiliitis.

The aim of this study was to investigate if patients who fulfil the mNY criteria also fulfil the ASAS criteria for r-axSpA and vice- versa. The second objective was to investigate reasons for disagreement.

METHODS

Patients diagnosed with axSpA who had back pain for at least 3 months and definite radiographic sacroiliitis based on local reading, according the mNY radiographic criterion (#4a or 4b in table 1) were selected from eight cohorts (ASAS, Esperanza, GErman SPondyloarthritis Inception Cohort (GESPIC), Outcome in Ankylosing Spondylitis International Study (OASIS), Reuma.pt, Swiss Clinical Quality Management (SCQM), SPondyloArthritis Caught Early cohort (SPACE), and University of California San Francisco (UCSF) axSpA cohort[1,3–9]). The ASAS cohort included patients with undiagnosed axSpA irrespective of symptom duration, in 25 ASAS centres across 16 countries in Western-Europe, Turkey, Asia, Colombia and Canada between 2005 and 2009.[1] Esperanza is a Spanish national health programme for early SpA, which started inclusion in 2007.[6] GESPIC started in 2000 and consists of patients with axSpA and symptom duration of up to 10 years.[7] OASIS consists of Dutch, Belgian and French patients with established AS, which started in 1996.[8]

Table 1 Classification of axial spondyloarthritis with radiographic sacroiliitis using the mNY criteria for the classification of AS[17], and the ASAS criteria for the classification of r-axSpA[1].

mNY criteria for the classification of AS	**ASAS criteria** for the classification of radiographic axSpA
1. Low back pain and stiffness for at least 3 months, which improves with exercise and is not relieved by rest 2. Limitation of lumbar spine motion in the sagittal and frontal planes 3. Decreased chest expansion, compared to age- and sex-matched controls **4a**. Unilateral sacroiliitis grade 3 or 4 **4b**. Bilateral sacroiliitis grade 2 to 4	1. Back pain ≥3 months 2. Age at onset <45 years 3. Definite radiographic sacroiliitis according to mNY criteria 4. ≥1 SpA feature: – Inflammatory back pain – Arthritis – Enthesitis – Uveitis – Dactylitis – Psoriasis – Crohn's/colitis – Good response to NSAIDs – Family history for SpA – HLA-B27 positive – Elevated CRP (or ESR)
Definite AS if sacroiliitis as described in 4a or 4b and any of the clinical symptoms (1-3)	Definite r-axSpA if fulfilment of 1 and 2, sacroiliitis as described in 3 and at least one of the clinical SpA features as described in 4

AS, Ankylosing Spondylitis; ASAS, Assessment of SpondyloArthritis international Society; axSpA, axial Spondyloarthritis; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; HLA-B27, Human leukocyte antigen B27; mNY, modified New York; NSAID, nonsteroidal anti-inflammatory drugs; r-axSpA, radiographic axial spondyloarthritis

Since 2008, Reuma.pt started with the inclusion of Portuguese rheumatic patients of various diseases and disease stages in a national register; including patients with early and established axSpA.[3] The SCQM axSpA cohort started in Switzerland in 2005 including patients with early and established disease.[4] SPACE is an early chronic back pain cohort including European patients since 2009.[9] Patients in the UCSF axSpA cohort started enrolling in 2007; patients with early and established disease from the UCSF clinic are included.[5] Approval from the medical ethical committees was obtained per cohort, and for all patients written informed consent was obtained prior to inclusion.

For these cohorts, we calculated how many patients with SpA with radiographic sacroiliitis fulfil the mNY criteria (mNY+) and the ASAS r-axSpA criteria (ASAS+). Subsequently, we calculated the percentage of patients who fulfil the ASAS r-axSpA criteria within the group of patients who fulfil the mNY criteria. In six cohorts, we were also able to calculate the percentage of patients fulfilling the mNY criteria within the group fulfilling the ASAS r-axSpA criteria. For the Esperanza and OASIS cohorts, specific information on the individual items of the mNY clinical criteria was unavailable. Consequently, it was not possible to calculate the percentage of patients fulfilling the mNY criteria within the subgroup fulfilling the ASAS criteria. Flowcharts were used to visualise fulfilment of the criteria sets (online supplementary figure S1).

For the patients with axSpA with radiographic sacroiliitis, the first step was to determine whether a patient had inflammatory back pain (IBP). For the purpose of this study, the first clinical criterion of the mNY was equated to IBP according to the ASAS definition.[10] The second step was to determine the number of SpA features (<1 vs ≥1) as well as whether the patient had mobility restrictions. Mobility restrictions were defined using the age-adjusted fifth percentile scores of healthy individuals from Ramiro *et al*.[11] ; if the Schober's test and lateral spinal flexion were below the age-adjusted fifth percentile value or chest expansion was below the age-adjusted and height-adjusted fifth percentile value, mobility was considered restricted. The final step was to look at age at onset of back pain (<45 vs ≥45 years old).

RESULTS

A total of 7636 patients with a SpA diagnosis and back pain >3 months were included in these eight cohorts. Of these, 4041 patients had a diagnosis of axSpA with radiographic sacroiliitis and were available for analysis. In total, 3882 patients fulfilled the mNY criteria, of which 3607 (93%; range 88%–100%) also fulfilled the ASAS r-axSpA criteria (figure 1A). From the six cohorts (n=3721) in which the fulfilment of the mNY criteria in the subgroup of patients fulfilling the ASAS r-axSpA criteria (n=3434) could be analysed, 3300 (96%; range 84%–98%) also fulfilled the mNY criteria (figure 1B).

For all, 4041 patients with r-axSpA fulfilment of the criteria sets was determined (online supplementary tables S1-S3). In total, 3607 (89%) of patients fulfilled both criteria sets; 275 (7%) only the mNY criteria; 134 (3%) only the ASAS criteria and 25 (1%) neither set (table 2).

Table 2 Percentage of patients with axSpA with radiographic sacroiliitis fulfilling both sets of criteria, either criteria set or neither

	mNY+ ASAS+	mNY+ ASAS-	mNY- ASAS+	mNY- ASAS-	Total mNY+*	Total ASAS+ǂ
ASAS (n=114)	86% (98)	2% (3)	10% (11)	2% (2)	89% (101)	96% (109)
GESPIC (n=96)	81% (78)	12% (11)	6% (6)	1% (1)	93% (89)	88% (84)
Esperanza (n=109)	97% (106)	3% (3)	NA‡	NA‡	100% (109)	--
OASIS (n=211)	95% (201)	5% (10)	NA‡	NA‡	100% (211)	--
Reuma.pt (n=1320)	88% (1156)	7% (93)	4% (55)	1% (16)	95% (1249)	92% (1211)
SCQM (n=1806)	89% (1612)	8% (148)	2% (40)	0.3% (6)	97% (1760)	91% (1652)
SPACE (n=92)	84% (77)	0% (0)	16% (15)	0% (0)	84% (77)	100% (92)
UCSF (n=293)	95% (279)	2.5% (7)	2.5% (7)	0% (0)	98% (286)	98% (286)
Total (n=4041)	89% (3607)	7% (275)	3% (134)	1% (25)	96% (3882)	--

*The total percentage of patients who fulfil the mNY criteria per cohort and in total; ǂ The total percentage of patients who fulfil the ASAS r-axSpA criteria per cohort and in total.‡ Specific information on the individual items of the mNY clinical criteria was unavailable, it was therefore not possible to accurately calculate the number of patients fulfilling the mNY in the subgroup fulfilling the ASAS r-axSpA criteria. ASAS, Assessment of SpondyloArthritis international Society cohort; Esperanza, Spanish national health programme for early SpA; GESPIC, GErman SPondyloarthritis Inception Cohort; NA, not available; OASIS, Outcome in Ankylosing Spondylitis International Study; Reuma.pt, Portuguese Register for Rheumatic Diseases; SCQM, Swiss Clinical Quality Managementcohort; SPACE, SPondyloArthritis Caught Early cohort; UCSF, University of California San Francisco axSpA cohort; axSpA, axial spondyloarthritis; mNY, modified New York; r-axSpA, radiographic axial spondyloarthritis

The main difference between the two criteria sets was caused by the reported age at onset of back pain; 99.7% of the patients fulfilling the mNY criteria could potentially fulfil the ASAS criteria except for registered age at onset (online supplementary figure S4).

Out of the 275 mNY +patients not fulfilling the ASAS criteria (7% of all included patients), 265 (96%) cases were due to the age criterion and 10 (4%) due to the absence of SpA features including IBP (online supplementary table 1). These 10 patients had spinal mobility limitation as the only clinical feature. The 134 mNY-/ASAS+ did not have mobility restriction or IBP but another SpA feature instead.

For the cohorts that had data available (n=1833), the human leucocyte antigen B27 (HLA-B27) status was determined in each of the subgroups. In the mNY+/ASAS+ group, HLA-B27 positivity was 68%. In the mNY-/ASAS+ group, a similar percentage was found (72%), whereas in the mNY+/ASAS- group this percentage was only 46%, thus only slightly higher than the mNY-/ASAS- group (42%) (online supplementary table S2).

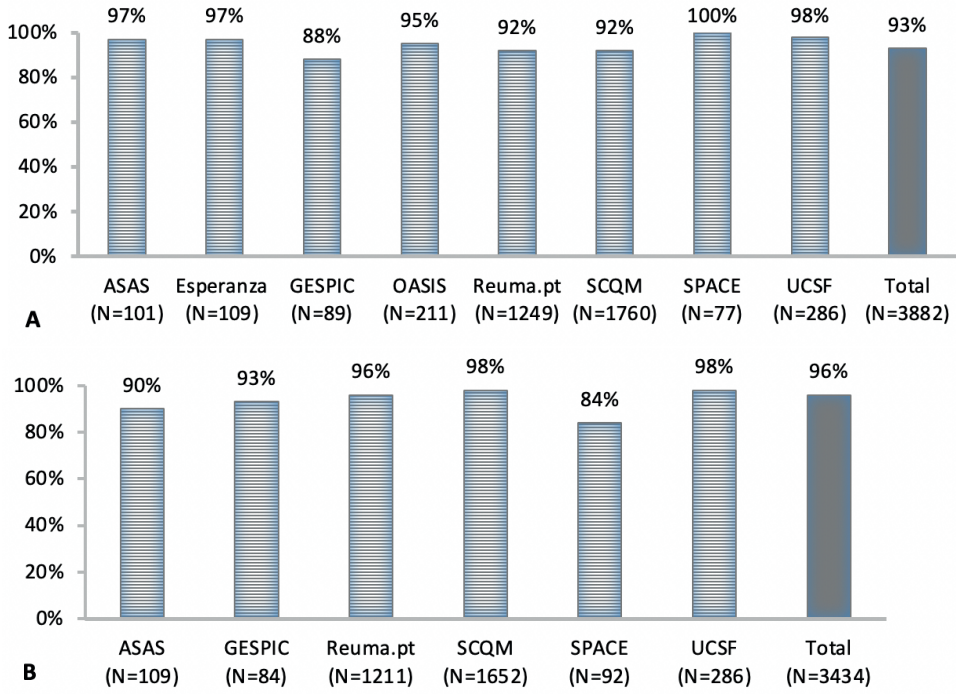


Figure 1: Percentage of patients fulfilling ASAS r-axSpA within subgroup fulfilling mNY criteria (3607/3882) (A) and percentage of patients fulfilling mNY criteria within subgroup fulfilling ASAS r-axSpA (3300/3434) (B), per cohort and overall.
ASAS, (Assessment of SpondyloArthritis international Society cohort); Esperanza, Spanish national health programme for early SpA; Esperanza (Spanish national health programme for early SpA); GESPIC, (GErman SPondyloarthritis Inception Cohort); mNY, modified New York; OASIS, (Outcome in Ankylosing Spondylitis International Study); r-axSpA, radiographic axial spondyloarthritis; Reuma.pt, Portuguese Register for Rheumatic Diseases; Reuma.pt (Portuguese Register for Rheumatic Diseases); SCQM, (Swiss Clinical Quality Management cohort); SPACE, (SPondyloArthritis Caught Early cohort); UCSF, (University of California San Francisco axSpA cohort)).

DISCUSSION

'"Classification criteria are standardised definitions that are primarily intended to create well-defined, relatively homogeneous cohorts of patients for clinical research; they are not intended to capture every single patient but rather to capture the majority of patients who share key features of the condition"'.[12] Patients with axSpA with radiographic sacroiliitis are traditionally classified according to the mNY criteria and more recently according to the ASAS criteria. The data presented in this study show that patients with axSpA classified as AS according to mNY criteria and those classified as r-axSpA according to ASAS criteria are mostly the same. Nonetheless, there is minor disagreement, mainly due to age at onset of back pain. The latter is reported by patients at the time of diagnosis in almost all cohorts and therefore susceptible to recall bias, a valid concern especially for the cohorts

containing patients with a long disease duration and long gap between symptom onset and diagnosis. The age criterion was introduced with the implementation of the ASAS criteria in 2009; this was mainly based on data from Feldtkeller *et al,*.[13] which showed that 95% of AS patients reported an age of onset <45 years. Based on this fact, one would expect around 5% of the patients fulfilling the mNY criteria not to fulfil the ASAS criteria. In this study, this percentage is 7%.

Due to the nature of the data and the slight differences between the two criteria sets some assumptions had to be made, which is a limitation to this study. The first assumption concerns IBP; in general, the ASAS definition of IBP[10] was used. However, if this was unavailable (and could not be defined from individual components of IBP), the rheumatologist's assessment as provided in the dataset was used instead. The second assumption regards mobility limitations; according to the mNY criteria, mobility limitations are to be identified based on age-adjusted and gender-adjusted comparisons; however, in the original publication no reference values were provided. Therefore, reference values resulting from the MOBILITY study[11] were used. If information on mobility was unavailable, the rheumatologist's judgement of 'restricted mobility' as provided in the dataset was used. Both assumptions may have influenced the proportion of patients fulfilling either of the criteria sets.

As shown in the HLA-B27 analysis, the mNY+/ASAS- group showed a lower percentage of HLA-B27 positives. HLA-B27 positivity is associated with earlier disease onset,[13–15] which may explain the low percentage of HLA-B27+ inthe mNY+/ASAS- group (48%) and is in line with the highest HLA-B27 positivity (72%) in the mNY-/ASAS+ group. An alternative explanation may be that patients in the mNY+/ASAS- group are misclassified as having r-axSpA as a higher HLA-B27 percentage is expected in mNY+ patients. The overall percentage of HLA-B27 found in this study is relatively low, which may be due to the local readings of the radiographs that may have resulted in false classifications for both classification sets.[16]

In conclusion, this study found that agreement between the mNY and ASAS r-axSpA criteria is very high, which supports the interchangeable use of the terms AS and r-axSpA. This has important implications for the axSpA research field, since older literature used mNY and AS, whereas more recent literature often uses ASAS criteria and r-axSpA. Acknowledging that both criteria sets identify the same patients implies that older literature on AS and newer literature on r-axSpA can be directly compared.

SUPPLEMENTARY DATA

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

REFERENCES

1. Rudwaleit M, van der Heijde D, Landewe R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68(6):777-83.
2. Malaviya AN, Rawat R, Agrawal N, et al. The Nonradiographic Axial Spondyloarthritis, the Radiographic Axial Spondyloarthritis, and Ankylosing Spondylitis: The Tangled Skein of Rheumatology. *Int J Rheumatol* 2017;2017:1824794.
3. Canhao H, Faustino A, Martins F, et al. Reuma.pt - the rheumatic diseases portuguese register. *Acta Reumatol Port* 2011;36(1):45-56.
4. Ciurea A, Scherer A, Exer P, et al. Tumor necrosis factor alpha inhibition in radiographic and nonradiographic axial spondyloarthritis: results from a large observational cohort. *Arthritis Rheum* 2013;65(12):3096-106.
5. Lee W, Reveille JD, Davis JC, Jr., et al. Are there gender differences in severity of ankylosing spondylitis? Results from the PSOAS cohort. *Ann Rheum Dis* 2007;66(5):633-8.
6. Munoz-Fernandez S, Carmona L, Collantes E, et al. A model for the development and implementation of a national plan for the optimal management of early spondyloarthritis: the Esperanza Program. *Ann Rheum Dis* 2011;70(5):827-30.
7. Rudwaleit M, Haibel H, Baraliakos X, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. *Arthritis Rheum* 2009;60(3):717-27.
8. Spoorenberg A, van der Heijde D, de Klerk E, et al. Relative value of erythrocyte sedimentation rate and C-reactive protein in assessment of disease activity in ankylosing spondylitis. *J Rheumatol* 1999;26(4):980-4.
9. van den Berg R, de Hooge M, van Gaalen F, et al. Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. *Rheumatology* 2013;52(8):1492-9.
10. Sieper J, van der Heijde D, Landewe R, et al. New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the Assessment of SpondyloArthritis international Society (ASAS). *Ann Rheum Dis* 2009;68(6):784-8.
11. Ramiro S, van Tubergen A, Stolwijk C, et al. Reference intervals of spinal mobility measures in normal individuals: the MOBILITY study. *Ann Rheum Dis* 2015;74(6):1218-24.
12. Aggarwal R, Ringold S, Khanna D, et al. Distinctions between diagnostic and classification criteria? *Arthritis Care Res* 2015;67(7):891-7.
13. Feldtkeller E, Khan MA, van der Heijde D, et al. Age at disease onset and diagnosis delay in HLA-B27 negative vs. positive patients with ankylosing spondylitis. *Rheumatol Int* 2003;23(2):61-6.
14. Bennett AN, McGonagle D, O'Connor P, et al. Severity of baseline magnetic resonance imaging-evident sacroiliitis and HLA-B27 status in early inflammatory back pain predict radiographically evident ankylosing spondylitis at eight years. *Arthritis Rheum* 2008;58(11):3413-8.
15. Endo Y, Fujikawa K, Koga T, et al. Characteristics of late-onset spondyloarthritis in Japan: A retrospective cohort study. *Medicine* 2019;98(7):e14431
16. van den Berg R, Lenczner G, Feydy A, et al. Agreement between clinical practice and trained central reading in reading of sacroiliac joints on plain pelvic radiographs. Results from the DESIR cohort. *Arthritis Rheum* 2014;66(9):2403-11.
17. 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.

CHAPTER 3

AGE AT ONSET IN AXIAL SPONDYLOARTHRITIS AROUND THE WORLD: DATA FROM THE ASSESSMENT IN SPONDYLOARTHRITIS INTERNATIONAL SOCIETY PERIPHERAL INVOLVEMENT IN SPONDYLOARTHRITIS STUDY

Anne Boel, Clementina López-Medina,
Désirée van der Heijde, Floris van Gaalen

Rheumatology (Oxford). 2022 Apr 11;61(4):1468-1475

ABSTRACT

Background

Age at onset is useful in identifying chronic back patients at an increased risk of axial SpA (axSpA). However, the majority of data on which the criterion of age at onset <45 years is based originates from Europe. Therefore it is unknown if this criterion applies in other parts of the world. We aimed to assess age at onset of axSpA and its relationship with HLA-B27 and gender across the world.

Methods

Analyses were applied to patients from 24 countries across the world with an axSpA diagnosis and known age at onset of axial complaints. Cumulative probability plots were used to display the cumulative distribution of age at onset of axial symptoms. Linear regression models were built to assess the effect of HLA-B27 and gender on age at onset of axial symptoms.

Results

Of 2579 axSpA patients, 92% had an age at onset of axial symptoms <45 years, with only small variations across the geographical regions [Asia, n=574 (94%); Europe and North America, n=988 (92%); Latin America, n=246 (89%); Middle East and North Africa, n=771 (91%)]. Age at onset of axial symptoms was consistently lower in HLA-B27-positive patients {median 25 years [interquartile range (IQR) 19–32] vs 31 [IQR 22–39]} and male patients [median 25 years (IQR 19–33) vs 28 (IQR 21–37)], but in multivariable models an additional statistically significant effect of male gender independent of HLA-B27 was only found in Asia.

Conclusion

Around the world, the great majority of axSpA patients had an age at onset of axial disease of <45 years, with HLA-B27 and male gender associated with earlier disease onset.

INTRODUCTION

Axial spondyloarthritis (axSpA) is a chronic, inflammatory disease predominantly affecting the sacroiliac joints and spine. HLA-B27 is the most important genetic risk factor for axSpA and has been reported to be associated with earlier onset of disease[1-4]. Data regarding the association between gender and age at onset of disease are ambiguous[5], even though there is a known difference in disease severity and disease expression between male and female patients[2, 4, 6].

AxSpA usually starts in the second or third decade of life[7, 8]. Age at onset of axSpA after 50 years appears to be uncommon[9], thus age at onset can be very useful in identifying chronic back pain patients suspected of axSpA[2], as it is an easy and accessible piece of information that can be used in the first selection of patients.

Previous research has shown that the vast majority of axSpA patients develop back pain before the age of 45 years[1, 10, 11], which formed the basis for the Assessment of Spondyloarthritis international Society (ASAS) definition of inflammatory back pain (IBP)[12] and the prominent place of age at onset in the current ASAS classification criteria for axSpA[13]. In fact, the criterion of onset before the age of 45 is an important difference between the modified New York (mNY) criteria for classification of AS[14] and the ASAS classification criteria and is even the main cause for discrepancy between the two criteria sets in classifying patients with radiographic axSpA (r-axSpA)[15]. Since the publication of the ASAS criteria for axSpA, some data have become available on the age at onset of axSpA patients in Brazil[3 16] and China[17], but the majority of the data originates from Western Europe.

Given that both prevalence of axSpA[18] and its main genetic risk factor HLA-B27[19] vary considerably throughout the world, a similar distribution in age at onset to the patients in the Feldtkeller study[1] in other parts of the world is not a given. Then again, since age at onset plays an important role in diagnosing patients with axSpA as well as in the classification of patients, the age at onset criterion should be representative of patients all around the world. Hence, the aim of this study was to assess age at onset of axSpA as well as its relationship with HLA-B27 and gender in various regions of the world, using data from the Assessment in SpondyloArthritis international Society peripheral involvement in Spondyloarthritis (ASAS-PerSpA) study[20].

MATERIALS AND METHODS

This study was conducted using data from the ASAS-PerSpA dataset, which has been described elsewhere[20]. In brief, ASAS-PerSpA was a multicentre observational study with a cross-sectional design, in which a total of 24 countries participated. Its main aim was to investigate clinical peripheral rheumatologic features in consecutively included SpA patients and evaluate the validity of existing outcome measures of peripheral rheumatological features.

Patients

In the ASAS-PerSpA study, patients with a diagnosis of spondyloarthritis (n=4465) were included between July 2018 and February 2020, representing 24 countries in four geographical regions. The study was approved by the ethical committees in all countries (complete list available in Supplementary Data S1, available at Rheumatology online), and written informed consent was obtained from participants prior to inclusion. For this analysis, only patients with a definite diagnosis of axSpA were included, which was defined as axSpA and either r-axSpA or non-radiographic axSpA (nr-axSpA) as a disease subgroup.

Outcomes

The primary outcome of interest was the age at onset of axial symptoms across all patients with a diagnosis of axSpA and stratified by geographical region. Age at onset was ascertained from the date of first axial symptoms, as reported by the rheumatologist, and the study date. Negative values for age at onset of axial symptoms were recoded to missing values (n=3).

Additional outcomes of interest were the association between HLA-B27 and age at onset of axial symptoms and the association between gender and age at onset of axial symptoms in the total included axSpA population and each of the geographical regions.

Analyses

Analyses were restricted to patients with a known age at onset of axial complaints. Categorical variables were reported as frequencies (proportions) and continuous variables as mean and S.D. in case of normally distributed data and as median and interquartile range (IQR) in case of non-normally distributed data.

Cumulative probability plots were used to display the cumulative distribution in age at onset of axial symptoms. Mann–Whitney U tests were used to compare the median age at onset of axial symptoms between groups stratified for HLA-B27 status or gender.

Linear regression models were built to assess the association between HLA-B27 status or gender and age at onset of axial symptoms with HLA-B27 status or gender as the independent variable and age at onset as a dependent variable. Finally, a multivariable linear regression model including both HLA-B27 status and gender as covariates was built to assess whether the association between HLA-B27 and age at onset was different for male and female patients.

Data were analysed using Stata SE version 16 (StataCorp, College Station, TX, USA). P-values <0.05 were considered statistically significant.

RESULTS

A total of 2579 patients had a definite diagnosis of axSpA and a known age at onset of axial complaints. Patients were grouped in four previously defined geographical regions: Asia (n=574), Europe and North America (n=988), Latin America (n=246) and the Middle East and North Africa (n=771) (Supplementary Table S1, available at Rheumatology online). Overall there was only a small percentage of missing data (<5% unless indicated otherwise), with the exception of HLA-B27 status and MRI of the pelvis, where information was unavailable for a larger proportion of patients, which was especially apparent in the Middle East and North Africa population.

Across the board, 69% of included patients were male, 79% were HLA-B27 positive, the vast majority (94%) had IBP according to the ASAS definition[21], the majority (78%) had r-axSpA and the level of confidence regarding the diagnosis axSpA was high, with very small variations between geographical regions (Table 1). Asian patients had a somewhat lower median age and shorter median symptom duration. Latin American patients more frequently had peripheral symptoms, as shown by the higher percentages of peripheral arthritis, enthesitis and dactylitis; uveitis was also more common compared with patients from the other geographical regions. Noticeably, biological DMARD use was much higher in Latin America compared with the other regions.

Table 1 Characteristics of the axSpA patients from the ASAS-PerSpA study analysed in this study, stratified by geographical region

	Total *n=2,579*	**Asia** *n=574*	**Europe & North America** *n=988*	**Latin America** *n=246*	**Middle East & North Africa** *n=771*
Gender, male	69%	79%	65%	70%	65%
Age, median (IQR)	40 (31-51)	34 (27-45)	44 (35-53)	42 (34-53)	39 (32-49)
Symptom duration (yrs), median (IQR)	11 (5-19)	8 (4-15)	13 (7-24)	12 (5-20)	10 (5-16)
HLA-B27 positive	79%**	89%*	79%**	81%**	67%***
IBP ASAS definition[†]	94%	91%	95%	96%	95%
Positive family history	34%	30%	38%	27%	36%
Peripheral arthritis	44%	52%	38%	72%	36%
Enthesitis	45%	53%	37%	71%	42%
Dactylitis	6%	7%	5%	16%	3%
Psoriasis	8%	4%	13%	4%	5%
IBD	5%	1%	7%	4%	6%
Acute anterior uveitis	22%	24%	25%	31%	15%
Elevated CRP	70%	74%	66%	77%	70%
Sacroiliitis on radiographs[‡]	78%	85%	73%	75%*	79%
Sacroiliitis on MRI[‡]	82%***	78%***	77%***	81%***	93%***
Number of SpA features[§], mean (SD)	4 (2)	4 (1)	4 (2)	4 (2)	3 (2)
Use of bDMARD	33%*	39%**	25%	58%	31%*
Use of NSAID	99%*	99%**	99%	99%	98%*
LoC regarding axSpA, mean (SD)	8 (3)	7 (3)	8 (3)	7 (4)	9 (2)

[†] 4 out of 5 of the following features: onset before the age of 40, insidious onset, improvement with exercise, no improvement with rest, pain at night[21]. [‡] Based on reading of local radiologists. [§] Excluding HLA-B27 status and sacroiliitis on imaging. * 5-10% missing values, ** 10-20% missing values, ***20-40% missing values **ASAS**, Assessment of SpondyloArthritis international Society; **axSpA**, axial spondyloarthritis; **bDMARD**, biological Disease Modifying Anti-Rheumatic Drug; **CRP**, C-reactive protein; **HLA-B27**, Human Leucocyte Antigen B27; **IBD**, Inflammatory Bowel Disease; **IBP**, Inflammatory Back Pain; **IQR**, interquartile range; **LoC**, Level of Confidence regarding the diagnosis; **MRI**, Magnetic Resonance Imaging; **NSAIDs**, Non-Steroidal Anti Inflammatory Drugs; **SpA**, Spondyloarthritis.

Age at onset of axial symptoms

The median age at onset of axial symptoms in all included patients with axSpA was 26 years (IQR 20–34), with the lowest age at onset in Asia [24 (19–31)] followed by Europe and North America [26 (20–35)], Latin America [27 (21–40)] and Middle East and North Africa [27 (21–35)] (Fig. 1). The majority (92%) of patients with axSpA had an age at onset of axial symptoms <45 years, with only a small variation across the various geographical regions (Fig. 1). This finding was even more pronounced in the HLA-B27-positive subgroup (Fig. 1 and Supplementary Table S2, available at Rheumatology online) in which 94% of patients had an age at onset of axial symptoms <45 years. Additionally, only in a very small proportion (4%) of patients did the axial complaints start after the age of 50 years. Cumulative distribution plots showed that among all axSpA patients, 95% developed axial complaints before the age of 48 years and this was before the age of 46, 47, 51 and 48 years for the Asian, European and North American, Latin American and Middle Eastern and North African populations respectively (Fig. 1). Patients with an onset of axial complaints

at the age of ≥45 years were less often male, had a shorter median symptom duration, were less often HLA-B27 positive and had IBP less often compared with patients with an age at onset <45 years (Table 2). Elevated CRP and sacroiliitis on radiographs were also less frequent in patients with an age at onset ≥45 years.

Association between gender and age at onset

In the total included axSpA population, the median age at onset of axial symptoms of male patients [25 years (IQR 19–33)] was significantly lower than that of female patients [28 years (IQR 21–37)] (P<0.001). This difference was seen in Asia [23 years (IQR 19–31) vs 28 (21–37), P=0.015], Latin America [26 (19–34) vs 34 (22–43), P¼0.002] and the Middle East and North Africa [26 (20–33) vs 29 (23–37), P<0.001], but was less pronounced in Europe and North America [26 (20–34) vs 28 (20–36), P=0.053] (Fig. 2). Linear regression models showed a significant effect of gender on the age at onset of axial symptoms in the total study population (P<0.001) and the Asian (P=0.010), Latin American (P=0.001) and Middle Eastern and North African (P<0.001) populations, but just missed the significance level in the European and North American population (P=0.054).

Association between HLA-B27 and age at onset

In the total included axSpA population, the median age at onset of axial symptoms of HLA-B27-positive patients was significantly lower than of HLA-B27-negative patients [25 years (IQR 19–32) vs 31 (22–39); P<0.001]. This difference was found in each of the geographical regions: Asia 23 years (IQR 19–30) vs 28 (20–36), P=0.009; Europe and North America 25 (19–32) vs 33 (22–40), P<0.001; Latin America 26 (19–36) vs 40 (26– 44), P<0.001; Middle East and North Africa 25 (19–32) vs 29 (22–39), P=0.008 (Fig. 1).

Linear regression models showed a significant effect of HLA-B27 status on the age at onset of axial symptoms in the total study population (P<0.001) and all geographical regions (Asia, P=0.006; Europe and North America, P<0.001; Latin America, P<0.001; Middle East and North Africa, P=0.005).

Multivariable model

First, we tested whether there was collinearity between gender and HLA-B27 status, which was not the case, meaning gender and HLA-B27 did not have a linear relationship and could both be included in the linear regression model. In the multivariable model in the total included axSpA population, both HLA-B27 and male gender were associated with earlier disease onset. However, when stratified by region, an additional statistically significant effect of male gender independent of HLA-B27 was only found in Asia (Table 3), but a similar trend could be observed in all regions.

Table 2 Characteristics of the axial spondyloarthritis patients from the ASAS-PerSpA study analysed in this study, stratified by age at onset.

	Total *N=2,579*	**Age at onset <45** *N=2,368*	**Age at onset ≥45** *N=211*
Gender, male	69%	70%	51%
Age, median (IQR)	40 (31-51)	39 (31-48)	58 (53-64)
Symptom duration (yrs), median (IQR)	11 (5-19)	11 (5-20)	6 (3-11)
HLA-B27 positive	79%**	80%**	60%***
IBP ASAS definition[†]	94%	95%	87%
Positive family history	34%	35%	25%
Peripheral arthritis	44%	43%	50%
Enthesitis	45%	45%	49%
Dactylitis	6%	6%	7%
Psoriasis	8%	8%	10%
IBD	5%	5%	7%
Acute anterior uveitis	22%	23%	17%
Elevated CRP	70%	71%	61%
Sacroiliitis on pelvic radiographs[‡]	78%	79%	68%
Sacroiliitis on pelvic MRI[‡]	82%***	83%***	76%***
Number of SpA features[§], mean (SD)	4 (2)	4 (2)	4 (2)
Use of bDMARD	33%*	33%*	30%*
Use of NSAID	99%*	99%*	97%*
Radiographic axSpA	79%	80%	73%
LoC regarding axSpA diagnosis, mean (SD)	8 (3)	8 (3)	7 (3)

* 5-10% missing values, ** 10-20% missing values, ***20-40% missing values

[†] 4 out of 5 of the following features: onset before the age of 40, insidious onset, improvement with exercise, no improvement with rest, pain at night[21]

[‡] Based on reading of local radiologists

[§] Excluding HLA-B27 status and sacroiliitis on imaging

ASAS, Assessment of SpondyloArthritis international Society; **axSpA**, axial spondyloarthritis; **bDMARD**, biological Disease Modifying Anti-Rheumatic Drug; **CRP**, C-reactive protein; **HLA-B27**, Human Leucocyte Antigen B27; **IBD**, Inflammatory Bowel Disease; **IBP**, Inflammatory Back Pain; **IQR**: Inter-Quartile Range; **LoC**, Level of Confidence regarding the diagnosis; **MRI**, Magnetic Resonance Imaging; **NSAIDs**, Non-Steroidal Anti Inflammatory Drugs; **SpA**, Spondyloarthritis.

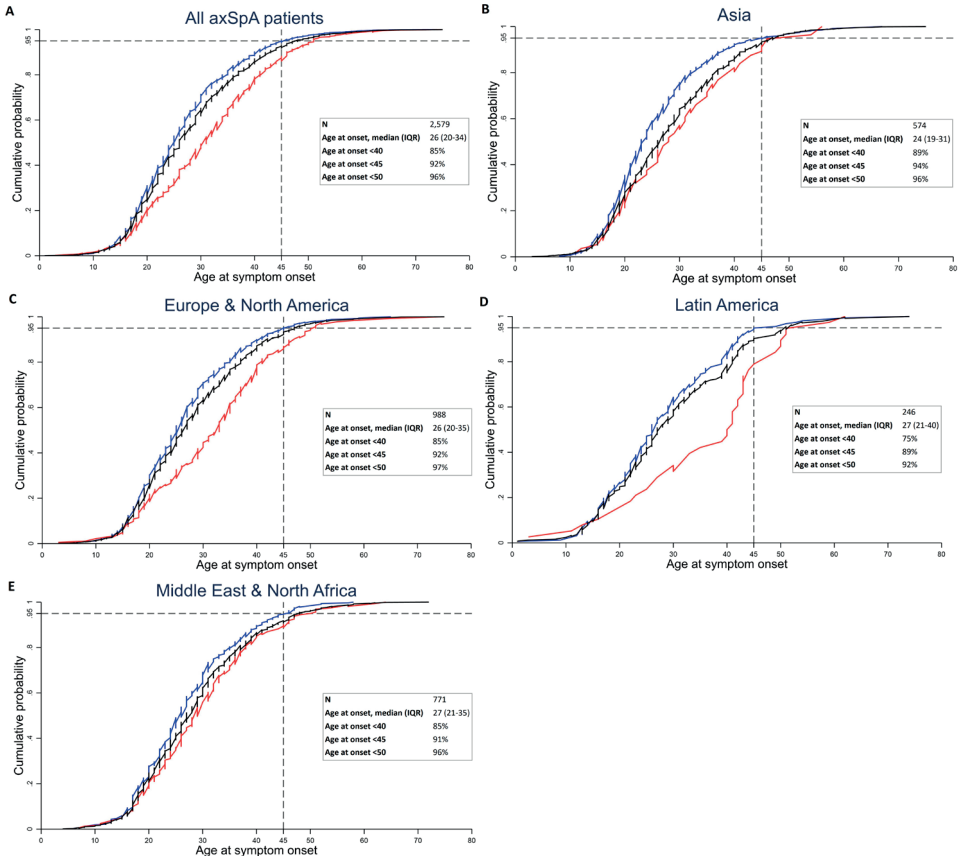


Figure 1 Cumulative distribution of the age at onset of axial symptoms, stratified by HLA-B27 status. **A** all included axial spondyloarthritis patients; **B** Asia; **C** Europe & North America; **D** Latin America; and **E** Middle East & North Africa. The black lines represent all patients in each region, the blue lines represent HLA-B27 positive patients, and the red lines represent the HLA-B27 negative patients. The horizontal dashed line represents the 95% point, and the vertical dashed line represents an age at onset of 45 years. IQR: Inter-Quartile Range

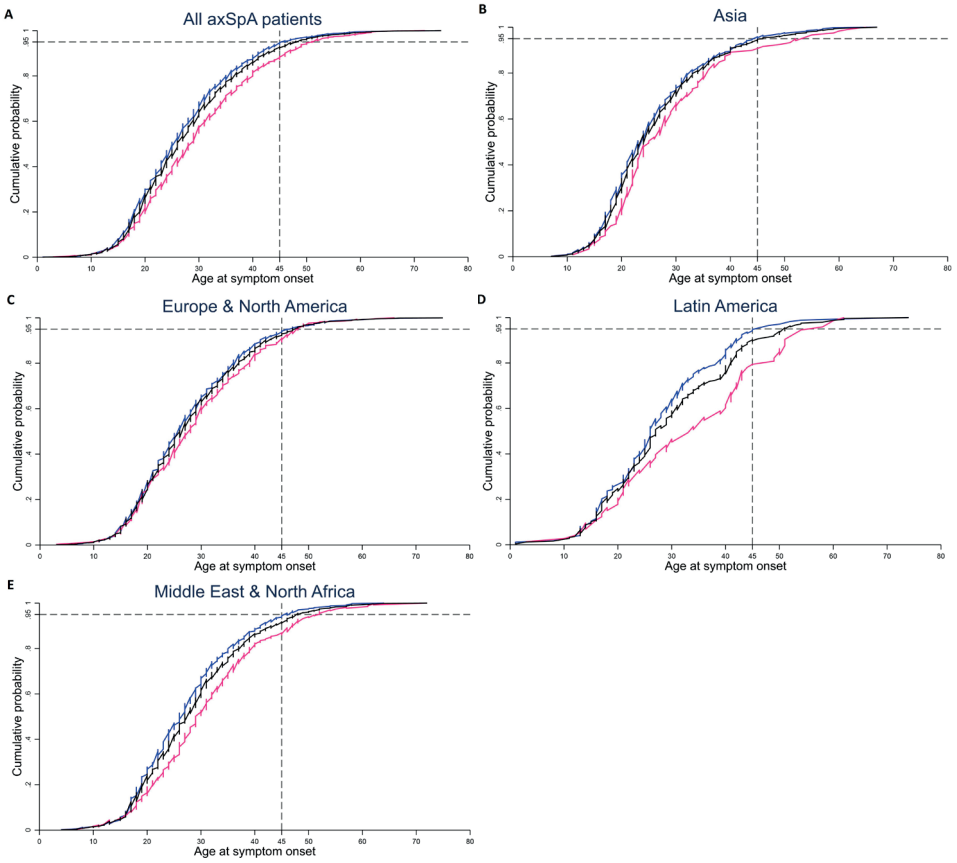


Figure 2 Cumulative distribution of the age at onset of axial symptoms, stratified by gender **A** all included axial spondyloarthritis patients; **B** Asia; **C** Europe & North America; **D** Latin America; and **E** Middle East & North Africa. The black lines represent all patients in each region, the blue lines represent male patients, and the pink lines represent the female patients. The horizontal dashed line represents the 95% point, and the vertical dashed line represents an age at onset of 45 years.

Table 3 Multivariable models assessing the effect of HLA-B27 & gender on age at onset of axial symptoms

	Multivariable linear regression	
	β (95% CI)	**p-value**
Total study population (n=2,063)		
HLA-B27		
Negative	*Ref.*	
Positive	**-4.35 (-5.45 ; -3.25)**	**<0.001**
Gender		
Female	*Ref.*	
Male	**-1.71 (-2.69 ; -0.74)**	**0.001**
Asia (n=525)		
HLA-B27		
Negative	*Ref.*	
Positive	**-3.68 (-6.44 ; -0.92)**	**0.009**
Gender		
Female	*Ref.*	
Male	**-2.23 (-4.34 ; -0.11)**	**0.039**
Europe & North America (n=862)		
HLA-B27		
Negative	*Ref.*	
Positive	**-5.18 (-6.88 ; -3.48)**	**<0.001**
Gender		
Female	*Ref.*	
Male	-0.92 (-2.38 ; 0.53)	0.215
Latin America (n=195)		
HLA-B27		
Negative	*Ref.*	
Positive	**-7.30 (-11.59 ; -3.00)**	**0.001**
Gender		
Female	*Ref.*	
Male	-3.44 (-7.11 ; 0.23)	0.066
Middle East & North Africa (n=481)		
HLA-B27		
Negative	*Ref.*	
Positive	**-2.44 (-4.36 ; -0.53)**	**0.013**
Gender		
Female	*Ref.*	
Male	-1.58 (-3.50 ; 0.34)	0.106

CI: confidence interval; Statistically significant associations are printed in **bold**

DISCUSSION

This study provides the first cumulative distribution of age at onset of axial symptoms in axSpA patients across the globe, showing that the vast majority of patients with axSpA have an age at onset before the age of 45 years in all parts of the world, which is consistent with the ASAS classification criteria for axSpA.

This study adds an important global perspective to what has been previously reported[1, 2, 4, 5]. Akin to what has been shown in previous studies[1, 2, 4], we found that patients with HLA-B27-negative disease had a significantly higher age at symptom onset than those with HLA-B27-positive disease and this held true in all geographical regions.

Contrary to Feldtkeller et al.[1], we showed a higher age at onset of axial symptoms in female patients compared to their male counterparts, which is in line with findings from other studies[2, 4, 5, 22, 23]. This difference may be partly explained by the fact that female patients were underrepresented in the study conducted by Feldtkeller et al.[1], possibly as a result of underdiagnosis of r-axSpA in women in the past[24]. Similar to Chung et al.[2], we found an additional effect of male gender and HLA-B27 on age at onset in multivariable analysis in the total included axSpA population, indicating a different association between HLA-B27 and age at onset for male and female patients. However, in multivariable analysis stratified by geographical region, an additional effect of male gender and HLA-B27 was only found in Asia. The current study adds important information to the work previously published, as the data presented in this study include patients with axSpA from across the globe. Also, patients had a rheumatologist-confirmed diagnosis rather than a self-reported diagnosis.
The precise pathophysiological mechanisms underlying axSpA remain unclear, but as different types of HLA-B27 are found in different parts of the world (e.g. HLAB*27:05 in Europe and HLA-B*27:04 in Asia) and the association between HLA-B27 and axSpA varies between races[19], one might have expected to find more variation in age at onset and its association with HLA-B27 across geographical regions. Race was unavailable in the ASAS-PerSpA dataset, yet we expect the majority of the patients included in each geographical region to identify with its most prominent race, hence a clear effect of race would have been seen in the data. Additionally, many other factors are thought to have an influence on the occurrence of axSpA, such as other genetic factors and differences in the human microbiome and environmental factors, such as smoking[5, 25], which makes the relative consistency in age at onset all the more intriguing.

A potential limitation of this study is the fact that data were collected cross-sectionally based on patient records and patient-reported information, which has resulted in some missing data, especially regarding HLA-B27 status, as this was not specifically analysed for this study. Nonetheless, all geographical regions contained both patients with HLA-B27-positive and-negative disease and patients whose HLA-B27 status was unknown were not different than those with non-missing data (data not shown).

CONCLUSION

Irrespective of geographical region, the majority of axSpA patients had an age at onset of axial disease before the age of 45 years and being an HLA-B27 carrier and male gender were associated with earlier disease onset around the globe, yet an independent effect of male gender on top of HLA-B27 was only found in Asian patients. These results provide crucial data for diagnosis, classification and policies aimed at improving recognition of axSpA.

