Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/36591 holds various files of this Leiden University dissertation.

Author: Navarro Compan, Maria Victoria
Title: Aspects of methodology in assessing inflammation and damage in rheumatoid arthritis and axial spondyloarthritis
Issue Date: 2015-12-03

Aspects of Methodology in Assessing Inflammation and Damage in Rheumatoid Arthritis and Axial Spondyloarthritis

Victoria Navarro Compán

2015

ISBN: 978-94-6233-091-7

Copyright © 2015 by Victoria Navarro Compán.

All rights reserved. No part of this book may be reproduced in any form without written permission of the author or, when appropriate, of the publishers of the publications.

Victoria Navarro Compán was supported by the Spanish Society of Rheumatology (SER) and the Assessment of SpondyloArthritis international Society (ASAS).

Printing of thesis was financially supported by the Association of Rheumatology of University Hospital La Paz (AREPAZ), the Dutch Arthritis Foundation, AbbVie B.V., Pfizer B.V. and UCB Pharma B.V., which is gratefully acknowledged.

Cover design and thesis layout: Midas Mentink (www.midasmentink.nl)
Printing: Gildeprint - Enschede (www.gildeprint.nl)

Aspects of Methodology in Assessing Inflammation and Damage in Rheumatoid Arthritis and Axial Spondyloarthritis

Proefschrift

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,

volgens besluit van het College voor Promoties

te verdedigen op donderdag 3 december 2015

klokke 16:15 uur

door

María Victoria Navarro Compán

geboren te Sevilla (Spanje)

in 1979

PROMOTIECOMMISSIE

Promotores:

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

Prof. dr. R.B.M. Landewé
AMC, Universiteit van Amsterdam

Overige leden:

Prof. dr. T.W.J. Huizinga

Prof. dr. M. Boers
VUmc, Vrije Universiteit van Amsterdam

Prof. dr. J. Sieper
Charité- Universitätsmedizin, Freie Universität zu Berlin

Dr. A.H.M. van der Helm-van Mil

Dr. M. Reijnierse

“Every effort has its rewards”

To Ignacio

CONTENT

Chapter 1	General introduction	9
Chapter 2	Relation between disease activity indices and their individual components and radiographic progression in rheumatoid arthritis: a systematic literature review. Rheumatology (Oxford) 2015;54(6):994-1007	19
Chapter 3	Relationship between types of radiographic damage and disability in patients with rheumatoid arthritis in the EURIDISS cohort: a longitudinal study. Rheumatology (Oxford) 2015;54(1):83-90	49
Chapter 4	Rate of adjudication of radiological progression in rheumatoid arthritis randomized controlled trials depending on preset limits of agreement: a pooled analysis from 15 randomized trials. Rheumatology (Oxford) 2013;52(8):1404-7	65
Chapter 5	Measurement error in the assessment of radiographic progression in rheumatoid arthritis clinical trials: the smallest detectable change revisited. Ann Rheum Dis 2014;73(6):1067-70	75
Chapter 6	Spondyloarthritis features forecasting the presence of HLA-B27 or sacroiliitis on magnetic resonance imaging in patients with suspected axial spondyloarthritis: results from a cross-sectional study in the ESPeranza cohort. Arthritis Res Ther (accepted for publication)	89
Chapter 7	Value of high-sensitivity C-reactive protein for classification of early axial spondyloarthritis: results from the DESIR cohort. Ann Rheum Dis 2013;72(5):785-6	103
Chapter 8	Calculating the ankylosing spondylitis disease activity score if the conventional C-reactive protein level is below the limit of detection or if high-sensitivity C-reactive protein is used: an analysis in the DESIR cohort. Arthritis Rheumatol 2015;67(2):408-13	109

Chapter 9	Disease activity is longitudinally related to sacroiliac inflammation on MRI in male patients with axial spondyloarthritis: 2-year of the DESIR cohort. Ann Rheum Dis (accepted for publication)	125
Chapter 10	Summary and conclusions	139
Chapter 11	Samenvatting en conclusies	151
	List of publications	161
	Curriculum Vitae	165
	Acknowledgements	166

1

General Introduction

INTRODUCTION

Rheumatoid arthritis (RA) and axial spondyloarthritis (axSpA) are both chronic systemic inflammatory diseases that may affect many tissues and organs, but the hallmark is pain and inflammation of the joints [1]. Over time, and in particular if this inflammation persists, it frequently leads to destruction of joints and periarticular structures [2]. This structural damage is usually visible on conventional radiographs and, importantly, the amount of radiographic damage is directly associated with the level of functional disability and quality of life [3,4]. Consequently, structural damage is nowadays one of the major long-term outcomes in RA and axSpA and, accordingly, slowing or prevention of structural damage is distinguished as one of the claims for the approval of therapies by the regulatory authorities.