SUPPLEMENTARY DATA

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

REFERENCES

1. Feldtkeller E, Khan MA, van der Heijde D, et al. Age at disease onset and diagnosis delay in HLA-B27 negative vs. positive patients with ankylosing spondylitis. *Rheumatol Int* 2003;23(2):61-6.
2. Chung HY, Machado P, van der Heijde D, et al. HLA-B27 positive patients differ from HLA-B27 negative patients in clinical presentation and imaging: results from the DESIR cohort of patients with recent onset axial spondyloarthritis. *Ann Rheum Dis* 2011;70(11):1930-36.
3. Skare TL, Leite N, Bortoluzzo AB, et al. Effect of age at disease onset in the clinical profile of spondyloarthritis: a study of 1424 Brazilian patients. *Clin Exp Rheumatol* 2012;30(3):351-7.
4. Rudwaleit M, Haibel H, Baraliakos X, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. *Arthritis Rheum* 2009;60(3):717-27.
5. Ciurea A, Scherer A, Weber U, et al. Age at symptom onset in ankylosing spondylitis: is there a gender difference? *Ann Rheum Dis* 2014;73(10):1908-10.
6. Rusman T, van Vollenhoven RF, van der Horst-Bruinsma IE. Gender Differences in Axial Spondyloarthritis: Women Are Not So Lucky. *Curr Rheumatol Rep* 2018;20(6):35.
7. Brophy S, Calin A. Ankylosing spondylitis: interaction between genes, joints, age at onset, and disease expression. *J Rheumatol* 2001;28(10):2283-8.
8. Braun J, Sieper J. Classification, diagnosis, and referral of patients with axial spondyloarthritis. *Rheum Dis Clin North Am* 2012;38(3):477-85.
9. Olivieri I, Salvarani C, Cantini F, et al. Ankylosing spondylitis and undifferentiated spondyloarthropathies: a clinical review and description of a disease subset with older age at onset. *Curr Opin Rheumatol* 2001;13(4):280-4.
10. van der Linden SM, Valkenburg HA, de Jongh BM, et al. 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(3):241-9.
11. Said-Nahal R, Miceli-Richard C, Berthelot JM, et al. The familial form of spondylarthropathy: a clinical study of 115 multiplex families. Groupe Français d'Etude Génétique des Spondylarthropathies. *Arthritis Rheum* 2000;43(6):1356-65.
12. Rudwaleit M, Metter A, Listing J, et al. Inflammatory back pain in ankylosing spondylitis: a reassessment of the clinical history for application as classification and diagnostic criteria. *Arthritis Rheum* 2006;54(2):569-78.
13. Rudwaleit M, van der Heijde D, Landewé R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68(6):777-83.
14. 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.
15. Boel A, Molto A, van der Heijde D, et al. Do patients with axial spondyloarthritis with radiographic sacroiliitis fulfil both the modified New York criteria and the ASAS axial spondyloarthritis criteria? Results from eight cohorts. *Ann Rheum Dis* 2019;78(11):1545-49.
16. Bendahan LT, Machado NP, Mendes JG, et al. Performance of the classification criteria in patients with late-onset axial spondyloarthritis. *Mod Rheumatol* 2018;28(1):174-81.
17. Chen HA, Chen CH, Liao HT, et al. Clinical, functional, and radiographic differences among juvenile-onset, adult-onset, and late-onset ankylosing spondylitis. *J Rheumatol* 2012;39(5):1013-8.
18. Dean LE, Jones GT, MacDonald AG, et al. Global prevalence of ankylosing spondylitis. *Rheumatology (Oxford)* 2014;53(4):650-7.
19. Khan MA. Polymorphism of HLA-B27: 105 subtypes currently known. *Curr Rheumatol Rep* 2013;15(10):362.
20. López-Medina C, Molto A, Sieper J, et al. Prevalence and distribution of peripheral musculoskeletal manifestations in spondyloarthritis including psoriatic arthritis: results of the worldwide, cross-sectional ASAS-PerSpA study. *RMD Open* 2021;7(1):e001450.
21. Sieper J, van der Heijde D, Landewé R, et al. New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the Assessment of SpondyloArthritis international Society (ASAS). *Ann Rheum Dis* 2009;68(6):784-88.
22. Ortolan A, van Lunteren M, Ramiro S, et al. Are gender-specific approaches needed in diagnosing early axial spondyloarthritis? Data from the SPondyloArthritis Caught Early cohort. *Arthritis Res Ther* 2018;20(1):218.
23. Tournadre A, Pereira B, Lhoste A, et al. Differences between women and men with recent-onset axial spondyloarthritis: results from a prospective multicenter French cohort. *Arthritis Care Res (Hoboken)* 2013;65(9):1482-9.
24. Feldtkeller E, Bruckel J, Khan MA. Scientific contributions of ankylosing spondylitis patient advocacy groups. *Current opinion in rheumatology* 2000;12(4):239-47.

25. de Koning A, Schoones JW, van der Heijde D, et al. Pathophysiology of axial spondyloarthritis: Consensus and controversies. *Eur J Clin Invest* 2018;48(5):e12913.

CHAPTER 4

GEOGRAPHICAL PREVALENCE OF FAMILY HISTORY IN PATIENTS WITH AXIAL SPONDYLOARTHRITIS AND ITS ASSOCIATION WITH HLA-B27 IN THE ASAS-PERSPA STUDY

Anne Boel*, Miranda van Lunteren*, Clementina López-Medina, Joachim Sieper, Désirée van der Heijde, Floris Alexander van Gaalen

* Shared co-first authorship: Anne Boel and Miranda van Lunteren contributed equally

RMD Open. 2022 Mar;8(1):e002174

ABSTRACT

Background

A positive family history (PFH) of spondyloarthritis (SpA) consists of five SpA-related entities, of which a PFH of axial spondyloarthritis (axSpA) is most common in European patients with axSpA. Moreover, a PFH of axSpA is associated with human leucocyte antigen B27 (HLA-B27) positivity in these patients. It is unknown if this holds true in patients with axSpA in other parts of the world.

Objective

To investigate the geographical prevalence of a PFH of SpA and its association with HLA-B27 positivity in patients with axSpA worldwide.

Methods

Cross-sectional analyses included patients from the ASAS peripheral involvement in Spondyloarthritis (PerSpA) study from 24 countries worldwide with an axSpA diagnosis, known HLA-B27 status and family history. Logistic regression models were built to assess the effect of HLA-B27 status on the occurrence of PFH. This was repeated for each of the five SpA entities in a PFH.

Results

Among 2048 patients, axSpA was the most common SpA entity in a PFH in all geographical regions (Asia 28%, Europe and North America 27%, Latin America 20%, Middle East and North Africa 41%). A PFH of axSpA was associated with HLA-B27 positivity in Asia (OR 4.19), Europe and North America (OR 2.09) and Latin America (OR 3.95), but not in the Middle East and North Africa (OR 0.98), which has a lower prevalence of HLA-B27 positivity. A PFH of other SpA entities was less prevalent and not consistently associated with HLA-B27 positivity.

Conclusion

In patients with axSpA worldwide, axSpA was consistently the most common SpA entity in a family history and was associated with HLA-B27 positivity in all geographical regions but one.

INTRODUCTION

The Assessment in SpondyloArthritis international Society (ASAS) has defined a positive family history (PFH) of spondyloarthritis (SpA) as a family history of axial spondyloarthritis (axSpA), psoriasis, reactive arthritis (ReA), acute anterior uveitis (AAU) or inflammatory bowel disease (IBD) in a first-degree or second-degree relative,1 and in this definition no distinction has been made between a PFH of axSpA or a PFH of another SpA entity (ie, psoriasis, ReA, AAU or IBD).

PFH can be used for different purposes: (1) as one of the clinical criteria in the ASAS classification criteria; (2) as a proxy for human leucocyte antigen B27 (HLA-B27) positivity in situations where HLA-B27 testing is not useful or not possible; and (3) as a risk factor for the development of axSpA.

The definition of PFH was based on consensus of experts and was not tested nor validated prior to its use as one of the SpA features and inclusion in the ASAS classification criteria for axSpA. Data from predominantly Western European cohorts have shown that in patients with axSpA the most common SpA entity in a family history is axSpA. Additionally, an association between PFH and HLA-B27 positivity was found. This association was driven by a PFH of axSpA and possibly AAU, but not by other SpA entities.2,3 While PFH has been regarded useful for identifying patients with back pain at risk of axSpA, research has shown that the diagnostic value of PFH is limited once the HLA-B27 status is known.4 However, these studies were limited to mostly Western European patients.

These findings suggest that, in the currently used classification criteria for axSpA, PFH may be overvalued—as both HLA-B27 and PFH are similarly weighted—and that its definition is likely too broad by including five SpA entities when classifying patients for scientific research. Furthermore, as axSpA was the most prevalent SpA entity and the association between HLA-B27 and PFH was driven by axSpA, the current definition of PFH might be too broad for identifying at-risk patients. Should the definition of PFH be revised in the future, a new definition must be applicable to patients all around the world. Additionally, it should be tested in different settings, as a different definition may be required for the different purposes described above. As PFH was defined based on experts' opinion, the first step is to investigate the prevalence of PFH in each geographical region. Thereafter, it should be investigated whether the same association between PFH and HLA-B27 positivity exists in other parts of the world (eg, Latin America or Africa).

The ASAS peripheral involvement in Spondyloarthritis (ASAS-PerSpA) study was conducted in 24 countries around the world and provides a unique opportunity to verify the findings of previous research in populations outside of Western Europe. The availability of worldwide

data on PFH provides a unique opportunity to assess the prevalence of PFH in various geographical regions, including populations outside of Western Europe.

Furthermore, the worldwide representation of patients in the ASAS-PerSpA study allows us to investigate the association between PFH and HLA-B27 across the globe. With this study, we aim to assess whether the association between PFH and HLA-B27 that was found in previous research applies to patients outside of Western Europe. Additionally, we aim to investigate whether this association is driven by a specific SpA entity.

MATERIAL AND METHODS

Data from the ASAS-PerSpA study were used in this study, of which detailed information can be found in the original publication.[5] In brief, ASAS-PerSpA was an observational study conducted in 24 countries around the world, in which data were collected cross-sectionally between 1 March 2018 and 29 February 2020. The countries were grouped in four geographical regions identical to the original publication: (1) Asia, (2) Europe and North America, (3) Latin America and (4) Middle East and North Africa (see online supplemental table S1). Its primary aim was to assess the prevalence of clinical peripheral rheumatological features in consecutively included patients with a diagnosis of SpA worldwide and to evaluate the validity of outcome measures of peripheral rheumatological features.[5]

Patients

For our analyses only patients with a definite axSpA diagnosis, as defined by the treating rheumatologist, from the ASAS-PerSpA study were included (radiographic or non-radiographic axSpA).

Patients included in this study had to have a known family history according to the ASAS definition1 and a known HLA-B27 status. If information on family history of more than two of the five SpA entities was missing for both first-degree and second-degree relatives, or if family history in first-degree relatives was missing for all SpA entities, family history was considered unknown and patients were excluded (online supplemental table S2 provides insight into the proportion of missing information for each SpA entity). However, if a complete family history of all five SpA entities was available for first-degree relatives only, patients were not excluded. Written informed consent was obtained from participants prior to inclusion.

Outcomes

The primary analysis of this study was to assess family history in all included patients diagnosed with axSpA and stratified by geographical region. The ASAS definition was used to define a PFH of SpA1; the presence of axSpA, AAU, psoriasis, IBD and/or ReA in first-degree or second-degree relatives was considered as a PFH. In this definition, parents, siblings and children are defined as first-degree relatives and maternal and paternal grandparents, aunt, uncle, niece and nephew as second-degree relatives. In this study, family history was analysed as a PFH of any SpA entity (ie, according to the ASAS definition) as well as a PFH of each specific SpA entity (ie, axSpA, AAU, psoriasis, IBD and ReA). Additionally, the association between a PFH of an extra-musculoskeletal manifestation (EMM) and presence (current or past) of the same EMM was also investigated. PFH was not split according to first-degree or second-degree relatives in the current study.

Other outcomes of interest were (1) the association between HLA-B27 positivity and family history; and (2) whether the association between HLA-B27 positivity and a family history of a specific SpA entity (eg, axSpA) was independent of a family history of the other four SpA entities.

Finally, we aimed to compare the prevalence of PFH in the PerSpA cohort with the prevalence of PFH in patients diagnosed with axSpA and patients with chronic back pain in the ASAS, DESIR DEvenir des Spondyloarthropathies Indifférenciées Récentes) and SPACE (SPondyloArthritis Caught Early) cohorts. Herein, descriptive statistics were used to compare the prevalence of PFH among HLA-B27-positive and HLA-B27-negative patients with an axSpA diagnosis in the PerSpA study with the prevalence of PFH in patients from the other (ax)SpA cohorts.[1 6–8]

Statistical analyses

The prevalence of PFH of all five SpA entities combined but also for each SpA entity separately was determined among all included patients, as well as stratified by HLA-B27 status per geographical region. Separate logistic regression models were used to assess the association between HLA-B27 status and each SpA entity in PFH in the total included axSpA population as well as stratified for each geographical region. Finally, multivariable logistic regression models were built to investigate if a family history of each specific SpA entity was associated with HLA-B27 positivity, independently of the other four SpA entities. These multivariable models were built for the total included study population as well as stratified per geographical region. In the models investigating the association of a PFH of axSpA and HLA-B27 positivity, a PFH of AAU, psoriasis, IBD and ReA was included as covariates.

RESULTS

Of the 2675 patients in the ASAS-PerSpA study diagnosed with radiographic or non-radiographic axSpA, 627 patients were excluded due to unknown HLA-B27 status (n=546) or unknown family history (n=81), as defined a priori. For the current analysis a total of 2048 patients from four geographical regions (Asia n=545, Europe and North America n=840, Latin America n=202 and Middle East and North Africa n=461) were included (online supplemental table S1). The median (IQR) age of the patients was 40 (31–50) years, 31% were female, with a median disease duration of 11 (5–20) years and a mean (SD) of 3 (2) SpA features (online supplemental table S3). Overall, there were only a small percentage of missing data (<5% unless indicated otherwise).

HLA-B27 positivity was 89% in Asia, 65% in the Middle East and North Africa, 78% in Europe and North America and 81% in Latin America. Patients in Latin America more often had concomitant peripheral symptoms—as shown by the higher proportions of peripheral arthritis, enthesitis and dactylitis compared with the other geographical regions—whereas psoriasis was more common in Europe and North America. Uveitis was the most common EMM and there was a significant association between a PFH of EMM and current or past presence of the same EMM in patients with axSpA: psoriasis: OR 4.95 (95% CI 3.43 to 7.13, p<0.001); AAU: OR 2.91 (95%CI 1.83 to 4.64, p<0.001); and IBD: OR 5.17 (95% CI 2.72 to 9.81, p<0.001).

Prevalence of family history by geographical region

A PFH of any of the SpA entities was most common in the Middle East and North Africa compared with the other geographical regions, which was largely due to a PFH of axSpA (figure 1). Across all geographical regions, a PFH of psoriasis was the second most common entity in a family history and a PFH of IBD, uveitis and ReA was uncommon.

Similar results were found when the data were stratified on HLA-B27 status (figure 2 and table 1); a PFH of axSpA was the most common regardless of HLA-B27 status in each geographical region, except for HLA-B27-negative patients in Europe and North America and Latin America in whom a PFH of psoriasis occurred the most. A PFH of psoriasis and IBD occurred more frequently in HLA-B27-negative patients than in HLA-B27-positive patients in all geographical regions.

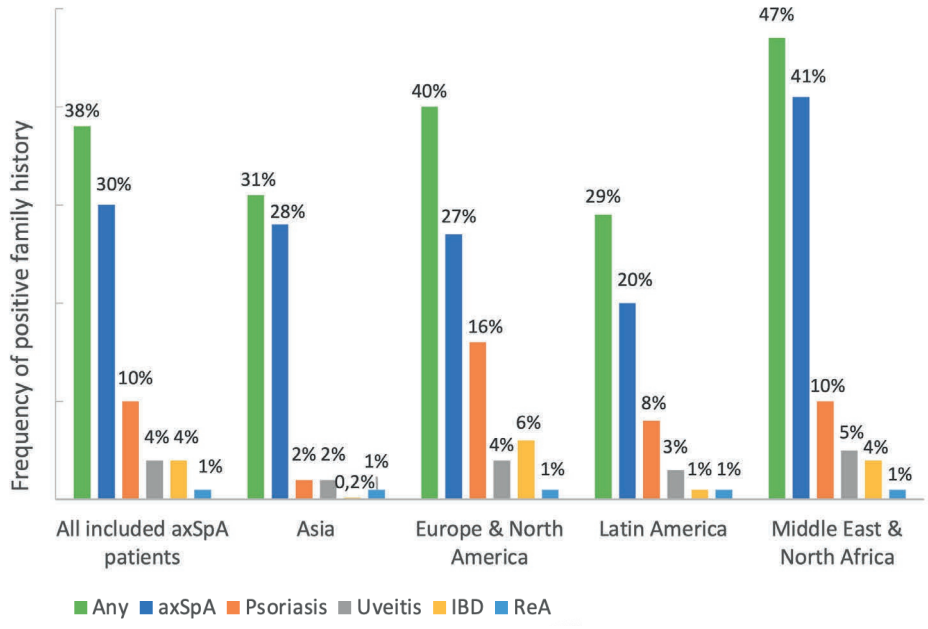


Figure 1 Frequency of a positive family history (any and for each SpA entity) stratified by geographical region **axSpA**, axial spondyloarthritis; **IBD**, Inflammatory Bowel Disease; **ReA**, reactive arthritis

Prevalence of a PFH compared with other cohorts

We compared the prevalence of PFH found in the ASAS-PerSpA cohort with those in HLA-B27-positive and HLA-B27-negative patients with an axSpA diagnosis and patients with chronic back pain (ie, no axSpA diagnosis) in other axSpA cohorts. Table 1 shows that a PFH of each specific SpA entity occurs in both HLA-B27-positive and HLA-B27-negative patients diagnosed with axSpA. This was not only apparent in the ASAS-PerSpA cohort, but also in the ASAS, DESIR and SPACE cohorts. Furthermore, across all cohorts the prevalence of a PFH of axSpA and uveitis was higher in HLA-B27-positive patients, whereas the prevalence of a PFH of psoriasis and IBD was higher in HLA-B27-negative patients. The ASAS and SPACE cohorts also show that a PFH of almost all SpA entities is present in similar frequencies among patients with chronic back pain suspected of axSpA who were eventually not diagnosed with axSpA.

Association between HLA-B27 positivity and family history

Univariable logistic regression models showed a positive association between HLA-B27 positivity and a PFH of axSpA in the total included axSpA population (OR 1.84) and when stratified by region, in Asia (OR 4.19), Europe and North America (OR 2.09) and Latin America (OR 3.95), but such an association was not found in the Middle East and North Africa (OR 0.98; table 2). A negative association with HLA-B27 was found for patients with a PFH of IBD in Europe and North America (OR 0.52) only; for patients with a PFH of

psoriasis, such an association was found in the Middle East and North Africa (OR 0.39) only. No associations between HLA-B27 positivity and a PFH of uveitis and ReA were found.

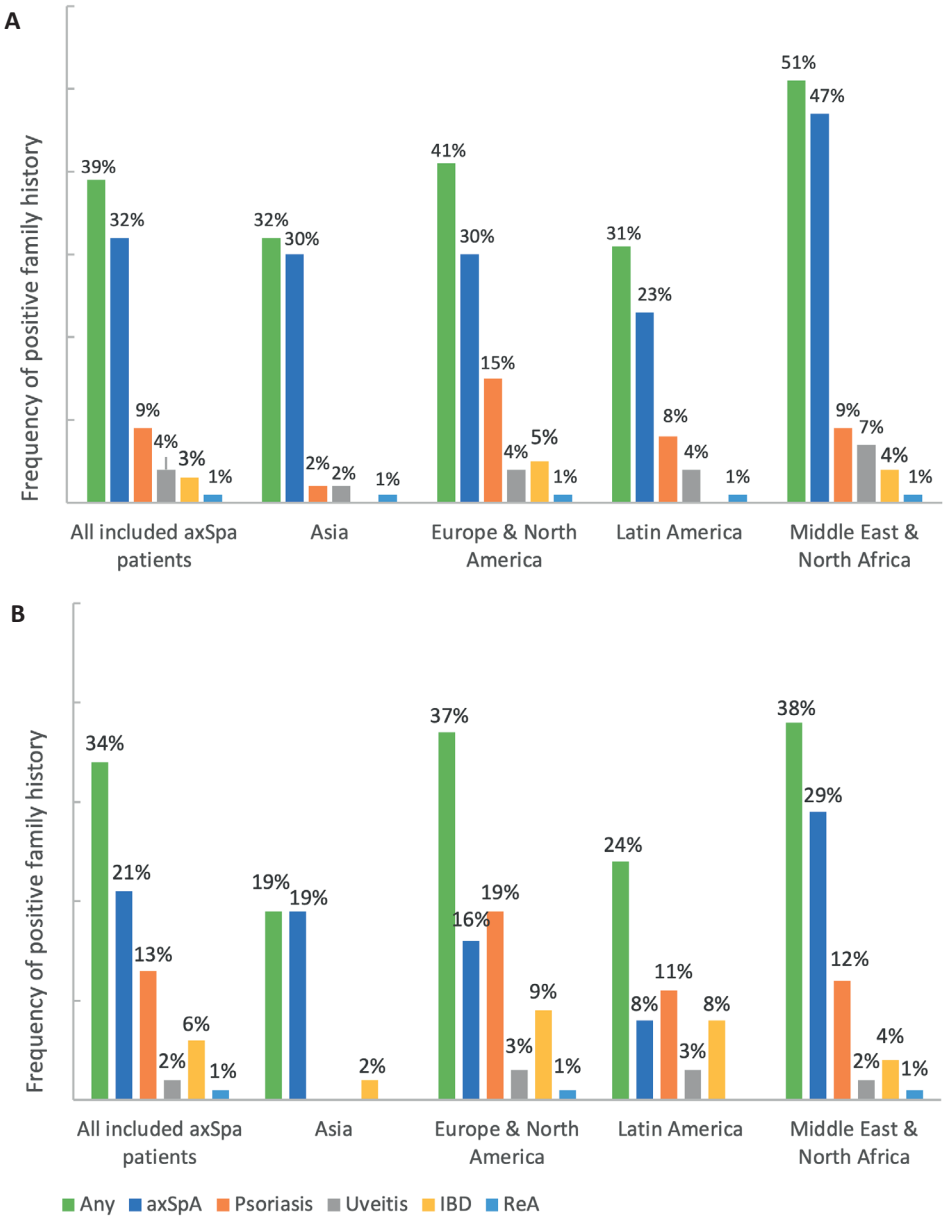


Figure 2 Frequency of a positive family history (any and per disease) stratified by geographical region for HLA-B27 positive (A) and negative (B) patients with axSpA.
axSpA, axial spondyloarthritis; **IBD**, Inflammatory Bowel Disease; **ReA**, Reactive Arthritis

Table 1 Frequency of positive family history among patients with axial spondyloarthritis in the ASAS, DESIR, ASAS-PerSpA and SPACE cohorts

	HLA-B27+ axSpA	HLA-B27- axSpA	No axSpA
ASAS-PerSpA cohort	***n=1609***	***n=439***	***not available***
Any positive family history	631 (39%)	149 (34%)	-
Family history axSpA	518 (32%)	90 (21%)	-
Family history uveitis	66 (4%)	10 (2%)	-
Family history psoriasis	148 (9%)	57 (13%)	-
Family history IBD	47 (3%)	27 (6%)	-
Family history ReA	14 (1%)	3 (1%)	-
ASAS cohort	***n=254***	***n=114***	***n=226***
Any positive family history	74 (29%)	17 (15%)	44 (19%)
Family history axSpA	61 (24%)	4 (4%)	22 (10%)
Family history uveitis	4 (2%)	-	3 (1%)
Family history psoriasis	12 (4%)	11 (10%)	13 (6%)
Family history IBD	1 (1%)	2 (2%)	9 (4%)
Family history ReA	3 (1%)	2 (2%)	3 (1%)
DESIR cohort	***n=410***	***n=297***	***not available***
Any positive family history	172 (42%)	101 (34%)	-
Family history axSpA	111 (27%)	28 (9%)	-
Family history uveitis	31 (8%)	1 (<1%)	-
Family history psoriasis	75 (18%)	66 (22%)	-
Family history IBD	17 (4%)	18 (6%)	-
Family history ReA	2 (<1%)	5 (2%)	-
SPACE cohort	***n=265***	***n=122***	***n=421***
Any positive family history	130 (49%)	45 (37%)	181 (44%)
Family history axSpA	76 (29%)	10 (8%)	81 (19%)
Family history uveitis	24 (9%)	3 (2%)	22 (5%)
Family history psoriasis	53 (20%)	33 (27%)	80 (19%)
Family history IBD	14 (5%)	9 (7%)	40 (10%)
Family history ReA	3 (1%)	1 (1%)	32 (8%)

Results are presented as n (%).
axSpA, axial spondyloarthritis; **HLA-B27**, Human Leucocyte Antigen B27; **IBD**, Inflammatory Bowel Disease; **IBP**, Inflammatory Back Pain; **ReA,** reactive arthritis; **SpA**, Spondyloarthritis.

Multivariable model

Multivariable logistic regression models showed that a PFH of axSpA was positively associated with HLA-B27 positivity independent of the presence of a PFH of other SpA entities in all included patients with axSpA (OR (95% CI) 1.95 (1.49 to 2.55), p<0.001). This association was also found in all geographical regions (Europe and North America: OR 2.37 (1.52 to 3.70), p<0.001; Latin America: OR 5.00 (1.13 to 22.06), p=0.034; Middle East and North Africa: OR 2.17 (1.40 to 3.35), p<0.001)), except for Asia which did show the same trend (OR 1.90 (0.93 to 3.87), p=0.077).

An inverse association with HLA-B27 positivity was found for patients with a PFH of IBD and psoriasis independent of the presence of a PFH of any of the other SpA entities in the total included axSpA population (IBD: OR 0.37 (0.22 to 0.61), p<0.001; psoriasis: OR 0.66 (0.46 to 0.93), p=0.018). When stratified by geographical region, a similar association was

only found in Europe and North America for a PFH of IBD (OR 0.45 (0.24 to 0.84), p=0.012). In other geographical regions the same trend was seen, but a PFH of IBD and psoriasis occurred less frequently than in Europe and North America (data not shown).

In multivariable models, no association with HLA-B27 was found for uveitis and ReA in the total included axSpA population (uveitis: OR 1.59 (0.78 to 3.22), p=0.198; ReA: OR 1.45 (0.41 to 5.22), p=0.565).

Table 2 Univariable associations between HLA-B27 and a positive family history stratified by geographical region

	HLA-B27+ n=1,609	**HLA-B27- n=439**	**OR (95% CI)**	**p-value**
Positive family history for axSpA				
Total population				
Present	**518**	**90**	**1.84 (1.43-2.38)**	**<0.001**
Absent	1,087	348	Ref.	
Per geographical region				
Asia	**144/487**	**11/58**	**4.19 (2.24-7.83)**	**<0.001**
Europe & North America	**196/658**	**30/182**	**2.09 (1.40-3.13)**	**<0.001**
Latin America	**37/164**	**3/38**	**3.95 (1.21-12.89)**	**0.023**
Middle East & North Africa	141/300	46/161	0.98 (0.69-1.40)	0.917
Positive family history for uveitis				
Total population				
Present	66	10	1.81 (0.92-3.56)	0.084
Absent	1,537	422	Ref.	
Per geographical region				
Asia	12/487	0/58	n.a.	n.a.
Europe & North America	28/658	5/182	1.54 (0.59-4.01)	0.379
Latin America	6/164	1/38	1.65 (0.20-13.72)	0.644
Middle East & North Africa	20/300	4/161	1.37 (0.47-4.04)	0.565
Positive family history for ReA				
Total population				
Present	14	3	1.26 (0.36-4.40)	0.719
Absent	1,583	427	Ref.	
Per geographical region				
Asia	5/487	0/58	n.a.	n.a.
Europe & North America	5/658	1/182	1.35 (0.16-11.58)	0.785
Latin America	2/164	0/38	n.a.	n.a.
Middle East & North Africa	2/300	2/161	0.27 (0.04-1.92)	0.191
Positive family history for IBD				
Total population				
Present	**47**	**27**	**0.45 (0.28-0.74)**	**0.001**
Absent	1,551	403	Ref.	
Per geographical region				
Asia	0/487	1/58	n.a.	n.a.
Europe & North America	**34/658**	**17/182**	**0.52 (0.29-0.94)**	**0.030**
Latin America	0/164	3/38	n.a.	n.a.
Middle East & North Africa	13/300	6/161	0.56 (0.21-1.49)	0.247

Table 2 Continued

	HLA-B27+ n=1,609	HLA-B27- n=439	OR (95% CI)	p-value
Positive family history for psoriasis				
Total population				
Present	**148**	**57**	**0.68 (0.49-0.94)**	**0.020**
Absent	1,461	382	Ref.	
Per geographical region				
Asia	9/487	0/58	n.a.	n.a.
Europe & North America	98/658	34/182	0.75 (0.50-1.13)	0.172
Latin America	13/164	4/38	0.85 (0.28-2.62)	0.777
Middle East & North Africa	**28/300**	**19/161**	**0.39 (0.21-0.70)**	**0.002**

axSpA, axial spondyloarthritis; **CI**, confidence interval; **HLA-B27**, Human Leucocyte Antigen B27; **IBD**, Inflammatory Bowel Disease; **IBP**, Inflammatory Back Pain; **n.a.**, not available **OR**, odds ratio; **ReA,** reactive arthritis; **Ref.**, reference category; **SpA**, Spondyloarthritis.

DISCUSSION

Across all geographical regions, axSpA was the most common SpA entity in a PFH while a PFH of ReA was rare, which is in line with previous findings from the ASAS, DESIR and SPACE cohorts.[2-4] In univariable stratified analyses, an association with HLA-B27 was apparent for a PFH of axSpA in all geographical regions except for the Middle East and North Africa. The absence of an association between HLA-B27 and a PFH of axSpA in the Middle East and North Africa may be caused by the high prevalence of a PFH of axSpA in HLA-B27-negative patients in that region combined with a lower prevalence of HLA-B27-positive disease.[9-11] Multivariable models showed that the association between a PFH of axSpA and HLA-B27 positivity was independent of the presence of a PFH of other SpA entities in all included patients diagnosed with axSpA in each geographical region. These findings confirm that the association between PFH and HLA-B27 status is largely driven by a PFH of axSpA, as was also shown in previous research, but is now confirmed in various regions worldwide.

In our study the associations between HLA-B27 positivity and a PFH of IBD and psoriasis were found to be of opposite direction, and an association was found only in Europe and North America and the Middle East and North Africa, respectively. Similar trends were found in the ASAS, DESIR and SPACE cohorts.[2-4] In a large cross-disease genetic study among chronic immune-mediated diseases, several SNPs (Single-Nucleotide Polymorphisms) were found with opposite directions of effect for AS, IBD and psoriasis analogous to the association of HLA-B27 with a PFH of axSpA, IBD and psoriasis in our study.[12] It is important to note that in the current study only patients with predominantly axial symptoms were investigated. We would like to emphasise that these results are therefore not applicable to patients with predominantly peripheral symptoms. In these patients a PFH of a different SpA entity such as psoriasis is likely more relevant.[13]

A predecessor of the ASAS classification criteria was the European Spondylarthropathy Study Group criteria,[14] where axSpA was considered part of a group of inflammatory diseases then known as spondylarthropathies. These included ReA, psoriatic arthritis, arthritis associated with IBD, a subgroup of juvenile chronic arthritis and ankylosing spondylitis.[14] This criteria set included PFH as a criterion and thus included all these individual SpA entities combined into a single feature for ease of use (ie, reduction of variables). This expert definition was not tested nor validated prior to inclusion in the ASAS classification criteria for axSpA but included on the basis of its performance as a criterion in previous studies.[14-21] In these studies, not only patients diagnosed with axSpA but also patients with other forms of SpA were analysed, and in almost all studies only the combined definition of a PFH of SpA (ie, in which a PFH of all SpA entities was combined) was studied.

Building on this, we emphasise that it is important to differentiate between the settings in which PFH is used. Settings can vary from using PFH for identifying patients at risk of axSpA before referral to using PFH for diagnosis and prognosis and to classification of patients with axSpA. The results from our study, but also the other cohorts, point towards adaptations of the definition of PFH in order to improve the sensitivity and/or specificity of PFH. Once this has been achieved, the role of PFH in the classification criteria should be re-evaluated too. Based on our data and three large axSpA cohorts (ASAS, DESIR, SPACE), we propose to investigate redefining a PFH of SpA to a PFH of axSpA (but not the other SpA entities) in a first-degree or second-degree family member to improve its sensitivity and/or specificity and to improve its position in the context of axSpA classification criteria.

The CLASSIC study (Classification of Axial Spondyloarthritis Inception Cohort) provides a unique opportunity to investigate the sensitivity and specificity of the classification criteria for axSpA, including this redefined definition, as this is a large worldwide prospective study which has been initiated by ASAS and SPARTAN (SPondyloArthritis Research and Treatment Network) to reassess the performance of the ASAS classification criteria for axSpA.

A major strength of this study is its worldwide character, which enabled us to investigate patients from various geographical regions. This study included patients from Asia, Latin America, and Middle East and North Africa, populations which have been largely neglected in previous research. Another strength is that family history was reported in detail, which allowed us to investigate PFH of each of the five SpA entities separately.

A major limitation is that family history is patient-reported, which could lead to an underestimation and overestimation for obvious reasons (eg, ReA in a distant relative). It requires specific follow-up questions from the healthcare professional collecting the information (eg, uncle with axSpA is only relevant if this is a blood relative). Nonetheless,

this is congruent to what is collected in clinical practice and the way the information of PFH is used. The only other way to collect information on SpA entities in the family is extensive family research, which is generally not feasible. Finally, data were collected cross-sectionally and only available data were used in this study. This resulted in a percentage of patients for whom HLA-B27 status was unknown (20%). However, this mirrors clinical practice worldwide too, as these were all patients visiting the rheumatological outpatient clinics where testing HLA-B27 is not always deemed informative or feasible.

CONCLUSION

In conclusion, across the globe, axSpA is the most common entity of SpA in a family history, expanding what was found in the ASAS, DESIR and SPACE cohorts to a more global perspective. In all geographical regions except the Middle East and North Africa, a PFH of axSpA was associated with HLA-B27 positivity in patients with axSpA. Although the prevalence of HLA-B27 positivity is relatively low in the Middle East and North Africa compared with other geographical regions, a PFH of axSpA was the most common form of PFH in this region identical to the other geographical regions.

Given the consistent findings from this study and other cohorts, the current expert definition of a PFH of SpA may be redefined to a PFH of axSpA, including only the presence of axSpA. This new definition should be re-evaluated by assessing if this definition improves the sensitivity and/or specificity of PFH and its role in the classification criteria for axSpA.

Given the similar pattern of PFH around the world, it is expected that a refined definition will be applicable to all parts of the world.

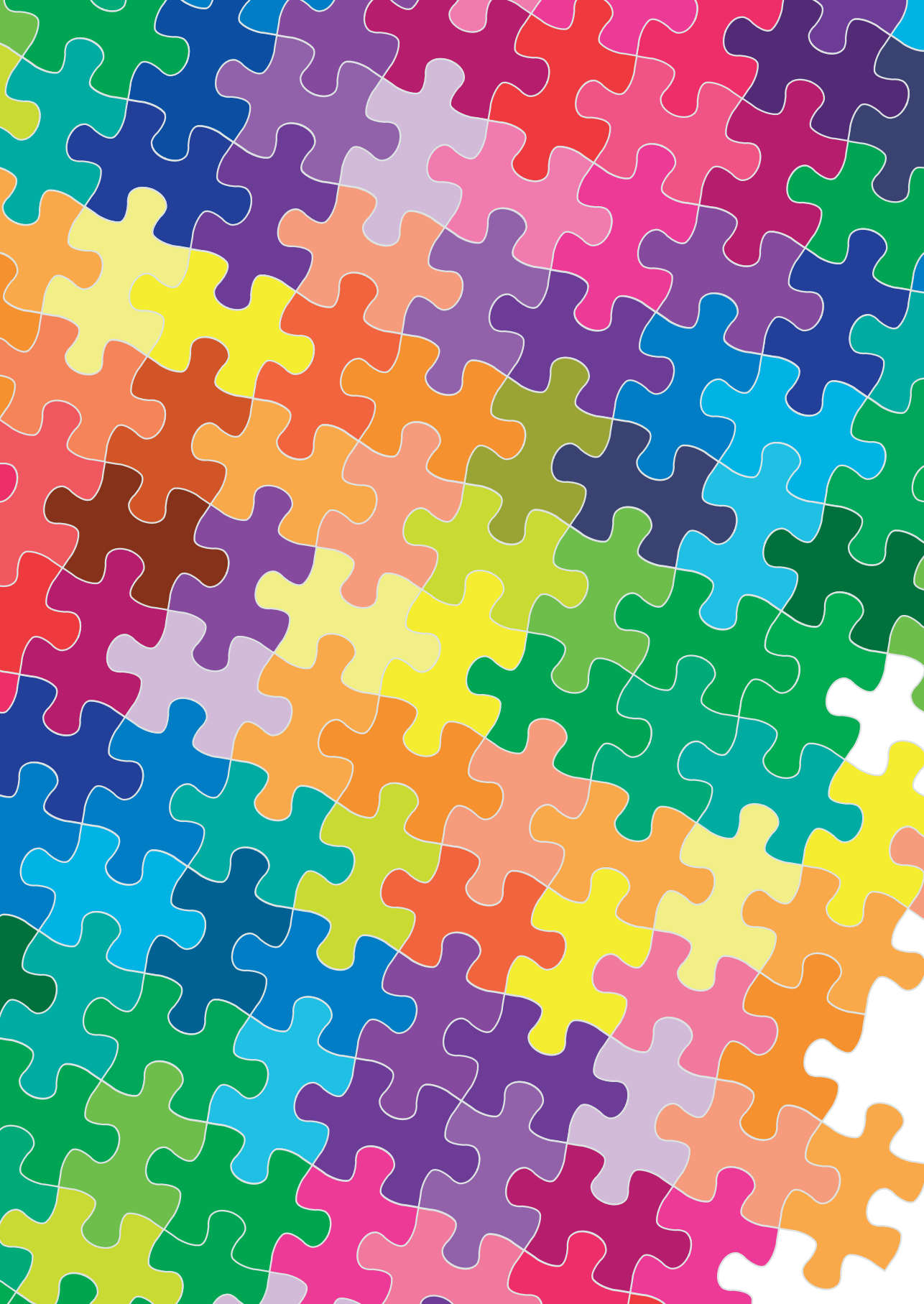
SUPPLEMENTARY DATA

Supplementary data are published online on the website of RMD Open.

REFERENCES

1. Rudwaleit M, van der Heijde D, Landewé R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68:777-83.
2. van Lunteren M, Sepriano A, Landewé R, Sieper J, Rudwaleit M, van der Heijde D, et al. Do ethnicity, degree of family relationship, and the spondyloarthritis subtype in affected relatives influence the association between a positive family history for spondyloarthritis and HLA-B27 carriership? Results from the worldwide ASAS cohort. *Arthritis Res Ther* 2018;20:166.
3. Ez-Zaitouni Z, Hilkens A, Gossec L, Berg IJ, Landewé R, Ramonda R, et al. Is the current ASAS expert definition of a positive family history useful in identifying axial spondyloarthritis? Results from the SPACE and DESIR cohorts. *Arthritis Res Ther* 2017;19:118.
4. van Lunteren M, van der Heijde D, Sepriano A, Berg IJ, Dougados M, Gossec L, et al. Is a positive family history of spondyloarthritis relevant for diagnosing axial spondyloarthritis once HLA-B27 status is known? *Rheumatology* 2019;58:1649-54.
5. López-Medina C, Molto A, Sieper J, Duruöz T, Kiltz U, Elzorkany B, et al. Prevalence and distribution of peripheral musculoskeletal manifestations in spondyloarthritis including psoriatic arthritis: results of the worldwide, cross-sectional ASAS-PerSpA study. *RMD Open* 2021;7:e001450.
6. Dougados M, d'Agostino MA, Benessiano J, Berenbaum F, Breban M, Claudepierre P, et al. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. *Joint Bone Spine* 2011;78:598-603.
7. van den Berg R, de Hooge M, van Gaalen F, Reijnierse M, Huizinga T, van der Heijde D. Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. *Rheumatology* 2013;52:1492-9.
8. Rudwaleit M, Landewé R, van der Heijde D, Listing J, Brandt J, Braun J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Ann Rheum Dis* 2009;68:770-6.
9. Mustafa KN, Hammoudeh M, Khan MA. HLA-B27 Prevalence in Arab Populations and Among Patients with Ankylosing Spondylitis. J Rheumatol 2012;39:1675-7.
10. Stolwijk C, Boonen A, van Tubergen A, Reveille JD. Epidemiology of spondyloarthritis. *Rheum Dis Clin North Am* 2012;38:441-76.
11. Ziade N, Abi Karam G, Merheb G, Mallak I, Irani L, Alam E, et al. HLA-B27 prevalence in axial spondyloarthritis patients and in blood donors in a Lebanese population: Results from a nationwide study. *Int J Rheum Dis* 2019;22:708-14.
12. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* 2016;48:510-8.
13. Rudwaleit M, van der Heijde D, Landewé R, Akkoc N, Brandt J, Chou CT, et al. The Assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. *Ann Rheum Dis* 2011;70:25-31.
14. 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.
15. Boyer GS, Templin DW, Goring WP. Evaluation of the European Spondylarthropathy Study Group preliminary classification criteria in Alaskan Eskimo populations. *Arthritis Rheum* 1993;36:534-8.
16. Collantes-Estevez E, Cisnal del Mazo A, Muñoz-Gomariz E. Assessment of 2 systems of spondyloarthropathy diagnostic and classification criteria (Amor and ESSG) by a Spanish multicenter study. European Spondyloarthropathy Study Group. *J Rheumatol* 1995;22:246-51.
17. Hukuda S, Minami M, Saito T, Mitsui H, Matsui N, Komatsubara Y, et al. Spondyloarthropathies in Japan: nationwide questionnaire survey performed by the Japan Ankylosing Spondylitis Society. *J Rheumatol* 2001;28:554-9.
18. Cury SE, Vilar MJ, Ciconelli RM, Ferraz MB, Atra E. Evaluation of the European Spondylarthropathy Study Group (ESSG) preliminary classification criteria in Brazilian patients. *Clin Exp Rheumatol* 1997;15:79-82.
19. Baddoura R, Awada H, Okais J, Habis T, Attoui S, Abi Saab M. Validation of the European Spondylarthropathy Study Group and B. Amor criteria for spondylarthropathies in Lebanon. *Rev Rhum Engl Ed* 1997;64:459-64.

20. Sadowska-Wróblewska M, Filipowicz A, Garwolinska H, Michalski J, Rusiniak B, Wróblewska T. Clinical symptoms and signs useful in the early diagnosis of ankylosing spondylitis. *Clin Rheumatol* 1983;2:37-43.
21. Sieper J, van der Heijde D, Landewé R, Brandt J, Burgos-Vagas R, Collantes-Estevez E, et al. New criteria for inflammatory back pain in patients with chronic back pain: a real patient exercise by experts from the Assessment of SpondyloArthritis international Society (ASAS). *Ann Rheum Dis* 2009;68:784-8.



PART II

ASAS/OMERACT CORE SET FOR AXIAL SPONDYLOARTHRITIS

CHAPTER 5

DOMAINS TO BE CONSIDERED FOR THE CORE OUTCOME SET OF AXIAL SPONDYLOARTHRITIS: RESULTS FROM A 3-ROUND DELPHI SURVEY.

Anne Boel, Victoria Navarro-Compán, Annelies Boonen, Philip Mease, Uta Kiltz, Maxime Dougados, Robert Landewé, Désirée van der Heijde

J Rheumatol 2021; 48(12):1810-1814

ABSTRACT

Background

Advances in the field of axial spondyloarthritis (axSpA) and the methodology to develop core sets have led the Assessment of SpondyloArthritis international Society (ASAS) group to update the ASAS–Outcomes in Rheumatology (OMERACT) core set. An important aspect was to ensure it would be applicable to the entire spectrum of axSpA. The first step was to define the most relevant disease domains.

Methods

A 3-round Delphi survey was conducted to gather opinions of 188 patients and 188 axSpA experts to define the most relevant disease domains to be included in the core set. The Delphi survey evaluated 2 separate research settings: (1) studies assessing symptom-modifying therapies; and (2) studies evaluating disease-modifying therapies. Importance of the domains was rated on a 1–9 Likert scale. A domain was considered for inclusion if, for both stakeholder groups, ≥ 70% of participants scored the domain as critical (7–9) and ≤ 15% scored it as not important (1–3) after 3 rounds.

Results

A total of 132 (70%) patients and 135 (72%) experts completed at least 1 round. After 3 rounds, 7 domains (pain, physical function, stiffness, disease activity, mobility, overall functioning and health, peripheral manifestations) were selected for the symptom-modifying therapies setting. For the disease-modifying therapies setting, 6 domains (physical function, disease activity, mobility, structural damage, extra-musculoskeletal manifestations, peripheral manifestations) were selected. All domains selected by experts were also selected by patients. Patients selected all offered domains except emotional function.

Conclusion

This study provides the domains selected by patients and axSpA experts that should be considered for the core set for axSpA.

INTRODUCTION

The Assessment of SpondyloArthritis international Society (ASAS) collaborated with Outcome Measures in Rheumatology (OMERACT) to develop a core outcome set for ankylosing spondylitis (AS) in 1999[1]. The core set has been well implemented in the field in the past 20 years[2]. Nevertheless, since the development of the original core set, it has become apparent that AS belongs to the broader disease spectrum of axial spondyloarthritis (axSpA), which consists of 2 subtypes: radiographic axSpA (also known as AS) and non-radiographic axSpA[3]. Further, there have been major advances in outcome instruments in the field of axSpA, such as the use of magnetic resonance imaging[4], and the development of the Ankylosing Spondylitis Disease Activity Score (ASDAS)[5], validated enthesitis scores[6], the ASAS-health index[7] and the ASAS-flare definition[8].

In addition, the methodology to develop core outcome sets has improved. Although there is no gold standard for the development or update of a core set, in the last few years OMERACT and the Core Outcome Measures in Effectiveness Trials (COMET) have worked exhaustively to provide specific guidance on how a core set should be developed (e.g. OMERACT handbook[9] and OMERACT Filter 2.0[10], COMET handbook[11] and Core Outcome Set-Standards for Development[12]). Because of all these advances, the ASAS group decided it was necessary to update the original ASAS-OMERACT core set for AS according to the current recommended methodology, to ensure the core set will be applicable to the entire spectrum of axSpA.

An important step in the process of updating the core outcome set was determining which disease domains (outcomes) are relevant. In order to establish these, a 3-round Delphi survey was employed to gather opinions from relevant stakeholders. The results of this 3-round Delphi survey formed the basis of the proposal for a final core set according to the new format of the OMERACT Onion[13]. Subsequently, the proposal was presented to OMERACT to seek endorsement for the proposed core domain set. A detailed description of the entire process that led to the selection of domains for the updated core set will be published separately. The methods used to compose and execute the Delphi survey, as well as its results, are described in the current paper. The aim of this study was to select the domains that should be considered for inclusion in the core set for axSpA.

MATERIALS AND METHODS

Preparation of the Delphi survey

The original core set[14] was developed for 3 different scenarios: 1) therapies that improve the symptoms and clinical features of inflammatory manifestations of the disease (symptom-modifying antirheumatic therapy [SMART]; this includes physical therapy); 2) therapies that change the course of disease by decreasing inflammatory manifestations (thereby improving function) and by preventing or decreasing structural damage (disease-modifying antirheumatic drugs [DMARDs]); and 3) clinical record keeping in daily practice, to facilitate uniform clinical record keeping to enable research from clinical records and to monitor patient care in a standardized way.

The core set update focused only on the first 2 scenarios. Thus, the Delphi survey consisted of 2 separate sections: one focused on the outcomes to be included in the core set for studies assessing symptom-modifying therapies, the other on the outcomes to be included in the core set for studies evaluating disease-modifying therapies.

A list of candidate domains to include in the Delphi survey was computed using 3 sources: 1) the current core set for AS[14]; 2) all domains assessed in studies evaluating pharmacological and nonpharmacological interventions identified in the systematic literature review (SLR) that assessed the implementation of the original core set[2] (to ensure the most recent studies were included, the search strategy from the SLR was used to identify studies published thereafter, i.e., between 2011 and 2018); and 3) information collected on the qualitative studies and patient focus group interviews conducted as part of the development of the ASAS/World Health Organization Comprehensive and Brief Core sets of the International Classification of Functioning, Disability and Health (ICF) for AS[15,16]. All aspects of health identified in this process were considered when defining candidate domains for the core set for axSpA.

After eliminating duplicates, the list of candidate domains was grouped and finalized by 3 of the authors (DvdH, VNC, AB) and later agreed on by the steering committee. The first round of the Delphi survey contained 11 candidate domains for symptom-modification therapies and 12 candidate domains for disease-modification therapies (the same domains with 1 additional domain representing structural damage). For this first round, participants had the opportunity to suggest additional domains.

Participants

The invited participants were divided in 2 main stakeholder groups: 1 group consisted of patients with axSpA and the other group consisted of a variety of expert stakeholders (all ASAS members, including rheumatologists, other healthcare professionals, methodologists,

and researchers, as well as representatives from the pharmaceutical industry and drug regulatory agencies), labelled as axSpA experts. The ASAS members were informed they would be invited to participate in the Delphi survey to update the current core set in an annual meeting prior to commencement of the project. Representatives of the pharmaceutical industry and drug regulatory agencies were informed of the project by email and invited to participate prior to commencement of the project. Patients were recruited through 3 national patient societies (Spondylitis Association of America, National Ankylosing Spondylitis Society, and Canadian Spondylitis Association) and were eligible to participate if they were aged ≥ 18 years and had a diagnosis of axSpA from their rheumatologist. Information regarding the Delphi survey and its purpose was posted on the websites of each of the organizations, and patients were invited to participate by email through their associations. Recruitment ceased once the group of patients was equal in size to the group of experts (n = 188). Ethical approval and consent to participate in the Delphi survey was not required based on the Dutch Medical Research Involving Human Subjects Act (WMO).

Content of the Delphi survey

An explanatory text was provided at the beginning of the survey in each round, containing information on the purpose of the Delphi and relevant information to fully understand the content and scoring system. This information was adapted per stakeholder group, using lay wording and more extensive explanations for the patients.

The main objective of this Delphi survey was to select the most relevant disease domains to be included in the core set for axSpA. Simultaneously, this survey was used to investigate the effect of invitation approach on the response rate and final outcome of a Delphi survey. The methods and results of this experiment are published separately[17]. In summary, the participants were not aware of the experiment and received identical information regarding the Delphi survey. All participants knew from the start that this was a 3-round Delphi but did not know that for half of the participants, an invitation for the second and third rounds was conditional on responding to the first round. The experiment on the 2 different ways of inviting participants showed no effect on the final results of the Delphi survey[17]. For the purpose of the domain selection, it was predetermined that the information from all participants regardless of invitation approach would be used. Here we present the results of the Delphi for the 2 different stakeholders that will be used for the core set.

The Delphi survey was split according to the 2 established scenarios (i.e., SMART and DMARD) and grouped by domain (i.e., participants who were invited to vote on the relevance of a specific domain in symptom-modifying therapies first and immediately thereafter on the same domain in disease-modifying therapies). This procedure was

maintained for all domains except structural damage, which was offered for voting only in the disease-modifying therapies section of the survey. A definition was provided for each domain in all rounds, including a brief explanation and examples (Supplementary Table 1, available with the online version of this article).

In each round, the participants received summarized information of the previous round, including their individual score and aggregated scores from their respective stakeholder group. Participants who responded for the first time to the invitation for round 2 received only aggregated scores of the first round, and the same procedure applied to round 3.

Each round was open for 2–3 weeks and a single reminder was sent after 1 week to those who did not yet complete the round. Data were collected online using SurveyMonkey (www.surveymonkey.com) between November 2 and December 30, 2018.

Domain selection

To identify the importance of each of the domains for the core set, each participant was asked to provide 1 score per domain using a 9-point Likert scoring system. Domains were graded according to their level of importance. Following the OMERACT handbook, a score of 1–3 signified an outcome as not important, 4-6 as important but not critical, and 7–9 as critical.9 The aggregated scores per domain were analyzed separately for each of the stakeholder groups. If a domain was scored as critical by ≥ 80% of the participants in a stakeholder group, the domain was selected for consideration in the core set and was not offered for voting in subsequent rounds for this stakeholder group. If a domain did not achieve this score, the predefined criteria to include a domain in the next round of the Delphi per stakeholder group were as follows: ≥ 50% of the participants scored the domain as critical; and ≤ 20% scored the domain as not important.

Finally, a domain was considered for inclusion in the core set if, for both stakeholder groups (experts and patients), ≥ 70% of participants scored the domain as critical and ≤ 15% scored it as not important after 3 rounds; this is in line with the guidelines provided in the OMERACT handbook[9].

Statistical analysis

For the purpose of this study, we used descriptive statistics to present the data. To determine which domains fulfilled the criteria to be considered for inclusion, the proportion of participants voting critical, important but not critical and not important were calculated.

RESULTS

In total, 376 participants were invited to participate: 188 patients and 188 axSpA experts. Patients were from 3 countries in 2 continents, and axSpA experts were from 41 countries from 5 continents (Supplementary Table 2, available with the online version of this article). The axSpA experts who completed at least 1 round consisted of 123 rheumatologists (of whom 10 were also methodologists and 2 were also patient representatives), 4 physiotherapists, 4 representatives from pharmaceutical companies, 2 radiologists, and 2 researchers.

Participants

The overall response rate was 49% for the patients, and 58% for the axSpA experts after the final round. In addition, round 1 and 2 were completed by 63% and 52% of patients and 60% and 55% of axSpA experts respectively.

Content of the Delphi survey

In round 1, stiffness was mentioned by multiple axSpA experts and was therefore added to the list of domains from round 2 onwards for both scenarios (i.e. symptom and disease modifying therapies). Supplementary Table 3 (available with the online version of this article) provides an overview of the domains that were offered for voting in each round for each of the stakeholder groups.

Domain selection

Table 1 and Table 2 present the proportion of critical votes per domain after the final round (split by stakeholder group) for the symptom- and disease-modifying therapy scenarios, respectively, wherein the domains voted as critical by ≥ 70% and not important by ≤ 15% are printed in bold. Supplementary Tables 4 and 5 (available with the online version of this article) present additional information on the proportions of critical, important but not critical, and not important votes per domain per round.

For the symptom-modifying therapies, 7 domains were voted as critical by ≥ 70% of patients and axSpA experts after 3 rounds. These were as follows: disease activity, pain, overall functioning and health, physical function, mobility, peripheral manifestations, and stiffness (Table 1). An additional 4 domains were voted as critical by ≥ 70% of patients only; in fact, the domain emotional function was the only domain voted critical by < 70% of patients. There were no domains voted as critical by ≥ 70% of axSpA experts only.

For the disease-modifying therapies, 6 domains were selected by ≥ 70% of patients and axSpA experts after the final round. These were as follows: disease activity, physical function, mobility, peripheral manifestations, extra-musculoskeletal manifestations, and structural

damage (Table 2). An additional 6 domains were voted as critical by ≥ 70% of patients only, who selected all domains except emotional function. Identical to the symptom-modifying therapies scenario, there were no domains voted as critical by experts only.

The domains that were voted as critical by ≥ 70% and voted not important by ≤ 15% in both stakeholder groups are presented in Figure 1.

Table 1 Proportion of critical votes per domain after the final round for the symptom modifying therapies scenario, split by stakeholder group.

	Patients			axSpA experts		
	N	Count	%	N	Count	%
Symptom modifying therapies scenario						
Disease activity	97	85	**88**	113	110	**97**
Pain	119	115	**97**	113	98	**87**
Fatigue	119	99	**83**	109	56	51
Sleep	119	96	**81**	109	22	20
Overall functioning & health	119	96	**81**	103	89	**86**
Physical function	119	109	**92**	113	98	**87**
Emotional function	93	57	61	103	13	13
Work & Employment	93	72	**77**	109	34	31
Mobility	119	104	**87**	109	81	**74**
Peripheral manifestations	119	98	**82**	109	90	**83**
Extra-musculoskeletal manifestations	119	99	**83**	109	74	68
Stiffness	97	87	**90**	109	94	**86**

Domains voted critical by ≥70% and not important by ≤15% of participants are printed in **bold**. **axSpA,** axial spondyloarthritis

Table 2 Proportion of critical votes per domain after the final round for the disease modifying therapies scenario, split by stakeholder group.

	Patients			axSpA experts		
	N	Count	%	N	Count	%
Disease modifying therapies scenario						
Disease activity	119	106	**89**	113	99	**88**
Pain	119	113	**95**	109	71	65
Fatigue	97	87	**90**	113	40	35
Sleep	93	72	**77**	113	18	16
Overall functioning & health	119	102	**86**	109	73	67
Physical function	119	109	**92**	103	90	**87**
Emotional function	93	52	56	113	12	11
Work & Employment	93	68	**73**	109	31	28
Mobility	119	105	**88**	109	88	**81**
Peripheral manifestations	119	98	**82**	109	78	**72**
Extra-musculoskeletal manifestations	119	102	**86**	109	77	**71**
Structural damage	119	102	**86**	113	95	**84**
Stiffness	97	87	**90**	109	53	49

Domains voted critical by ≥70% and not important by ≤15% of participants are printed in **bold**. **axSpA,** axial spondyloarthritis

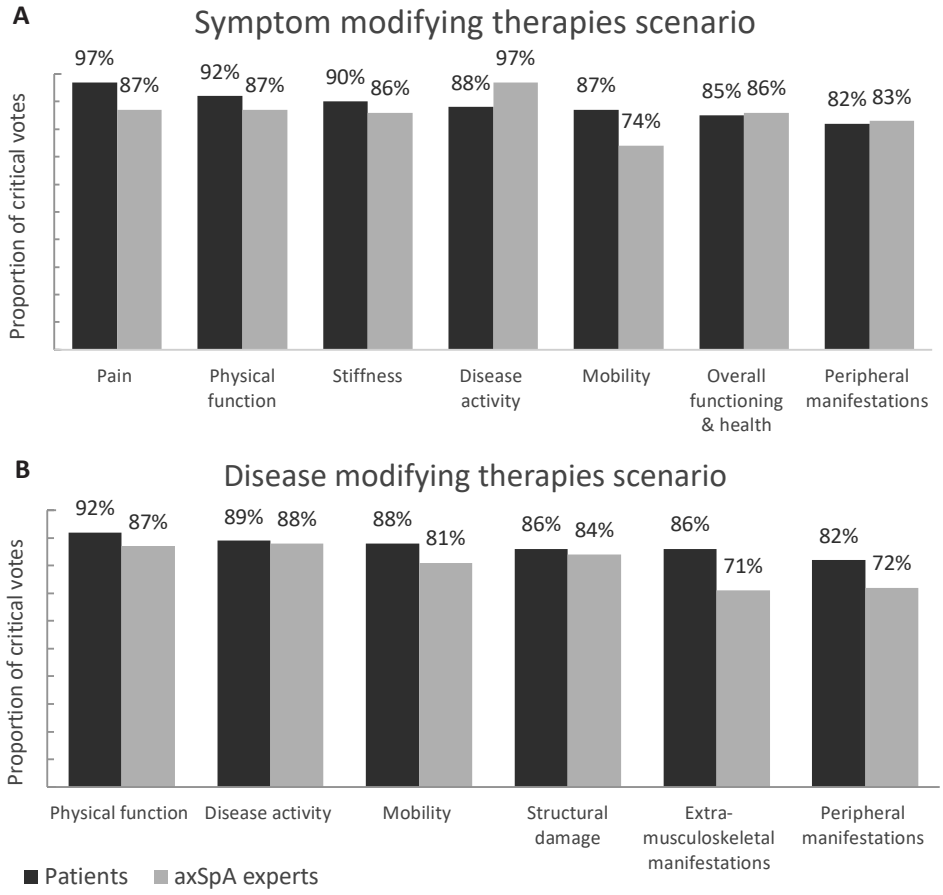


Figure 1 Domains selected after 3 rounds by patients (dark) and axSpA experts (light) in the setting assessing (**A**) symptom modifying therapies and (**B**) disease modifying therpies, including the percentage of critical votes. **axSpA,** axial spondyloarthritis

DISCUSSION

This 3-round Delphi survey was an important step in the process to update the core outcome set and aimed to determine which domains should be considered for inclusion according to patients and axSpA experts. Only 1 additional domain (stiffness) was added from round 2 onward, indicating that the candidate domains identified in the preparatory steps were a good representation of the domains of interest in the field.

In our study, patients selected more domains to be included in the core set compared with the axSpA experts. Specifically, domains such as fatigue, sleep, and work and employment were deemed very important by patients, but less so by experts. These domains have a major effect on the daily life of the patient, but are not necessarily specific to the disease; this could explain the difference in importance between patients and experts. In general, axSpA experts deemed the more objectively measurable domains such as structural damage and mobility as most critical to be measured in settings investigating disease-modifying therapies, whereas the more subjective domains such as pain, stiffness, and overall functioning and health were limited to the settings investigating symptom-modifying treatments.

The domain with the highest percentage of critical votes in the group of axSpA experts in both settings was disease activity, indicating that this domain is most important to measure in all trials investigating therapies for axSpA according to experts; in patients, however, the highest percentage of critical votes in both settings was for pain. There was a noticeable difference in the voting for the domain pain in the disease-modifying therapies setting, wherein 95% of patients voted it as critical, yet only 65% of the experts deemed this domain important enough to be measured in all trials investigating DMARDs.

A large panel of international axSpA experts and patients were invited to participate in this study. The use of an electronically distributed Delphi ensured no travel was required and anonymity was guaranteed. Further, no public speaking was required, which is known to increase patient participation[18]. Despite these measures, not all continents were equally represented, as the majority of axSpA experts who responded were from Europe and America, and invitations to patients were restricted to native English speakers to ensure understanding of the survey and its components. Nevertheless, all stakeholder groups who will benefit from an updated core outcome set were included in its development, which we hope will increase uptake. Finally, OMERACT and COMET methodology were followed as closely as possible.

CONCLUSION

This Delphi survey study identified 7 domains that should be considered for the core set evaluating the efficacy of symptom-modifying therapies, and 6 domains that should be considered for the core set investigating disease-modifying therapies, according to patients and axSpA experts. The results from this study will be used to compose the core outcome set for axSpA, in which a distinction will be made for the domains mandatory for studies assessing symptom-modifying therapies and studies evaluating disease-modifying therapies. After finalizing the core outcome set, the next step for ASAS will be to identify appropriate instruments to measure the chosen domains.

SUPPLEMENTARY DATA

Supplementary data are published online on the website of the Journal of Rheumatology

REFERENCES

1. van der Heijde D, van der Linden S, Dougados M, et al. Ankylosing spondylitis: plenary discussion and results of voting on selection of domains and some specific instruments. *J Rheumatol* 1999;26(4):1003-5.
2. Bautista-Molano W, Navarro-Compán V, Landewé RB, et al. How well are the ASAS/OMERACT Core Outcome Sets for Ankylosing Spondylitis implemented in randomized clinical trials? A systematic literature review. *Clin Rheumatol* 2014;33(9):1313-22.
3. 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.
4. Rudwaleit M, Jurik AG, Hermann K-GA, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by ASAS/OMERACT MRI group. *Ann Rheum Dis* 2009;68(10):1520-27.
5. Lukas C, Landewé R, Sieper J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68(1):18-24.
6. Heuft-Dorenbosch L, Spoorenberg A, van Tubergen A, et al. Assessment of enthesitis in ankylosing spondylitis. *Ann Rheum Dis* 2003;62(2):127-32.
7. Kiltz U, van der Heijde D, Boonen A, et al. Development of a health index in patients with ankylosing spondylitis (ASAS HI): final result of a global initiative based on the ICF guided by ASAS. *Ann Rheum Dis* 2015;74(5):830-35.
8. Gossec L, Portier A, Landewé R, et al. Preliminary definitions of ‘flare’ in axial spondyloarthritis, based on pain, BASDAI and ASDAS-CRP: an ASAS initiative. *Ann Rheum Dis* 2016;75(6):991-96.
9. Boers M, Kirwan JR, Tugwell P. OMERACT Handbook, 2018.
10. Boers M, Kirwan JR, Gossec L, et al. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves filter 2.0. *J Rheumatol* 2014;41(5):1025-30.
11. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials* 2017;18(3):280.
12. Kirkham JJ, Davis K, Altman DG, et al. Core Outcome Set-STAndards for Development: The COS-STAD recommendations. *PLoS Med* 2017;14(11):e1002447.
13. Maxwell LJ, Beaton DE, Shea BJ, et al. Core Domain Set Selection According to OMERACT Filter 2.1: The OMERACT Methodology. *J Rheumatol* 2019;46(8):1014-20.
14. van der Heijde D, van der Linden S, Bellamy N, et al. 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(4):945-47.
15. Boonen A, Braun J, van der Horst Bruinsma IE, et al. ASAS/WHO ICF Core Sets for ankylosing spondylitis (AS): How to classify the impact of AS on functioning and health. *Ann Rheum Dis* 2010;69(1):102-7.
16. Boonen A, van Berkel M, Kirchberger I, et al. Aspects relevant for functioning in patients with ankylosing spondylitis according to the health professionals: a Delphi study with the ICF as reference. *Rheumatology* 2009;48(8):997-1002.
17. Boel A, Navarro-Compán V, Landewé R, et al. Two different invitation approaches for consecutive rounds of a Delphi survey led to comparable final outcome. *J Clin Epidemiol* 2020;129:31-39.
18. Khodyakov D, Grant S, Denger B, et al. Practical Considerations in Using Online Modified-Delphi Approaches to Engage Patients and Other Stakeholders in Clinical Practice Guideline Development. T*he patient* 2020;13(1):11-21

CHAPTER 6

TWO DIFFERENT INVITATION APPROACHES FOR CONSECUTIVE ROUNDS OF A DELPHI SURVEY LED TO COMPARABLE FINAL OUTCOME.

Anne Boel, Victoria Navarro-Compán, Robert Landewé,
Désirée van der Heijde

J Clin Epidemiol 2021;129:31-39

ABSTRACT

Objective

There are two different approaches to involve participants in consecutive rounds Delphi survey: 1) Invitation to every round independent of response to the previous round ('all-rounds'); 2) Invitation only when responded to the previous round ('respondents-only'). This study aimed to investigate the effect of invitation approach on the response rate and final outcome of a Delphi survey.

Study Design and Setting

Both experts (n=188) and patients (n=188) took part in a Delphi survey to update the core outcome set (COS) for axial spondyloarthritis. A study with 1:1 allocation to two experimental groups (i.e. 'all-rounds' [n=187] and 'respondents-only' [n=189]) was built in.

Results

The overall response rate was lower in the 'respondents-only group' (46%) compared to the 'all-rounds group' (61%). All domains that were selected for inclusion in the COS by the 'respondents-only group' were also selected by the 'all-rounds group'. Additionally, the four most important domains were identical between groups after the final round; with only minor differences in the other domains.

Conclusion

Inviting panel members who missed a round to a subsequent round will lead to a better representation of opinions of the originally invited panel and reduces the chance of false consensus, while it does not influence the final outcome of the Delphi.

INTRODUCTION

The Delphi technique is a structured forecasting method based on the presumption that combining the opinion of a group of experts will result in a more accurate prediction of the truth than relying on the opinion of a most knowledgeable single individual[1]. Responses can be altered between rounds, based on the aggregated information of peers from the previous round[2]. An additional benefit of the Delphi is that participants tend to perceive ownership of the results due to their participation in the process. In turn, this perceived ownership improves the acceptance of the findings among those who participated[3]. As the participants are a reflective sample of the end-users, their involvement in the development-stage increases implementation in the field[4]. The Delphi process ends when (the predefined level of) consensus is achieved, or when the prespecified number of rounds has been completed[5].

Common applications of the Delphi technique in health care settings are the selection of outcomes for a core set and the identification of research priorities [1 6]. Even though the Delphi technique is often used, there is hardly any guidance on the methodology underlying the Delphi technique[5-7], which results in large variability in its execution. Research on methodological guidance of the Delphi technique in the development of core outcome sets (COS) is slowly increasing[8-11], but a lot of the methodology remains unclear to this day. Guidance on which participants to invite to consecutive rounds has not yet been described in existing literature. There are two options: 1) Invite only participants that have completed the previous round for the consecutive round. This approach ensures participants provide their own authentic opinion in the first round and are challenged to rethink their own response in light of the responses of others each round. Hence, this approach increases engagement in the decision-making process and the final outcome of the Delphi will be an accurate representation of the opinions of those who participated. 2) Invite every participant for all consecutive rounds irrespective of whether they have responded or not. This approach decreases the chance of non-random loss of opinions which could lead to false consensus[6], as it considers the opinion of every participant who completed one or more rounds, and the final outcome may therefore be a better representation of the opinions of the entire panel that was invited to partake. Scientific evidence to guide Delphi researchers on whether panel members who miss a round can be included in a subsequent round is sparse. Yet, if the results are consistent with the conventional approach of excluding these experts from subsequent rounds, the final outcome may be a better reflection of the opinions of the originally invited panel and false consensus caused by drop-out of those with a different opinion may be reduced.

The objective of this study is to investigate two different approaches of inviting participants to consecutive rounds in a 3-round Delphi survey and their effect on the final result of the Delphi and the (overall) response rate.

MATERIALS AND METHODS

Design and population

Two stakeholder groups were invited by email to partake in two separate Delphi surveys, as part of a larger project to update the ASAS/OMERACT COS for axial spondyloarthritis (axSpA)[12]. One group consisted of patients with axSpA and the other group consisted of axSpA experts, including rheumatologists, other health care professionals, methodologists and other stakeholders. The axSpA experts were all ASAS members, who were informed they would be invited to partake in the Delphi survey to update the current COS in an annual meeting prior to the commencement of the project. The patients with axSpA were recruited through three national patient societies (SAA (Spondylitis Association of America), NASS (National Ankylosing Spondylitis Society [UK]), and CSA (Canadian Spondylitis Association)) and eligible to partake if they had a diagnosis of axSpA from their rheumatologist. Patients were contacted by their associations via email, in which the study was explained and patients were asked to participate. Additionally, information was placed on the websites of each of the organizations. Patients could either send an email directly to the researcher in charge of sending the Delphi invitations (AB) or their respective organization if they were interested in partaking in the Delphi survey. Recruitment ceased once the group of patients was equal in size to the group of experts (n=188). This was an opportunistic sample and no sample size calculations were performed upfront. The main objective of this Delphi survey was to define which are the most relevant disease domains (outcomes) for all stakeholders to be included in the updated COS. We did not specifically ask consent for the experiment, since knowledge about the assignment would have biased the results.

For each separate stakeholder group, the invited participants were randomly allocated 1:1 to two experimental groups to ensure an even distribution of stakeholders in each experimental group. Experimental group 1 was labelled as 'respondents-only group'; and experimental group 2 as 'all-rounds group'. Two randomization sequences (one for patients, one for experts) were created, using a computer-generated schedule (developed by a member of the data management team of the rheumatology department in the LUMC). Randomization was performed by a researcher (AB) after the recruitment of patients was complete. Only the researcher in charge of sending the Delphi and collecting the data (AB) was aware of the group-allocation of each participant.

Even though the results of the Delphi were analyzed separately for each of the stakeholder groups (i.e. patients and experts) to update the COS for axSpA; the stakeholder groups were not the focus of this study. The aim of this experiment was solely to compare the responses of the 'all-rounds group' with those of the 'respondents-only group' to investigate whether there is an effect of invitation-procedure on the final result of the Delphi survey. Therefore, all data in this manuscript focused only on the differences between the 'respondents-only' and 'all-rounds' groups.

The participants in the 'respondents-only group' received an invitation for the second round only if they completed the first round; and only received an invitation for the third round if they completed the second round. The participants in the 'all-rounds group' received an invitation for each round irrespective of response to any of the previous rounds. In each round, the participants received summarized information of the previous round, including aggregated scores from their respective stakeholder group. Those participants who partook in the previous round received their individual score of the previous round as well. Participants in the 'all-rounds group' who responded for the first time to the invitation for the second round received only aggregated scores of the first round, and the same procedure applied to round three.

The participants were not aware of the experiment and received identical information regarding the Delphi survey. All participants knew from the start that this was a three-round Delphi, but did not know that an invitation for the second and third round was conditional on responding to the first round.

Each round was open for 2-3 weeks and a single reminder was sent after one week to those who did not yet complete the round. Data were collected online using SurveyMonkey between November 2nd and December 30th 2018.

Survey Questionnaires

The Delphi survey consisted of two separate sections, one focused on the outcomes to be included in the core set for studies assessing a symptom-modifying drug; the other on the outcomes to be included in the core set for studies evaluating a disease modifying drug. In this manuscript the results for the survey on symptom modification will be described in detail and the results on disease modification only in the appendix.

The survey on symptom modification contained 11 candidate domains for the core set in the first round, and participants had the opportunity to suggest additional domains in this round. This led to the addition of one more domain from round 2 onwards, which brought the total number of domains to 12. It was decided upfront that the survey on disease modification would contain the same domains as the survey on symptom modification.

To identify importance of each of the domains for the core set, each participant was asked to provide one score per domain using a 9-point Likert scoring system. Domains were graded in accordance to their level of importance. Following the Outcome Measures in Rheumatology (OMERACT) handbook, a score of 1 to 3 signified an outcome as not important; 4 to 6 as important but not critical; and 7 to 9 as critical[13]. The criteria to include a domain in the next round of the Delphi were: at least 50% of the participants scored the domain as critical; and: 15% or less scored the domain as not important. If a domain was scored as critical by ≥80% in the specific stakeholder group, the domain was considered selected for the core set, and not offered for voting in subsequent rounds within that stakeholder group. The aggregated scores per domain were analyzed separately for each of the stakeholder groups.

A domain was considered for inclusion in the update of the core set if ≥70% of participants scored the domain as critical; and: ≤15% scored not important after the final round, which was in line with the guidelines provided in the OMERACT handbook for the development of COS[13]. This was the first step in the update of the COS for axSpA and explains the availability of the data per round. More detailed information about the selection of domains is beyond the scope of this article. In addition to the 9-point Likert scoring grade, all participants were asked to select six domains from the list of possible domains in each Delphi round. The chosen domains were those with the highest priority to be included in the COS, there was no further ranking within the six chosen domains.

Outcomes

The main outcomes of this experiment were: 1) the response rates after each round and the final overall response rate; and 2) the finally selected domains for the core set at the end of round three in the 'all-rounds group' versus the 'respondents-only group'. Secondary aspects of interest were differences between groups regarding: 1) the choice of the 'top-six' domains; 2) changes in the 'top-six' domains across rounds. Additionally, the design of this study enabled us to study if the results of the Delphi survey are similar when randomly selecting two independent samples. As the experiment started after round 1, we had the ability to compare domains between two panels ('respondents-only' and 'all-rounds') who completed an identical survey (i.e. round 1).

For the purpose of this study, we used descriptive statistics to describe the data, using mean (SD) scores; statistical testing of between group differences was not performed. For the 9-point Likert scale scores per domain, means and standard deviations were used to describe the data. For the current analysis, the last available scores were used to compare the 'respondents-only' and 'all-rounds' groups if a domain was selected before the final round (e.g. if a domain was selected after round 2, the last available mean is the mean of round 2). The top-six domains were presented as percentages, and the change between rounds in proportion of participants showing change. Similar to the mean scores, if a domain was selected before the final round the percentage of critical votes from the last available round were used to compare the groups.

RESULTS

A total of 376 participants were invited by email to partake in this Delphi survey. They were randomized into a 'respondents-only group' (n=187; 93 in the patient survey and 94 in the axSpA expert survey) and 'all-rounds group' (n=189; 95 in the patient survey and 94 in the axSpA expert survey).

Response rates

The overall response rate after 3 rounds was lower in the 'respondents-only group' compared to the 'all-rounds group' (46% [86/187] vs. 61% [116/189]). The response rate in the 'respondents-only group' increased per each additional round (from 65% [122/187] to 91% [86/95]), while the response rate in the 'all-rounds group' varied only slightly between rounds (from 56% [110/189] to 61% [116/189]). The retention rate was similar for patients and experts (online appendix table A.1) and there was no difference between participants and non-participants in the continent of residence (online appendix table A.2).

Table 1 Response rates per group for each round of the Delphi survey

	'Respondents-only group' (n=187)	**'All-rounds group'** (n=189)
Round 1	Invited: 187 Completed: 122 **Response rate: 65%**	Invited: 189 Completed: 110 **Response rate: 58%**
Round 2	Invited: 122 Completed: 95 [response rate (78%)] **Overall response rate: 51%**	Invited: 189 Completed: 105 **Response rate: 56%**
Round 3	Invited: 95 Completed: 86 [response rate (91%)] **Overall response rate: 46%**	Invited: 189 Completed: 116 **Response rate: 61%**

Domains selected after final round

There was no difference in mean (SD) scores between the 'respondents-only' and 'all-rounds' groups for any of the domains (figure 1).

Figure 2 depicts the percentage of participants that voted for 'critical' or 'not important' per domain for each of the experimental groups. The vertical lines at 15% and 70% represent the cut-offs as described in section 2.2. This figure shows that both groups selected the same domains, apart from extra-musculoskeletal manifestations, which was only selected by the 'all-rounds group'.

The corresponding results from the disease modification survey can be found in online appendix figures A.2 and A.3.

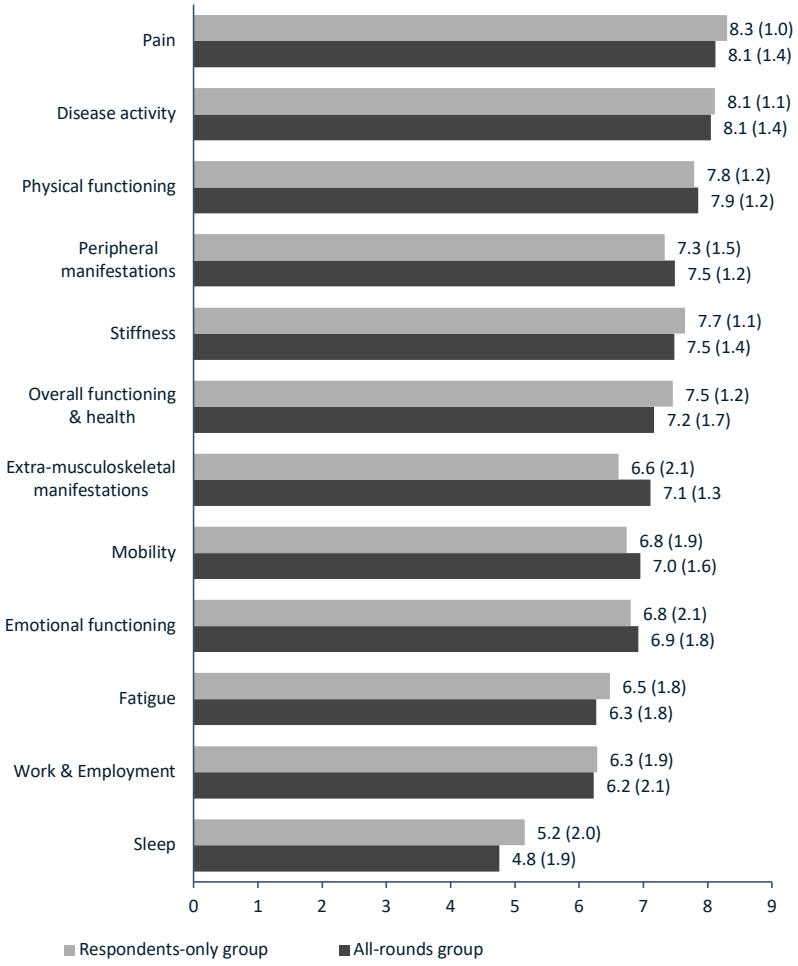


Figure 1 Mean score (standard deviation) per domain for the 'respondents-only group' (light) and 'all-rounds group' (dark) from the round when the domain was selected (i.e. the last available scores).

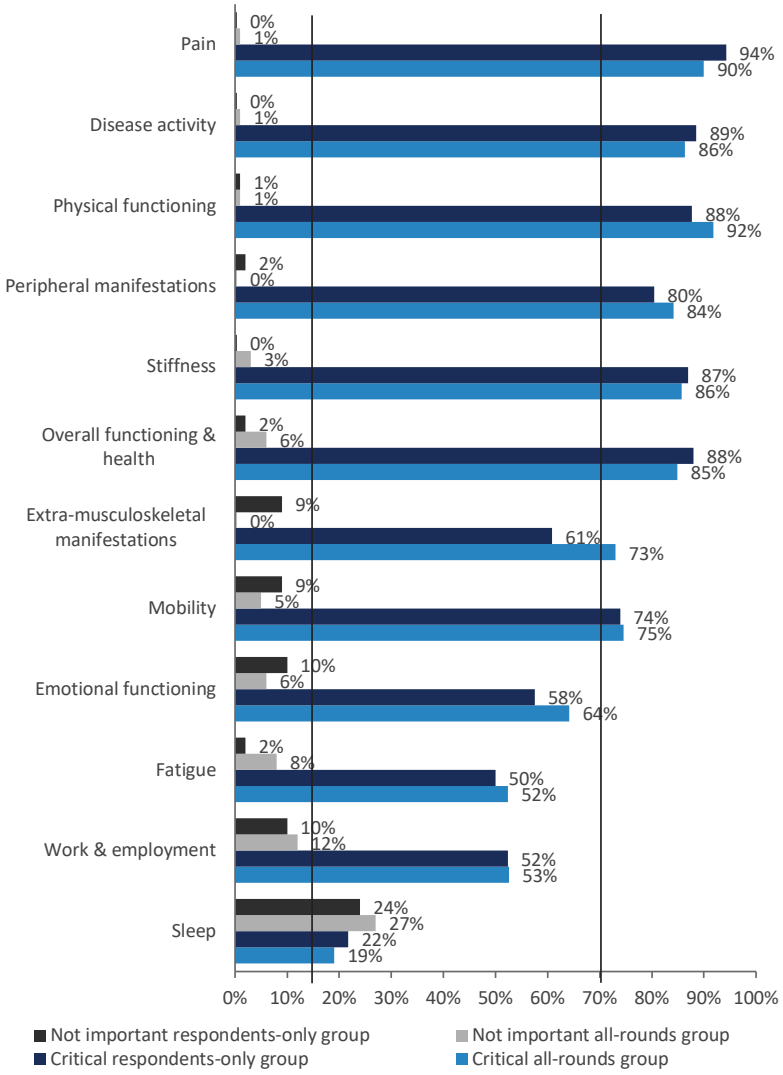


Figure 2 Percentage of patients that voted not important (in gray) and critical (in blue) for the 'respondents-only group' (dark color) and 'all-rounds group' (light color) from the round when the domain was selected (i.e. the last available scores); the vertical lines at 15% and at 70% represent the cut-offs for inclusion of domains.

Percentage of respondents regarding the top-six domains

For each of the domains in the Delphi survey, table 2 shows the percentage of participants that voted a domain as one of their six most important domains after the final round. The domains were ranked in descending order based on selection by the 'respondents-only group' and matched with the same domain in the 'all-rounds group'. Domains in *italic* represent the top 6 of the 'respondents-only group' and in **bold** of the 'all-rounds group'.

Table 2 Most important domains after round 3 for the 'respondents-only group' and 'all-rounds group' ranked in descending order, based on the selection by the 'respondents-only group' and matched with the same domain in the 'all-rounds group'; and the difference in percentage of votes between the groups per domain. Domains in *italic* represent the top 6 in the 'respondents-only group' and in **bold** in the 'all-rounds group'.

'Respondents-only group' n=86		**'All-rounds group'** n=116		**Difference between groups**	
Pain	*95%*	**Pain**	**91%**	Pain	4%
Stiffness	*62%*	**Stiffness**	**62%**	Stiffness	0%
Physical functioning	*62%*	**Physical functioning**	**61%**	Physical functioning	1%
Mobility	*59%*	**Mobility**	**54%**	Mobility	5%
Disease activity	*55%*	Disease activity	49%	Disease activity	6%
Fatigue	*50%*	**Fatigue**	**55%**	Fatigue	-5%
Overall functioning & health	47%	**Overall functioning & health**	**57%**	Overall functioning & health	-10%
Extra-musculoskeletal manifestations	44%	Extra-musculoskeletal manifestations	45%	Extra-musculoskeletal manifestations	-1%
Peripheral manifestations	34%	Peripheral manifestations	44%	Peripheral manifestations	-10%
Sleep	30%	Sleep	26%	Sleep	4%
Work & Employment	21%	Work & Employment	23%	Work & Employment	-2%
Emotional functioning	16%	Emotional functioning	18%	Emotional functioning	-2%

After the final round, the four outcomes with highest voting rates were the same in both groups, with only small differences between groups. The domain 'disease activity' was voted in the top-six of the 'respondents-only group' (55%) but not in the top-six of the 'all-rounds group' (49%), where it was replaced by the domain 'overall functioning and health' (57%).The differences were small; the maximum difference between groups was 10% for the domains 'overall functioning and health' and 'peripheral manifestations'.

The corresponding results from the disease modification survey can be found in online appendix table A.5.

Changes in the top-six domains across rounds

Per individual we determined the number of domains that changed in their top-six ranking across rounds. This was done from round 1 to round 2 and from round 2 to round 3 (figure 3). For the 'all-rounds group' various combinations of completion were possible, these can be found in online appendix table A.3. Those participants that responded to rounds 1 and 2 but not to round 3 (n=9) were only included in the change between the rounds they completed, and the same applied to those who completed rounds 2 and 3 but not round 1 (n=13). There were 12 participants in the 'all-rounds group' who only missed round 2. For these 12 participants the change between rounds 1 and 3 was calculated. Finally, those participants that only responded to a single round were excluded from this analysis (n=38). Figure 3 shows that between rounds 1 and 2, 49% of the 'respondents-only group' changed at least one domain, whereas this was 38% in the 'all-rounds group'. In both groups hardly anyone did not change a single domain (1% in the 'respondents-only' versus 2% in the 'all-rounds' group). A larger proportion of participants in the 'all-rounds group'

changed 2 domains (42%) or more than 2 domains (18%), compared to the 'respondents-only group' (37% and 13% respectively). Between round 2 and 3, 45% of the participants in the 'respondents-only group' changed 1 domain, whereas this was only 31% in the 'all-rounds group'. Contrary, in the 'all-rounds group' 22% changed more than 2 domains, whereas this was only 7% in the 'respondents-only group'. The proportion of participants changing no domains, or 2 domains were similar between groups.

The corresponding results from the disease modification survey can be found in online appendix figure A.4.

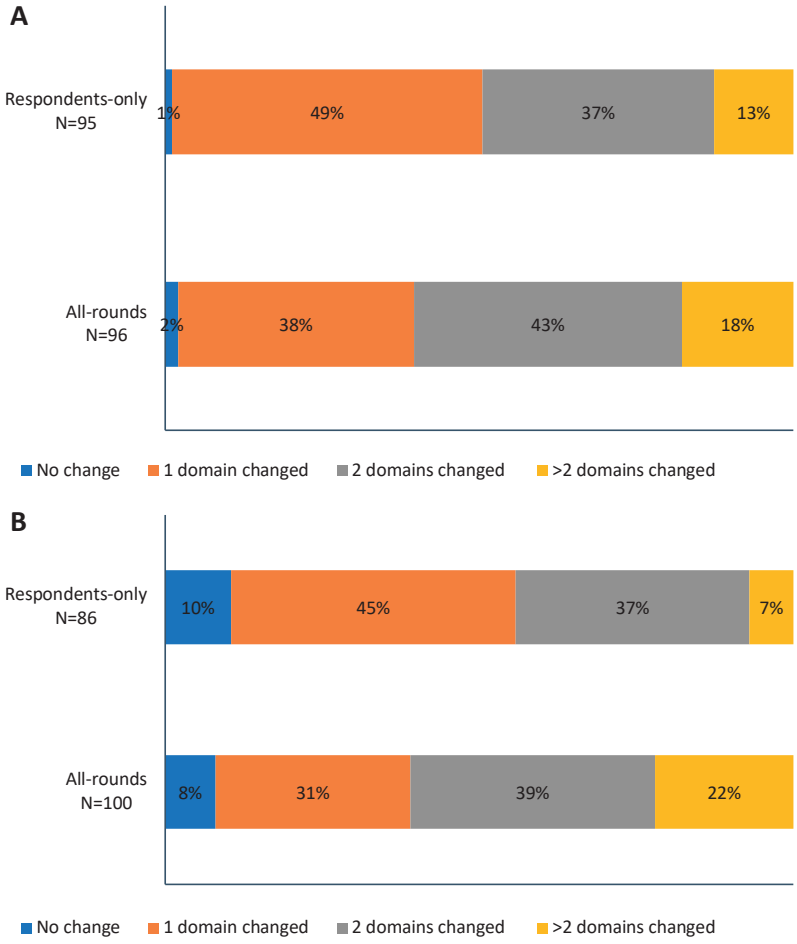


Figure 3 Changes in the 'top six' domains per group, from round 1 to round 2 (A) and round 2 to round 3 (B), presented as the proportion of participants over four change categories (no change, 1 domain changed, 2 domains changed and >2 domains changed).