Pain and inflammation of the joints are typically the initial manifestations both for patients with RA and axSpA, but the target joints of these diseases are different. In RA, inflammation occurs mainly in the peripheral joints (particularly in hands and feet) while axSpA predominantly affects the sacroiliac and spinal joints. These differences in localization have consequences both for making a diagnosis or classification of the disease and for monitoring disease activity and damage in clinical practice as well as in research. In patients with RA, it is common that physicians and even patients can detect swelling of peripheral joints at first sight as a sign of inflammation. In patients with axSpA, inflammation is far less obvious, because of the localization of the spinal and sacroiliac joints. Tests that may reflect the presence of inflammation in the sacroiliac joints and the spine such as acute phase reactants (APRs) and magnetic resonance imaging (MRI) are therefore relevant in patients with axSpA. Another obvious difference between RA and axSpA is the nature of structural damage: in RA the process is destructive and includes loss of bone and cartilage, while in axSpA new bone formation is predominant.

Although the relationship between disease activity, radiographic damage and disability is well established in patients with RA, some fundamental aspects of this relationship are not yet fully understood. Also in axSpA, a clear relationship between disease activity, structural damage, functional disability and quality of life has been established [5]. However, it has been proven only recently that there is an unequivocal longitudinal relationship between disease activity and structural damage in axSpA [4].

Against this background, the objectives of this thesis were twofold:

- First (pertaining to rheumatoid arthritis):
    a. To increase the understanding of the relationship between disease activity, radiographic damage and disability.
    b. To provide guidance for assessing of and reporting on radiographic damage.

- Second (pertaining to axial spondyloarthritis):

    c. To optimize the usage of diagnostic tests for clinical practice.
    d. To investigate the yield of additional tests reflecting inflammation in order to classify patients and to monitor disease activity.

These points will be addressed in this thesis in two dedicated parts: Part I focuses on radiographic damage in patients with RA; Part II focuses on the use of supplementary tests in patients with axSpA.

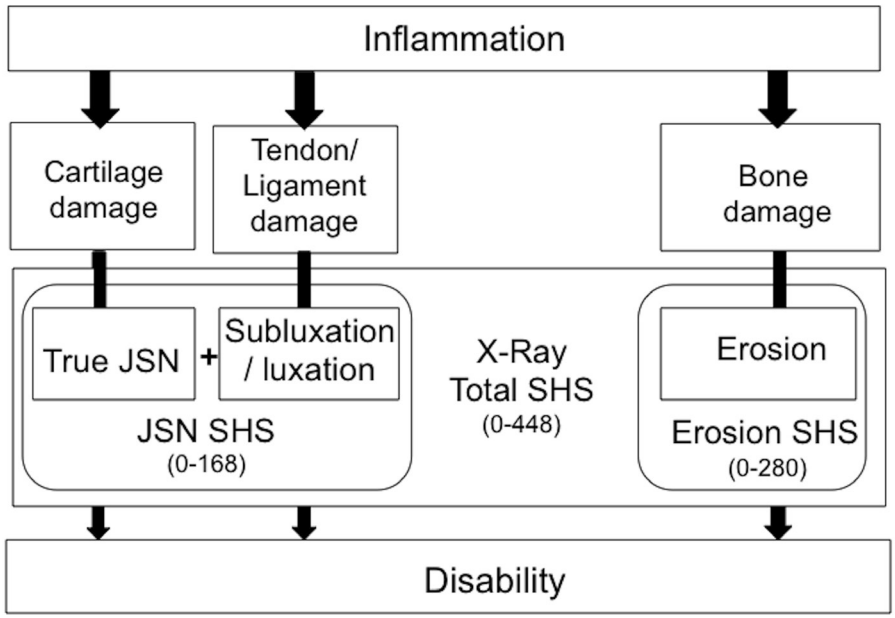
RHEUMATOID ARTHRITIS

Relationship between disease activity measures and radiographic damage

Disease activity in RA usually implies stiffness, swelling and tenderness of peripheral joints, as well as elevated APRs [6]. These manifestations can be measured separately using instruments or scales or in different combinations by composite disease activity indices (DAIs). As part of the treat-to-target recommendations for RA in clinical practice, DAIs are prioritized over separate instruments, since they are more reliable, generelisable and responsive to change [2]. Nevertheless, some experts consider the assessment of only patient-reported outcomes sufficient, and more feasible in a clinical setting because a physical examination or laboratory test is not required [5]. In several studies, it has been shown that some patients may still develop significant radiographic joint damage despite having low disease activity. In these patients, international recommendations ordain to intensify or adjust the treatment, which means a challenge for some tools used to monitor disease activity [7]. Based on this principle, the best tool to monitor disease activity in patients with RA is the tool that best reflects the association between disease activity and radiographic progression. However, the relationship of purely patient-reported measures for disease activity and radiographic progression is yet incompletely understood. By systematic literature review, in **chapter 2**, we investigate the relationship between the different DAIs and their individual components and radiographic progression in patients with RA.