Comparison of round 1 results

Mean (SD) scores where very similar between the 'all-rounds' and 'respondents-only' groups for all of the domains after the first round of the Delphi survey (figure 4). Additionally, the proportions of 'not important votes' and 'critical votes' were similar between the 'all-rounds' and 'respondents-only' groups after the first round (online appendix figure A.1). Furthermore, there were no differences in the top-six domains after the first round (online appendix table A.4).

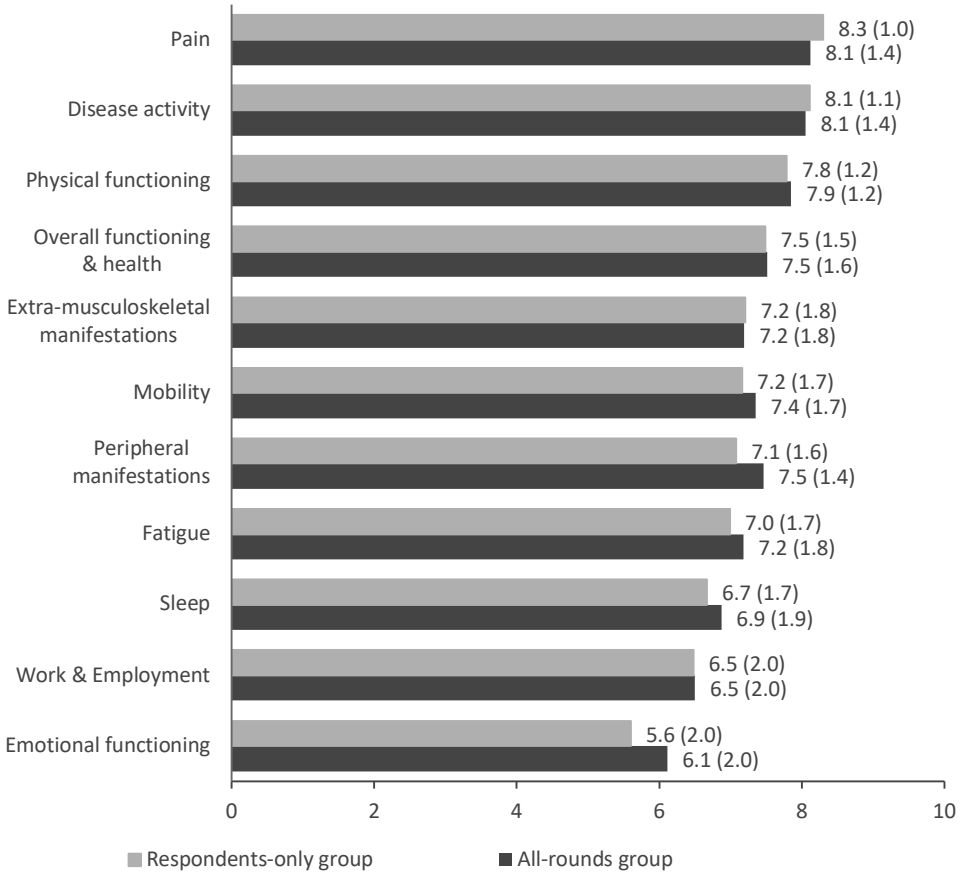


Figure 4 Mean score (standard deviation) per domain for the 'respondents-only group' (light) and 'all-rounds group' (dark) after round 1.

The corresponding results from the disease modification survey can be found in online appendix table A.6 and online appendix figures A.5 and A.6.

DISCUSSION

This study showed no differences between the 'respondents-only' and 'all-rounds' groups in mean (SD) scores, nor in the percentage of critical votes for any of the domains after the final round. These results showed that the same domains were selected by the 'all-rounds group' and the 'respondents-only group'. Invitation-approach had no impact on which domains were selected after the final round. Additionally, differences between experimental groups regarding the domains in the top-six were small and hardly influenced the order of importance, suggesting that the invitation approach does not influence the outcome of a Delphi exercise at all. Thus, it may be preferential to invite panel members who missed a round to a subsequent round, as this approach is less sensitive to the non-random loss of opinions that could lead to false consensus[6].

However, we found a difference in the overall response rate, which was higher in the 'all-rounds group' (61%) compared to the 'respondents-only group' (46%). This was expected, since the numbers of persons invited for rounds 2 and 3 were larger in the 'all-rounds group' than in the 'respondents-only group'.

The proportion of participants that changed one domain between rounds was larger in the 'respondents-only group'. Contrary, the proportion of participants that changed more than two domains between rounds was larger in the 'all-rounds group'. From all change categories, the no-change category had the smallest proportion of participants across all rounds and in both groups. These results indicate that information of peers from previous rounds is taken into account when evaluating the domains which are deemed most important.

This study may have a few limitations. Since the data used was from a true Delphi experiment, there was no complete data on mean scores and percentages of critical votes for all domains for every round, as selected and excluded domains were not offered in the next round. We attempted to solve this by using the last available data for each of the domains (i.e. from the round when the domain was selected). Nonetheless, this may have influenced the comparison between the 'all-rounds' and 'respondents-only' groups.

In the current study it was decided upfront to do a 3-round Delphi survey, which appears sufficient to achieve consensus[14-16]. Furthermore, determining the number of rounds upfront may actually be preferential to continue until consensus is reached, as attrition rates increase with each additional round and those with a very different opinion may drop-out, causing false consensus[6]. Due to this decision, we cannot be sure whether these results can be extrapolated to Delphi surveys consisting of more than three rounds.

Using an electronically distributed Delphi ensured involvement of international experts and patients as no travel is required, anonymity is guaranteed and no public speaking is required, which increases patient participation[17]. Although all patients included in this study were native English speakers, as this ensured a good understanding of the content, they did represent three different countries. Furthermore, the inclusion of patients; rheumatologists; other health care professionals; policy makers; and representatives of pharmaceutical companies, resulted in a sample that is reflective of the population who will use the updated COS.

A strength of this study was the random selection of two independent samples, namely the 'respondents-only' and 'all-rounds' groups. As the experiment started after round 1, we had the ability to compare results of the two panels who completed an identical first-round survey. Little has been published on the agreement between multiple independent panels going through an identical survey. Previous research showed high correlations between endorsement frequencies in a replication of a Delphi survey, even with a gap of several years between replications and use of a different expert panel[18]. In more recent work groups had been randomized to assess the effect of feedback provided between rounds, and showed high agreement in selected items between different randomization groups[10]. Here we add information to this topic by showing that the results after the first round were similar between two randomly selected independent panels who completed an identical survey.

CONCLUSION

This study showed that the content of the outcome of this 3-round Delphi survey was similar regardless of using data from all persons invited to the first round; or of those persons only who participated in all rounds. We therefore conclude that invitation approach does not seem to influence the final results of a Delphi survey.

SUPPLEMENTARY DATA

Supplementary data are published online on the website of the Journal of Clinical Epidemiology.

REFERENCES

1. Keeney S, McKenna H, Hasson F. The Delphi technique in nursing and health research: John Wiley & Sons 2010.
2. Walker A, Selfe JJBJoT, Rehabilitation. The Delphi method: a useful tool for the allied health researcher. 1996;3(12):677-81.
3. McKenna HPJJoan. The Delphi technique: a worthwhile research approach for nursing? 1994;19(6):1221-25.
4. Jorm AF. Using the Delphi expert consensus method in mental health research. *Aust N Z J Psychiatry* 2015;49(10):887-97.
5. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. 2000;32(4):1008-15.
6. Humphrey-Murto S, de Wit M. The Delphi method—more research please. *J Clin Epidemiol* 2019;106:136-39.
7. Sinha IP, Smyth RL, Williamson PR. Using the Delphi Technique to Determine Which Outcomes to Measure in Clinical Trials: Recommendations for the Future Based on a Systematic Review of Existing Studies. *Plos Med* 2011;8(1)
8. Brookes ST, Macefield RC, Williamson PR, et al. Three nested randomized controlled trials of peer-only or multiple stakeholder group feedback within Delphi surveys during core outcome and information set development. *Trials* 2016;17(1):409.
9. De Meyer D, Kottner J, Beele H, et al. Delphi procedure in core outcome set development: rating scale and consensus criteria determined outcome selection. *J Clin Epidemiol* 2019;111:23-31.
10. MacLennan S, Kirkham J, Lam TBL, et al. A randomized trial comparing three Delphi feedback strategies found no evidence of a difference in a setting with high initial agreement. *J Clin Epidemiol* 2018;93:1-8.
11. Turnbull AE, Dinglas VD, Friedman LA, et al. A survey of Delphi panelists after core outcome set development revealed positive feedback and methods to facilitate panel member participation. *J Clin Epidemiol* 2018;102:99-106.
12. van der Heijde D, van der Linden S, Dougados M, et al. Ankylosing spondylitis: plenary discussion and results of voting on selection of domains and some specific instruments. *J Rheumatol* 1999;26(4):1003-5.
13. Boers M, Kirwan J, Tugwell P, et al. The OMERACT Handbook. 2015.
14. Boulkedid R, Abdoul H, Loustau M, et al. Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. *PLOS ONE* 2011;6(6):e20476.
15. Rowe G, Wright G. Expert Opinions in Forecasting: The Role of the Delphi Technique. In: Armstrong JS, ed. Principles of Forecasting: A Handbook for Researchers and Practitioners. Boston, MA: Springer US 2001:125-44.
16. Belton I, MacDonald A, Wright G, et al. Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process. *Technol Forecast Soc Change Change* 2019;147:72-82.
17. Khodyakov D, Grant S, Denger B, et al. Practical Considerations in Using Online Modified-Delphi Approaches to Engage Patients and Other Stakeholders in Clinical Practice Guideline Development. *The patient* 2020;13(1):11-21.
18. Ross AM, Kelly CM, Jorm AF. Re-development of mental health first aid guidelines for suicidal ideation and behaviour: a Delphi study. *BMC Psychiatry* 2014;14:241.

CHAPTER 7

THE ASAS-OMERACT CORE DOMAIN SET FOR AXIAL SPONDYLOARTHRITIS

Victoria Navarro-Compán, **Anne Boel**, Annelies Boonen, Philip Mease, Robert Landewé, Uta Kiltz, Maxime Dougados, Xenofon Baraliakos , Wilson Bautista-Molano, Hilde Carlier, Praveena Chiowchanwisawakit, Hanne Dagfinrud, Natasha de Peyrecave, Bassel El-Zorkany, Lana Fallon, Karl Gaffney, Marco Garrido-Cumbrera, Lianne Gensler, Nigel Haroon, Yu Heng Kwan, Pedro Machado, Walter Maksymowych, Denis Poddubnyy, Mikhail Protopopov, Sofia Ramiro, Bev Shea, In Ho Song, Salima van Weely, Désirée van der Heijde

Seminars in Arthritis and Rheumatism, 2021 Dec;51(6):1342-1349a

ABSTRACT

Background

The current core outcome set for ankylosing spondylitis (AS) has had only minor adaptations since its development 20 years ago. Considering the significant advances in this field during the preceding decades, an update of this core set is necessary.

Objective

To update the ASAS-OMERACT core outcome set for AS into the ASAS-OMERACT core outcome set for axial spondyloarthritis (axSpA).

Methods

Following OMERACT and COMET guidelines, an international working group representing key stakeholders (patients, rheumatologists, health professionals, pharmaceutical industry and drug regulatory agency representatives) defined the core domain set for axSpA. The development process consisted of: 1) Identifying candidate domains using a systematic literature review and qualitative studies; 2) Selection of the most relevant domains for different stakeholders through a 3-round Delphi survey involving axSpA patients and axSpA experts; 3) Consensus and voting by ASAS; 4) Endorsement by OMERACT. Two scenarios are considered based on the type of therapy investigated in the trial: symptom modifying therapies and disease modifying therapies.

Results

The updated core outcome set for axSpA includes 7 mandatory domains for all trials (disease activity, pain, morning stiffness, fatigue, physical function, overall functioning and health, and adverse events including death). There are 3 additional domains (extra-musculoskeletal manifestations, peripheral manifestations and structural damage) that are mandatory for disease modifying therapies and important but optional for symptom modifying therapies. Finally, 3 other domains (spinal mobility, sleep, and work and employment) are defined as important but optional domains for all trials.

Conclusion

The ASAS-OMERACT core domain set for AS has been updated into the ASAS-OMERACT core domain set for axSpA. The next step is the selection of instruments for each domain.

INTRODUCTION

The management of axial spondyloarthritis (axSpA) has come a long way in the last two decades[1,2]. The development of new therapeutic options, especially pharmaceutical drugs, covering the entire spectrum of the disease has been a major advance[3,4]. This progress should go hand-in-hand with updating outcome measures, so that all studies consistently assess the most relevant domains and instruments for axSpA.

Clinical trials seek to evaluate whether an intervention is effective and safe. This is determined by comparing the effects of a specific intervention on selected outcomes versus a control to identify the possible beneficial or harmful effects of the intervention. Therefore, the careful selection of appropriate outcomes is crucial when designing clinical trials and other clinical studies. To avoid selective reporting of outcomes and to facilitate comparison of results across trials, it is important to use standardised outcomes[5]. Moreover, it is important to use outcomes that are relevant to all stakeholders. Such issues can be addressed with the development and application of an agreed standardised set of outcomes for all clinical trials, which is defined as the core outcome set for a specific health condition, population and setting[6].

The core outcome set represents the minimum that should be measured and reported in all clinical trials. Nevertheless, this does not imply that the outcomes in a particular study should be restricted to those in the core outcome set[7]. Rather, there is an expectation that the core outcomes will be collected and reported to allow the results of trials and other studies to be compared, contrasted and combined as appropriate. Therefore, the use of a core outcome set may reduce heterogeneity of outcomes between studies in axSpA, will lead to research that is more likely to have measured relevant outcomes, and is of potential value to use in clinical audit and meta-analyses. Also, it enhances the value of evidence synthesis by reducing the risk of outcome reporting bias and ensuring that all trials contribute relevant information[5,6].

Although the core outcome sets are essential, not many have been developed according to the highest standard and/or have been implemented adequately. The most notable work relating to outcome standardisation has been conducted by the Outcomes Measures in Rheumatology (OMERACT) collaboration, which is an independent initiative of international multi-stakeholders interested in outcome measures in rheumatology, integrating patient, clinician, trialist, methodological and industry perspective.

OMERACT had its first meeting and definition of a core outcome set in 1992[8]. This successful initiative was followed by a more global group also addressing other fields outside of Rheumatology, set up as the Core Outcome Measures in Effectiveness Trials

(COMET) Initiative in 2010. The aim of COMET is to promote the development of core sets and bring together researchers interested in the development and application of core outcome sets[9].

The Assessment of SpondyloArthritis international Society (ASAS) is an international group of experts in the field of spondyloarthritis (SpA), with the ultimate goal to improve the overall health and outcome of patients with SpA[10,11]. Outcome assessment has always been the focus of ASAS, similar to OMERACT, and both organisations have collaborated closely. In fact, the development of the ASAS-OMERACT core set for outcome measures in ankylosing spondylitis (AS) was the first activity undertaken by ASAS after its launch in 1995. The first preliminary ASAS core set for AS was published in 1997[12]. This was followed by a publication in 1999 on the selection of the instruments for each outcome in the core set[13]. And finally, the core set was endorsed by OMERACT in 1999[14,15]. In 2007 minor changes in relation to a few selected instruments were implemented by a consensus process by ASAS[16].

As shown by a recent systematic literature review, the ASAS-OMERACT core set for AS was well implemented after its original publication two decades ago[17]. However, since then, there have been major advances in the field of SpA as well as in the methodology to develop core sets, which may have an impact on the agreed outcomes two decades ago. Main accomplishments in the field of axSpA outcomes include the use of magnetic resonance imaging (MRI), the development of the Ankylosing Spondylitis Disease Activity Score (ASDAS)[18], validated enthesitis scores[19], and the ASAS Health Index[20,21]. With regards to the methodology to develop core sets, there is no gold standard yet but during the last years OMERACT and COMET have intensively worked to provide specific guidance about how this should be done, e.g. OMERACT handbook and Filter 2.0, COMET handbook and Core Outcome Set-STAndards for Development (COS-STAD)[5,7,22,23].

Moreover, there have been developments with respect to the definition of the disease. The presence of definite sacroiliitis on radiographs is mandatory to define AS. With the availability of MRI became evident that there are also forms without radiographic sacroiliitis. This so-called non-radiographic axSpA (nr-axSpA) together with AS, also known as radiographic axSpA (r-axSpA) defines the entire spectrum of the disease, called axSp[1,24]. The new classification thus also requires an update of the ASAS-OMERACT core outcome set for axSpA.

The ASAS group decided to update the original version into the ASAS-OMERACT core outcome set for axSpA and started working on this process in 2018 according to the currently accepted methodology. The first step of this project is the selection of *what to measure* (core domain set). Thereafter, it needs to be defined *how to measure* each of the

chosen domains -selecting instruments or tools (core measurement set). Both, *what to measure* and *how to measure* will form the final core outcome set. Here we present the results of the first step.

METHODS

For this project, OMERACT and COMET guidelines were followed[5,22,25-27], but taking into account that the goal of this process was an update of an existing core set and not a completely new one. The main phases of the development process for a core set are summarized in Figure 1.

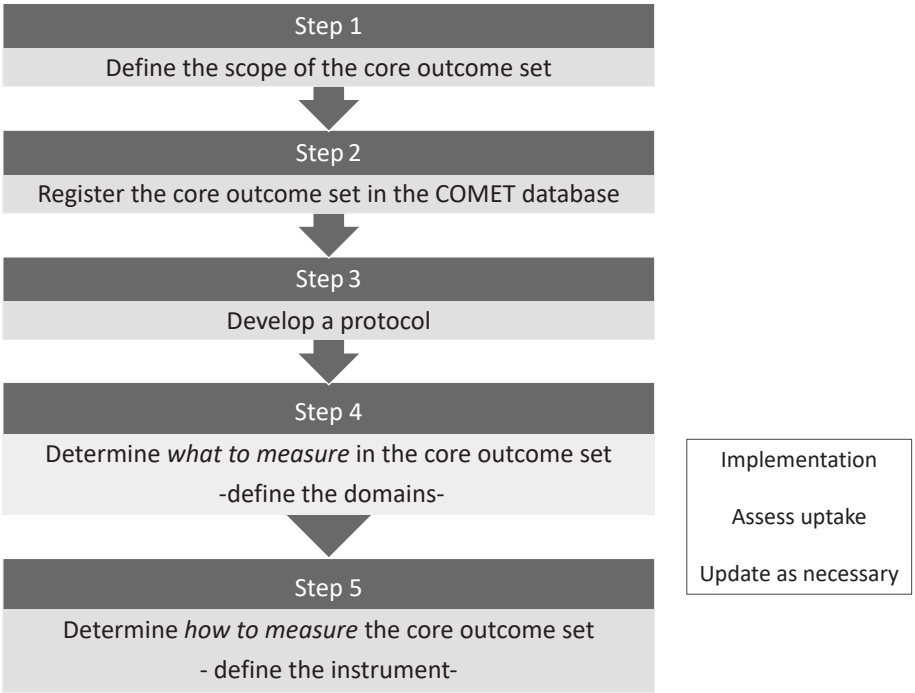


Figure 1: The core outcome set development process. Adapted from Williamson PR et al. Trials. 2017;18(Suppl 3):280. COMET: Core Outcome Measures in Effectiveness Trials.

Define the scope

First of all, the steering committee of the project defined the scope of the core set, which was established as follows:

Health condition

axSpA, with or without peripheral rheumatological manifestations (arthritis, enthesitis and dactylitis) and with or without extra-musculoskeletal manifestations (uveitis, inflammatory bowel disease and psoriasis). Pure peripheral SpA was excluded.

Population

Patients 18 years or older with axSpA, covering the whole spectrum of the disease including nr-axSpA and r-axSpA, early disease and established disease. The lower limit of the age range (18 years) was based on ethical considerations arguments, as this is the common limit required to include patients in interventional studies.

Types of intervention

Pharmacological and non-pharmacological interventions, excluding surgery. According to the type of intervention, two main scenarios are considered: 1) Symptom modifying antirheumatic therapies (SMART). This type of therapy improves the symptoms and clinical features of inflammatory manifestations and include non-pharmacological treatment (e.g. physical exercise) and symptom modifying antirheumatic drugs (SMARD) such as non-steroidal anti-inflammatory drugs (NSAIDs). 2) Disease modifying antirheumatic drugs (DMARDs).This type of intervention changes the course of the disease by a) improving and sustaining functioning and overall health and b) preventing or significantly decreasing structural damage (e.g. cytokine inhibitors).

Settings

Two main settings are described: 1) Research: clinical trials and longitudinal observational studies (including registries); and 2) Clinical practice. Nevertheless, due to the known differences in the development process between the different settings, ASAS decided to work first on a core set for the research setting and later develop a core set for the clinical practice setting.

Register in the COMET database

The COMET Initiative database is a repository of studies relevant to the development of core outcome sets. At the beginning of the project, the steering committee checked in this database that no other group was working on the update of this core set. Once this was confirmed, the project was registered in the COMET database on 19th of March 2018. Further details are available at COMET website[28].

A detailed protocol of the project was written by two of the cochairs (VN-C and DvdH) and reviewed by all members of the steering committee. OMERACT and COMET guidelines were considered for this purpose.

Working group

First, a steering committee was formed. This consisted of the four co-chairs of the project (DvdH, VN-C, AB and PM), two additional ASAS members with expertise in OMERACT and COMET methodology (RL, MD), one patient representative (UK) and one fellow (AB). The steering committee invited the members of the axSpA working group based on their background, geographical region, knowledge, experience with trials and the stakeholder group to which they belong. Potential conflicts of interest of the invited members were listed and discussed by the steering committee. The working group was formed at the beginning of 2018 involving a total of 28 participants (including the steering committee), representing those stakeholders who will use the core set in research, including rheumatologists and methodologists [17], healthcare professionals [2], patient research partners [3], representatives from pharmaceutical companies[4] and drug regulatory agencies [1], and a research-fellow[1,29].

OMERACT workshop application

In December 2018 the steering committee submitted an application for having an axSpA workshop to vote on the core domains at the OMERACT 2020 meeting, initially scheduled for April in Colorado. This application was accepted in February 2019. Nevertheless, due the COVID-19 pandemic the face-to-face meeting was postponed and eventually replaced by a virtual workshop in November 2020.

Identify all candidate and relevant domains for stakeholders

Fig. 2 shows a summary of the different phases of the process to identify the possible domain candidates and to select the final set of core domains by means of reducing the extensive list to a concise set. This part has been published in detail in a separate manuscript[30,31]. Briefly, a list of the candidate domains was identified using three different sources and later two groups of stakeholders (patients and experts) selected the domains that should be considered for inclusion in the core set via two identical but separate Delphi surveys, which were launched between November 2nd and December 30th 2018.

Working group consensus

The working group met twice during the update process. The first meeting took place in January 2019 in Amsterdam and the second virtually in November 2019. The views from all key stakeholder groups were considered. The purpose of these meetings was to provide all stakeholders the opportunity to discuss the results of the Delphi survey and to agree on a proposal for a final core set according to the new format of the OMERACT

onion [25]. As shown in Fig. 3, this follows a structure in which the domains are placed in concentric spheres by decreasing importance classifying the outcomes in three categories: 1) mandatory, i2 optional but important and 3) for research agenda.

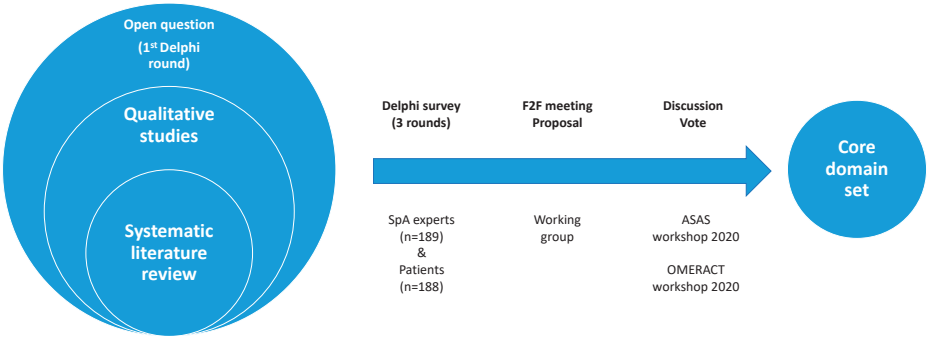


Figure 2 Development process to determine the core domain set.
SpA, Spondyloarthritis; **F2F**, face to face; **ASAS**, Assessment of SpondyloArthritis international Society; **OMERACT**, Outcomes Measures in Rheumatology.

ASAS consensus

After discussion with the working group, the results of the Delphi survey were presented and discussed with all ASAS members in a plenary session during the ASAS annual workshop 2019 in Amsterdam. By consensus, the following decisions were made:

- If a domain was included in the original core set, there should be a strong reason for excluding the domain in the updated core set.
- If a domain had been selected for the SMART scenario, this should be selected for the DMARD scenario too.

This thinking is in line with registration of drugs: drugs can show disease modification in addition to relieving of signs and symptoms. No registered treatment for axSpA has been shown to only impact structural damage progression, and even in such a trial, signs and symptoms should be assessed to know if an effect on these is lacking.

Finally, the agreed domains by the working group in the virtual meeting were presented to all the ASAS members in a plenary session during the annual ASAS workshop, in January 2020 in Houston. After discussion, each full ASAS member voted anonymously using a digital voting system (engagenow.live) on agreement with the final proposed set of domains by answering the following question “Do you agree with the proposed onion of domain core set”? The predefined requirement to accept the proposed outcomes was that at least 50% of the members voted positively.

OMERACT endorsement

Finally, the ASAS proposal of the core domain set was presented at a specific OMERACT 2020 virtual meeting, which took place on November 13th. In total, 125 participants recruited by ASAS and OMERACT attended the meeting in two different time zone sessions to ensure that participants around the world could partake. Pre-reading material was sent to all participants, which included a whiteboard video (accessible at https://omeract.org/working-groups/axial-spa), one-pager with the definitions for each of the selected domains (shown in Table 1) and a lay summary. Each meeting lasted for 90 minutes and included a plenary session, 5-7 breakout sessions (with a facilitator, a content expert, a reporter, at least one patient research partner and 5 representatives from other stakeholders) and a final voting session. All participants were asked to vote anonymously on two questions using the Zoom polling feature for meetings: 1) Can you accept the proposed set as mandatory domains for all trials? And 2) Can you accept the proposed additional domains as mandatory for disease modifying drug trials? The results were summarised in two groups: patient research partners and other stakeholders. The predefined requirement to endorse the core set was that at least 70% of the participants in each group accepted the proposal.

RESULTS

Relevant domains for stakeholders

As mentioned, the results for the selected domains to be considered for inclusion in the final core outcome set have been published in detail separately[31]. In summary, the selected domains required to be voted as critical by ≥70% of participants and not important by ≤15% of participants for both stakeholder groups, separately. After the three Delphi-rounds, a total of 7 domains (pain, physical function, stiffness, disease activity, mobility, overall functioning and health, and peripheral manifestations) were selected to be considered for inclusion in the SMART setting.

For the DMARD setting, 6 domains (physical function, disease activity, mobility, structural damage, extra-musculoskeletal manifestations, peripheral manifestations) were selected. All domains selected by experts were also selected by patients. Patients selected all offered domains except 'emotional function', including fatigue, work and employment and sleep for both settings in addition to the selected domains.

Working group proposal

After the virtual meeting in November 2019, the working group agreed on a proposal for the core domains, distributed across the OMERACT onion (Fig. 3), which took into account the two decisions previously taken (i.e., only delete a previous domain for strong reasons

and all mandatory domains for the SMART setting should also be mandatory for the DMARD setting). This proposal included 7 mandatory domains for all trials independently of the therapy investigated. These mandatory domains were: disease activity, pain, morning stiffness, fatigue, physical function, overall functioning and health, and adverse events including death. In addition, 3 extra domains (extra-musculoskeletal manifestations, peripheral manifestations, and structural damage) were included as mandatory for DMARDs, leaving them as optional but important for SMART. As a clarification, structural damage was included as a mandatory domain for at least one trial during the development program of a specific DMARD but not in every trial on that DMARD. Finally, 3 other domains (spinal mobility, sleep, and work and employment) were included as important but optional for all trials. No domain was included in the research agenda layer.

Research agenda		
Important but optional domains		• Spinal mobility • Work and Employment • Sleep
Mandatory domains	**Disease modifying drugs***	• Extra-musculoskeletal manifestations** • Peripheral manifestations*** • Structural damage
	All trials	• Disease activity • Physical functioning • Pain • Overall functioning and Health • Morning stiffness • Adverse Events including death • Fatigue

*Important but optional for trials for interventions other than DMARDs
** Uveitis, inflammatory bowel disease, psoriasis
*** Arthritis, enthesitis, dactylitis

Figure 3 Update core domain set for axial spondyloarthritis presented according to the OMERACT onion**. OMERACT,** Outcomes Measures in Rheumatology; **DMARDs,** disease modifying anti-rheumatic drugs

ASAS voting

In total, 92% (n=57) of ASAS full members participating in the annual workshop voted to accept this proposal. Furthermore, three other aspects related to the domains included in the final onion were voted on. Most members agreed that the most appropriate term when referring to inflammatory bowel disease, uveitis and psoriasis in patients with axSpA is 'extra-musculoskeletal manifestations (EMMs)'. In addition, the assessment of this domain should include the three mentioned manifestations. The domain "peripheral manifestations" should include arthritis, enthesitis and dactylitis. The working group proposal for the onion was slightly adjusted to include these points.

OMERACT endorsement

The ASAS proposal for the core domains is depicted in Fig. 3 and the definition for each of the domains is provided in Table 1 and 2. The proposal was broadly accepted. Combining the results of the two sessions, 100% (n=18) patient research partners and 99% (n=95) representatives of other stakeholders voted to accept the 7 mandatory domains set for all trials. Furthermore, 95% (n=17) patient research partners and 99% (n=97) representatives of other stakeholders accepted to include the three additional mandatory domains for DMARDs. Finally, some minor edits proposed by OMERACT participants were implemented in the final version of the onion.

Table 1 Definitions of domains included in the OMERACT onion. Mandatory domains for all trials.

Disease activity
The domain 'disease activity' covers the level of activity of the disease including signs and symptoms but also objective inflammation that can be assessed by imaging or in the lab.
Pain
Pain, includes overall pain, peripheral pain (pain in the hands and feet, wrists, elbows, shoulders, ankles and knees) and/or spinal pain (pain in the neck and spine) experienced throughout the day as well as pain at night. The sensation of pain (sensation of unpleasant feeling indicating potential or actual damage to some body part or throughout the body) as well as pain intensity (how much pain) and duration are included in this domain.
Morning stiffness
A feeling of stiffness in the back upon getting up in the morning, which influences the ability to move about.
Fatigue
Fatigue describes the overall feeling of tiredness and/or lack of energy; inability to optimally use mental or physical capacity.
Physical function
Physical functioning is defined as one's ability to carry out various activities that require physical capability, ranging from self-care (activities of daily living) to more vigorous activities that require increasing degrees of mobility, strength, or endurance. An important aspect in this domain is physical difficulty: any problems with physical activity resulting from impairment, any activity limitations and participation restrictions; and the ability to transfer oneself from one place to another (i.e. walking, cycling).
Overall functioning and health
In general, overall functioning and health is the perceived quality of an individual's daily life, that is, an assessment of their well-being or lack thereof. This includes all emotional, social and physical aspects of the individual's life. Overall functioning and health is an assessment of how the individual's well-being may be affected over time by a disease, disability or disorder Participation at work, at home and leisure, overall well-being, daily function, social support from family and friends, interpersonal relationships and social roles are all included in overall functioning and health. Also included in this domain are any impairments experienced during the day as a result of sleep problems.
Adverse events
An unexpected medical problem that happens during treatment with a drug or other therapy. Adverse events may be mild, moderate, or severe, and may be caused by something other than the drug or therapy being given.

Table 2 Additional mandatory domains for trials investigating the effect of disease modifying drugs.

Extra-musculoskeletal manifestations (uveitis, inflammatory bowel disease, psoriasis)
Extra-musculoskeletal manifestations include uveitis, inflammatory bowel disease (Crohn's disease and Ulcerative Colitis) and psoriasis. These are frequently occurring in patients with axial spondyloarthritis and belong to the disease spectrum. Other extra-musculoskeletal manifestations that occur more frequently than in the healthy population but do not belong to the disease spectrum are problems with cardiovascular and pulmonary functioning. · Uveitis is a form of eye inflammation. It affects the middle layer of tissue in the eye wall (uvea), hence its name uveitis and occurs in attacks. · Inflammatory bowel disease (IBD) is an umbrella term used to describe disorders that involve chronic inflammation of your digestive tract. Types of IBD include Crohn's disease and Ulcerative Colitis · Psoriasis: a common chronic, inflammatory skin disease characterized by redness of the skin and small dry pieces of skin across the body.
Peripheral manifestations (arthritis, enthesitis, dactylitis)
Peripheral manifestations include enthesitis, dactylitis and arthritis · Enthesitis is the term used to describe inflammation at tendon, ligament or joint capsule insertions. A common location for enthesitis is at the heel, particularly the Achilles tendon. · Dactylitis is severe inflammation of the finger or toe joints. The puffy nature of the inflammation can make your digits look like sausages, which is why they are sometimes called sausage fingers or toes · Arthritis: Inflammation of a joint. When joints are inflamed, they can develop stiffness, warmth, swelling, redness and pain.
Structural damage
Structural damage, determined by any method (e.g. imaging), including structural damage to the spine, peripheral joints (hands and feet, elbows, wrists, ankles, and knees), and root joints (shoulders and hips). Damage to the organs is another manifestation of 'structural damage'.

DISCUSSION

The definition of the core domain set responds to one of the relevant unmet needs in the field of axSpA[32]. The original core set was developed more than 20 years ago and was well implemented[12,17]. However, after more than two decades this core set became outdated and required revision to address all the advances achieved recently in the field of axSpA and to address the current recommended methodology for development of a core outcome set[33]. This manuscript presents the result of a crucial collaborative initiative between ASAS and OMERACT to update the ASAS-OMERACT core outcome set for AS into the ASAS-OMERACT core outcome set for axSpA.

Compared to the original core set, the updated core set for axSpA represents a substantial advance both in content and in the methodology employed. The most recent guidelines for development of a core set were followed as closely as possible. In this sense, the OMERACT and COMET handbooks have been the basis for updating the core set to the highest possible quality[5,22]. The procedure associated with these guidelines is extensive and meticulous. An important aspect of this procedure is the working group and stakeholders participating in the selection of the domains. The updated core set involved all key stakeholders. Furthermore, the number and heterogeneity of participants also increased. While the original core set involved approximately 40 participants the update of the core set involved 376 participants in total, with 50% experts (from more than 40

countries worldwide) and 50% patients, representing both genders equally and covering the entire spectrum of the disease.

Importantly, it should be stressed that the updated core set is meant to be employed in a research setting (i.e. studies evaluating the effect of therapies) but not necessarily in all observational studies or clinical practice. These two latter settings require a different methodology to the one followed in this procedure. Similar to the original core set, the updated core set applies to two scenarios depending on the type of intervention investigated in the trial, splitting the core domains in those that should apply for all trials and those that are mandatory only for DMARDs, while still considered to be important but optional for SMARTs. Like the original core set, the following four domains remained mandatory for all trials: pain, morning stiffness, fatigue and physical function. However, there are some differences between the core sets. The original core set included as mandatory domains for all trials the patient global assessment and spinal mobility. For the updated core set the patient global assessment was removed as this is not really a domain but an instrument, while mobility was moved to being optional but an important domain for all trials. Reasons for this change are lack of standardisation and poor reliability and sensitivity to change[21]. Additionally, overall functioning and health is now included as mandatory for all trials.

This domain was considered relevant when the original core set was defined (at that moment called quality of life); however, the lack of an appropriate instrument to assess this domain in axSpA drove the decision to leave it out. Over time several instruments were developed to assess overall functioning and health[20,34], which led to the inclusion of this domain as mandatory for the updated core set. Furthermore, the original core set also includes two domains as optional but important for all trials, which are sleep and work and employment. Over the last decades, it was shown that sleep disorders and the impact on work and employment are important aspects for patients with axSpA[35-37]. Two new domains have been added as mandatory for all trials in the updated core set. One of them is included in all OMERACT core sets, which is death and adverse events[25]. The other one is disease activity. This was not included as a specific domain in the original set but several instruments assessing this domain such as patient global assessment and acute-phase reactants were included, which reflects that this was already considered relevant[3,38]. The importance of objective measures to assess disease activity such as imaging and serological acute phase reactants was stressed in the breakout sessions, but this will be further discussed during the selection of instruments for this domain.

Importantly, the update of the core outcome set for axSpA is not final. After deciding *what to measure* (core domain set) the next step is deciding *how to measure* the domains by selecting instruments or tools for each domain[5,22]. An important aspect of this step is the

assessment of the measurement properties of candidate instruments. The working group is currently working on this. With this information, the selection of the most appropriate instruments will be achieved by consensus of the key stakeholders. Moreover, we cannot forget one of the most important steps in the development of a core set, which is its implementation. The original core set was successfully implemented[17]. For the update we will design strategies for a broad dissemination and implementation. We are convinced that having the support from ASAS and OMERACT will help in this process. A few potential limitations should be considered. First, the working group followed as closely as possible the current guidelines to develop a core outcome set. Even so, minor modifications had to be made as this process was an update of a previously developed core set and no specific guidelines are currently available to update a core outcome set. Another possible limitation is that instead of running specific qualitative studies to update the core outcome set, we employed the data from the qualitative studies to develop the ASAS/World Health Organisation (WHO) Comprehensive and Brief Core sets of the International Classification of Functioning, Disability and Health (ICF)[39]. These data were used only to identify the candidate domains. After this, all participating stakeholders could add extra domains during the first round of the Delphi survey if they thought these were missing. Hence, we do not think this has influenced the outcome of the process.

CONCLUSION

This manuscript presents the updated ASAS-OMERACT core domain set for axSpA, which is an essential tool for research in this disease. This core set includes the minimum but mandatory set of domains that should be assessed in all clinical trials and longitudinal observational studies evaluating a therapy in patients with axSpA. As this is a minimum, it does not exclude that other domains may be additionally assessed within specific trials. This core set will contribute to ensure that the most relevant aspects of the disease are assessed in all studies and that this is done in a standardised and homogeneous way that will allow comparisons of results across studies

SUPPLEMENTARY DATA

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

REFERENCES

1. Sieper J, Poddubnyy D. Axial spondyloarthritis. *Lancet* 2017;390(10089):73–84.
2. Maguire S, Sengupta R, O'Shea F. The future of axial spondyloathritis treatment. *Rheum Dis Clin North Am* 2020;46(2):357–65.
3. van der Heijde D, Ramiro S, Landewée R, Baraliakos X, Van den Bosch F, Sepriano A, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis* 2017;76(6):978–91.
4. Ward MM, Deodhar A, Gensler LS, Dubreuil M, Yu D, Khan MA, et al. Update of the American college of rheumatology/spondylitis association of america/spondyloarthritis research and treatment network recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. *Arthritis Rheumatol.* 2019 2019;71(10):1599–613.
5. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: version 1.0. *Trials* 2017;18(Suppl 3):280.
6. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;13:132.
7. Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core outcome set-standards for development: the COS-STAD recommendations. *PLoS Med* 2017;14(11):e1002447.
8. Fried BJ, Boers M, Baker PR. A method for achieving consensus on rheumatoid arthritis outcome measures: the OMERACT conference process. *J Rheumatol* 1993;20(3):548–51.
9. Prinsen CA, Vohra S, Rose MR, King-Jones S, Ishaque S, Bhaloo Z, et al. Core Outcome Measures in Effectiveness Trials (COMET) initiative: protocol for an international Delphi study to achieve consensus on how to select outcome measurement instruments for outcomes included in a 'core outcome set. *Trials* 2014;15:247.
10. Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. *Ann Rheum Dis* 2009;68(2):ii1–44 Suppl.
11. [Available from: https://www.asas-group.org.] 2021
12. 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 SpondylitisWorking Group. *J Rheumatol* 1997;24(11):2225–9.
13. van der Heijde D, Calin A, DougadosM, KhanMA, 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(4):951–4.
14. 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(4):945–7.
15. van der Heijde D, van der Linden S, Dougados M, Bellamy N, Russell AS, Edmonds J. Ankylosing spondylitis: plenary discussion and results of voting on selection of domains and some specific instruments. J *Rheumatol* 1999;26(4):1003–5.
16. Zochling J, Sieper J, van der Heijde D, Braun J. Assessment in Ankylosing Spondylitis International Working G. Development of a core set of domains for data collection in cohorts of patients with ankylosing spondylitis receiving anti-tumor necrosis factor-alpha therapy. *J Rheumatol* 2008;35(6):1079–82.
17. Bautista-Molano W, Navarro-Compan V, Landewe RB, Boers M, Kirkham JJ, van der Heijde D. How well are the ASAS/OMERACT core outcome sets for ankylosing spondylitis implemented in randomized clinical trials? A systematic literature review. *Clin Rheumatol* 2014;33(9):1313–22.
18. Lukas C, Landewe R, Sieper J, Dougados M, Davis J, Braun J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68(1):18–24.
19. Maksymowych WP, Mallon C, Morrow S, Shojania K, Olszynski WP, Wong RL, et al. Development and validation of the Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index. *Ann Rheum Dis* 2009;68(6):948–53.
20. Kiltz U, van der Heijde D, Boonen A, Cieza A, Stucki G, Khan MA, et al. Development of a health index in patients with ankylosing spondylitis (ASAS HI): final result of a global initiative based on the ICF guided by ASAS. *Ann Rheum Dis* 2015;74 (5):830–5.
21. Ogdie A, Duarte-Garcia A, Hwang M, Navarro-Compan V, van der Heijde D, Mease P. Measuring outcomes in axial spondyloarthritis. *Arthritis Care Res* 2020;72(10):47–71 Suppl.
22. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L,

d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol* 2014;67(7):745–53.
23. Boers M, Idzerda L, Kirwan JR, Beaton D, Escorpizo R, Boonen A, et al. Toward a generalized framework of core measurement areas in clinical trials: a position paper for OMERACT 11. *J Rheumatol* 2014;41(5):978–85.
24. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new criteria? *Arthritis Rheum* 2005;52 (4):1000–8.
25. Maxwell LJ, Beaton DE, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Core domain set selection according to OMERACT Filter 2.1: The OMERACT methodology. *J Rheumatol* 2019;46(8):1014–20.
26. 2021 [Available from: https://www.dropbox.com/s/fd3673fsma45qe0/OMERACT%20Handbook%20Chapter%204%20Apr%2016%202019.pdf?dl=0.]
27. Boers M, Beaton DE, Shea BJ, Maxwell LJ, Bartlett SJ, Bingham CO, et al. OMERACT filter 2.1: elaboration of the conceptual framework for outcome measurement in health intervention studies. *J Rheumatol* 2019;46(8):1021–7.
28. [Available from: https://www.comet-initiative.org/Studies/Details/1132.] 2021
29. [Available from: https://omeract.org/working-groups/axial-spa.] 2021
30. Boel A, Navarro-Compan V, Landewe R, van der Heijde D. Two different invitation approaches for consecutive rounds of a Delphi survey led to comparable final outcome. *J Clin Epidemiol* 2021;129:31–9.
31. Boel A, Navarro-Compan V, Boonen A, Mease P, Kiltz U, Dougados M, Landewé R, van der Heijde D, et al. Domains to be considered for the core outcome set of axial spondyloarthritis: results from a 3-round Delphi survey. *J Rheumatol* 2021
32. Winthrop KL, Weinblatt ME, Bathon J, Burmester GR, Mease PJ, Crofford L, et al. Unmet need in rheumatology: reports from the Targeted Therapies meeting 2019. *Ann Rheum Dis* 2020;79(1):88–93.
33. Andreasen RA, Kristensen LE, Baraliakos X, Strand V, Mease PJ, de Wit M, et al. Assessing the effect of interventions for axial spondyloarthritis according to the endorsed ASAS/OMERACT core outcome set: a meta-research study of trials included in Cochrane reviews. *Arthritis Res Ther* 2020;22(1):177.
34. Doward LC, Spoorenberg A, Cook SA, Whalley D, Helliwell PS, Kay LJ, et al. Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis. *Ann Rheum Dis* 2003;62(1):20–6.
35. Boonen A, Sieper J, van der Heijde D, Dougados M, Bukowski JF, Valluri S, et al. The burden of non-radiographic axial spondyloarthritis. *Semin Arthritis Rheum* 2015;44(5):556–62.
36. Reilly MC, Gooch KL, Wong RL, Kupper H, van der Heijde D, Validity, reliability and responsiveness of the work productivity and activity impairment questionnaire in ankylosing spondylitis. *Rheumatology* 2010;49(4):812–9.
37. Leverment S, Clarke E, Wadeley A, Sengupta R. Prevalence and factors associated with disturbed sleep in patients with ankylosing spondylitis and nonradiographic axial spondyloarthritis: a systematic review. *Rheumatol Int* 2017;37(2):257–71.
38. Smoln JS, Schols M, Braun J, Dougados M, FitzGerad O, Gladman DD, et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. *Ann Rheum Dis* 2018;77(1):3–17.
39. Boonen A, Braun J, van der Horst Bruinsma IE, Huang F, MaksymowychW, Kostanjsek N, et al. ASAS/WHO ICF Core Sets for ankylosing spondylitis (AS): how to classify the impact of AS on functioning and health. *Ann Rheum Dis* 2010;69(1):102–7.

CHAPTER 8

TEST–RETEST RELIABILITY OF OUTCOME MEASURES: DATA FROM THREE TRIALS IN RADIOGRAPHIC AND NON-RADIOGRAPHIC AXIAL SPONDYLOARTHRITIS

Anne Boel, Victoria Navarro-Compán, Désirée van der Heijde

RMD Open. 2021 Dec;7(3):e001839

ABSTRACT

Background

Aim of this study was to assess test–retest reliability of candidate instruments for the mandatory domains of the Assessment of Spondyloarthritis international Society (ASAS)-Outcome Measures in Rheumatology core set for axial spondyloarthritis (axSpA).

Methods

Screening and baseline data from COAST-V, COAST-X and RAPID-axSpA was used to evaluate test-retest reliability of each candidate instrument for the mandatory domains (disease activity, pain, morning stiffness, fatigue, physical function, overall functioning and health). A maximum time interval of 28 days between both visits was used for inclusion in this study. Test–retest reliability was assessed by intraclass correlation coefficient (ICC). Bland and Altman plots provided mean difference and 95% limits of agreement, which were used to calculate the smallest detectable change (SDC). Data were analysed for radiographic and non-radiographic axSpA separately.

Results

Good reliability was found for Ankylosing Spondylitis Disease Activity Score (ICC 0.79, SDC 0.6), C reactive protein (ICC 0.72–0.79, SDC 12.3–17.0), Bath Ankylosing Spondylitis Functional Index (ICC 0.87, SDC 1.1) and 36-item Short-Form Health Survey (ICC Physical Component Summary 0.81, SDC 4.7, Mental Component Summary 0.80, SDC 7.3). Moderate reliability was found for Bath Ankylosing Spondylitis Disease Activity Index (ICC 0.72, SDC 1.1), patient global assessment (ICC 0.58, SDC 1.5), total back pain (ICC 0.64, SDC 1.3), back pain at night (ICC 0.67, SDC 1.3), morning stiffness (ICC 0.52–0.63, SDC 1.5–2.2), fatigue (ICC 0.65, SDC 1.3) and ASAS-Health Index (ICC 0.74, SDC 2.5). Reliability and SDC for the radiographic and non-radiographic axSpA subgroups were similar.

Conclusion

Overall reliability was good, and comparable levels of reliability were found for patients with radiographic and non-radiographic axSpA, even though most instruments were developed for radiographic axSpA. Composite measures showed higher reliability than single-item measures in assessing disease activity in patients with axSpA.

INTRODUCTION

Uniformity in reporting primary outcomes of clinical trials allows for a direct comparison between studies investigating different therapies in the same patient population. Herein, there is an essential role for core outcome sets (COS), which contain the mandatory outcomes (domains) that should be assessed and reported as a minimum in all trials[1,2]. Over time, new instruments to assess these domains may be developed and also more data may become available regarding measurement properties of already existing instruments, underlining the need to periodically review COS. Currently, the Assessment of Spondyloarthritis international Society (ASAS) is working on an update of the original ASAS/Outcome Measures in Rheumatology (OMERACT) core set for ankylosing spondylitis (AS) of which the domains have been selected and endorsed[3,4]. An important aspect that led to this decision was that AS belongs to a broader disease spectrum, axial spondyloarthritis (axSpA), which includes two forms—that can also be regarded as two stages- of the same disease: radiographic axSpA (r-axSpA, traditionally known as AS, that is, axSpA with definite sacroiliitis according to the modified New York (mNY) criteria[5]) and non-radiographic axSpA (nr-axSpA, that is, axSpA without definite sacroiliitis on radiographs[6]). Even though both nr-axSpA and r-axSpA are now considered part of the same disease spectrum, most instruments used to assess effectiveness of treatment were developed for and tested only in patients with r-axSpA.

The updated COS should be applicable to all patients with axSpA. Therefore, all instruments should have good psychometric properties for patients in both disease subgroups (i.e., r-axSpA and nr-axSpA) to be included as mandatory instruments[1,2]. The psychometric properties include truth (domain match, face and content validity), feasibility, construct validity and discrimination (test-retest reliability, responsiveness, clinical trial discrimination and thresholds of meaning)[7]. In this manuscript, we evaluate only one aspect in detail, namely test–retest reliability. Reliability is an important psychometric property, as it informs users whether the same result will be obtained if assessed twice in a situation where there is no change. Hence, the aim of this study was to assess test-retest reliability of the candidate instruments for the selected mandatory domains of the core set that should be assessed in all trials evaluating a new treatment in patients with r-axSpA and nr-axSpA[4].

MATERIALS AND METHODS

Patients

For this study, we used screening and baseline data from three large samples in axSpA: data from COAST-V and COAST-X (initiated by Eli Lilly and Company and registered with ClinicalTrials.gov as NCT02696785 and NCT02757352 respectively) and RAPID-axSpA (initiated by UCB Pharma and registered with ClinicalTrials.gov as NCT01087762). These randomised controlled trials (RCTs) are described in detail elsewhere[8-10]. In brief, all RCTs included patients aged ≥18 years who fulfilled ASAS criteria for axSpA[11] and had an inadequate response to NSAIDs or a history of intolerance to NSAIDs. COAST-V included patients with r-axSpA[8] (i.e., with sacroiliitis according to the mNY criteria[5]) while COAST-X included patients with nr-axSpA[9]; and RAPID-axSpA comprised patients with either r-axSpA or nr-axSpA[10]. As these patients were entering an RCT, they needed to have active disease at screening and baseline, defined as a Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)[12] score of ≥4 and total back pain in the past week ≥4 (on a 0–10 Numeric Rating Scale (NRS)).

Outcomes

The ASAS-OMERACT core domain set for axSpA[4] describes the domains that should be measured in axSpA trials investigating symptom modifying and disease-modifying therapies. Seven domains are mandatory in all axSpA trials: disease activity, pain, morning stiffness, fatigue, physical function, overall functioning and health and adverse events. Information from all the instruments (n=13) employed to assess these domains -with the exception of adverse events- at both screening and baseline in COAST-V, COAST-X and RAPID-axSpA was used to evaluate test–retest reliability of each instrument.

Four instruments that could be used to assess the domain disease activity were available: the Ankylosing Spondylitis Disease Activity Score (ASDAS) -specifically ASDAS-C reactive protein (CRP)[13], the BASDAI using NRS answer modalities[12], the patient global assessment (PtGA) using an NRS[14] and CRP, measured in mg/L. Two of the instruments used to assess pain were available: 0–10 NRS for total back pain in the past week and 0–10 NRS for pain at night in the past week[14]. Questions 5 (How would you describe the overall level of morning stiffness you have had from the time you wake up?) and 6 (How long does your morning stiffness last from the time you wake up?) of the BASDAI and a composite score of questions 5 and 6 ((Q5 +Q6)/2) were the instruments available to evaluate morning stiffness. The one instrument available to estimate fatigue was question 1 of the BASDAI. To evaluate physical function, one instrument was present: the Bath Ankylosing Spondylitis Functional Index (BASFI)[15].

Two of the instruments that could survey overall functioning and health were available: the ASAS-Health Index (ASAS-HI)[16] and Medical Outcomes Study 36-item Short-Form Health Survey (SF-36)[17]. All these instruments are commonly used in trials assessing treatment effect in axSpA and have shown content, face and construct validity[18].

Spinal mobility was considered an important but optional domain in the axSpA ASAS/OMERACT domain core set[4]. Nonetheless, it was included in this study as it is often assessed in clinical trials and daily practice. One composite instrument and two additional single measures that can be used to evaluate spinal mobility were evaluated: the Bath Ankylosing Spondylitis Metrology Index (BASMI) linear[19] (including modified Schober, lateral spinal flexion, tragus-to-wall distance, cervical rotation, intermalleolar distance) and chest expansion and occiput-to-wall distance[14].

Analyses

Test–retest reliability was assessed by intraclass correlation coefficient (ICC) (two-way random effect model with absolute agreement[20,21]). An ICC >0.9 was an indication of excellent reliability, >0.75 to 0.9 of good reliability, 0.5 to 0.75 of moderate reliability and ICC <0.5 of poor reliability[21]. Bland and Altman plots were created for each instrument to assess mean difference and 95% limits of agreement and to evaluate homoscedasticity. Measurement error as a measure of the scale was assessed by analysing the smallest detectable change (SDC) based on the 95% limits of agreement using the formula: SDC=1.96×SD of the mean difference of the two assessments/(√2 x √2)[22]. The SDC corresponds to the minimum change beyond measurement error that can be detected in an individual patient over time with 95% likelihood. Calculation of the limits of agreement (and the SDC) assumed that reliability was homoscedastic.

In this study, we operated under an a priori assumption underlying the test–retest experiments, namely that in truth the scores for all instruments do not change over the limited period of time between assessments (i.e., there is no systematic error). This assumption of no change has been proven by the Bland and Altman plots, which demonstrated that the mean difference between test and retest was always (very close to) zero, indicating that the no systematic error assumption holds.

As there was a large variation in the number of days between screening and baseline assessments in both datasets, it was decided to use a maximum time interval of 28 days between both visits as a cut-off for inclusion in this study.

Unfortunately, in the RAPID-axSpA dataset the PtGA was only assessed at baseline, and the baseline values were used to calculate ASDAS both at screening and baseline. As the ASDAS is calculated from the PtGA, questions 2, 3 and 6 from the BASDAI and CRP[13], the results of this dataset should be interpreted with caution, as variability in patient global was not considered and as a result the reliability of the ASDAS may be artificially improved. However, the values in the COAST trials were very similar.

Results were bundled per domain and presented for all axSpA patients, followed by information per disease subgroup (i.e., r-axSpA and nr-axSpA). Data from both COAST datasets were combined to assess test–retest reliability of the instruments in axSpA patients.

RESULTS

A total of 341 r-axSpA patients in the COAST-V dataset, 302 nr-axSpA patients in the COAST-X dataset and 326 patients (177 r-axSpA and 149 nr-axSpA) in the RAPID-axSpA dataset had data available at screening and baseline. From these, 104 r-axSpA patients from COAST-V, 104 nr-axSpA patients from COAST-X and 221 patients from RAPID-axSpA (119 r-axSpA and 102 nr-axSpA) who had both measurements for at least one of the assessed instruments within a time frame of 28 days were included in this analysis.

Of the included r-axSpA patients from COAST-V 81% were male median (IQR) age was 39 (34–47) and mean (SD) symptom duration 15.1 (9.9) years. The selection of nr-axSpA patients from COAST-X included 55% male patients, with a median age of 38 (27–49) and mean symptom duration of 9.9 (8.8) years. In RAPID-axSpA 62% of the included patients were male (74% in r-axSpA, 49% in nr-axSpA), the median age range was 31–35 years (46–50 in r-axSpA, 31–35 in nr-axSpA) and mean symptom duration was 6.0 (6.9) years (7.4 (7.6) in r-axSpA, 4.3 (5.6) in nr-axSpA).

The mean symptom duration in the patient selection included in this study was somewhat shorter than the mean symptom duration of the entire study populations (COAST-V 16.1 (10.9); COAST-X 10.7 (9.7); RAPID-axSpA 6.7 (7.4)). Median age and the percentage of female patients were similar to the original study populations[8-10].

The number of days between assessments ranged between 8 and 28 days in COAST-V, between 9 and 28 days in COAST-X and between 2 and 28 days in RAPID-axSpA; the mean (SD) number of days between assessments were 22 (5) in COAST-V, 21 (5) in COAST-X and 18 (7) days in RAPID-axSpA. The proportion of missing data varied somewhat between measurements and datasets, but was always very small (<5%). Participants with missing

data for an instrument at either screening or baseline were excluded from analysis for that specific instrument. The number of available data per instrument is provided in table 1. Information available from the literature regarding reliability of the instruments included in the current study is presented in table 1[23-36].

Detailed results from all trials and subgroups are provided in tables 1 and 2. In the text, reliability per domain is described only for the total axSpA group in the COAST datasets, as these included most instruments. Only if reliability varied considerably between subgroups or trials, reliability of these groups is discussed additionally.

Regarding the four instruments assessing disease activity: good reliability was found for ASDAS (ICC 0.79, SDC 0.6) and CRP in COAST (ICC 0.79, SDC 12.3), whereas reliability for CRP in the RAPID-axSpA dataset was slightly lower (ICC 0.72, SDC 17.0) (table 1). Reliability was moderate for BASDAI (ICC 0.72, SDC 1.1); and for the PtGA reliability was moderate (ICC 0.58, SDC 1.5) too, except for the r-axSpA group, for which reliability was poor (ICC 0.48, SDC 1.6). The two instruments used to evaluate pain showed moderate reliability (NRS total back pain (ICC 0.64, SDC 1.3); NRS back pain at night (ICC 0.67, SDC 1.3)). Moderate reliability was found for the instruments used to assess morning stiffness (ICC 0.52–0.63, SDC 1.5–2.2) as well. The instrument used to determine fatigue showed moderate reliability (ICC 0.65, SDC 1.3). The data showed good reliability (ICC 0.87, SDC 1.1) for the BASFI, used to measure physical function. For the two instruments used to survey overall functioning and health, good reliability was found for the Physical Component Summary (ICC 0.81, SDC 4.7) and Mental Component Summary (ICC 0.80, SDC 7.3) subscales of the SF-36, and the ASAS-HI had moderate reliability (ICC 0.74, SDC 2.5), except for the nr-axSpA subgroup in which reliability was good (ICC 0.77, SDC 2.5). In the domain spinal mobility, reliability was excellent (ICC 0.93, SDC 0.6) for BASMI in RAPID-axSpA. Tragus-to-wall and occiput-to-wall distance showed excellent reliability, except for the nr-axSpA subpopulation, for which the reliability was good. For all other mobility measures reliability was good (table 2)[37-43].

Bland and Altman plots showed a reasonably homoscedastic variation for all measurement instruments, with the exception of CRP where the variation was more pronounced in the lower end of the range (online supplemental figures 1–27).

Table 1 Test-retest data of assessed instruments in COAST (combined data COAST-V & COAST-X) and RAPID-axSpA,28-day interval

Data source		N	Screening mean (SD)	Baseline mean (SD)	Mean difference (95% CI)	ICC (95% CI)	SDC	*Data from literature*	
								ICC	MCID/MCII/ SDC
Disease activity								*Disease activity*	
	ASDAS (0.6 to 8)							**ASDAS**	
	axSpA	204	3.8 (.9)	3.8 (.8)	.08 (.00 to .16)	.79 (.73 to .84)	.6	ICC: 0.95[32]	SDC range 1.01-1.18; MCII: 1.1[29]
COAST	r-axSpA	103	3.7 (.9)	3.7 (.8)	.03 (-.08 to .13)	.80 (.71 to .86)	.6		
	nr-axSpA	101	4.0 (.9)	3.8 (.9)	.13 (.02 to .24)	.78 (.69 to .85)	.6		
	ASDAS (0.6 to 8)*								
	axSpA	215	4.0 (.8)	4.0 (.9)	.01 (-.06 to .08)	.79 (.73 to .83)	.5		
RAPID-axSpA	r-axSpA	118	4.1 (.7)	4.0 (.9)	.01 (-.09 to .12)	.75 (.66 to .82)	.6		
	nr-axSpA	97	3.9 (.8)	3.9 (.9)	.01 (-.09 to .11)	.83 (.76 to .88)	.5		
	BASDAI (0 to 10)							**BASDAI**	
	axSpA	208	6.9 (1.5)	6.9 (1.4)	.05 (-.10 to .20)	.72 (.65 to .78)	1.1	ICC range 0.87-0.94[23,30,34]	MCID: 1.3[34]; MCII: 1.1-1.2[27] SDC: 0.9[30]
COAST	r-axSpA	104	6.6 (1.4)	6.6 (1.3)	.05 (-. 16 to .26)	.67 (.55 to .77)	1.1		
	nr-axSpA	104	7.2 (1.5)	7.2 (1.5)	.05 (-.16 to .26)	.74 (.64 to .82)	1.1		
	BASDAI (0 to 10)								
	axSpA	217	6.5 (1.5)	6.6 (1.5)	-.13 (-.31 to .05)	.62 (.53 to .70)	1.3		
RAPID-axSpA	r-axSpA	119	6.5 (1.5)	6.6 (1.6)	-.11 (-.36 to .15)	.61 (.49 to .71)	1.3		
	nr-axSpA	98	6.6 (1.4)	6.7 (1.5)	-.17 (-.42 to .09)	.64 (.50 to .74)	1.2		
	Patient global (0 to 10)							**Patient global**	
	axSpA	208	7.0 (1.7)	7.1 (1.6)	-.05 (-.25 to .16)	.58 (.48 to .66)	1.5	ICC range 0.91-0.93[23,31]	MCII: 1.4[33] SDC†: 1.8[23]
COAST	r-axSpA	104	6.7 (1.6)	6.8 (1.5)	-.13 (-.44 to .19)	.48 (.32 to .61)	1.6		
	nr-axSpA	104	7.3 (1.6)	7.3 (1.7)	.03 (-.24 to .30)	.64 (.51 to .74)	1.3		
	CRP (mg/dL)								
	axSpA	204	16.7 (21.7)	13.8 (17.8)	2.93 (1.20 to 4.67)	.79 (.73 to .84)	12.3		
COAST	r-axSpA	103	15.7 (19.0)	14.3 (17.2)	1.39 (-.46 to 3.24)	.86 (.80 to .90)	9.3		
	nr-axSpA	101	17.8 (24.2)	13.3 (18.6)	4.51 (1.56 to 7.46)	.75 (.63 to .83)	14.6		
	CRP (mg/dL)								
	axSpA	219	20.6 (20.9)	20.4 (25.6)	.18 (-2.17 to 2.53)	.72 (.65 to .78)	17.0		
RAPID-axSpA	r-axSpA	118	21.1 (18.8)	22.2 (28.1)	-1.17 (-5.12 to 2.77)	.60 (.47 to .70)	21.0		
	nr-axSpA	101	20.1 (23.1)	18.4 (22.2)	1.76 (-.37 to 3.90)	.89 (.84 to .92)	10.5		

Table 1 Continued

Data source		N	Screening mean (SD)	Baseline mean (SD)	Mean difference (95% CI)	ICC (95% CI)	SDC	*Data from literature*	
								ICC	MCID/MCII/ SDC
Pain								*Pain*	
	Total back pain (0 to 10)							**Total back pain**	
	axSpA	208	7.1 (1.6)	7.2 (1.5)	-.15 (-.32 to .03)	.64 (.56 to .72)	1.3	ICC range 0.86-0.9[23,34]	MCID: 1.6 (range 1.5-1.6)[34] SDC: 1.8[30]
COAST	r-axSpA	104	6.9 (1.5)	7.0 (1.3)	-.13 (-.38 to .13)	.58 (.43 to .69)	1.3		
	nr-axSpA	104	7.2 (1.6)	7.4 (1.6)	-.17 (-.42 to .07)	.69 (.57 to .78)	1.3		
	Night pain (0 to 10)							**Night pain**	
	axSpA	208	7.0 (1.8)	7.1 (1.6)	-.07 (-.26 to .12)	.67 (.59 to .74)	1.3	ICC range 0.83-0.92[23,34]	MCID: 1.8 (range 1.5-2.1)[34]
COAST	r-axSpA	104	6.9 (1.8)	6.8 (1.6)	.05 (-.22 to .32)	.65 (.53 to .75)	1.3		
	nr-axSpA	104	7.2 (1.8)	7.3 (1.7)	-.19 (-.47 to .08)	.69 (.56 to .78)	1.4		
Morning stiffness								*Morning stiffness*	
	BASDAI Q5: Morning stiffness severity (0 to 10)							**BASDAI Q5: Morning stiffness severity**	
	axSpA	208	7.3 (1.9)	7.2 (1.7)	.13 (-3.02 to 3.29)	.63 (54 to .70)	1.5	ICC 0.85[36]	SDC†: 1.4[36]
COAST	r-axSpA	104	7.2 (1.8)	6.9 (1.6)	.26 (-2.61 to 3.13)	.64 (.51 to .74)	1.4		
	nr-axSpA	104	**7.5 (2.1)**	7.5 (1.8)	.01 (-3.40 to 3.42)	.62 (.48 to .72)	1.7		
	BASDAI Q6: Morning stiffness duration (0 to 10)‡								
	axSpA	208	6.2 (2.4)	6.2 (2.2)	-.01 (-4.55 to 4.52)	.52 (.41 to .61)	2.2		
COAST	r-axSpA	104	6.0 (2.4)	5.8 (2.2)	.13 (-4.43 to 4.68)	.51 (.35 to .64)	2.2		
	nr-axSpA	104	6.4 (2.4)	6.5 (2.3)	-.15 (-4.68 to 4.37)	.52 (.37 to .65)	2.2		
	BASDAI Morning stiffness (0 to 10) composite [Q5+Q6/2]							**BASDAI Morning stiffness composite**	
	axSpA	208	6.7 (1.8)	6.7 (1.7)	.06 (-.15 to .27)	.63 (.55 to .71)	1.5	ICC range 0.85-0.91[34,36]	MCID: 1.7 (range 1.0-2.7)[34]
COAST	r-axSpA	104	6.6 (1.8)	6.4 (1.6)	.19 (-.11 to .49)	.58 (.43 to .69)	1.5		
	nr-axSpA	104	6.9 (1.9)	7.0 (1.8)	-.07 (-.37 to .22)	.67 (.55 to .77)	1.5		
	BASDAI Morning stiffness (0 to 10) composite [Q5+Q6/2]								
	axSpA	217	6.3 (2.6)	6.3 (2.4)	.01 (-.28 to .32)	.60 (.51 to .68)	2.2		
RAPID-axSpA	r-axSpA	119	6.5 (2.5)	6.4 (2.4)	.09 (-.31 to .50)	.60 (.47 to .70)	2.2		
	nr-axSpA	89	6.2 (2.7)	6.3 (2.4)	-.08 (-.53 to .37)	.61 (.47 to .72)	2.2		

Table 1 Continued

Data source		N	Screening mean (SD)	Baseline mean (SD)	Mean difference (95% CI)	ICC (95% CI)	SDC	*Data from literature*	
								ICC	MCID/MCII/ SDC
Fatigue								*Fatigue*	
	BASDAI Q1: Fatigue (0 to 10)							**BASDAI Q1**	
	axSpA	208	7.2 (1.7)	7.1 (1.6)	.08 (-.10 to .27)	.65 (.57 to .72)	1.3	ICC: 0.60-0.85[34,35]	MCID: 1.1 (range 1.0-1.5)[34]
COAST	r-axSpA	104	6.8 (1.6)	6.8 (1.6)	.01 (-.27 to .29)	.59 (.45 to .71)	1.4		SDC: 1.7[30]
	nr-axSpA	104	7.5 (1.7)	7.4 (1.6)	.16 (-.09 to .41)	.68 (.57 to .77)	1.3		
	BASDAI Q1: Fatigue (0 to 10)								
	axSpA	217	6.5 (1.9)	6.8 (1.9)	-.29 (-.54 to-.03)	.53 (.42 to .62)	1.8		
RAPID-axSpA	r-axSpA	119	6.6 (1.9)	6.8 (2.0)	-.13 (-.48 to .21)	.54 (.40 to .65)	1.8		
	nr-axSpA	98	6.4 (1.9)	6.9 (1.8)	-.47 (-.84 to-.10)	.51 (.35 to .64)	1.8		
Physical function								*Physical function*	
	BASFI (0 to 10)							**BASFI**	
	axSpA	208	6.2 (2.1)	6.3 (2.0)	-.13 (-.27 to .02)	.87 (.83 to .90)	1.1	ICC range 0.92-0.94[15,30]	MCID: 1.1 (range 1.0-1.1)[34]
COAST	r-axSpA	104	6.0 (2.1)	6.0 (2.1)	-.08 (-.27 to .11)	.89 (.84 to .92)	.9		MCII: 0.6-1.1[27,33]
	nr-axSpA	104	6.4 (2.0)	6.5 (2.0)	-.17 (-.39 to .05)	.84 (.77 to .89)	1.1		SDC: 0.7[30]
Overall functioning & health								*Overall functioning & health*	
	ASAS Health Index (0-17)							**ASAS Health Index**	
	axSpA	208	8.6 (3.6)	8.7 (3.6)	-.07 (-.43 to .28)	.74 (.68 to .80)	2.5	ICC range 0.84-0.98[24-26,28]	SDC: 3.0[25]
COAST	r-axSpA	104	7.7 (3.4)	7.7 (3.3)	-.05 (-.57 to .48)	.68 (.56 to .77)	2.6		
	nr-axSpA	104	9.5 (3.7)	9.8 (3.7)	-.10 (-.59 to .39)	.77 (.68 to .84)	2.5		
	SF-36 PCS (0-100)							**SF-36 PCS**	
	axSpA	208	34.0 (7.9)	34.9 (7.8)	-.89 (-1.55 to-.23)	.81 (.75 to .85)	4.7		MCID: 3.8[34]
COAST	r-axSpA	104	35.7 (8.0)	36.3 (7.4)	-.63 (-1.49 to .23)	.83 (.76 to 88)	4.3		
	nr-axSpA	104	32.4 (7.6)	33.5 (8.1)	-1.15 (-2.17 to-.14)	.77 (.68 to .84)	5.1		
	SF-36 MCS (0-100)							**SF-36 MCS**	
	axSpA	208	47.8 (11.7)	48.0 (11.8)	-.15 (-1.17 to .87)	.80 (.75 to .84)	7.3		MCID: 2.4[34]
COAST	r-axSpA	104	50.5 (10.3)	50.4 (10.5)	.08 (-1.26 to 1.42)	.78 (.69 to .85)	6.7		
	nr-axSpA	104	45.2 (12.6)	45.6 (12.5)	-.37 (-1.93 to 1.18)	.80 (.72 to .86)	7.8		

* PtGA was only assessed at baseline, and baseline values were used to calculate ASDAS at both timepoints, meaning variability in PtGA was not considered and reliability of the ASDAS may be artificially improved. † Calculated from the SDD using the formula SDC=1.96×SDD/(√2 x √2) ‡ score range 0-10, with three anchors: "0 hours" (score 0), "1 hour"

(score 5), and "2 or more hours" (score 10)

ASAS-HI, Assessment of Spondyloarthritis international Society Health Index [scores range from 0 to 17, higher scores indicate worse health status]; **ASDAS**, Ankylosing Spondylitis Disease Activity Score [scores range from 0.6 to 8, determined by the level of CRP or ESR (8 is the approximate maximum, given a CRP of 200), higher scores signify higher disease activity]; **axSpA**, axial spondyloarthritis; **BASDAI**, Bath Ankylosing Spondylitis Disease Activity Index [scores range from 0 to 10, higher scores signify higher disease activity]; **BASFI**, Bath Ankylosing Spondylitis Functional Index [scores range from 0 to 10, higher scores signify greater impairment]; **CRP**, C-reactive protein [measured in mg/dL]; **ICC,** intraclass correlation coefficient; **MCID,** minimal clinically important difference; **MCII,** minimal clinically important improvement; **Night pain**, Nocturnal back pain in the past week measured using a NRS scale [scores range from 0 to 10, higher scores indicating more pain]; **nr-axSpA**, non-radiographic axial spondyloarthritis; **Patient global**, Patient global assessment of disease activity measured using an NRS scale [scores range from 0 to 10, higher scores indicate worse health]; **r-axSpA**, radiographic axial spondyloarthritis; **SDC,** smallest detectable change; **SDD,** standard deviation of the mean difference of the two assessments; **SF-36 PCS**, Physical Component Summary of the Medical Outcomes Study 36-item Short-Form Health Survey [scores range from 0 to 100, higher scores indicate better health]; **SF-36 MCS**, Mental Component Summary of the Medical Outcomes Study 36-item Short-Form Health Survey [scores range from 0 to 100, higher scores indicate better health]; **Total back pain**, in the past week measured using a NRS scale [scores range from 0 to 10, higher scores indicating more pain].

Table 2 Test-retest data of spinal mobility instruments measured in RAPID-axSpA, 28-day interval

	N	Screening mean (SD)	Baseline mean (SD)	Mean difference (95% CI)	ICC (95% CI)	SDC	*Data from literature* ICC	MCID/MCII/SDC
BASMI (0 to 10)							**BASMI**	
axSpA	218	3.6 (1.5)	3.7 (1.6)	-.12 (-.20 to-.05)	.93 (.91 to .95)	.6	ICC rage 0.91-0.97[37,39,41]	SDC range 0.82-0.95[37,41]
r-axSpA	117	4.2 (1.5)	4.4 (1.5)	-.19 (-.31 to-.08)	.91 (.87 to .94)	.6		
nr-axSpA	101	3.0 (1.3)	3.0 (1.4)	-.05 (-.15 to .06)	.93 (.90 to .95)	.5		
Modified Schober (cm)							**Modified Schober**	
axSpA	216	3.9 (2.2)	3.8 (2.0)	.12 (-.03 to .27)	.86 (.82 to .89)	1.1	ICC inter-observer range 0.75-0.96[38,40,42,43]; ICC intra-observer range 0.63-0.94[38,40,43]	SDC 1.4[40]
r-axSpA	116	3.5 (2.1)	3.3 (2.0)	.15 (-.05 to .36)	.85 (.79 to .90)	1.1		
nr-axSpA	100	4.4 (2.3)	4.3 (1.8)	.09 (-.14 to .31)	.85 (.78 to .90)	1.1		
Lateral spinal flexion (cm)							**Lateral spinal flexion**	
axSpA	216	12.6 (6.6)	12.1 (6.0)	.39 (-.02 to .79)	.80 (.75 to .85)	2.9	ICC inter-observer range 0.-0.77-0.98[38,40,42,43] ; ICC intra-observer range 0.65-0.98[38,40,43]	SDC 5.1[40]
r-axSpA	116	11.6 (7.3)	10.3 (5.6)	.98 (.46 to 1.50)	.75 (.65 to .82)	2.8		
nr-axSpA	100	13.8 (5.6)	14.1 (5.8)	-.30 (-.91 to .32)	.86 (.79 to .90)	3.0		
Tragus-to-wall distance (cm)							**Tragus-to-wall distance**	
axSpA	218	13.5 (4.9)	13.7 (5.0)	-.19 (-.46 to .07)	.92 (.90 to .94)	1.9	ICC inter-observer range 0.85-0.98[38,42,43] ; ICC intra-observer range 0.94-0.98[38,43]	
r-axSpA	117	14.8 (5.8)	14.9 (5.7)	-.14 (-.45 to .17)	.96 (.94 to .97)	1.6		
nr-axSpA	101	12.1 (3.0)	12.4 (3.7)	-.25 (CI-.70 to .19)	.78 (.70 to .85)	2.2		
Cervical rotation (degrees)							**Cervical rotation**	
axSpA	218	55.9 (20.9)	54.7 (21.4)	1.20 (-.38 to 2.79)	.85 (.80 to .88)	11.5	ICC inter-observer range 0.69-0.94[38,40,42] ; ICC intra-observer range 0.56-0.95[38,40]	SDC 12.2[40]
r-axSpA	117	48.5 (19.7)	46.6 (19.7)	1.89 (-.62 to 4.41)	.76 (.67 to .83)	13.4		
nr-axSpA	101	64.4 (19.2)	64.0 (19.3)	.40 (-1.40 to 2.21)	.89 (.84 to .92)	8.9		
Intermalleolar distance (cm)							**Intermalleolar distance**	
axSpA	216	98.1 (25.8)	96.6 (27.3)	1.53 (-.10 to 3.16)	.89 (.87 to .92)	11.7	ICC inter-observer 0.93[42]; ICC intra-observer 0.72[40]	SDC 20.2[40]
r-axSpA	116	94.5 (25.9)	93.2 (28.2)	1.40 (-1.06 to 3.85)	.88 (.83 to .92)	12.9		
nr-axSpA	100	102.2 (25.2)	100.6 (25.7)	1.69 (-.39 to 3.76)	.92 (.88 to .94)	10.2		
Chest expansion (cm)							**Chest expansion**	
axSpA	217	3.7 (2.2)	3.7 (2.0)	-.06 (-.22 to .10)	.78 (.72 to .83)	1.1	ICC inter-observer range 0.55-0.85[40,42,43] ; ICC intra-observer range 0.63-0.95[40,43]	SDC 2.2[40]
r-axSpA	118	3.6 (2.5)	3.4 (2.0)	.07 (-.14 to .29)	.76 (.68 to .83)	1.1		
nr-axSpA	99	3.9 (1.8)	4.2 (2.0)	-.22 (-.46 to .01)	.81 (.72 to .87)	1.1		

Table 2 Continued

	N	Screening mean (SD)	Baseline mean (SD)	Mean difference (95% CI)	ICC (95% CI)	SDC	*Data from literature*	
							ICC	**MCID/MCII/SDC**
Occiput-to-wall distance (cm)							**Occiput-to-wall distance**	
axSpA	218	3.7 (5.6)	3.6 (5.5)	.06 (-.19 to .30)	.95 (.93 to .96)	1.8	ICC inter-observer range 0.84-0.89[40,42,43] ; ICC intra-observer range 0.49-0.94[40,43]	SDC 0.9[40]
r-axSpA	118	4.5 (6.3)	4.7 (6.4)	-.11 (-.41 to .19)	.97 (.96 to .98)	1.6		
nr-axSpA	100	2.6 (4.5)	2.4 (3.7)	.25 (-.14 to .65)	.88 (.83 to .92)	1.9		

axSpA, axial spondyloarthritis; **BASMI**, Bath Ankylosing Spondylitis Metrology Index [scores range from 0 to 10, higher scores signify greater impairment]; **ICC,** intraclass correlation coefficient; **MCID,** minimal clinically important difference; **MCII,** minimal clinically important improvement; **nr-axSpA**, non-radiographic axial spondyloarthritis; **r-axSpA**, radiographic axial spondyloarthritis; **SDC,** smallest detectable change; **SDD,** standard deviation of the mean difference of the two assessments

DISCUSSION

The results from this study showed that the test-retest reliability of the investigated instruments was moderate to excellent and similar in the axSpA group and each of the disease subgroups r-axSpA and nr-axSpA. Furthermore, for those instruments where data was available from the COAST and RAPID-axSpA studies, levels of reliability were comparable between datasets as well. Finally, we found ICCs were higher for multi-item instruments compared with single-item instruments in the same domain. This is reasonable as the impact of variance caused by measurement error in the individual items of a multi-item instrument is reduced when they are combined into a single score, resulting in a more precise score for a multi-item instrument compared with its single-item counterparts[44-46]. For all instruments assessed in this study, ICCs were somewhat lower than those previously reported in the literature, with the exception of the spinal mobility measures. This is not unexpected as all patients included in this study had high disease activity, which resulted in less variability in scores between patients for the investigated instruments (e.g., BASDAI and total back pain had a possible range of 4–10 instead of 0–10). It has been shown that reduced variability in scores decreases ICCs in case of unchanged number of observations and measurement error[21,46]. This might explain why for almost all measurement instruments the reliability found in this study was somewhat lower than those reported previously. Other characteristics, such as the proportion of female patients, age and symptom duration of the patients included in this study were comparable to the populations included in previous studies investigating reliability[23,25,27,29,30,32-35]. The decreased variability in scores has an opposite effect on the SDCs, as the mean difference between two assessments (and its SD) is expected to be smaller when the scoring range is reduced, this applies to scores between patients as well as between two measurements within the same patient. An SDC represents a minimum that can be observed reliably based on measurement error. This can be compared with a minimal clinically important improvement (MCII, defined in relation to an external standard for an individual patient) and minimal clinically important difference (MCID, defined by an external standard between (groups of) patients). We compared the observed SDCs with the published SDCs, MCIIs and MCIDs in the literature. The SDCs for ASDAS found in this study were indeed lower than the MCII defined in the literature[29], while SDCs for BASDAI, PtGA and BASFI found in these datasets were similar to the previously reported MCIIs[27,33]. Based on the data analysed in this study, we can conclude ASDAS has the best reliability and smallest SDC of the instruments used to assess disease activity.

For total back pain and pain at night in the past week, SDCs were smaller than the MCID defined in the literature[34], and ICCs were comparable for both instruments. The data for the fatigue and stiffness questions of the BASDAI was inconclusive. In the COAST-X and COAST-V datasets SDCs were similar to the reported MCIDs[34,47-49]. Conversely, measurement

error in the RAPID-axSpA was somewhat larger, complicating detection of the MCID. Comparing the ICCs and SDCs of the various instruments used to assess morning stiffness in the COAST datasets, duration of morning stiffness seems slightly less reliable compared with severity of morning stiffness and the composite score. Finally, the SDC for the ASAS-HI was slightly smaller than previously reported[25], which could be the result of the afore mentioned limited range in disease activity in the current study populations.

Compared with the SF-36, the SDC of the ASAS-HI was higher (12% vs 5%–7% of the total score range) and the ICC slightly lower, indicating the SF-36 might have better reliability. However, the ASAS-HI is a disease-specific instrument, whereas the SF-36 is a general instrument, thus other measurement properties are vital for a final conclusion. Before a definite decision can be made regarding which instrument is best to assess each domain, the other measurement properties will have to be collected too.

This study used data from three recent trials in axSpA, which ensured all instruments currently used in clinical trials were represented. All patients included in these datasets had active disease and were candidate to receive a disease-modifying therapy, which matches the target group of the ASAS-OMERACT core outcome set[4]. As the core outcome set will be used in clinical trials assessing the effect of treatment in axSpA and RCTs in principle require patients with active disease, the data from this study provide valuable information on the reliability of measurement instruments in this patient group. Furthermore, an equal number of patients with r-axSpA and nr-axSpA were included, thereby representing all patients with axSpA disease. Nonetheless, there were limitations to this study, the most important one being the relatively long time-interval used in the current study to ensure the sample sizes would be large enough, which might explain some of the differences found between the literature and the results in this study. Based on the data from this study and information available in the literature, ASDAS, BASDAI, PtGA and CRP are reliable measures to assess disease activity in all patients with axSpA, both total back pain and pain at night in the past week could be considered reliable in assessing pain, questions 5 and 6 of the BASDAI can be used to reliably assess morning stiffness, BASDAI question 1 can reliably evaluate fatigue, BASFI was found reliable to investigate physical functioning, ASAS-HI and SF-36 were found reliable to survey overall functioning & health, and BASMI and its components as well as chest expansion can be used to reliably assess spinal mobility. Further research will have to focus on collecting information on the other psychometric properties before a definite decision can be made regarding the best instrument for each domain.

CONCLUSION

The results from this study showed overall reliability was good and levels of reliability were comparable for patients with r-axSpA and nr-axSpA, indicating ASDAS, BASDAI, PtGA, CRP, NRS total back pain, NRS back pain at night, BASFI, ASAS-HI, SF-36 and BASMI are reliable measures for all patients with axSpA, even though most instruments were developed for r-axSpA. Composite measures showed higher reliability than single-item measures in assessing disease activity and spinal mobility in patients with axSpA and may therefore be preferred over single-item instruments for this aspect of the OMERACT filter.

SUPPLEMENTARY DATA

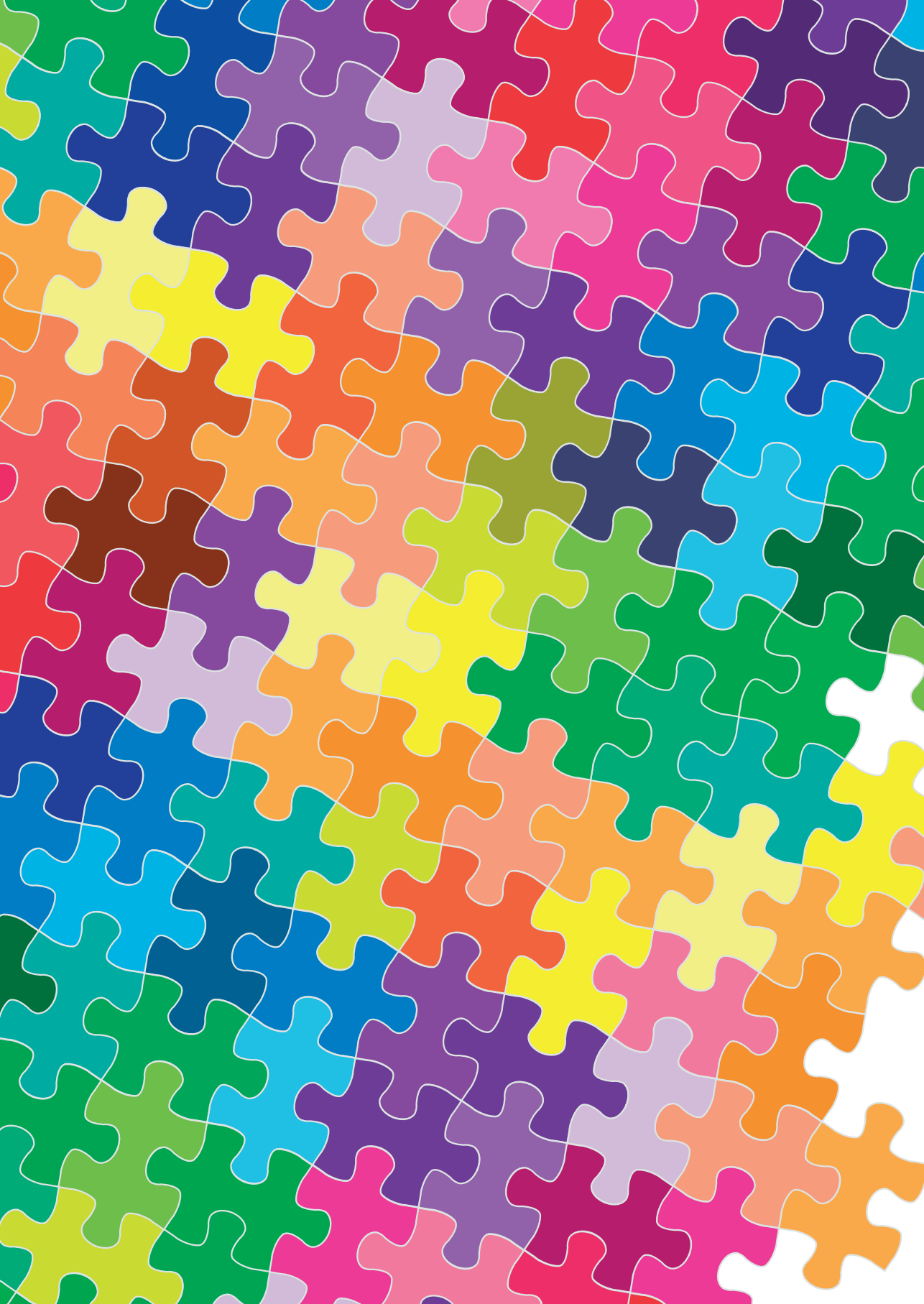
Supplementary data are published online on the website of RMD Open.

REFERENCES

1. Boers M, Kirwan JR, Tugwell P. OMERACT Handbook, 2018.
2. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials* 2017;18(Suppl 3):280.
3. Boel A, Navarro-Compán V, Boonen A, et al. Domains to Be Considered for the Core Outcome Set of Axial Spondyloarthritis: Results From a 3-round Delphi Survey. *J. Rheumatol* 2021
4. Navarro-Compán V, Boel A, Boonen A, et al. The ASAS-OMERACT core domain set for axial spondyloarthritis. *Semin Arthritis Rheum* 2021
5. 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.
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):1000-08.
7. Beaton DE, Maxwell LJ, Shea BJ, et al. Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. *J. Rheumatol* 2019;46(8):1028-35.
8. van der Heijde D, Cheng-Chung Wei J, Dougados M, et al. Ixekizumab, an interleukin-17A antagonist in the treatment of ankylosing spondylitis or radiographic axial spondyloarthritis in patients previously untreated with biological disease-modifying anti-rheumatic drugs (COAST-V): 16 week results of a phase 3 randomised, double-blind, active-controlled and placebo-controlled trial. *Lancet* 2018;392(10163):2441-51.
9. Deodhar A, van der Heijde D, Gensler LS, et al. Ixekizumab for patients with non-radiographic axial spondyloarthritis (COAST-X): a randomised, placebo-controlled trial. *Lancet* 2020;395(10217):53-64.
10. Landewé R, Braun J, Deodhar A, et al. Efficacy of certolizumab pegol on signs and symptoms of axial spondyloarthritis including ankylosing spondylitis: 24-week results of a double-blind randomised placebo-controlled Phase 3 study. *Ann Rheum Dis* 2014;73(1):39-47.
11. Rudwaleit M, van der Heijde D, Landewé R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68(6):777-83.
12. Garrett S, Jenkinson T, Kennedy LG, et al. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. *J.Rheumatol* 1994;21(12):2286-91.
13. Lukas C, Landewé R, Sieper J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68(1):18-24.
14. Landewé R, van Tubergen A. Clinical Tools to Assess and Monitor Spondyloarthritis. *Curr Rheumatol Rep* 2015;17(7):47.
15. 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(12):2281-5.
16. Kiltz U, van der Heijde D, Boonen A, et al. Development of a health index in patients with ankylosing spondylitis (ASAS HI): final result of a global initiative based on the ICF guided by ASAS. *Ann Rheum Dis* 2015;74(5):830-5.
17. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.
18. Ogdie A, Duarte-García A, Hwang M, et al. Measuring Outcomes in Axial Spondyloarthritis. *Arthritis Care & Research* 2020;72(S10):47-71.
19. Jones SD, Porter J, Garrett SL, et al. A new scoring system for the Bath Ankylosing Spondylitis Metrology Index (BASMI). *J. Rheumatol* 1995;22(8):1609.
20. Qin S, Nelson L, McLeod L, et al. Assessing test–retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and documenting the analytical formula. *Qual Life Res* 2019;28(4):1029-33.
21. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15(2):155-63.
22. Navarro-Compán V, van der Heijde D, Ahmad HA, et al. Measurement error in the assessment of radiographic progression in rheumatoid arthritis (RA) clinical trials: the smallest detectable change (SDC) revisited. *Ann Rheum Dis* 2014;73(6):1067-70.
23. Auleley G-R, Benbouazza K, Spoorenberg A, et al. Evaluation of the smallest detectable difference in outcome or process variables in ankylosing spondylitis. 2002;47(6):582-87.
24. Bautista-Molano W, Landewé RBM, Kiltz U, et al. Validation and reliability of translation of the ASAS Health Index in a Colombian Spanish-speaking population with spondyloarthritis. *Clinical Rheum* 2018;37(11):3063-68.

25. Kiltz U, van der Heijde D, Boonen A, et al. Measurement properties of the ASAS Health Index: results of a global study in patients with axial and peripheral spondyloarthritis. 2018;77(9):1311-17.
26. Kiltz U, Winter J, Schirmer M, et al. [Validation of the German translation of the ASAS health index : A questionnaire to assess functioning and health in patients with spondyloarthritis]. *Zeitschrift fur Rheumatologie* 2019;78(4):352-58.
27. Kviatkovsky MJ, Ramiro S, Landewé R, et al. The Minimum Clinically Important Improvement and Patient-acceptable Symptom State in the BASDAI and BASFI for Patients with Ankylosing Spondylitis. *J Rheumatol* 2016;43(9):1680-6.
28. Kwan YH, Aw FF, Fong W, et al. Validity and reliability of the Assessment of Spondyloarthritis International Society Health Index in English-speaking patients with axial spondyloarthritis in Singapore. 2019;22(9):1644-51.
29. Machado P, Landewé R, Lie E, et al. Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. 2011;70(1):47-53.
30. Madsen OR, Rytter A, Hansen LB, et al. Reproducibility of the Bath Ankylosing Spondylitis Indices of disease activity (BASDAI), functional status (BASFI) and overall well-being (BAS-G) in anti-tumour necrosis factor-treated spondyloarthropathy patients. *Clinical Rheum* 2010;29(8):849-54.
31. Ozer HT, Sarpel T, Gulek B, et al. Evaluation of the Turkish version of the Bath Ankylosing Spondylitis Patient Global Score (BAS-G). *Clinical Rheum* 2006;25(2):136-9.
32. Salaffi F, Gasparini S, Ciapetti A, et al. Usability of an innovative and interactive electronic system for collection of patient-reported data in axial spondyloarthritis: comparison with the traditional paper-administered format. *Rheumatology* 2013;52(11):2062-70.
33. Tubach F, Ravaud P, Martin-Mola E, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. *Arthritis Care Res* 2012;64(11):1699-707.
34. van Tubergen A, Black PM, Coteur G. Are patient-reported outcome instruments for ankylosing spondylitis fit for purpose for the axial spondyloarthritis patient? A qualitative and psychometric analysis. *Rheumatology* 2015;54(10):1842-51.
35. van Tubergen A, Coenen J, Landewé R, et al. Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis. *Arthritis Rheum* 2002;47(1):8-16.
36. van Tubergen A, Debats I, Ryser L, et al. Use of a numerical rating scale as an answer modality in ankylosing spondylitis–specific questionnaires. 2002;47(3):242-48.
37. Garrido-Castro JL, Curbelo R, Mazzucchelli R, et al. High Reproducibility of an Automated Measurement of Mobility for Patients with Axial Spondyloarthritis. *J Rheumatol* 2018;45(10):1383-88.
38. Haywood KL, Garratt AM, Jordan K, et al. Spinal mobility in ankylosing spondylitis: reliability, validity and responsiveness. *Rheumatology* 2004;43(6):750-7.
39. Maksymowych WP, Mallon C, Richardson R, et al. Development and validation of the Edmonton ankylosing spondylitis metrology index. 2006;55(4):575-82.
40. Marques ML, Ramiro S, Goupille P, et al. Measuring spinal mobility in early axial spondyloarthritis: does it matter? *Rheumatology* 2019;58(9):1597-606.
41. Martindale JH, Sutton CJ, Goodacre L. An exploration of the inter- and intra-rater reliability of the Bath Ankylosing Spondylitis Metrology Index. *Clinical Rheum* 2012;31(11):1627-31.
42. Ramiro S, van Tubergen A, Stolwijk C, et al. Reference intervals of spinal mobility measures in normal individuals: the MOBILITY study. *Ann Rheum Dis* 2015;74(6):1218-24.
43. Viitanen JV, Heikkilä S, Kokko ML, et al. Clinical assessment of spinal mobility measurements in ankylosing spondylitis: a compact set for follow-up and trials? *Clinical Rheum* 2000;19(2):131-7.
44. Frost MH, Reeve BB, Liepa AM, et al. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 2007;10 Suppl 2:S94-s105.
45. Landewé RBM, van der Heijde D. Use of multidimensional composite scores in rheumatology: parsimony versus subtlety. *Ann Rheum Dis* 2020
46. Lee KM, Lee J, Chung CY, et al. Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopaedic research. *Clin Orthop Surg* 2012;4(2):149-55.
47. Chen M-H, Lee M-H, Liao H-T, et al. Health-related quality of life outcomes in patients with rheumatoid arthritis and ankylosing spondylitis after tapering biologic treatment. *Clinical Rheum* 2018;37(2):429-38.
48. Davis JC, Jr., Revicki D, van der Heijde DM, et al. Health-related quality of life outcomes in patients with active ankylosing spondylitis treated with

adalimumab: results from a randomized controlled study. *Arthritis Rheum* 2007;57(6):1050-7.
49. van der Heijde D, Deodhar A, Braun J, et al. The Effect of Golimumab Therapy on Disease Activity and Health-related Quality of Life in Patients with Ankylosing Spondylitis: 2-year Results of the GO-RAISE Trial. 2014:jrheum.131003.



PART III

PATIENT REPORTED OUTCOMES IN EARLY AXIAL SPONDYLOARTHRITIS

CHAPTER 9

PATIENTS WITH EARLY AXIAL SPONDYLOARTHRITIS HAVE BETTER WORK AND ACTIVITY OUTCOMES AND HEALTH-RELATED QUALITY OF LIFE COMPARED TO CHRONIC BACK PAIN PATIENTS WITHOUT SPONDYLOARTHRITIS AT TWO YEARS: RESULTS FROM THE SPONDYLOARTHRITIS CAUGHT EARLY COHORT.

Anne Boel, Miranda van Lunteren, Karen Fagerli, Roberta Ramonda,
Sofia Exarchou, Marleen van de Sande, Désirée van der Heijde, Floris van Gaalen

ABSTRACT

Background

As with other causes of chronic back pain (CBP), axial spondyloarthritis (axSpA) negatively affects health-related quality of life (HRQoL) and work outcomes. The aim of this study was to compare HRQoL and work and activity outcomes between patients with and without an axSpA diagnosis over two years in routine care.

Methods

Two-year follow-up data from the Spondyloarthritis Caught Early cohort was used. CBP patients were allocated to the axSpA or no-axSpA group, based on the rheumatologist's diagnosis at two-year follow-up. HRQoL was assessed by the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36); work and activity outcomes by the Work Productivity and Activity Impairment (WPAI). Outcome measures at two-year follow-up were compared between groups using linear regression models, corrected for baseline values, NSAID-use over time, gender and age. Wilcoxon signed-rank tests were used to investigate change within groups.

Results

In total, results from 337 CBP patients (209 axSpA and 128 no-axSpA) were analysed. Physical Component Summary scores were significantly higher (better) in the axSpA group (40(±SD12) vs 35(15), $p<0.001$), and levels of all WPAI outcomes were significantly lower (better) in the axSpA group at two years (presenteeism: 20(25) vs 30(28), $p=0.029$, absenteeism: 3(1) vs 8(2), $p=0.041$, WPL: 21(26) vs 34(32), $p=0.012$, activity impairment: 23(25) vs 31(27), $p=0.030$), after correction for gender, age, NSAID-use over time and baseline values. There was no difference between groups regarding the mental component summary score ($p=0.272$).

Conclusion

After two years axSpA patients in routine care had significantly better outcomes compared to patients without axSpA.

INTRODUCTION

Axial spondyloarthritis (axSpA) is an inflammatory arthritis of the spine. The disease can be subdivided into two subtypes: non-radiographic axSpA (nr-axSpA) and radiographic axSpA (r-axSpA) (also known as ankylosing spondylitis(AS)). In the latter radiographic abnormalities consistent with sacroiliitis on plain radiographs are present, in the former the abnormalities are not (yet) present.

Chronic back pain (CBP) is the hallmark of axSpA, but also a symptom of many other diseases. This contributes to the substantial diagnostic delay in axSpA. Given that effective treatment is available for axSpA including nr-axSpA, experts have designated improving identification of early SpA and early referral to rheumatologists as important unmet needs in the clinical care of spondyloarthritis[1].

As with other causes of CBP, axSpA negatively affects health-related quality of life (HRQoL) and work outcomes, with an increasing impact with increasing severity of CBP[2-4]. Work productivity loss (WPL)-mainly caused by a decrease in work productivity while being at work (so called presenteeism)- contributes to the substantial societal costs of axSpA[5].

Early axSpA cohorts have reported improvement in HRQL and WPL following diagnosis, suggesting a beneficial effect of early diagnosis and subsequent treatment[5-7]. However, lack of a comparator group makes these results difficult to interpret and it is particularly difficult to attribute the observed improvement to axSpA treatment. Ideally, a study should be performed where immediately after diagnosis patients are randomized to receive either routine treatment or no treatment. Apart from ethical issues of withholding recommended treatment[8], such a study would be challenging to execute given that nonsteroidal anti-rheumatic drugs (NSAIDs)-which is the first-line pharmacological treatment of axSpA- are available as *over the counter* medications in most countries.

The Spondyloarthritis Caught Early (SPACE) cohort commenced in 2009 with the aim of identifying early axSpA in patients presenting with back pain of short duration. For this purpose, the SPACE cohort started off with a single inclusion criterion: chronic back pain present for at least three months, not exceeding two years with an onset before the age of 45. All included patients were followed for at least two years, at which point the baseline diagnosis had to be confirmed or rejected. By design, this resulted in a group of patients with an axSpA diagnosis with high certainty at 2-year follow-up, and a group of patients with a diagnosis of 'no axSpA' at 2-year follow-up. This set-up provides a unique opportunity to compare patient reported outcomes including HRQoL and work and activity outcomes between CBP patients with and without a diagnosis of axSpA in a daily practice setting during the first two years after diagnosis.

METHODS

The SPACE cohort has been described in detail previously; in brief, patients over 16 years of age referred to the rheumatology outpatient clinic-in the Netherlands, Italy, Norway and Sweden- with CBP (duration of back pain ≥ 3 months and <2 years) starting before the age of 45 were included in the SPACE cohort. Follow-up was performed only in patients with at least two SpA features or one feature with a positive likelihood ratio for axSpA ≥6.4[9]. Using information on all SpA features and imaging, the treating rheumatologists provided a preliminary baseline diagnosis (axSpA or no axSpA) and a definite 2-year diagnosis, as well as the level of confidence (LoC) regarding this diagnosis on a 0-10 scale (0, not confident; 10, very confident).

Patients

Patients were allocated to one of two groups based on the 2-year diagnosis from the rheumatologist and the LoC regarding that diagnosis. The first group consisted of all patients with a diagnosis of axSpA with a LoC of at least 7: the axSpA group; the second group consisted of all patients without a diagnosis of axSpA, as well as those with a diagnosis of axSpA with a LoC of 6 or smaller: the no axSpA group. For this study, patients were included if they completed 2 years of protocolised follow-up, meaning a diagnosis and accompanying LoC and MRI had to be available at two-year follow-up. Additionally, there had to be complete clinical data and data for and at least one questionnaire (i.e. HRQoL or work outcomes) at both timepoints.

Outcomes

Health-related quality of life was assessed using the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) version 1[10]. Age-, sex- and country-weighted scale scores were created for each of the 8 subscales of the SF-36[11-13]. Numeric scores ranged from 0 (worst health) to 100 (best health), after recoding and recalibration. In absence of Italian age- and sex-matched scores, Dutch age-and sex-matched scores were used for the Italian patients (n=46). The physical (PCS) and mental component summary (MCS) scores were calculated from the adjusted scores on each of the respective subscales and transformed to enable comparison to the general population mean of 50. Higher scores indicated better HRQoL[14]. A few cases (n=17) had a negative PCS, these were set to 0[6]. Additionally, the proportion of patients with an improvement or worsening of the PCS and MCS above the minimal clinically important difference (MCID) was assessed. We applied the MCID commonly used in clinical trials with biological disease modifying anti-rheumatic drugs (bDMARD) in axSpA of 5 points for the PCS and MCS[15-17].

Work productivity and activity impairment were assessed using the Work Productivity and Activity Impairment (WPAI) questionnaire[18,19]. Consisting of 6 items, the WPAI assesses the

impact of chronic back pain complaints on presenteeism, absenteeism and work productivity loss (WPL). Presenteeism reflects the reduction in performance while at work due to disease; presenteeism was calculated from the influence of disease on work productivity as reported by the patient (on a 0-10 scale). Absenteeism indicated the hours missed from work due to disease; absenteeism was calculated by taking the reported number of hours missed at work due to disease and dividing this number by ten. WPL was derived from presenteeism and absenteeism (WPL=absenteeism+((1-absenteism)×(presenteeism))) and provided an indication of the total loss of work productivity due to disease. Finally, activity impairment was defined as impairment due to disease in all non-work-related activities; activity impairment was calculated from the reported influence of disease on regular daily activities. All WPAI outcomes were presented as percentages between 0-100; higher scores implied greater impairment. Additionally, the proportion of patients with any (>0%) absenteeism, presenteeism, WPL and activity impairment were assessed.

Assessment of presenteeism, absenteeism and WPL was restricted to the working population, which was defined as those with paid work at baseline and 2-year follow up. Activity impairment was assessed for the entire study population. In addition, the proportion of employed patients was assessed for baseline and two-year follow-up; and expressed as the percentage of the employable population (defined as everyone of working age (>16), who was not a fulltime student). Patients were considered employed if they reported to have worked for at least one hour in the previous week or had a permanent job during the previous week[20].

Analyses

The database was locked on January 1st 2020, at that time a total of 807 patients were included in the SPACE cohort, of whom 468 completed at least 2 years of follow-up. 396 of these patients had a MRI and a diagnosis with corresponding LoC available at 2-year follow-up. Of these 396 patients, 337 patients with complete clinical data and data at both timepoints for at least one questionnaire (i.e. SF-36 or WPAI) were available for analysis (Supplementary figure S.1).Categorical variables were reported as frequencies (proportions) and continuous variables as means and standard deviation (SD). Wilcoxon signed-rank tests were used to compare data within groups over time. Linear regression models were built for the SF-36 (PCS and MCS) and WPAI outcomes (presenteeism, absenteeism, WPL and activity impairment) with diagnosis at 2 years as the independent variable, the respective outcomes as dependent variable and the baseline value of the respective outcome as covariate to compare 2-year outcomes between groups.

Age at baseline, gender and NSAID-use over time were added to the models as potential confounders. Age was considered as axSpA might have a different impact on the QoL and work-related outcomes in those just starting their working life than in those who

have been working for a while[21,22]. Gender was considered as it is known that women experience a larger impact of axSpA on their QoL and work productivity[6,21]. NSAID-use over time was considered since efficacious treatment is known to improve QoL and work-related outcomes[23].

As treatment with bDMARDs was only available to those patients who got a diagnosis of axSpA, it was decided it would be worthwhile to perform sensitivity analyses. In these sensitivity analyses the patient population was restricted to patients not using bDMARDs at any point during the 2-year period of follow-up to ensure potential differences between groups could not be explained by the availability of treatment.

Data was analysed using STATA SE V.16 (Statacorp). P-values < 0.05 were considered statistically significant.

RESULTS

In total 209 patients with a diagnosis of axSpA and 128 patients without a diagnosis (no axSpA) were analysed in this study. Patients with an axSpA diagnosis were more often male and HLA-B27 positive, had a slightly lower age at baseline, and a higher number of SpA features than the no axSpA patients, whereas no axSpA patients more often had a positive family history (Table 1). Furthermore, sacroiliitis on MRI and radiographs according to the local radiologist was frequent in the axSpA group, but uncommon in the no axSpA group. Contrary to use of NSAIDs -which was high in both groups-, use of biological DMARDs (bDMARD) was very limited at baseline: 1 patient in the no axSpA group used a bDMARD for concomitant inflammatory bowel disease (IBD); 8 patients in the axSpA group used a bDMARD at baseline, of whom 5 had psoriasis and 1 uveitis. At two-year follow-up 56 patients in the axSpA group were using a bDMARD; in the no axSpA group there were 4 bDMARD-users, 3 as treatment for IBD and 1 for psoriasis.

Mean total back pain was significantly lower at two-year follow-up compared to baseline in both groups (axSpA 3.1 (±SD 2.5) vs. 4.5 (2.5), p<0.01; no axSpA (4.2 (3.0) vs. 5.3 (2.5),p<0.01).

Table 1 Baseline characteristics of CBP patients included in the SPACE cohort stratified by two-year clinical diagnosis.

Characteristic	Diagnosis axSpA (n=209)	CBP (n=128)
Male, n(%)	117 (56)	38 (30)
Age (years), mean (SD)	29 (7)	31 (8)
Symptom duration (months), mean (SD)	13 (7)	13 (7)
HLA-B27 positive, n(%)	157 (75)	50 (39)
IBP, n(%)	174 (83)	92 (72)
Good response to NSAIDs[†], n(%)	100 (48)	48 (38)
Positive family history of SpA, n(%)	94 (45)	80 (63)
Past history or current symptoms*		
Peripheral arthritis, n(%)	55 (26)	14 (11)
Enthesitis, n(%)	68 (33)	21 (16)
Dactylitis, n(%)	25 (12)	2 (2)
Psoriasis, n(%)	35 (17)	12 (9)
IBD, n(%)	11 (5)	9 (7)
Acute anterior uveitis, n(%)	25 (12)	8 (6)
Elevated CRP/ESR, n(%)	96 (46)	24 (19)
Sacroiliitis radiographs[‡], n(%)	54 (26)	3 (2)
Sacroiliitis MRI[‡], n(%)	149 (71)	9 (7)
Number of SpA features[§], mean (SD)	3 (2)	2 (1)
Use of bDMARDs, n(%)	8 (4)	1 (1)
Use of NSAIDs, n(%)	158 (75)	88 (69)

*Past or present condition, either diagnosed or confirmed by a physician

[†] Back pain no longer present or much better 24–48 hours after a full dose of NSAID

[‡] Based on reading of local radiologists

[§] Excluding HLA-B27 status and sacroiliitis on imaging

axSpA, axial Spondyloarthritis; **bDMARD**, biological Disease Modifying Anti-Rheumatic Drug; **CBP**, Chronic Back Pain; **CRP**, C-reactive protein; **ESR**, Erythrocyte Sedimentation Rate; **HLA-B27**, Human Leucocyte Antigen B27; **IBD**, Inflammatory Bowel Disease; **IBP**, Inflammatory Back Pain; **MRI**, Magnetic Resonance Imaging; **NSAIDs**, Non-Steroidal Anti Inflammatory Drugs; **SpA**, Spondyloarthritis.

SF-36

At baseline the mean PCS score was comparable between the axSpA and no axSpA groups (28(14) vs 27(13)). In both groups the mean PCS score significantly improved over two years. However, the PCS was higher in the group with an axSpA diagnosis compared to the no axSpA group at two-year follow-up (40(12) vs 35(15)). In the linear regression analysis, a diagnosis of axSpA was an independent predictor of better PCS scores at two-year follow-up after correction for baseline PCS scores, NSAID-use over time, gender and age ($p<0.001$)(Table 2). Despite the improvements over time, PCS scores were still well below the general population mean of 50 in both groups at two-year follow-up.

The MCS scores were also comparable between the axSpA and no axSpA groups at baseline (47(14) vs 47(12)). Mean MCS scores did not significantly change over time within the groups, nor were MCS scores significantly different between groups at follow-up ($p=0.272$). Moreover, MCS scores in both groups were close to the general population mean of 50, especially at two-year follow-up.

Table 2 Health-related quality of life measured by the SF-36 in CBP patients stratified by two year clinical diagnosis

	axSpA (n=205)		**No axSpA** (n=125)		**p-value** between groups at 2 years
	Baseline	**2 years**	**Baseline**	**2 years**	
PCS, mean (SD)	28 (14)	40 (12)†	27 (13)	35 (15)†	p<0.001*
% Improvement >MCID		67		58	
% Worsening >MCID		11		14	
MCS, mean (SD)	47 (14)	48 (12)‡	47 (12)	49 (11)‡	p=0.272
% Improvement >MCID		34		36	
% Worsening >MCID		30		22	

* Significant difference between groups at two years; after correction for baseline values, gender, age and NSAID use over time (p<0.05)
†Signed-rank test: significant improvement within group over time (p<0.05)
‡Signed-rank test: not significant
axSpA, Axial Spondyloarthritis; **CBP**, Chronic Back Pain; **MCID**, Minimal Clinically Important Difference; **MCS**, Mental Component Summary; **PCS**, Physical Component Summary; **SF-36**, Short-Form Health Survey

Although the proportion of patients with an improvement above the MCID of 5 points was higher in the axSpA group, over half of the patients in both groups (axSpA 67%, no axSpA 58%) had such an improvement of their PCS score. The proportion of patients whose PCS score worsened more than 5 points was low in both groups (axSpA 11%, no axSpA 14%). For the MCS scores the proportion of patients with an improvement and worsening of more than 5 points was less distinct. Approximately one-third had an improvement above the MCID in both groups (axSpA 34%, no axSpA 36%), in the axSpA group the proportion of patients with a worsening of more than 5 points was 30%, this was slightly lower in the no axSpA group (20%).

WPAI

The working population (paid work both at baseline and 2-year follow-up) consisted of 141 patients (69%) in the axSpA group and 87 patients (71%) in the no axSpA group, for these patients presenteeism, absenteeism and WPL were assessed. At baseline, presenteeism was lower in the axSpA group (32(28)% vs 41(27)%), absenteeism was lower in the axSpA group too (7(2)% vs 9(2)%), thus WPL was also lower in the axSpA group (34(29)% vs 44(28)%). In both groups mean percentage of WPL was significantly lower at two-year follow-up, the same applied to presenteeism, yet mean percentages of presenteeism (20(25) vs 30(28)) and WPL (21(26) vs 34(32)) were better at two-year follow-up in the axSpA group (Table 3). For absenteeism, only the axSpA group showed a significant reduction over two years and mean percentages of absenteeism were significantly lower (3(1) vs 8(2)) at two-year follow-up in the axSpA group (Table 3).

Activity impairment could be assessed for all patients, and the mean percentage of activity impairment was lower in the axSpA group (38(27)% vs 48(25)%). At two-year follow-up activity impairment improved significantly in both groups, nevertheless, mean percentages of activity impairment (23(25) vs 31(27)) were better at two-year follow-up in the axSpA group (Table 3).

In linear regression analysis, a diagnosis of axSpA was an independent predictor of better presenteeism (p=0.029), absenteeism (p=0.041), WPL (p=0.012) and activity impairment(p=0.030) at two-year follow-up after correction for baseline scores, NSAID-use over time, gender and age (table 3).

Although the proportion of patients with any WPL and any activity impairment decreased over time, over half of the patients in both groups still experience productivity loss at work (55% in axSpA and 68% in no axSpA) and impairment in non-work related activities (66% in axSpA and 72% in no axSpA).

We found an increase in the employable population over time in both groups (from 89% to 95% in the axSpA group and from 91% to 98% in the no axSpA group). This could be explained by the fact that there were quite a few students (15 in the axSpA group and 10 in the no axSpA group) in the SPACE cohort who completed their studies and found a job in the first two years of study follow-up. Even though the employable population increases, the proportion of patients with paid work remains similar in the axSpA group (from 86 to 89%), which indicated an increase in the number of patients with paid work over time in this group. In the no axSpA group the proportion of patients with paid work slightly decreases (from 84% to 82%) as the employable population increased.

Table 3 Results from the work productivity and activity impairment questionnaire in CBP patients stratified by two-year clinical diagnosis

	axSpA		**no axSpA**		**p-value** between
	Baseline	**2 years**	**Baseline**	**2 years**	groups at 2 years
Working population	*n=141*		*n=87*		
Presenteeism, mean % (SD)	32 (28)	20 (25)†	41 (27)	30 (28)†	p=0.029*
% Presenteeism present	74	56	91	69	
Absenteeism, mean % (SD)	7 (2)	3 (10)†	9 (20)	8 (20)‡	p=0.041*
% Absenteeism present	21	8	25	16	
Work productivity loss, mean % (SD)	34 (29)	21 (26)†	44 (28)	34 (32)†	p=0.012*
% Work productivity loss present	73	55	91	68	
Total population	*n=204*		*n=123*		
Activity impairment, mean % (SD)	38 (27)	23 (25)†	48 (25)	31 (27)†	p=0.030*
% Activity impairment	86	66	96	72	

* Significant difference between groups at two years; after correction for baseline values, gender, age and NSAID use over time (p<0.05)
†Signed-rank test: significant improvement within group over time (p<0.05)
‡Signed-rank test: not significant

Sensitivity analyses

To ensure the differences found between groups were not only due to a difference in the availability of biological treatment, we performed sensitivity analyses in which we included only patients who did not use a bDMARD during the 2 years of follow-up.

For the PCS there was still a highly significant difference (42 (12) vs 36 (14),p<0.001) between those with and without a diagnosis of axSpA at two-year follow-up (supplementary table S.1), when the analysis was restricted to patients not using biologicals. The MCS remained comparable between groups (50 (11) vs 49(11), p=0.655), the subgroup of patients with an axSpA diagnosis not using bDMARDs actually reached an MCS equal to the population mean.

For the WPAI variables, the differences between those with and without a diagnosis became even more apparent for presenteeism (16 (22) vs 30 (28), p=0.002), WPL (17 (23) vs 34(32), p=0.008) and activity impairment (19 (22) vs 31 (27), p<0.001) when analyses were restricted to patients not on bDMARD therapy (supplementary table S.2). However, for absenteeism, the difference between groups was no longer present (2 (12) vs 7 (20), p=0.174).

DISCUSSION

The SPACE cohort-an inception cohort of back pain patients suspected of axSpA- provided a unique opportunity to compare HRQoL and work and activity outcomes between CBP patients with and without a diagnosis of axSpA in a daily practice setting during the first two years after diagnosis. The performed analyses showed an improvement over time in physical HRQoL, WPL, presenteeism, absenteeism and activity impairment over two years of protocolised follow-up in all patients with chronic back pain complaints, regardless of diagnosis. Improvement in both groups suggests that some improvements are due to regression to the mean, with complaints being most severe at the first visit to the rheumatology outpatient clinic regardless of diagnosis. Nonetheless, we showed that a diagnosis of axSpA was an independent predictor of better PCS and WPAI scores at two-year follow-up, emphasizing the value of the comparator group available in this study.

One of the differences between those who get a diagnosis of axSpA versus those who do not get diagnosed is the availability of treatment. At two years NSAID-use was higher in axSpA patients (11% on full-dose) compared to the no axSpA group (5% on full-dose). Moreover, treatment with biologicals is solely available to those diagnosed with axSpA. At two years the number of axSpA patients treated with biological was about a quarter. Therefore, sensitivity analyses were performed to investigate the role of biological therapy in the improvement of the outcomes. These showed that the differences in outcomes between those with and without a diagnosis of axSpA remained when analyses were restricted to patients not on biological therapy, indicating that treatment with biologicals did not explain the differences between the groups.

Another possible explanation for the difference between patients with and without a diagnosis could be their illness perceptions and subsequent influence on coping. Compared to non-specific back pain axSpA has a much clearer pathophysiological framework, and there is a better understanding of what causes the complaints of these patients. For example, in the current ASAS-EULAR management recommendation for axSpA, the primary goal of treating patients with axSpA is to maximise long-term HRQoL through control of-among others- inflammation[8]. Through patient education such a relatively clear conceptual framework could increase patients' understanding of disease and influence illness perceptions, which will be investigated in the future. Finally, there are numerous active patient societies for axSpA patients to turn to for information and support. These factors combined may enhance acceptance of a chronic disease. In future research we will investigate the role of illness perceptions and coping strategies and whether these might help explain the differences between groups found in this study.

A potential limitation of the study is the use of patient reported outcomes for all primary outcomes. An alternative in assessing work productivity loss would have been to use absenteeism numbers reported by employers to get a more objective measure of the hours lost due to disease instead of relying on patient reported information. However, the major cause of work productivity loss was presenteeism and not absenteeism, and this would have been missed by relying solely on employer reported absenteeism.
By design, the patients who did not get diagnosed with axSpA were excluded from the SPACE cohort after two-year follow-up, which meant the maximum follow-up time for which a control group was available was two years. In the cohort, axSpA patients are followed beyond those two years and this will allow us to continue monitoring if the observed improvements in health-related quality of life and work productivity are maintained by using the national population mean of the SF-36 and the European Working Conditions surveys by Eurofound[24] as comparators.

CONCLUSION

In chronic back pain patients suspected of axSpA, we found significant improvements in physical functioning and work-related outcomes over two years of protocolised follow-up. Nonetheless, axSpA patients had significantly better outcomes in physical functioning and work-related outcomes compared to patients with chronic back pain without axSpA.

SUPPLEMENTARY DATA

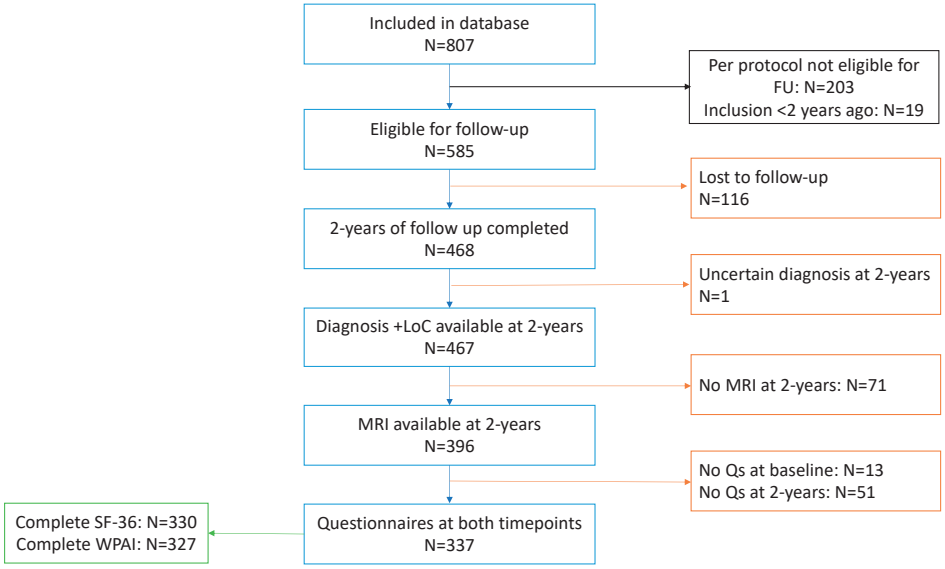


Figure S.1 Flowchart of inclusion

Table S.1 Health-related quality of life measured by the SF-36 in CBP patients stratified by two year clinical diagnosis, restricted to those patients not using bDMARDs.

	axSpA (n=147)		**No axSpA** (n=121)		**p-value** between groups at 2 years
	Baseline	**2 years**	**Baseline**	**2 years**	
PCS, mean (SD)	29 (15)	42 (12)†	27 (13)	36 (14)†	p<0.001*
MCS, mean (SD)	48 (14)	50 (11)‡	47 (12)	49 (11)‡	p=0.655

* Significant difference between groups at two years; after correction for baseline values, gender, age and NSAID use over time (p<0.05); †Signed-rank test: significant improvement within group over time (p<0.05); ‡Signed-rank test: not significant

axSpA, Axial Spondyloarthritis; **CBP**, Chronic Back Pain; **MCID**, Minimal Clinically Important Difference; **MCS**, Mental Component Summary; **PCS**, Physical Component Summary; **SF-36**, Short-Form Health Survey

Table S.2 Results from the work productivity and activity impairment questionnaire in CBP patients stratified by two-year clinical diagnosis, restricted to those patients not using bDMARDs.

	axSpA		**no axSpA**		**p-value** between groups at 2 years
	Baseline	**2 years**	**Baseline**	**2 years**	
Working population	*n=92*		*n=83*		
Presenteeism, mean % (SD)	27 (25)	16 (22)†	41 (27)	30 (28)†	p=0.002*
Absenteeism, mean % (SD)	3 (9)	2 (12)‡	9 (20)	7 (20)‡	p=0.174
Work productivity loss, mean % (SD)	27 (25)	17 (23)†	44 (28)	34 (32)†	p=0.008*
Total population	*n=148*		*n=117*		
Activity impairment, mean % (SD)	37 (28)	19 (22)†	48 (25)	31 (27)†	p<0.001*

REFERENCES

1. Winthrop KL, Weinblatt ME, Crow MK, et al. Unmet need in rheumatology: reports from the Targeted Therapies meeting 2018. *Ann Rheum Dis* 2019;78(7):872-78.
2. Husky MM, Ferdous Farin F, Compagnone P, et al. Chronic back pain and its association with quality of life in a large French population survey. *Health Qual Life Outcomes* 2018;16(1):195.
3. Sadosky AB, Taylor-Stokes G, Lobosco S, et al. Relationship between self-reported low-back pain severity and other patient-reported outcomes: results from an observational study. *J Spinal Dis Tech* 2013;26(1):8-14.
4. Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014;73(6):968-74.
5. van Lunteren M, Ez-Zaitouni Z, Fongen C, et al. Disease activity decrease is associated with improvement in work productivity over 1 year in early axial spondyloarthritis (SPondyloArthritis Caught Early cohort). *Rheumatology* 2017;56(12):2222-28.
6. van Lunteren M, Ez-Zaitouni Z, de Koning A, et al. In Early Axial Spondyloarthritis, Increasing Disease Activity Is Associated with Worsening of Health-related Quality of Life over Time. *J Rheumatol* 2018;45(6):779-84.
7. López-Medina C, Dougados M, Collantes-Estévez E, et al. Adherence to recommendations for the use of anti-tumour necrosis factor and its impact over 5 years of follow-up in axial spondyloarthritis. *Rheumatology* 2018;57(5):880-90.
8. van der Heijde D, Ramiro S, Landewé R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis* 2017;76(6):978-91.
9. Rudwaleit M, van der Heijde D, Khan MA, et al. How to diagnose axial spondyloarthritis early. *Ann Rheum Dis* 2004;63(5):535-43.
10. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical care* 1992;30(6):473-83.
11. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998;51(11):1055-68.
12. Loge JH, Kaasa S. Short form 36 (SF-36) health survey: normative data from the general Norwegian population. *Scand. J. Soc. Med* 1998;26(4):250-8.
13. Sullivan M, Karlsson JJJoce. The Swedish SF-36 Health Survey III. Evaluation of criterion-based validity: results from normative population. 1998;51(11):1105-13.
14. Ware JE, Kosinski M. Interpreting SF-36 summary health measures: a response. *Qual Life Res* 2001;10(5):405-13; discussion 15-20.
15. Davis JC, Jr., Revicki D, van der Heijde DM, et al. Health-related quality of life outcomes in patients with active ankylosing spondylitis treated with adalimumab: results from a randomized controlled study. *Arthritis Rheum* 2007;57(6):1050-7.
16. Chen MH, Lee MH, Liao HT, et al. Health-related quality of life outcomes in patients with rheumatoid arthritis and ankylosing spondylitis after tapering biologic treatment. *Clinical Rheum* 2018;37(2):429-38.
17. van der Heijde D, Deodhar A, Braun J, et al. The effect of golimumab therapy on disease activity and health-related quality of life in patients with ankylosing spondylitis: 2-year results of the GO-RAISE trial. *J Rheumatol* 2014;41(6):1095-103.
18. Reilly MC, Gooch KL, Wong RL, et al. Validity, reliability and responsiveness of the Work Productivity and Activity Impairment Questionnaire in ankylosing spondylitis. *Rheumatology* 2010;49(4):812-19.
19. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *PharmacoEconomics* 1993;4(5):353-65.
20. OECD. OECD Employment Outlook 2019. 2019.
21. Boonen A, Chorus A, Miedema H, et al. Withdrawal from labour force due to work disability in patients with ankylosing spondylitis. *Ann Rheum Dis* 2001;60(11):1033-9.
22. van Lunteren M, Landewé R, Fongen C, et al. Do Illness Perceptions and Coping Strategies Change Over Time in Patients Recently Diagnosed With Axial Spondyloarthritis? *J Rheumatol* 2020
23. Kroon FP, van der Burg LR, Ramiro S, et al. Nonsteroidal Antiinflammatory Drugs for Axial Spondyloarthritis: A Cochrane Review. *J Rheumatol* 2016;43(3):607-17.
24. Eurofound. [Available from: https://www.eurofound.europa.eu/surveys/european-working-conditions-surveys-ewcs accessed 09-11 2020.

CHAPTER 10

SUMMARY AND GENERAL DISCUSSION

In the first part of this thesis, we provided an international perspective on the characterisation of patients with axial spondyloarthritis (axSpA). We investigated the similarities and differences between the modified New York (mNY) criteria for ankylosing spondylitis and the ASAS classification criteria for radiographic axSpA, and studied the distribution in age at onset and prevalence of a positive family history of axSpA outside of Europe. In the second part of this thesis, we described the process of the development of the core set for axSpA (i.e. the minimum and mandatory set of outcomes to be assessed in every trial) by updating the domains of the ASAS-OMERACT core set for ankylosing spondylitis. In the third part of this thesis, we increased knowledge on work and activity outcomes and health-related quality of life over time in chronic back pain patients with a diagnosis of axSpA or a suspicion thereof.

In this final chapter I will summarize the main findings of the studies presented here within, elaborate on the impact of these findings, and highlight remaining knowledge gaps and future perspectives.

INTERNATIONAL CHARACTERISATION OF AXIAL SPONDYLOARTHRITIS

In the first section of this thesis, we aimed to provide an international perspective on the characterisation of patients with axSpA. Several aspects of the classification criteria for axial spondyloarthritis were closely inspected for this purpose, starting with the nomenclature used to describe patients with axSpA with radiographic damage to the sacroiliac joints in **chapter 2**, for which we included data from several European cohorts[1-6], an American cohort[7] and a cohort including data from Asian, American and European patients[8,9]. Traditionally, patients with axSpA with definite structural changes on conventional radiographs were classified according to the mNY criteria as ankylosing spondylitis. However, an alternative was provided by the more recent ASAS axSpA criteria, wherein these patients could be classified as radiographic axSpA. Critics doubted whether both would classify the same patients[10], but this had never been assessed. In **chapter 2** we concluded that almost all patients with axSpA with radiographic sacroiliitis who met the mNY criteria also met the ASAS criteria for radiographic axSpA and vice-versa, which supported the interchangeable use of the terms ankylosing spondylitis and radiographic axSpA. Thus far, no consensus has been reached on whether one definition should be preferred over the other. Nonetheless, as we move towards one diagnosis (i.e. axial spondyloarthritis) with two subgroups for classification purposes only (i.e. radiographic and non-radiographic), in my opinion it would be desirable to use radiographic axSpA to describe these patients.

Furthermore, it is acknowledged nowadays that non-radiographic and radiographic axSpA have an equal disease burden and treatment with biological disease-modifying antirheumatic drugs (bDMARDs) is effective in both subgroups. Therefore, using axSpA to describe all patients with a diagnosis and radiographic vs non-radiographic to provide additional information on the expression of disease seems most appropriate.

The main cause for disagreement between patients classified by the axSpA criteria and mNY criteria was found to be the age at onset of back pain, which was introduced with the implementation of the ASAS criteria in 2009 and mainly based on data from Feldtkeller et al[11], which showed that onset after the age of 45 was seen in only 5% of the patients. Other studies performed since then have shown similar distributions regarding the age at onset[1,12-14], yet were also based on mainly European data. Therefore, **chapter 3** aimed to provide a worldwide perspective on the age at first symptom onset, to confirm whether the distribution of age at onset of axial symptoms was similar across the globe. Using data from the ASAS-PerSpA study we were able to confirm a similar distribution in age at first symptom onset in various geographical regions, and confirmed the vast majority of patient with axSpA indeed experienced their first symptoms before the age of 45 years. Compared to Feldtkeller et al.[11] and van der Linden et al.[15], we found a slightly lower percentage of patients with age at onset <45 years in this study, which might be explained by the fact that-contrary to these studies- patients with non-radiographic axSpA were also included in the current study. Another important finding of this study was the fact that HLA-B27 carriership was consistently associated with a younger age of symptom onset across the globe, as was male gender.

The third and final aspect of the classification criteria under review in this thesis was the positive family history of spondyloarthritis. The value of a positive family history in its current form has been questioned previously, as the definition was not tested nor validated prior to inclusion in the ASAS classification criteria. **Chapter 4** described the prevalence of a positive family history of spondyloarthritis in various geographical regions and its relationship with HLA-B27 carriership. We found that axSpA is the most common entity of spondyloarthritis in a positive family history, and the association between a positive family history of axSpA and HLA-B27 carriership was independent of a positive family history of other SpA entities. These findings confirmed that the association between a positive family history and HLA-B27 status is largely driven by a positive family history for axSpA in a worldwide cohort, which was previously shown in cohorts that included mostly European and some Asian patients[16-18]. However, we were not able to show an association between a positive family history of axSpA and HLA-B27 carriership in the Middle East & North Africa. This may be explained by the fact that the prevalence of HLA-B27 positive disease was much lower compared to the other regions, which is in line with other research performed in the Middle East[19,20]. Additionally, the Middle East & North Africa showed a high prevalence

of a positive family history of axSpA in HLA-B27 negative patients compared to the other regions, where there was a very low prevalence of a positive family history of axSpA in HLA-B27 negative patients. Nonetheless, a positive family history of axSpA was the most common amongst patients who reported a positive family history in this region, identical to the other geographical regions.

Implications from these findings

Acknowledging that ankylosing spondylitis and radiographic axSpA are interchangeable -as confirmed in **chapter 2-** increases comparability between studies, since both terms describe the same patients. This also ensures that research performed in ankylosing spondylitis cohorts can be compared to more recently published articles on radiographic axSpA cohorts. This is of tremendous importance for medication trials. If the effectiveness of a given medication has been proven in the past, they are not subjected to further randomised clinical trials assessing its effectiveness, as it would be unethical to withhold effective medication from patients. Including data from r-axSpA in meta-analyses allow for comparisons taking into account all treatment types, including those that have been investigated in ankylosing spondylitis. This means treatment can be initiated without the need for conducting a trial first.

The data described in **chapter 3** showed that the age at symptom onset was similar in all investigated geographical regions and the age at symptom onset was consistently lower in HLA-B27 positive patients compared to their HLA-B27 negative counterparts, and also consistently lower in male compared to female patients. This data was long overdue, as it confirms the age at onset criterion can be applied to patients anywhere in the world, rather than just the European, North American and Asian patients in which it was developed. Further, these data imply axSpA manifests at an earlier age in HLA-B27 positive and male patients. Thus, it appears that age at onset is a helpful tool in identifying those at risk of axSpA in the group of patients who present to the rheumatologists with chronic back pain complaints. Given that only a very small proportion of patients develop symptoms after the age of 45 years, it is very unlikely that a patient above this age will be diagnosed with axSpA, which is important knowledge for clinical practice. Even though symptoms occur at a somewhat earlier age in HLA-B27 positive patients and male patients, this does not imply a diagnosis of axSpA in HLA-B27 negative patients and female patients should not be considered. Being aware that symptom onset may be somewhat later in HLA-B27 negative and female patients might result in earlier consideration of an axSpA diagnosis in these 'less typical' patients, which might subsequently reduce diagnostic delay. As for the classification criteria, the entry criterion of an age at onset <45 years seems valid, as the vast majority of patients with axSpA developed symptoms before this age across the globe. As classification criteria are aimed at creating a homogenous group of patients, the age at onset criterion seems a useful tool in excluding the less typical patients.

The findings presented in **chapter 4** of this thesis combined with previous research on family history (in the ASAS, DESIR and SPACE cohorts[16-18]) suggest it is time to critically re-evaluate this criterion. All available data show that axSpA is the most common entity as part of a positive family history and the association between a positive family history and HLA-B27 status is largely driven by a positive family history for axSpA. Given the consistent findings across studies and across the globe, it should be investigated whether the current expert definition of a positive family history in the classification criteria may be redefined to only include the presence of a positive family history of axSpA. The criterion for a positive family history is not only present in the classification criteria for axSpA, but also in the classification criteria for peripheral SpA. As these are used to classify a different subset of patients, it is very likely the definition of a positive family history will have different implications in this patient population. In the event of a redefinition of the family history criterion for axSpA, it would be expected the performance of the family history criterion should be assessed for peripheral SpA too.

Further discussion and future perspectives

Uniformity in classification and reporting aids global communication of scientific, clinical, and epidemiological findings which enhances understanding of the pathogenesis and treatment of axSpA. The fact that it was not clear if two major elements in the ASAS classification criteria (i.e. age at onset and positive family history) applied to all axSpA patients worldwide, points to a flaw in the scientific process. As patients from different continents may vary in their disease presentation[21], sufficient patients from all over the world should be included from the get-go, to ensure the classification criteria are representative for all patients across the globe. The ASAS-PerSpA cohort[21] provided proof for the feasibility of such a study: through international collaboration and smart use of an electronic data collection system it was possible to collect data in 24 countries across the globe. By decreasing the start-up costs of a study (e.g. by providing an electronic data collection system), the threshold to partake will be lowered for countries with less financial funds.

In the past, the validity of the ASAS classification criteria for axSpA has been questioned, as it has been argued that its complex two-arm selection design and its broad spectrum may lead to differences among the composition of patients in different studies[10]. Others have stated that the increased sensitivity of the two-arm design compared to a classification set that included only the imaging arm (82.9% compared to 66.2%) at the cost of specificity (84.4% compared to 97.3%)is a cause for concern, as classification criteria are aimed at creating homogeneous study populations and should therefore aim for the highest possible specificity[22]. Thus, critics emphasize the importance of revision of the ASAS axSpA classification criteria in order to improve specificity and reduce heterogeneity within the group of axSpA patients classified using these criteria[10,22]. Nonetheless, the imaging arm in its current form is not perfect either, as it is well-known

that there is no perfect agreement between readers, and inflammatory lesions on MRI can also be found in healthy controls[23,24].

In 2019 ASAS and SPARTAN initiated the CLASSIC study (CLassification of Axial Spondyloarthritis Inception Cohort) with the aim to re-evaluate the sensitivity and specificity of the ASAS classification criteria for axSpA and provide training in assessment of imaging and diagnosis of patients with axSpA worldwide. For this purpose, patients with chronic back pain suspected of axSpA are included, and a diagnosis of axSpA or no axSpA is made after careful evaluation of clinical, laboratory and imaging results. By design, the CLASSIC study includes patients with and without a diagnosis of axSpA, allowing for assessment of the sensitivity and specificity not only of the classification set as a whole, but also its individual components. In principle, the classification criteria will remain unchanged if they show ≥75% sensitivity and ≥90% specificity in CLASSIC. Nonetheless, for some of the individual components it may be advisable to assess whether they should remain unchanged, regardless of whether the sensitivity and specificity of the classification set as a whole is reached.

One of the individual components that should be investigated is the value of the definition of a positive family history. Herein, it should be taken into account that a positive family history can be used for different purposes. The first application of the family history is as one of the clinical criteria in the ASAS classification criteria. To determine its value for this purpose, the original definition should be compared with a redefined definition (i.e. a positive family history that solely includes axSpA in a first- or second-degree family member). Additionally, it should be investigated whether its weight is appropriate, and whether it should remain an independent SpA feature in addition to HLA-B27. This should be assessed separately for the ASAS criteria for axSpA and peripheral SpA.

Secondly, family history can be used as a proxy for HLA-B27 positivity. This is particularly relevant in situations where HLA-B27 testing is not useful (e.g. in general practice where the axSpA prevalence is low) or not possible (e.g. high costs in countries with lower funds). As shown in **chapter 4** of this thesis, in the Middle East & North Africa there was a higher percentage of patients for whom HLA-B27 was not available compared to the other geographical regions. Furthermore, HLA-B27 is less prevalent in the Middle East & North Africa[19,20], indicating that in this region information on family history may be especially valuable.

Finally, family history can be used as a risk factor for the development of axSpA. Notably, it has been shown that as soon as HLA-B27 status is known, a positive family history does not contribute to the likelihood of an axSpA diagnosis[17]. Therefore, its use as a risk factor for the development of axSpA seems limited to situations in which HLA-B27

is unavailable. The predictive value of the individual SpA entities in a family history for the risk of developing axSpA have not been assessed in the past. However, based on the fact that an association with HLA-B27 was only found for axSpA in the PerSpA and ASAS cohorts[16] and for axSpA and uveitis in the DESIR and SPACE cohorts[18], it is very likely that the definition of a positive family history should be redefined for this purpose too. Of note, the fact that an association between HLA-B27 and uveitis was only found in two European cohorts, emphasises yet again the importance of including patients from around the world when a decision is made regarding a changed definition, which should be applicable to all patients worldwide.

As classification criteria are used in the selection of patients for clinical studies, their most important aim should remain to create a homogeneous group of patients, yet the nature of the disease for which they are employed should be taken into account. AxSpA is a disease with a great variation in symptom expression, hence the classification criteria cannot be too restrictive, as it will cut out a chunk of patients with 'less typical' symptoms. The CLASSIC study provides a unique opportunity to assess different combinations of criteria, or more likely, different weights for the various symptoms.

In conclusion, the first part of this thesis has emphasised the importance of including patients from various countries with various ethnic backgrounds. Future research should focus on further enhancing the ASAS classification criteria -and its components- by including patients from across the globe.

ASAS/OMERACT CORE SET FOR AXIAL SPONDYLOARTHRITIS

The first section of this thesis reviewed the importance of global applicability of classification criteria, as they ensure the same patients are selected for participation in clinical trials worldwide, allowing for direct comparisons between studies executed in different geographical regions. Likewise, standardised assessment and reporting of results allows for direct comparisons between studies investigating different treatments, or identical treatments in populations from a different ethnicity or background, which is debated in the second part of this thesis.

A core outcome set describes the minimum and mandatory set of instruments that should be assessed and reported all clinical studies of a specific health condition, population and setting[25,26]. A core outcome set consists of domains (*what to measure*) and instruments (*how to measure*). The core outcome set currently used in axSpA is the ASAS-OMERACT core outcome set for ankylosing spondylitis[27,28]. Since the development of the original core outcome set over two decades ago, it has become apparent that axSpA is in fact a

disease spectrum that consists of two subtypes: radiographic axSpA and non-radiographic axSpA[29]. Additionally, major advances have occurred in the outcome assessments in the field of axSpA, such as the use of magnetic resonance imaging[30], the development of the Ankylosing Spondylitis Disease Activity Score (ASDAS)[31], validated enthesitis scores[32], and the ASAS-health index[33] . Finally, progress regarding the methodology surrounding the development of core sets has occurred, all of which made ASAS decide it was time to update the core outcome set.

The first step towards an updated core outcome set was to assess whether the domains represented in the original core set are still relevant. In order to gather information from patients with and experts in axSpA regarding the importance of outcomes to be assessed in all trials in axSpA, a Delphi survey was conducted. From the results of this Delphi survey -described in **chapter 5** of this thesis- we learned patients with axSpA had a different opinion regarding the outcomes that have to be assessed in all trials investigating therapies in axSpA than the experts involved in their treatment. Patients preferred an all-inclusive approach, whereas experts turned out to make a distinction between outcomes of critical importance for different treatment settings. According to the experts in the Delphi exercise, more objectively measurable domains such as structural damage and mobility were most critical to be measured in settings investigating disease modifying therapies, whereas the importance of assessing the more subjective domains such as pain, stiffness, and overall functioning & health was considered limited to the settings investigating symptom modifying therapies.

The Delphi survey is a common method used to gather opinions from a large group of participants, either to prioritise research topics, for importance ratings or to reduce item lists. For something used so often, guidance on the fundamentals of the methodology is scarce. In **chapter 6** we aimed to provide insight in the effect of choosing a certain invitation technique on the outcome of the Delphi, by comparing two often used invitation approaches: 1) Invite all participants to subsequent rounds, irrespective of response to the previous round; or 2) Invite only those participants who completed the previous round to subsequent rounds.

We found there is no effect on the final outcome of the Delphi, but argued it may be preferential to invite participants who missed a round to subsequent rounds, as this approach is less sensitive to the non-random loss of opinions that could lead to false consensus. Additionally, this approach ensures the end-result displays the opinion of all those invited.

The Delphi survey was a small part of a much larger effort to update the core outcome set for ankylosing spondylitis, to ensure applicability to all patients with axSpA. In **chapter 7** we described the process that led to the ASAS-OMERACT core domain set for axSpA. At

its core, the resulting core domain set is similar to the original core set for ankylosing spondylitis[34,35], in that both include the domains physical function, morning stiffness (called spinal stiffness in the core set for ankylosing spondylitis), pain, fatigue and disease activity (represented by patient global assessment in the core set for ankylosing spondylitis) in their core and structural damage (referred to as spine radiograph in the core set for ankylosing spondylitis) for DMARD settings only. The first noticeable difference is the addition of overall functioning and health in the core of the updated core domain set for axSpA, representing the impact of axSpA on other aspects of life and how this has received increased attention in the past years. Secondly, the removal of spinal mobility from the core is remarkable, which was caused by the lack of standardisation and poor reliability and sensitivity to change of spinal mobility outcomes[36,37]. Now the domains have been established, appropriate instruments need to be selected to measure these domains.

As a preparatory step towards the selection of instruments for the core set for axSpA, **chapter 8** described the test-retest reliability of the outcomes assessed in three recent randomised controlled trials in axSpA. From this study we concluded that even though most instruments were developed for radiographic axSpA they were also found reliable for non-radiographic axSpA. Furthermore, this study provided evidence in favour of multi-item instruments, as they were found to be more robust against measurement error.

Implications from these findings

Using the data collected in the Delphi survey-described in **chapter 5**-, we were able to compose a concise list of domains of which all stakeholders agree these are the most important domains to assess in all trials of axSpA. These data provided the basis for the development of the updated core domain set for axSpA. **Chapter 7** revealed the process underlaying the development of and emanating endorsement of the core domain set for axSpA. The changes made compared to the original core set for ankylosing spondylitis will have implications for future research, as it requires the assessment and reporting of slightly different outcomes than have been done previously. Unfortunately, for some aspects of disease (e.g. spinal mobility) this reduces comparability with older studies, yet at the same time ensures more comparability in future trials due to increased clarity on which outcomes should be measured and reported. Furthermore, all stakeholders who will benefit from an updated core set were involved in its development, which will increase the uptake. Next steps include the selection of the best instruments to assess the selected domains for which careful consideration includes assessment of measurement properties, feasibility and usability of the candidate instruments. The assessment of test-retest reliability and measurement error was described in **chapter 8** of this thesis, which provided proof that the assessed instruments are reliable for all patients with axSpA. Further research will have to investigate the other measurement properties (such as construct validity, and discrimination) before a final decision can be made on which

instruments are best suitable to measure the endorsed domains. Finally, **chapter 6** filled a knowledge gap regarding who to invite when conducting a Delphi survey. With this paper we provided a first piece of methodological guidance regarding the Delphi survey.

Further discussion and future perspectives

The core set for ankylosing spondylitis[34,35] which is currently used to determine which outcomes should be assessed as minimum in each clinical trial conducted in axSpA does not contain an instrument for each domain. No specific instruments were defined for the assessment of fatigue or enthesitis, because at the time of development of the core set there was no validated instrument available[28]. As a result, various instruments have been used to assess these domains, impeding comparisons between trials that are ever so important for assessing treatment efficacy. Therefore, it is vital the updated core set will advise one specific instrument for each domain (with the potential of adding more) that will not only have to be measured, but more importantly will have to be reported in each trial, enabling one-on-one comparison of trials and the development and update of treatment recommendations. The original core outcome set was well-implemented[38], indicating the implementation of a core set leads to structured collection of information for the endorsed domains in clinical trials. However, there was quite some variation in the instruments used to collect the information, emphasising the importance of recommending one specific instrument for each domain. Moreover, the review found that not all information that was collected was also reported[38]. For example, BASDAI includes a measure of fatigue, but frequently this was not reported separately and therefore no conclusion could be drawn on the effect of the investigated therapy on fatigue based on the presented data.

The main aim of a core outcome set is to provide the minimum and mandatory set of domains and instruments that has to be assessed in every trial. As our understanding of a disease increases, or the course of disease changes as a result of earlier recognition and effective therapy, this may lead to the development of new instruments (e.g. the ASAS-HI) or validation of existing instruments. There is a fair chance that these new instruments outperform existing instruments, and become the preferred instrument to measure a given domain, which would require an update of the core outcome set. Unfortunately, updating a core set is a lengthy and time-consuming process, and one might wonder whether the process surpasses its goal and whether there would be more core sets (i.e. more standardised measurements) if the process was more user-friendly. One option to simplify the process might be the regular review of a core set (e.g. every 10 years), in which it can be decided to replace an instrument if the new instrument has been shown to outperform the existing one, without having to go through all the steps required for the development of a new core set. In order to do so, a complete comparison on all psychometric properties between the new and existing instrument would be a pre-requisite.

For all its flaws, core outcome sets are valuable tools in research, as they allow for more transparency in the drug registration process (due to direct comparisons with previously registered drugs) and to a better acceptance of new treatments in the field (because a direct comparison of outcome measures shows its performance in relation to previously accepted/more familiar drugs). The past has taught us it is important to specifically define which instruments should be used to assess each domain, and which can be optionally added. Additionally, the use of an instrument does not guarantee all collected data is reported too. Hence, in addition to providing the domains and instruments to be assessed in each trial, the core set for axSpA should provide specific instructions for the reporting of data as well.

A Delphi survey was used to gain consensus among experts in and patients with axSpA. One can argue this is not the best way to collect information regarding the importance of domains in axSpA, as its main aim is to strive for consensus among participants. Though consensus tends to dilute the strength of less favoured opinions, and replaces individuality by group opinion, it ensures all the involved can accept the final outcome. An additional benefit of the Delphi survey is that it can be completed online (i.e. no space- and time constraints) and anonymously, thereby providing a safe environment to express what could be considered a less favourable opinion. One major drawback with the taken approach is the lack of standardisation in the methodology to execute Delphi surveys, which decreases its validity as a tool for such important aspects of research. Within this thesis we made an effort to critically assess one aspect in the methodology of the Delphi survey. Providing clear methodological instructions for the use of Delphi surveys will improve its validity and more importantly, the validity of its results. Future research should focus on other elements of the methodology, such as how feedback should be provided between rounds, or how many panellists and stakeholder groups should be invited. Within these methodological recommendations, separate instructions should be provided for the various applications of a Delphi survey (e.g. a preference of care evaluation requires a different approach than the selection of domains for a core set), such that clear guidance will become available for researchers who wish to perform a Delphi survey.

In conclusion, core sets are a valuable tool in outcome assessment, but future research should investigate whether their development process can be smoothened. As for Delphi surveys, there is no doubt they have proven their worth as a means of gathering opinions and reaching consensus, yet the lack of methodological guidance should be addressed in the future.

PATIENT REPORTED OUTCOMES IN EARLY AXIAL SPONDYLOARTHRITIS

In the final part of this thesis, we discussed health-related quality of life and work and activity impairments in axSpA. In **chapter 9** we showed work and activity outcomes as well as health-related quality of life improved over two years of protocolized follow-up in patients with chronic back pain suspected of axSpA. This improvement was shown for patients with and without a diagnosis of axSpA, yet patients diagnosed with axSpA showed a larger improvement compared to those without a diagnosis of axSpA.

Implications from these findings

The results from **chapter 9** have taught us we can expect some improvement with time in all chronic back pain patients suspected of axSpA-regardless of diagnosis-, which apart from treatment may relate to the fact that at their first visit to the rheumatologist a patient's complaints are at their most severe and thus might improve naturally with time (i.e. regression to the mean). Furthermore, as there was a control group available in this study (i.e. those without a diagnosis) we were able to conclude a diagnosis of axSpA is an independent predictor of improvement in health-related quality of life and work- and activity outcomes in patients with chronic back pain complaints. Importantly, sensitivity analyses showed that the differences in outcomes between those with and without a diagnosis of axSpA remained when analyses were restricted to patients not on biological therapy, indicating that treatment with biologicals did not explain the differences between the groups. Importantly, despite the improvements over time, outcomes were still impaired compared to the general population. These results emphasise the importance of optimising long-term health-related quality of life and social participation of patients with axSpA, which is also described in the current axSpA treatment guidelines as one of the primary treatment goals[39].

Further discussion and future perspectives

Chapter 9 showed that patients who were diagnosed with axSpA after two years of protocolised follow-up showed a larger improvement in their health-related quality of life, work productivity loss and activity impairment than chronic back pain patients who did not get diagnosed with axSpA. In **chapter 9** we mentioned that this might be due to a difference in available treatment options, but this might also be explained by a difference in illness perceptions and subsequent coping mechanisms. One of the questions we asked ourselves is whether being diagnosed could have an impact on how a patient perceives his/her illness and subsequently influence coping mechanisms. Future research should focus on getting more insight in the psychological effects of getting a diagnosis, whether simply knowing 'what is wrong with you' has an impact on how complaints are perceived, and whether quality of life might improve even further if patients are taught adequate

coping mechanisms. Simultaneously, adequate coping mechanisms might have an impact on work-outcomes. Work productivity loss is based on presenteeism (reduced ability to perform one's job adequately) and absenteeism (the hours missed from work due to disease), if effective treatment combined with adequate coping mechanisms can reduce presenteeism, subsequently work productivity loss will improve. As axSpA affects the lives of people that are in the prime of their life, there is an immense value in increasing our understanding of effective therapies -either medicinal, educational or psychological- that have a positive effect on quality of life and work- and activity outcomes.

Another aspect that deserves attention is the use of generic versus disease-specific questionnaires to assess health-related quality of life. The main advantage of using a generic questionnaire is that the scores can be compared to scores from patients diagnosed with other (chronic) diseases or healthy controls, augmenting societal value. Contrary, disease specific questionnaires (such as the ASAS-HI and Ankylosing Spondylitis Quality of Life survey) pertain more disease specific questions, providing a clearer insight in the effect of disease on quality of life of patients with axSpA and could be considered of higher scientific value. In **chapter 9** the use of the SF-36 allowed for a direct comparison of patients with and without a diagnosis of axSpA and also a comparison with the general population. Directly comparing those who did and did not receive a diagnosis of axSpA showed us that in fact there is a difference in improvement over time between these groups, a fact that would have gone unnoticed had we used a disease-specific questionnaire. Nonetheless, using a disease specific questionnaire could have led to insight in which aspects (if any) contributed to the improvement in quality of life over time. Therefore, a disease-specific questionnaire may be preferred in a longitudinal cohort of patients with a chronic disease. As it is very likely health-related quality of life will remain impaired in these patients compared to the general population, insight in aspects that contribute to improvement or worsening of quality of life over time might be more valuable, as this could bring about new treatment goals. In the end, research will always demand making choices and finding the optimal balance between cost and reward, hence, what constitutes as 'the right choice' will depend on the question at hand, the available data and many other variables that are beyond our control.

REFERENCES

1. Rudwaleit M, Haibel H, Baraliakos X, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. *Arthritis Rheum* 2009;60(3):717-27.
2. Muñoz-Fernández S, Carmona L, Collantes E, et al. A model for the development and implementation of a national plan for the optimal management of early spondyloarthritis: the Esperanza Program. *Ann Rheum Dis* 2011;70(5):827-30.
3. Spoorenberg A, van Tubergen A, Landewé R, et al. Measuring disease activity in ankylosing spondylitis: patient and physician have different perspectives. *Rheumatology* 2005;44(6):789-95.
4. Canhão H, Faustino A, Martins F, et al. Reuma.pt - the rheumatic diseases portuguese register. *Acta Reumatol Port* 2011;36(1):45-56.
5. Ciurea A, Scherer A, Exer P, et al. Tumor necrosis factor α inhibition in radiographic and nonradiographic axial spondyloarthritis: results from a large observational cohort. *Arthritis Rheum* 2013;65(12):3096-106.
6. van den Berg R, de Hooge M, van Gaalen F, et al. Percentage of patients with spondyloarthritis in patients referred because of chronic back pain and performance of classification criteria: experience from the Spondyloarthritis Caught Early (SPACE) cohort. *Rheumatology* 2013;52(8):1492-9.
7. Lee W, Reveille JD, Davis JC, et al. Are there gender differences in severity of ankylosing spondylitis? Results from the PSOAS cohort. *Ann Rheum Dis* 2007;66(5):633-38.
8. Rudwaleit M, Landewé R, van der Heijde D, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part I): classification of paper patients by expert opinion including uncertainty appraisal. *Ann Rheum Dis* 2009;68(6):770-6.
9. Rudwaleit M, van der Heijde D, Landewé R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis* 2009;68(6):777-83.
10. van der Linden S, Akkoc N, Brown MA, et al. The ASAS Criteria for Axial Spondyloarthritis: Strengths, Weaknesses, and Proposals for a Way Forward. *Curr Rheumatol Rep* 2015;17(9):62.
11. Feldtkeller E, Khan MA, van der Heijde D, et al. Age at disease onset and diagnosis delay in HLA-B27 negative vs. positive patients with ankylosing spondylitis. *Rheumatol Int* 2003;23(2):61-6.
12. Ciurea A, Scherer A, Weber U, et al. Age at symptom onset in ankylosing spondylitis: is there a gender difference? *Ann Rheum Dis* 2014;73(10):1908-10.
13. Chung HY, Machado P, van der Heijde D, et al. HLA-B27 positive patients differ from HLA-B27 negative patients in clinical presentation and imaging: results from the DESIR cohort of patients with recent onset axial spondyloarthritis. *Ann Rheum Dis* 2011;70(11):1930-6.
14. Skare TL, Leite N, Bortoluzzo AB, et al. Effect of age at disease onset in the clinical profile of spondyloarthritis: a study of 1424 Brazilian patients. *Clin Exp Rheumatol* 2012;30(3):351-7.
15. van der Linden SM, Valkenburg HA, de Jongh BM, et al. 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(3):241-9.
16. van Lunteren M, Sepriano A, Landewé R, et al. Do ethnicity, degree of family relationship, and the spondyloarthritis subtype in affected relatives influence the association between a positive family history for spondyloarthritis and HLA-B27 carriership? Results from the worldwide ASAS cohort. *Arthritis Res Ther* 2018;20(1):166.
17. van Lunteren M, van der Heijde D, Sepriano A, et al. Is a positive family history of spondyloarthritis relevant for diagnosing axial spondyloarthritis once HLA-B27 status is known? *Rheumatology* 2019;58(9):1649-54.
18. Ez-Zaitouni Z, Hilkens A, Gossec L, et al. Is the current ASAS expert definition of a positive family history useful in identifying axial spondyloarthritis? Results from the SPACE and DESIR cohorts. *Arthritis Res Ther* 2017;19(1):118.
19. Mustafa KN, Hammoudeh M, Khan MA. HLA-B27 Prevalence in Arab Populations and Among Patients with Ankylosing Spondylitis. *J Rheumatol* 2012;39(8):1675-7.
20. Ziade N, Abi Karam G, Merheb G, et al. HLA-B27 prevalence in axial spondyloarthritis patients and in blood donors in a Lebanese population: Results from a nationwide study. *Int J Rheum Dis* 2019;22(4):708-14.
21. López-Medina C, Molto A, Sieper J, et al. Prevalence and distribution of peripheral musculoskeletal manifestations in spondyloarthritis including psoriatic arthritis: results of the worldwide, cross-sectional ASAS-PerSpA study. *RMD Open* 2021;7(1)
22. Akkoc N, Khan MA. ASAS classification criteria for axial spondyloarthritis: time to modify. *Clin Rheumatol* 2016;35(6):1415-23.
23. de Winter J, de Hooge M, van de Sande M, et al. Magnetic Resonance Imaging of the Sacroiliac

Joints Indicating Sacroiliitis According to the Assessment of SpondyloArthritis international Society Definition in Healthy Individuals, Runners, and Women With Postpartum Back Pain. *Arthritis Rheumatol* 2018;70(7):1042-48.
24. Weber U, Jurik AG, Zejden A, et al. Frequency and Anatomic Distribution of Magnetic Resonance Imaging Features in the Sacroiliac Joints of Young Athletes: Exploring "Background Noise" Toward a Data-Driven Definition of Sacroiliitis in Early Spondyloarthritis. *Arthritis Rheumatol* 2018;70(5):736-45.
25. Boers M, Kirwan JR, Tugwell P. OMERACT Handbook, 2018.
26. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials* 2017;18(3):280.
27. van der Heijde D, van der Linden S, Bellamy N, et al. 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(4):945-47.
28. van der Heijde D, van der Linden S, Dougados M, et al. Ankylosing spondylitis: plenary discussion and results of voting on selection of domains and some specific instruments. *J Rheumatol* 1999;26(4):1003-5.
29. 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.
30. Rudwaleit M, Jurik AG, Hermann K-GA, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. *Ann Rheum Dis* 2009;68(10):1520-27.
31. Lukas C, Landewé R, Sieper J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68(1):18-24.
32. Heuft-Dorenbosch L, Spoorenberg A, van Tubergen A, et al. Assessment of enthesitis in ankylosing spondylitis. *Ann Rheum Dis* 2003;62(2):127-32.
33. Kiltz U, van der Heijde D, Boonen A, et al. Development of a health index in patients with ankylosing spondylitis (ASAS HI): final result of a global initiative based on the ICF guided by ASAS. *Ann Rheum Dis* 2015;74(5):830-35.
34. van der Heijde D, Bellamy N, Calin A, et al. Preliminary core sets for endpoints in ankylosing spondylitis. Assessments in Ankylosing Spondylitis Working Group. *J Rheumatol* 1997;24(11):2225-9.
35. van der Heijde D, Calin A, Dougados M, et al. 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(4):951-4.
36. Marques ML, Ramiro S, Goupille P, et al. Measuring spinal mobility in early axial spondyloarthritis: does it matter? *Rheumatology* 2019;58(9):1597-606.
37. Ogdie A, Duarte-García A, Hwang M, et al. Measuring Outcomes in Axial Spondyloarthritis. *Arthritis Care & Research* 2020;72(S10):47-71.
38. Bautista-Molano W, Navarro-Compán V, Landewé RB, et al. How well are the ASAS/OMERACT Core Outcome Sets for Ankylosing Spondylitis implemented in randomized clinical trials? A systematic literature review. *Clin Rheumatol* 2014;33(9):1313-22.
39. van der Heijde D, Ramiro S, Landewé R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis* 2017;76(6):978-91

CHAPTER 11

NEDERLANDSE SAMENVATTING

Axiale spondyloartritis (axSpA) is een chronische reumatische ziekte die meestal ontstaat tussen het 20e en 30e levensjaar. AxSpA wordt veroorzaakt door chronische ontsteking in de wervelkolom en de belangrijkste symptomen zijn chronische rugpijn (aanwezig voor ten minste 3 maanden) en stijfheid van de rug.

In tegenstelling tot andere reumatische aandoeningen, zoals reumatoïde artritis waarbij ontsteking kan leiden tot botafbraak, wordt axSpA gekenmerkt door de vorming van nieuw bot op de plek van ontsteking. Dit kan resulteren in botvergroeiingen in de sacro-iliacale gewrichten (het bekken) en de wervelkolom. Deze vergroeiingen dragen bij aan beperkingen in mobiliteit en fysiek functioneren, waardoor veel activiteiten van het dagelijks leven negatief worden beïnvloed. Aangezien axSpA meestal op relatief jonge leeftijd optreedt, zullen patiënten het grootste deel van hun leven moeten omgaan met deze ziekte. Naast pijn en stijfheid ervaren veel patiënten vermoeidheid en slaapproblemen, die allemaal een grote invloed hebben op de kwaliteit van leven en hun vermogen om deel te nemen aan dagelijkse activiteiten, zoals (huishoudelijk) werk.

Er zijn twee subtypes van axSpA: 1) radiografische axSpA (r-axSpA, ook bekend als ankyloserende spondylitis (AS)), gekenmerkt door aanzienlijke structurele schade aan de sacro-iliacale gewrichten (het bekken) zichtbaar op röntgenfoto's; en 2) niet-radiografische axSpA (nr-axSpA), gekenmerkt door klinische symptomen van axSpA in afwezigheid van duidelijke schade zichtbaar op röntgenfoto's. Nr-axSpA wordt vaak beschouwd als een vroeg stadium van de ziekte, wat impliceert dat patiënten gaandeweg r-axSpA kunnen ontwikkelen. Er zijn bepaalde risicofactoren voor het ontwikkelen van r-axSpA, zoals mannelijk geslacht, aanwezigheid van HLA-B27 gen (een erfelijke factor, welke leidt tot een verhoogde kans op ontstekingsziekten zoals axSpA), hoge ontstekingsactiviteit (d.w.z. verhoogde ontstekingswaarden in het bloed of ontsteking zichtbaar op MRI), en roken. Progressie van nr-axSpA naar r-axSpA wordt gezien bij ongeveer 5-20% van de patiënten in een tijdsperiode van 2-5 jaar, terwijl een deel van de patiënten nooit r-axSpA ontwikkelt. Nr-axSpA is dus meer dan alleen een vroeg stadium van ziekte, het is ook een ziekte-expressie. De klachten en ziekteactiviteit van patiënten met nr-axSpA zijn even ernstig en beperkend als die van patiënten met r-axSpA.

Er is geen eenduidig patroon waar alle patiënten met axSpA aan te herkennen zijn. Ziektekenmerken kunnen zelfs zeer verschillen tussen patiënten, wat axSpA een zogenaamd heterogene ziekte maakt. Naast de kenmerkende rugklachten zijn er verschillende andere klinische kenmerken die veel voorkomen bij patiënten met axSpA, de zogenaamde spondyloartritis (SpA) kenmerken. De meest voorkomende is inflammatoire rugpijn. Van inflammatoire rugpijn is volgens huidige criteria sprake als ten minste vier van de volgende vijf factoren aanwezig zijn: 1) ontstaan van de rugklachten vóór de leeftijd van 40; 2) geleidelijk ontstaan; 3) verbetering met beweging; 4) geen verbetering met

rust; en 5) nachtelijke pijn met verbetering bij het opstaan. Andere SpA kenmerken zijn een goede reactie op pijnstillers met een ontstekingsremmende werking (zogenaamd NSAIDs zoals naproxen of ibuprofen) en het voorkomen van SpA bij een familielid. Indien een eerstegraads (ouder, broer/zus, kind) of tweedegraads (grootouder, oom/tante, neef/nicht) familielid de diagnose SpA heeft spreken we van een positieve familieanamnese.

Ook klachten zoals ontstekingen in gewrichten buiten bekken en wervelkolom (artritis), ontstekingen van peesaanhechtingen (enthesitis) en zwellingen die leiden tot worstvormige vingers en tenen (dactylitis) zijn SpA kenmerken. Ontstekingen buiten de gewrichten en rug, zoals de ogen (uveitis), huid (psoriasis) en darmen (ziekte van Crohn en colitis ulcerosa, IBD) worden aangeduid als extra-musculoskeletale aandoeningen (niet tot het bewegingsapparaat behorend), en zijn ook SpA kenmerken. Verder zijn verhoogde ontstekingswaarden in het bloed en de aanwezigheid van de genetische factor HLA-B27 SpA kenmerken.

Aangezien axSpA bij vrijwel alle patiënten in ieder geval het bekken aantast, speelt beeldvorming van de sacro-iliacale gewrichten een cruciale rol bij de diagnose en classificatie van axSpA. Röntgenfoto's en MRI zijn de meest gebruikte beeldvormende technieken in de klinische praktijk. Er zijn echter beperkingen aan het gebruik van röntgenfoto's van de sacro-iliacale gewrichten bij patiënten met een vroege ziekte, omdat structurele veranderingen over het algemeen jaren duren. Daarom kan MRI van de sacro-iliacale gewrichten waardevolle informatie opleveren, omdat het de identificatie van actieve ontsteking mogelijk maakt, evenals de aanwezigheid van structurele veranderingen die het gevolg zijn van ontstekingen. Alle SpA-kenmerken zijn zeer nuttig bij de diagnose van axSpA en bij de classificatie van patiënten voor klinische onderzoeken. Bovendien bieden deze kenmerken belangrijke informatie over de ziekteprognose.

In de reumatologie zijn classificatiecriteria bedoeld om goed gedefinieerde, relatief homogene groepen patiënten te creëren voor klinisch onderzoek. Gevalideerde classificatiecriteria zijn van groot belang voor de interpretatie van onderzoeksbevindingen en vergelijkingen van resultaten tussen onderzoeken. De classificatiecriteria omvatten niet het hele spectrum van mogelijke uitingen van een ziekte, maar moeten zeer specifiek zijn om het onjuist labelen als het hebben van een ziekte tot een minimum te beperken.

Aangezien reumatische aandoeningen heterogeen van aard zijn, kunnen classificatiecriteria er niet in slagen om alle patiënten met axSpA te identificeren. Dit komt doordat classificatiecriteria gericht zijn op een meer homogene populatie dan in de dagelijkse klinische praktijk wordt gezien. Classificatiecriteria mogen dus niet worden gebruikt om patiënten te diagnosticeren, maar uitsluitend om patiënten te includeren in klinische onderzoeken.

Classificatiecriteria zouden van toepassing moeten zijn op alle patiënten met axSpA waar dan ook ter wereld. Dit is nodig om te garanderen dat de patiënten die worden geselecteerd voor deelname aan klinische onderzoeken wereldwijd hetzelfde zijn. In dit proefschrift willen we een internationaal perspectief bieden op de karakterisering van patiënten met axSpA-specifiek met betrekking tot de leeftijd bij aanvang van de symptomen en positieve familieanamnese van axSpA-, om te onderzoeken of classificatiecriteria inderdaad wereldwijd toepasbaar zijn.

Verder is het belangrijk dat alle onderzoeken die in verschillende delen van de wereld worden uitgevoerd dezelfde uitkomstmaten meten en deze ook op een vergelijkbare manier rapporteren. Dit om te zorgen dat bijvoorbeeld gegevens uit Amerikaanse onderzoeken kunnen worden vergeleken met onderzoeken die in Azië zijn uitgevoerd. Uitkomstmaten omschrijven alle meetinstrumenten die worden gebruikt om data te verzamelen in klinische studies, zoals een vragenlijst over pijn, of een score voor gewrichtspijn. Door het gebruik van dezelfde uitkomstmaten voorafgaand aan de start van de behandeling en na afloop van de behandeling, kunnen we iets zeggen over het effect dat de behandeling heeft gehad, bijvoorbeeld op pijn. Hierin is een belangrijke rol weggelegd voor een standaard set uitkomstmaten, ook wel een 'core outcome set' genoemd. De standaard set beschrijft de gegevens die ten minste gemeten/verzameld en gerapporteerd moeten worden wanneer er een studie wordt uitgevoerd. Door standaardisatie van metingen en rapportage wordt het mogelijk om directe vergelijkingen te maken tussen klinische onderzoeken naar de effectiviteit en veiligheid van behandelingen en wordt voorkomen dat alleen de gunstige uitkomsten gerapporteerd worden.

De eerste stap die moet worden gezet bij het ontwikkelen of vernieuwen van een standaard set is bepalen wat er gemeten moet worden, dit wordt gedefinieerd in zogenaamde domeinen. Daarna moet worden gedefinieerd hoe elk van de gekozen domeinen moet worden gemeten door de selectie van instrumenten (zoals vragenlijsten of gewricht-scores). De uiteindelijke standaard set zal zowel de geselecteerde domeinen als instrumenten bevatten.

Voor veel uitkomstmaten -zoals pijn of kwaliteit van leven- maken reumatologen en onderzoekers gebruik van de subjectieve informatie die door de patiënt wordt verstrekt, omdat er geen objectieve metingen beschikbaar zijn. Vandaar dat een groot deel van de uitkomstmaten die vaak worden gebruikt in axSpA zogenaamde patiënt-gerapporteerde uitkomstmaten zijn. Daarnaast heeft eerder onderzoek aangetoond dat artsen en patiënten verschillende opvattingen hebben van ziekteactiviteit en fysiek functioneren, wat het belang van de patiënt-gerapporteerde uitkomstmaten verder benadrukt.

Ook hebben de patiënt-gerapporteerde uitkomstmaten een grote rol gespeeld bij de erkenning dat de ziektelast vergelijkbaar is tussen patiënten met en zonder schade aan het bekken (het sacro-iliacaal gewricht) op de röntgenfoto's.

Het is bekend dat axSpA een nadelige invloed kan hebben op de kwaliteit van leven. Om die reden is het optimaliseren van de kwaliteit van leven op de lange termijn benoemd als het belangrijkste behandeldoel bij axSpA. Eerder onderzoek heeft aangetoond dat de kwaliteit van leven al verminderd is bij patiënten in de vroege fase van axSpA en dat deze kan worden verbeterd door de ziekteactiviteit te verminderen met een effectieve behandeling.

Bovendien gaat axSpA gepaard met een groot risico op beperking van de arbeidsproductiviteit gedurende het leven van de patiënt, wat bijdraagt aan substantiële maatschappelijke kosten van axSpA. De werkloosheidscijfers en arbeidsongeschiktheidscijfers zijn aanzienlijk hoger in vergelijking met de algemene bevolking. Ook komt overstappen naar een fysiek minder veeleisende baan of vervroegde uittreding vaker voor bij patiënten met axSpA. Verminderd vermogen om het werk adequaat uit te voeren (presenteïsme) en een toename van het aantal gemiste werkuren door ziekte (absenteïsme) leiden tot verminderde arbeidsproductiviteit bij patiënten met axSpA. Omdat axSpA vaak in de meest productieve periode van iemands leven ontstaat (rond het 20^{e}-30^{e} levensjaar), is onderzoek naar de impact van de ziekte op het vermogen om te werken erg belangrijk.

Dit proefschrift is opgebouwd rondom drie thema's: 1) een internationaal perspectief bieden op de kenmerken van patiënten met axSpA; 2) beschrijven van het proces van de ontwikkeling van een standaard set uitkomstmaten voor axSpA; en 3) meer kennis vergaren over werkproductiviteit en kwaliteit van leven bij chronische rugpijnpatiënten met de diagnose axSpA of een vermoeden daarvan. In dit laatste hoofdstuk worden de bevindingen van dit proefschrift samengevat en in een breder perspectief geplaatst. Ten slotte zullen er aanbevelingen worden gedaan voor toekomstig onderzoek.

INTERNATIONALE CLASSIFICATIE VAN AXIALE SPONDYLOARTRITIS

In het eerste deel van dit proefschrift wilden we een internationaal perspectief bieden op de classificatie van patiënten met axSpA. Hiervoor zijn verschillende aspecten van de classificatiecriteria voor axSpA nauwkeurig onderzocht. Te beginnen met de naamgeving die wordt gebruikt om patiënten met axSpA met radiografische schade aan de sacro-iliacale gewrichten (het bekken) te beschrijven in **hoofdstuk 2**. Van oudsher werden patiënten met axSpA met onomkeerbare structurele veranderingen op röntgenfoto's geclassificeerd

volgens de modified New York (mNY) criteria als ankyloserende spondylitis (AS). Echter, met het ontstaan van de recentere ASAS axSpA-criteria is een alternatief ontstaan, waarmee deze patiënten konden worden geclassificeerd als radiografische axSpA. Critici betwijfelden of beide classificatiesets dezelfde patiënten zouden classificeren, maar dit was nooit onderzocht. In **hoofdstuk 2** concludeerden we dat bijna alle patiënten met axSpA met radiografische sacroiliitis die aan de mNY-criteria voldeden, ook voldeden aan de ASAS-criteria voor radiografische axSpA en vice versa. Dit betekent dat de termen AS en radiografische axSpA uitwisselbaar zijn. Tot nu toe is er geen overeenstemming bereikt of de ene definitie de voorkeur verdient boven de andere. Desalniettemin, aangezien we op weg zijn naar één diagnose (d.w.z. axiale spondyloartritis) met twee subgroepen enkel voor classificatiedoeleinden (d.w.z. radiografische en niet-radiografische axSpA), zou het naar mijn mening wenselijk zijn om radiografische axSpA te gebruiken om deze patiënten te beschrijven. Bovendien wordt tegenwoordig erkend dat niet-radiografische en radiografische axSpA een gelijke ziektelast hebben en behandeling met anti-reumatische geneesmiddelen (bDMARDs) is effectief gebleken in beide groepen patiënten. Daarom lijkt het gebruik van axSpA het meest geschikt om alle patiënten met een diagnose te beschrijven en de toevoeging radiografisch versus niet-radiografisch om aanvullende informatie te geven over de expressie van ziekte.

Erkennen dat de termen AS en radiografische axSpA uitwisselbaar zijn vergroot de vergelijkbaarheid tussen onderzoeken, aangezien beide termen dezelfde patiënten beschrijven. Dit zorgt er ook voor dat onderzoek uitgevoerd in cohorten van AS kan worden vergeleken met meer recent gepubliceerde artikelen over radiografische axSpA-cohorten. Dit is van enorm belang voor medicatieonderzoeken. Als de effectiviteit van een bepaald medicijn in het verleden is bewezen, worden ze niet onderworpen aan verdere gerandomiseerde klinische onderzoeken om de effectiviteit ervan te beoordelen, omdat het onethisch zou zijn om patiënten effectieve medicatie te onthouden. Om het effect van medicatie grootschalig te kunnen onderzoeken worden vaak zogenaamde meta-analyses uitgevoerd, hierin worden resultaten van een aantal vergelijkbare klinische studies gebundeld. Door gegevens van radiografische axSpA in meta-analyses op te nemen, kunnen vergelijkingen worden gemaakt waarbij rekening wordt gehouden met alle soorten behandelingen, inclusief diegene welke zijn onderzocht bij AS. Dit betekent dat de behandeling kan worden gestart zonder dat er eerst een nieuw onderzoek moet worden uitgevoerd.

De belangrijkste reden voor discrepanties tussen patiënten geclassificeerd volgens de axSpA-criteria en mNY-criteria bleek de leeftijd waarop rugpijn ontstaan is. De leeftijd waarop de rugpijn ontstaat werd geïntroduceerd in de ASAS-criteria in 2009 en is voornamelijk gebaseerd op gegevens van Feldtkeller en collega's gepubliceerd in 2003. Zij toonden aan dat het ontstaan van rugpijn na de leeftijd van 45 bij slechts 5%

van de patiënten voorkwam. Vergelijkbare verdelingen in het ontstaan van de rugpijn zijn gevonden in andere onderzoeken die sindsdien zijn uitgevoerd, maar ook deze waren gebaseerd op voornamelijk Europese gegevens. **Hoofdstuk 3** was erop gericht een wereldwijd perspectief te bieden op de leeftijd waarop de eerste symptomen zich voordeden. Dit om te bevestigen of de leeftijd bij het ontstaan van de rugklachten over de hele wereld gelijk was. In **hoofdstuk 3** hebben we bevestigd dat de leeftijd bij het ontstaan van de rugklachten vergelijkbaar was in verschillende geografische regio's. Ook konden we bevestigen dat de overgrote meerderheid van de patiënten met axSpA inderdaad hun eerste symptomen ervoeren vóór de leeftijd van 45 jaar. Verder toonde onze studie aan dat over de hele wereld de rugpijn op jongere leeftijd ontstaat in patiënten die het HLA-B27 gen dragen. Ook vonden we dat bij mannen de rugpijn op iets jongere leeftijd ontstaat dan bij vrouwen.

Deze gegevens zijn erg belangrijk, omdat het bevestigt dat het criterium leeftijd bij aanvang van rugpijn kan worden toegepast op patiënten overal ter wereld. Verder impliceren deze gegevens dat axSpA zich op jongere leeftijd manifesteert bij HLA-B27-positieve en mannelijke patiënten. Het lijkt er dus op dat de leeftijd bij het ontstaan van de rugklachten een heel nuttig hulpmiddel is bij het identificeren van personen met een risico op axSpA. Aangezien slechts een zeer klein deel van de patiënten symptomen ontwikkelt na de leeftijd van 45 jaar, is het zeer onwaarschijnlijk dat een patiënt die rugklachten krijgt op het moment dat hij of zij een stuk ouder is dan 45 jaar axSpA heeft, wat belangrijke kennis is voor de klinische praktijk omdat het onnodige diagnostiek kan voorkomen.

Hoewel de symptomen bij HLA-B27-positieve en mannelijke patiënten op wat jongere leeftijd optreden, betekent uiteraard niet dat de diagnose axSpA bij HLA-B27-negatieve en vrouwelijke patiënten niet moet worden overwogen. Het besef dat de aanvang van de symptomen gemiddeld iets later kan zijn bij HLA-B27-negatieve en vrouwelijke patiënten, is dus iets wat een arts mee moet nemen in zijn of haar overwegingen. Wat de classificatiecriteria betreft, lijkt het criterium van een ontstaan van klachten <45 jaar dus geldig, aangezien de overgrote meerderheid van de patiënten met axSpA symptomen ontwikkelden vóór deze leeftijd over de hele wereld. Aangezien classificatiecriteria zijn gericht op het creëren van een homogene groep patiënten, lijkt het leeftijd-criterium een nuttig hulpmiddel bij het uitsluiten van de minder typische patiënten.

Het derde en laatste aspect van de classificatiecriteria dat in dit proefschrift wordt besproken, is de positieve familieanamnese van spondyloartritis, oftewel het voorkomen van SpA bij een familielid. De waarde van een positieve familieanamnese in zijn huidige vorm is eerder in twijfel getrokken, omdat de definitie niet is getest of gevalideerd voordat deze in de ASAS-classificatie werd opgenomen. **Hoofdstuk 4** beschreef de prevalentie van een positieve familieanamnese van spondyloartritis (d.w.z. hoe vaak een positieve

familieanamnese voorkomt bij patiënten met axSpA) in verschillende geografische regio's en de relatie met het HLA-B27 gen. We ontdekten dat axSpA de meest voorkomende vorm van spondyloartritis is in een positieve familieanamnese. Verder bleek de relatie tussen een positieve familieanamnese van axSpA en HLA-B27-dragerschap onafhankelijk van een positieve familieanamnese voor andere vormen van spondyloartritis. Deze bevindingen bevestigen dat het verband tussen een positieve familieanamnese en HLA-B27-dragerschap grotendeels wordt veroorzaakt door een positieve familieanamnese voor axSpA in een wereldwijd cohort. Dit werd eerder al aangetoond in cohorten met voornamelijk Europese en enkele Aziatische patiënten.

De bevindingen gepresenteerd in **hoofdstuk 4** van dit proefschrift in combinatie met eerder onderzoek naar familieanamnese (in de ASAS-, DESIR- en SPACE-cohorten) suggereren dat het tijd is om dit criterium kritisch te evalueren. Alle beschikbare gegevens tonen aan dat axSpA de meest voorkomende vorm van spondyloartritis is in een positieve familieanamnese. Verder wordt het verband tussen een positieve familieanamnese en HLA-B27-dragerschap grotendeels bepaald door een positieve familieanamnese voor axSpA. Gezien de consistente bevindingen over de hele wereld, moet worden onderzocht of de huidige definitie van een positieve familieanamnese in de classificatiecriteria kan worden versmald naar een positieve familieanamnese van axSpA.

Verdere discussie en toekomstig onderzoek

Uniformiteit in classificatie en rapportage helpt de wereldwijde communicatie van wetenschappelijke en klinische bevindingen, wat het begrip van het ontstaan en beloop van axSpA en diens behandeling vergroot. Het feit dat het niet duidelijk was of twee belangrijke elementen in de ASAS-classificatiecriteria (d.w.z. leeftijd bij ontstaan klachten en positieve familieanamnese) van toepassing waren op alle axSpA-patiënten wereldwijd, wijst op een kwetsbaarheid in het wetenschappelijke proces. Aangezien patiënten uit verschillende continenten kunnen verschillen in hun ziektepresentatie, is het belangrijk dat er voldoende patiënten van over de hele wereld worden geïncludeerd om ervoor te zorgen dat de classificatiecriteria representatief zijn voor alle patiënten wereldwijd. Het ASAS PerSpA-cohort leverde het bewijs voor de haalbaarheid van een dergelijk onderzoek: door internationale samenwerking en slim gebruik te maken van een elektronisch dataverzamelingssysteem was het mogelijk om data te verzamelen in 24 landen over de hele wereld. Door de opstartkosten van een onderzoek te verlagen (bijvoorbeeld door een elektronisch gegevensverzamelingssysteem aan te bieden) en het aanbieden van een (bescheiden) vergoeding, wordt de drempel om deel te nemen verlaagd voor landen met minder financiële middelen.

De validiteit van de ASAS-classificatiecriteria voor axSpA zijn in het verleden in twijfel getrokken. Hierbij werd gezegd dat de classificatiecriteria te breed zijn, wat zou kunnen leiden

tot verschillen in de samenstelling van patiënten populaties in verschillende onderzoeken. Anderen benoemen dat de verhoogde sensitiviteit van de huidige ASAS criteria ten koste van specificiteit een reden tot bezorgdheid is, aangezien classificatiecriteria gericht zijn op het creëren van homogene onderzoekspopulaties en daarom moeten streven naar de hoogst mogelijke specificiteit. Critici benadrukken dus het belang van her-evaluatie van de ASAS axSpA-classificatiecriteria om de specificiteit te verbeteren en daarmee heterogeniteit te verminderen binnen de groep van axSpA-patiënten die volgens deze criteria zijn geclassificeerd.

In 2019 startten ASAS en SPARTAN de CLASSIC-studie (CLassification of Axial Spondyloarthritis Inception Cohort) met als doel de ASAS-classificatiecriteria voor axSpA opnieuw te evalueren. Voor dit doel worden patiënten met chronische rugpijn met verdenking op axSpA geïncludeerd, en een diagnose van axSpA of geen axSpA wordt gesteld na zorgvuldige evaluatie van klinische, laboratorium- en beeldvormingsresultaten. Doordat patiënten met en zonder een diagnose van axSpA worden geïncludeerd, is het niet alleen mogelijk om de sensitiviteit en specificiteit van de classificatiecriteria als geheel te onderzoek, maar ook de afzonderlijke componenten ervan.

In principe blijven de classificatiecriteria ongewijzigd als een sensitiviteit ≥75% en specificiteit ≥90% wordt behaald in de CLASSIC studie. Desalniettemin kan het noodzakelijk zijn voor sommige van de afzonderlijke componenten te beoordelen of deze ongewijzigd moeten blijven, ongeacht of de sensitiviteit en specificiteit van de classificatiecriteria als geheel wordt behaald.

Een van de afzonderlijke componenten die onderzocht moet worden, is de definitie van een positieve familieanamnese. Hierbij moet er rekening mee worden gehouden dat een positieve familieanamnese voor verschillende doeleinden kan worden gebruikt. De eerste toepassing van de familieanamnese is als een van de klinische criteria in de ASAS-classificatiecriteria. Om de waarde voor dit doel te bepalen, moet de oorspronkelijke definitie worden vergeleken met een herziene definitie (d.w.z. een positieve familieanamnese die alleen axSpA in een eerste- of tweedegraads familielid omvat). Daarnaast moet worden onderzocht of het gewicht geschikt is en of het een onafhankelijk SpA-kenmerk moet blijven naast HLA-B27 dragerschap.

Ten tweede kan familieanamnese worden gebruikt als een indicatie voor HLA-B27-positiviteit. Dit is met name relevant in situaties waarin HLA-B27-testen niet nuttig zijn (bijv. in de huisartsenpraktijk waar de axSpA-prevalentie laag is) of niet mogelijk (bijv. hoge kosten in landen met minder financiële middelen). Zoals getoond in **hoofdstuk 4** van dit proefschrift was er in het Midden-Oosten en Noord-Afrika een hoger percentage patiënten voor wie HLA-B27 niet beschikbaar was in vergelijking met de andere geografische regio's.

Bovendien komt HLA-B27 minder vaak voor in het Midden-Oosten en Noord-Afrika, wat aangeeft dat in deze regio informatie over familieanamnese bijzonder waardevol kan zijn. Ten slotte kan de familieanamnese worden gebruikt als een risicofactor voor de ontwikkeling van axSpA. Eerder onderzoek heeft aangetoond dat zodra de HLA-B27-status bekend is, een positieve familieanamnese niet bijdraagt aan de kans op een axSpA-diagnose. Daarom lijkt het gebruik ervan als een risicofactor voor de ontwikkeling van axSpA beperkt tot situaties waarin HLA-B27 niet beschikbaar is. De voorspellende waarde van de verschillende SpA-entiteiten in een familieanamnese voor het risico op het ontwikkelen van axSpA is in het verleden niet beoordeeld. Echter, gebaseerd op het feit dat een relatie met HLA-B27 dragerschap alleen werd gevonden voor axSpA in de PerSpA- en ASAS-cohorten en voor axSpA en uveitis in de DESIR- en SPACE-cohorten, is het zeer waarschijnlijk dat de definitie van een positieve familieanamnese ook voor dit doel opnieuw gedefinieerd dient te worden. Opmerkelijk is het feit dat een verband tussen HLA-B27 en uveitis slechts in twee Europese cohorten werd gevonden, wat nogmaals benadrukt hoe belangrijk het is om patiënten van over de hele wereld te includeren wanneer een beslissing wordt genomen over een gewijzigde definitie, die van toepassing zou moeten zijn op alle patiënten wereldwijd.

Aangezien classificatiecriteria worden gebruikt bij de selectie van patiënten voor klinische onderzoeken, moet hun belangrijkste doel blijven om een homogene groep patiënten te creëren. Echter, er moet rekening worden gehouden met de aard van de ziekte waarvoor ze worden gebruikt.

AxSpA is een ziekte met een grote variatie in symptoomexpressie, daarom kunnen de classificatiecriteria niet te restrictief zijn, omdat de patiënten met 'minder typische' symptomen dan niet geclassificeerd zullen worden. De CLASSIC studie biedt een unieke kans om verschillende combinaties van criteria te beoordelen.

Concluderend, in het eerste deel van dit proefschrift is het belang benadrukt van het opnemen van patiënten uit verschillende landen met verschillende etnische achtergronden. Toekomstig onderzoek moet zich richten op het verder verbeteren van de ASAS-classificatiecriteria -en de componenten ervan- door patiënten van over de hele wereld te includeren.

ASAS/OMERACT STANDAARD SET UITKOMSTMATEN VOOR AXIALE SPONDYLOARTRITIS

In het eerste deel van dit proefschrift werd het belang van wereldwijde toepasbaarheid van classificatiecriteria besproken. Deze zorgen ervoor dat over de hele wereld dezelfde patiënten worden geselecteerd voor deelname aan klinische onderzoeken, waardoor directe vergelijkingen mogelijk zijn tussen onderzoeken die zijn uitgevoerd in verschillende landen. Op eenzelfde manier maken gestandaardiseerde beoordeling en rapportage van resultaten directe vergelijkingen mogelijk tussen onderzoeken die verschillende therapieën onderzoeken, waarover in het tweede deel van dit proefschrift wordt gediscussieerd.

Een standaard set uitkomstmaten beschrijft de minimale en verplichte set uitkomstmaten die moeten worden beoordeeld en gerapporteerd in alle klinische onderzoeken van een specifiek ziektebeeld en/of patiëntpopulatie. Een standaard set uitkomstmaten bestaat uit domeinen ("wat te meten") en instrumenten ("hoe te meten"). De standaard set die momenteel in axSpA wordt gebruikt, is de ASAS-OMERACT standaard set uitkomstmaten voor ankyloserende spondylitis (AS). Sinds de ontwikkeling van de oorspronkelijke standaard set meer dan twee decennia geleden, is het duidelijk geworden dat axSpA in feite een ziektespectrum is dat uit twee subtypes bestaat: radiografische axSpA en niet-radiografische axSpA. Daarnaast hebben er grote ontwikkelingen plaatsgevonden in de meetinstrumenten die gebruikt worden in axSpA, zoals het gebruik van MRI, de ontwikkeling van de Ankylosing Spondylitis Disease Activity Score (ASDAS)-voor het meten van ziekteactiviteit-, en de ASAS Health Index-voor het meten van kwaliteit van leven en algehele gezondheid-. Ten slotte is er vooruitgang geboekt in de methodologie rond de ontwikkeling van standaard sets uitkomstmaten, waardoor ASAS besloot dat het tijd was om de standaard set te herzien.

De eerste stap naar een vernieuwde standaard set uitkomstmaten was om te beoordelen of de domeinen die in de oorspronkelijke standaard set vertegenwoordigd waren nog steeds relevant zijn. Om informatie te verzamelen over het belang van de uitkomstmaten van patiënten met axSpA en experts op het gebied van axSpA, werd een Delphi-enquête uitgevoerd. Uit de resultaten van deze Delphi-enquête -beschreven in **hoofdstuk 5** van dit proefschrift- leerden we dat patiënten met axSpA een andere mening hadden over de uitkomsten die in alle onderzoeken naar axSpA beoordeeld moeten worden dan de experts die betrokken zijn bij hun behandeling. Patiënten gaven de voorkeur aan een allesomvattende benadering, terwijl experts een onderscheid bleken te maken tussen uitkomsten die van cruciaal belang zijn voor verschillende typen behandelingen. Volgens de experts waren meer objectief meetbare domeinen, zoals structurele schade en mobiliteit, het meest cruciaal om te meten in onderzoeken waarin anti-reumatische behandelingen worden onderzocht. Dit zijn behandelingen waarbij men verwacht dat het beloop van de

ziekte fundamenteel beïnvloed wordt, d.w.z. meer dan alleen bestrijden van de klachten. De beoordeling van de meer subjectieve domeinen zoals pijn, stijfheid en algemeen functioneren en gezondheid werd vooral belangrijk bevonden voor onderzoeken die zich richten op symptoom bestrijding. Deze gegevens stelden ons in staat om een korte lijst samen te stellen met de belangrijkste domeinen om te beoordelen in alle studies naar axSpA. Deze gegevens vormden de basis voor de ontwikkeling van de vernieuwde standaard set uitkomstmaten voor axSpA.

De Delphi-enquête is een veelgebruikte methode die wordt ingezet om meningen van een grote groep deelnemers te verzamelen, ofwel om een prioriteitenlijst te maken omtrent onderzoeksonderwerpen, om itemlijsten in te korten, of voor het verzamelen van feedback. Voor een methode die zo vaak wordt gebruikt, is informatie over de methodologie opvallend schaars. In **hoofdstuk 6** wilden we inzicht geven in het effect van het kiezen van een bepaalde uitnodigingstechniek op de uitkomst van de Delphi, door twee veelgebruikte uitnodigingsbenaderingen te vergelijken: 1) Alle deelnemers uitnodigen voor volgende rondes, ongeacht of zij hebben geantwoord op de vorige ronde; of 2) Alleen de deelnemers uitnodigen voor volgende rondes die de vorige ronde hebben voltooid. We ontdekten dat er geen effect is op de uiteindelijke uitkomst van de Delphi. Wel stelden we dat het de voorkeur kan hebben om deelnemers die een ronde hebben gemist toch uit te nodigen voor volgende rondes, omdat deze benadering minder gevoelig is voor het niet-willekeurige verlies van meningen dat zou kunnen leiden tot valse overeenstemming. Bovendien zorgt deze aanpak ervoor dat het eindresultaat de mening weergeeft van iedereen die was uitgenodigd om deel te nemen.

De Delphi-enquête was een klein onderdeel van een veel grotere project om de standaard set uitkomstmaten voor AS te vernieuwen, zodat deze toepasbaar is op alle patiënten met axSpA. In **hoofdstuk 7** beschreven we het proces dat leidde tot de ASAS-OMERACT core outcome set voor axSpA. De domeinen van de vernieuwde standaard set uitkomstmaten voor axSpA zijn vergelijkbaar met de oorspronkelijke standaard set voor AS. Beiden bevatten de domeinen fysiek functioneren, ochtendstijfheid, pijn, vermoeidheid en ziekteactiviteit in hun kern. Daarnaast is het meten van structurele schade alleen verplicht voor studies die anti-reumatische therapieën onderzoeken, waarbij men verwacht dat het beloop van de ziekte fundamenteel beïnvloed wordt (meer dan alleen klachten). Het eerste opvallende verschil is de toevoeging van algemeen functioneren en gezondheid in de kern van de vernieuwde standaard set uitkomstmaten voor axSpA. Hiermee wordt de impact van axSpA op andere aspecten van het leven gemeten, wat de afgelopen jaren meer aandacht heeft gekregen. Ten tweede is het opmerkelijk dat mobiliteit van de wervelkolom niet langer deel uitmaakt van de kern in de standaard set voor axSpA. Dit is te verklaren door het gebrek aan standaardisatie, onvermogen verandering in mobiliteit van de wervelkolom goed weer te geven en de slechte test-hertest betrouwbaarheid. Test-

hertest betrouwbaarheid betekent het vermogen van een test om hetzelfde resultaat te geven als de test wordt herhaald op een ander moment, en de overige omstandigheden gelijk zijn gebleven. Nu de domeinen zijn vastgesteld, moeten geschikte meetinstrumenten worden geselecteerd om deze domeinen te meten.

De veranderingen ten opzichte van de oorspronkelijke standaard set voor AS zullen implicaties hebben voor toekomstig onderzoek, omdat er net andere domeinen beoordeeld moeten worden dan eerder is gedaan. Helaas vermindert dit voor sommige aspecten van ziekte de vergelijkbaarheid met oudere onderzoeken (bijv. mobiliteit van de wervelkolom), maar tegelijkertijd zorgt het voor een betere vergelijkbaarheid in toekomstige onderzoeken omdat er meer duidelijkheid is over welke uitkomstmaten moeten worden gemeten en gerapporteerd. Bovendien zijn alle belanghebbenden die baat hebben bij een vernieuwde standaard set uitkomstmaten betrokken bij de ontwikkeling ervan, wat de acceptatie zal vergroten. De volgende stap is om voor elk van de geselecteerde domeinen de beste meetinstrumenten te selecteren. Hierbij is een zorgvuldige afweging van de meeteigenschappen, haalbaarheid en bruikbaarheid van de meetinstrumenten van belang.

Een van de meeteigenschappen van de meetinstrumenten is de test-hertest betrouwbaarheid. In **hoofdstuk 8** wordt de test-hertest betrouwbaarheid van de meetinstrumenten die zijn gebruikt in drie recente klinische studies in axSpA onderzocht. Uit deze studie hebben we geconcludeerd dat hoewel de meeste meetinstrumenten zijn ontwikkeld voor radiografische axSpA, ze ook betrouwbaar zijn bevonden voor niet-radiografische axSpA. Bovendien liet deze studie zien dat meetinstrumenten met meerdere items robuuster bleken te zijn tegen meetfouten vergeleken met meetinstrumenten die slechts uit één item bestaan. Verder onderzoek zal de andere meeteigenschappen moeten onderzoeken voordat een definitief besluit kan worden genomen over welke instrumenten het beste geschikt zijn om de geselecteerde domeinen te meten.

Verdere discussie en toekomstig onderzoek

De standaard set uitkomstmaten voor AS die momenteel wordt gebruikt om te bepalen welke data minimaal moeten worden verzameld in elke klinische studie in axSpA, bevatte niet direct na ontwikkeling een instrument voor elk domein. Er waren niet direct specifieke instrumenten gedefinieerd voor de beoordeling van vermoeidheid of enthesitis (ontstekingen van peesaanhechtingen), omdat er op het moment van ontwikkeling van de standaard set uitkomstmaten geen gevalideerd instrument beschikbaar was. Als gevolg hiervan zijn verschillende instrumenten gebruikt om deze domeinen te beoordelen, waardoor vergelijkingen tussen onderzoeken worden belemmerd, terwijl die erg belangrijk zijn voor het beoordelen van de effectiviteit van de behandeling. Daarom is het van groot belang dat de vernieuwde standaard set één specifiek instrument voor elk

domein adviseert (met de mogelijkheid om er meer toe te voegen). Deze instrumenten dienen niet alleen te worden gemeten, maar, belangrijker nog, in elke studie moet de resultaten op gestandaardiseerde wijze worden gerapporteerd, wat één-op-één vergelijking van onderzoeken en de ontwikkeling van behandelaanbevelingen mogelijk maakt. De oorspronkelijke standaard set wordt veelvuldig gebruikt, wat aangeeft dat de beschikbaarheid van een standaard set leidt tot een gestructureerde verzameling van informatie in klinische onderzoeken. Er was echter nogal wat variatie in de instrumenten die werden gebruikt om de informatie te verzamelen, wat het belang benadrukt om voor elk domein één specifiek instrument aan te bevelen. Bovendien bleek dat niet alle verzamelde informatie ook werd gerapporteerd. Zo bevat BASDAI een maat voor vermoeidheid, maar deze werd vaak niet specifiek gerapporteerd en daarom kon op basis van de gepresenteerde gegevens geen conclusie worden getrokken over het effect van de onderzochte therapie op vermoeidheid.

Het belangrijkste doel van een standaard set uitkomstmaten is het beschrijven van de domeinen en instrumenten die ten minste in elke studie moeten worden beoordeeld. Naarmate ons begrip van een ziekte toeneemt, of het ziektebeloop verandert door eerdere herkenning en effectieve therapie, kan dit leiden tot de ontwikkeling van nieuwe instrumenten (bijvoorbeeld de ASAS Health Index) of validatie van bestaande instrumenten. Er is een redelijke kans dat deze nieuwe instrumenten beter presteren dan bestaande instrumenten, en daarmee de voorkeur krijgen om een bepaald domein te meten, wat het vernieuwen van een standaard set uitkomstmaten vereist. Helaas is het vernieuwen van een standaard set een langdurig en tijdrovend proces, en je kunt je afvragen of het proces zijn doel overstijgt en of er meer standaard sets (d.w.z. meer gestandaardiseerde metingen) zouden zijn als het proces gebruiksvriendelijker zou zijn. Een mogelijkheid om het proces te vereenvoudigen is het regelmatig herzien van een standaard set (bijv. om de 10 jaar), waarbij kan worden besloten een instrument te vervangen als is aangetoond dat het nieuwe instrument beter presteert dan het bestaande, zonder dat daarvoor alle stappen die nodig zijn voor de ontwikkeling van een nieuwe standaard set hoeven worden doorlopen. Om dit te doen, zou een volledige vergelijking van alle meeteigenschappen tussen het nieuwe en bestaande instrument een eerste vereiste zijn. Echter, het regelmatig veranderen van een standaard uitkomst set heeft het nadeel dat nieuwer en ouder onderzoek niet langer één-op-één te vergelijken is.

Ondanks de tekortkomingen zijn standaard sets waardevol in onderzoek, omdat ze zorgen voor meer transparantie in het geneesmiddelenregistratieproces door directe vergelijkingen met eerder geregistreerde geneesmiddelen. Ook kunnen zij bijdragen aan een betere acceptatie van nieuwe behandelingen in het veld, omdat een directe vergelijking van uitkomstmaten de prestaties van het nieuwe geneesmiddel laten zien in relatie tot eerder geaccepteerde/meer bekende medicijnen. Het verleden heeft ons

geleerd dat het belangrijk is om per domein concreet te definiëren welke instrumenten moeten worden gebruikt en welke optioneel kunnen worden toegevoegd. Bovendien garandeert het gebruik van een instrument niet dat alle verzamelde gegevens ook worden gerapporteerd. Daarom moet de standaard set uitkomstmaten voor axSpA niet alleen de domeinen en instrumenten bieden die in elke studie moeten worden beoordeeld, maar ook specifieke instructies voor het rapporteren van verzamelde data.

Concluderend, standaard sets uitkomstmaten zijn een waardevol hulpmiddel bij het beoordelen van resultaten, maar toekomstig onderzoek zou moeten bekijken of hun ontwikkelingsproces kan worden verbeterd.

PATIËNT GERAPPORTEERDE UITKOMSTMATEN IN VROEGE AXIALE SPONDYLOARTRITIS

In het laatste deel van dit proefschrift bespraken we gezondheids-gerelateerde kwaliteit van leven en beperkingen in werkproductiviteit in axSpA. In **hoofdstuk 9** lieten we zien dat werkproductiviteit evenals gezondheids-gerelateerde kwaliteit van leven verbeterden gedurende twee jaar follow-up bij patiënten met chronische rugpijn verdacht van axSpA. Deze verbetering werd aangetoond voor patiënten met en zonder een diagnose van axSpA, maar patiënten met axSpA vertoonden een grotere verbetering in vergelijking met rugpijnpatiënten zonder axSpA.

De resultaten uit **hoofdstuk 9** hebben ons geleerd dat we in de loop van de tijd enige verbetering kunnen verwachten bij alle chronische rugpijnpatiënten die verdacht worden van axSpA, ongeacht de diagnose. Dit kan naast de behandeling te maken hebben met het feit dat patiënten bij hun eerste bezoek aan de reumatoloog de meest ernstige klachten ervaren. De klachten kunnen in de loop van de tijd op natuurlijke wijze verbeteren (regressie naar het gemiddelde). Bovendien konden we concluderen dat een diagnose van axSpA een onafhankelijke voorspeller is van verbetering in gezondheids-gerelateerde kwaliteit van leven en verbetering in werkproductiviteit bij patiënten met chronische rugpijnklachten. De verschillen in uitkomsten tussen degenen met en zonder een diagnose van axSpA bleven bestaan wanneer de analyses beperkt werden tot patiënten die geen anti-reumatische medicatie kregen, wat aangeeft dat behandeling met deze medicatie de verschillen tussen de groepen niet verklaarde. Belangrijk is dat, ondanks de verbeteringen in de tijd, de kwaliteit van leven en werkproductiviteit nog steeds slechter waren in vergelijking met de algemene bevolking. Deze resultaten benadrukken het belang van het optimaliseren van de gezondheids-gerelateerde kwaliteit van leven en de sociale participatie van patiënten met axSpA op de lange termijn. Dit wordt ook beschreven in de huidige axSpA-behandelingsrichtlijnen als een van de belangrijkste behandeldoelen.

Verdere discussie en toekomstig onderzoek

In **hoofdstuk 9** hebben we gevonden dat patiënten met de diagnose axSpA een grotere verbetering in werkproductiviteit en gezondheids-gerelateerde kwaliteit van leven vertoonden vergeleken met diegenen zonder een diagnose van axSpA. We vermeldden dat dit mogelijk te verklaren is door een verschil in beschikbare behandelingsopties, maar dit kan ook worden verklaard door een verschil in ziektepercepties en daaropvolgende coping-mechanismen (d.w.z. de manier waarop men met de ziekte omgaat). Een van de vragen die we ons stelden is of de diagnose een invloed kan hebben op hoe een patiënt zijn of haar ziekte ervaart en dit vervolgens de coping-mechanismen beïnvloedt. Toekomstig onderzoek zou zich moeten richten op het verkrijgen van meer inzicht in de psychologische effecten van het stellen van een diagnose, of alleen weten 'wat er met je aan de hand is' invloed heeft op hoe klachten worden ervaren. Verder zou moeten worden onderzocht of de kwaliteit van leven nog verder kan verbeteren als patiënten geïnformeerd worden over mogelijke coping-mechanismen. Tegelijkertijd kunnen adequate coping-mechanismen een invloed hebben op de werkproductiviteit. Aangezien axSpA het leven beïnvloedt van jonge mensen, is er een enorme waarde in het uitbreiden van ons begrip van effectieve therapieën -hetzij medicamenteus, educatief of psychologisch- die een positief effect hebben op de kwaliteit van leven en werkproductiviteit.

Een ander aspect dat aandacht verdient, is het gebruik van algemene versus ziekte-specifieke vragenlijsten om kwaliteit van leven te beoordelen. Het belangrijkste voordeel van het gebruik van een algemene vragenlijst is dat de scores kunnen worden vergeleken met scores van patiënten met andere (chronische) ziekten of de algehele bevolking, wat de maatschappelijke waarde vergroot. Daarentegen hebben ziekte-specifieke vragenlijsten (zoals de ASAS Health Index) betrekking op meer ziekte-specifieke vragen, die een duidelijker inzicht geven in het effect van ziekte op de kwaliteit van leven van patiënten met axSpA, waardoor ze wellicht van hogere wetenschappelijke waarde zijn. In **hoofdstuk 9** maakte het gebruik van de SF-36 (een vragenlijst die wordt gebruikt om de kwaliteit van leven in kaart te brengen) een directe vergelijking mogelijk van patiënten met en zonder een diagnose van axSpA en er kon ook een vergelijking met de algemene bevolking worden gemaakt. Directe vergelijking van degenen die wel en geen diagnose van axSpA kregen, toonde ons dat er een verschil in verbetering in de tijd is tussen deze groepen, een feit dat onopgemerkt zou zijn gebleven als we een ziekte-specifieke vragenlijst hadden gebruikt. Desalniettemin had het gebruik van een ziekte-specifieke vragenlijst kunnen leiden tot inzicht in welke aspecten (indien aanwezig) hebben bijgedragen aan de verbetering van de kwaliteit van leven over tijd. Daarom kan een ziekte-specifieke vragenlijst de voorkeur hebben in het onderzoeken van patiënten met een chronische ziekte over de tijd. Aangezien het zeer waarschijnlijk is dat de gezondheids-gerelateerde kwaliteit van leven bij deze patiënten verminderd zal blijven in vergelijking met de algemene bevolking, zou inzicht in aspecten die bijdragen aan verbetering of verslechtering van de kwaliteit van leven in

de loop van de tijd waardevoller kunnen zijn, omdat dit tot nieuwe behandeldoelen zou kunnen leiden. Uiteindelijk zal onderzoek altijd vragen om het maken van keuzes en het vinden van de optimale balans tussen kosten en beloning, dus wat 'de juiste keuze' is, hangt af van de onderzoeksvraag, de beschikbare gegevens en de beschikbare middelen.

APPENDICES

CURRICULUM VITAE
LIST OF PUBLICATIONS
DANKWOORD

CURRICULUM VITAE

Anne Boel is geboren op 27 juli 1992 in Terheijden. In 2010 haalde ze haar gymnasium diploma op het Stedelijk Gymnasium te Breda, waarna ze Bewegingswetenschappen ging studeren aan de Vrije Universiteit (VU) te Amsterdam. Door haar interesse in de psychologische achtergrond van bewegen en de mens in zijn algemeenheid, besloot zij extra vakken in deze richting te volgen, zowel aan de VU als aan de Universiteit van Utrecht. In 2014 rondde ze de bachelor Bewegingswetenschappen af. In datzelfde jaar besloot ze naar Engeland te verhuizen om daar de research master Exercise and Sport Sciences te starten aan de University of Birmingham. Gedurende de research master groeide haar liefde voor onderzoek en na het behalen van de master titel in 2015 ging ze op zoek naar een promotieplek. Echter, er kwam een andere werkplek tussendoor: Kinetic Analysis, een start-up bedrijf wat zich o.a. bezig hield met contracted research. Ondanks dat hier ook onderzoek gedaan werd, bleef het idee van promotieonderzoek op de achtergrond aanwezig en in april 2018 is zij dan ook gestart met haar promotietraject in het Leids Universiteit Medisch Centrum (LUMC). Op de afdeling reumatologie in het LUMC voegde zij zich bij het onderzoeksteam van prof. Dr. D. van der Heijde en dr. FA. van Gaalen. Gedurende haar promotie werd zij ook begeleid door dr. V. Navarro-Compán. Momenteel werkt Anne als Medical Science Liaison reumatologie bij UCB Pharma.

LIST OF PUBLICATIONS

1. **Boel A**, van Lunteren M, López-Medina C, Sieper J, van der Heijde D, van Gaalen FA. Geographical prevalence of family history in patients with axial spondyloarthritis and its association with HLA-B27 in the ASAS-PerSpA study. *RMD Open.* 2022 Mar;8(1):e002174.
2. **Boel A**, Navarro-Compán V, van der Heijde D. Test-retest reliability of outcome measures: data from three trials in radiographic and non-radiographic axial spondyloarthritis. *RMD Open*. 2021 Dec;7(3):e001839.
3. **Boel A**, Navarro-Compán V, Boonen A, Mease P, Kiltz U, Dougados M, Landewé R, van der Heijde D. Domains to be considered for the core outcome set of axial spondyloarthritis: results from a 3-round Delphi survey. *J Rheumatol.* 2021 Dec;48(12):1810-1814.
4. **Boel A**, López-Medina C, van der Heijde DMFM, van Gaalen FA. Age at onset in axial spondyloarthritis around the world: data from the ASAS-PerSpA study. *Rheumatology (Oxford)*. 2022 Apr;61(4):1468-1475.
5. **Boel A**, Navarro-Compán V, Landewé R, van der Heijde D. Two different invitation approaches for consecutive rounds of a Delphi survey led to comparable final outcome. *J Clin Epidemiol.* 2021 Jan;129:31-39
6. **Boel A**, Molto A, van der Heijde D, Ciurea A, Dougados M, Gensler LS, Santos MJ, De Miguel E, Poddubnyy D, Rudwaleit M, van Tubergen A, van Gaalen FA, Ramiro S. Do patients with axial spondyloarthritis with radiographic sacroiliitis fulfil both the modified New York criteria and the ASAS axial spondyloarthritis criteria? Results from eight cohorts. *Ann Rheum Dis.* 2019 Nov;78(11):1545-1549
7. Navarro-Compán V, **Boel A**, Boonen A, Mease PJ, Dougados M, Kiltz U, Landewé RBM, Baraliakos X, Bautista-Molano W, Chiowchanwisawakit P, Dagfinrud H, Fallon L, Garrido-Cumbrera M, Gensler L, ElZorkany BK, Haroon N, Kwan YH, Machado PM, Maksymowych W, Molto A, de Peyrecave N, Poddubnyy D, Protopopov M, Ramiro S, Song IH, van Weely S, van der Heijde D. Instrument selection for the ASAS core outcome set for axial spondyloarthritis. *Ann Rheum Dis.* 2022 Jun 9:annrheumdis-2022-222747.
8. Navarro-Compán V, **Boel A**, Boonen A, Mease P, Landewé R, Kiltz U, Dougados M, Baraliakos X, Bautista-Molano W, Carlier H, Chiowchanwisawakit P, Dagfinrud H, de Peyrecave N, El-Zorkany B, Fallon L, Gaffney K, Garrido-Cumbrera M, Gensler LS, Haroon N, Kwan YH, Machado PM, Maksymowych WP, Poddubnyy D, Protopopov M, Ramiro S, Shea B, Song IH, van Weely S, van der Heijde D. The ASAS-OMERACT core domain set for axial spondyloarthritis. *Semin Arthritis Rheum*. 2021 Dec;51(6):1342-1349.
9. Toupin-April K, Décary S, de Wit M, Meara A, Barton JL, Fraenkel L, Li LC, Brooks P, Shea B, Stacey D, Légaré F, Lydiatt A, Hofstetter C, Proulx L, Christensen R, Voshaar M, Suarez-Almazor ME, Boonen A, Meade T, March L, Jull JE, Campbell W, Alten R, Morgan

EM, Kelly A, Kaufman J, Hill S, Maxwell LJ, Guillemin F, Beaton D, El-Miedany Y, Mittoo S, Westrich Robertson T, Bartlett SJ, Singh JA, Mannion M, Nasef SI, de Souza S, **Boel A**, Adebajo A, Arnaud L, Gill TK, Moholt E, Burt J, Jayatilleke A, Hmamouchi I, Carrott D, Blanco FJ, Mather K, Maharaj A, Sharma S, Caso F, Fong C, Fernandez AP, Mackie S, Nikiphorou E, Jones A, Greer-Smith R, Sloan VS, Akpabio A, Strand V, Umaefulam V, Monti S, Melburn C, Abaza N, Schultz K, Stones S, Kiwalkar S, Srinivasalu H, Constien D, King LK, Tugwell P. Endorsement of the OMERACT core domain set for shared decision making interventions in rheumatology trials: Results from a multi-stepped consensus-building approach. *Semin Arthritis Rheum*. 2021 Jun;51(3):593-600.

DANKWOORD

Velen hebben bijgedragen aan de totstandkoming van dit proefschrift. In het bijzonder dank ik graag de volgende betrokkenen.

Dank aan de patiënten die hebben deelgenomen aan de ASAS, ASAS-PerSpA, Esperanza, GESPIC, OASIS, Reuma.pt, SCQM, SPACE en UCSF axSpA cohorten. En aan alle artsen en onderzoekers die hebben bijgedragen aan het verzamelen van de data, zonder welke dit proefschrift niet had kunnen ontstaan.

A special word of thanks is reserved for my promotor prof. dr. van der Heijde and co-promotores dr. van Gaalen en dr. Navarro-Compán, thank you all for your knowledge, inspiration and support. You have made me into the researcher I am today. Beste prof. van der Heijde, beste Désirée, jouw liefde en passie voor de wetenschap zijn besmettelijk en halen het beste in je promovendi naar boven, often minste in mij. Dank ook voor je kritische vragen en stimulans om altijd dat stapje extra te doen. Dank zij jou heeft een volledige roodgekleurd document een positieve betekenis gekregen, daar je zei: “Als hetgeen wat er staat niet goed is, kan ik ook geen suggesties voor verbetering of verdere verdieping geven”. Beste dr. van Gaalen, beste Floris, dank voor je geduld, je oneindige vertrouwen en geruststellende woorden. Jij stimuleerde mij altijd de link te maken naar de klinische praktijk, en daagde me uit om vanuit een ander perspectief naar de resultaten te kijken. Dear dr. Navarro-Compán, dear Victoria, your work-ethos and warm personality are something to aspire, and I can only hope to have picked up some of that from our time working together. You have shown me there is always a silver lining, and always managed to simultaneously challenge and support me.

Prof. dr. Huizinga, beste Tom, wat heb ik me enorm welkom gevoeld op jouw afdeling, welke bruist van de wetenschappelijke kennis en ambitie. Dank voor de jaren dat ik hier deel van heb mogen uitmaken, en alles wat ik hier heb kunnen leren.

To all my co-authors, it has been a true privilege to work with you. I greatly appreciate all the feedback and insights you have provided over the years and the high quality articles we were able to publish because of it.

Dear SPACE-family, it was amazing to have been part of a group of fantastic physicians and researchers. Thank you for your continuous efforts to collect data, and collaborate on research projects.
Miranda, Roos en Queeny, ik had me geen betere SPACE-collega’s kunnen wensen. Dank voor de ‘klaag-uurtjes’ en alle leuke momenten die we hebben mogen delen. Dr. Ramiro, beste Sofia, dank voor alle goede adviezen en kritische blik. Het lijkt alsof jij

in wetenschappelijke discussies altijd 3 stappen verder vooruit denkt, en je vermogen om 'advocaat van de duivel' te spelen waardeer ik enorm, daar deze altijd zal leiden tot het best mogelijke resultaat.

Aan alle promovendi en postdocs die ik tijdens mijn traject heb mogen leren kennen, dank voor alle tips, suggesties, elk luisterend oor, elke kritische vraag, elke lach en alle gezelligheid.

Lieve paranimfen, Miranda en Sytske Anne, dank dat jullie mij willen ondersteunen in de voorbereidingen en de verdediging van mijn proefschrift.

Nancy, Sandra, Joyce, Jozé en Cedric, dank voor alle ondersteuning in praktische zaken en voor het feit dat jullie deur altijd open stond voor een praatje, waarna ik weer met nieuwe energie aan de slag kon.

Poli secretaresses, en Marie-José in het bijzonder, dank voor alle ondersteuning bij het plannen van patiëntvisites, jullie flexibiliteit en de gezellige wachtpraatjes.

Lieve familie en vrienden dank voor alle interesse en steun, een welkome afleiding als het even tegen zat.

Lieve Ernest en Mariëlle, pap en mam, dank voor jullie oneindige steun en onvoorwaardelijke liefde. Jullie hebben me altijd gestimuleerd om het beste uit mijzelf te halen, en om een toekomst achterna te jagen die bij mij past. Dankzij jullie steun en de kansen die jullie mij hebben geboden heb ik mij kunnen ontwikkelen tot de persoon, en de onderzoeker, die ik vandaag ben.

Tot slot, lieve Marco, ondanks dat je de helft van de tijd niet begreep waar ik het over had, mocht ik altijd mijn hart bij je luchten. Jouw steun en liefde maakte de mindere momenten beter en de leuke momenten een waar feestje.