Scoring of radiographic damage

The gold standard to assess structural damage in RA is conventional radiography but the lesions observed depend on the affected anatomical structure [8]; usually, cartilage damage is reflected by joint space narrowing (JSN), damage to tendons or ligaments may present as subluxation or complete luxation -(sub)luxation- and bone damage is visible as erosions (Figure).



Different scoring methods are available to quantify radiographic damage and the most frequently employed one is the Sharp-van der Heijde (SHS) method [9], which is also the recommended method for trials by the European Medicines Agency. This method assesses 32 joints in hands and 12 joints in feet for the presence of erosions (range 0-5) and JSN (range 0-4) separately, but the JSN score accounts for the true JSN and also for the presence of (sub)luxation.

Relationship between types of radiographic damage and disability

Several studies have demonstrated that the degree of radiographic damage is related to functional impairment in patients with RA [1,10]. What is still a matter of debate is whether the three types of structural damage contribute independently and similarly to disability or whether one or more of them mainly explain functional impairment in RA patients. In **chapter 3**, we evaluate if any of the three types of radiographic damage (true JSN, (sub)luxation and erosions) is preferentially related to disability in patients with RA.

Methodological aspects of assessment of radiographic damage

Precision of scoring (i.e. the degree to which the score can be consistently reproduced in different reads or by different readers) is an important psychometric attribute of the SHS method [11]. In clinical trials, the mean of two readers who score all radiographs provides a more precise estimate of the truth than the score of one reader, but inter-reader variability

may still affect the precision and therefore the results of the study. So, if the difference between readers exceeds a certain pre-set threshold, re-reading and/or adjudication by a third reader is usually applied to increase precision. Regulators have expressed concerns if an arbitrary 20% or more of all cases in a given clinical trial are refereed by adjudication. Having said this, it is clear that the choice of the threshold for adjudication (historically between 7 and 15 units) is entirely arbitrary, since no single study has ever provided data that give insight into the number of cases to be adjudicated at different thresholds. In **chapter 4**, we provide data on how a selected threshold for the difference in change scores between two readers results in a specific adjudication rate using the SHS method, and we also evaluate the influence of several factors on these results.

Furthermore, in order to be able to detect differences in radiological progression between treatment arms in clinical trials, reliability (i.e. the degree to which the measurement is free from measurement error) of the SHS method needs to be considered too. Nowadays, the preferred statistic to interpret changes at the individual patient level is the smallest detectable change (SDC) [12,13]. However, in complex databases such as those from clinical trials, several methodological reasons related to the complexity of databases make it challenging to obtain an SDC. In **chapter 5**, we address these issues and propose alternatives to simpler calculate SDC in complex databases.

AXIAL SPONDYLOARTHRITIS

Assessment of objective inflammation for the diagnosis

Magnetic resonance imaging of sacroiliac joints (MRI-SI) has an important role in the early diagnosis of axSpA, especially because of its ability to detect inflammatory lesions early in the course of the disease, when structural damage cannot yet be found on conventional radiographs [14,15]. Both MRI-SI and a test for the human leucocyte antigen B27 (HLA-B27) have a relatively high sensitivity, and especially a high specificity for axSpA [14]. This is why imaging (radiographs or MRI-SI) and HLA-B27 are entry criteria in the Assessment of SpondyloArthritis international Society (ASAS) classification algorithm for axSpA [16]. On the other hand, these supplementary tests are costly and time-consuming and consequently cannot be applied in all patients with chronic back pain (CBP) because CBP is one of the most prevalent symptoms in the general population and only a minority of patients with CBP ultimately has axSpA. It would therefore be very helpful for rheumatologists to identify which patients referred with CBP have highest likelihood of a positive MRI-SI or a positive HLA-B27 test. To answer this question, in **chapter 6** we investigate which spondyloarthritis (SpA) features may increase the likelihood of a positive test result for HLA-B27 or MRI-SI in patients with suspected axSpA.

Patient reported information plays an important role for the diagnosis of axSpA. Supplementary tests may serve to obtain more objective information about the level of inflammation. In this sense, the best tests are those that quantify the acute phase response (C-reactive protein –CRP- or erythrocyte sedimentation rate –ESR-) as well as MRI-SI.

APRs, serum biomarkers commonly employed by rheumatologists to trace inflammation, are elevated in only a part of the patients with axSpA whilst they may also be elevated in other conditions [17]. Elevated CRP, not explained by other reasons, in the context of a young patient with CBP, has been included as one of the SpA features in the ASAS classification criteria for axSpA [16,18]. Recently, more sensitive tests to measure CRP, so-called high sensitivity CRP (hsCRP), have been developed and can detect lower concentrations of CRP compared with the conventional methods [19]. The value of hsCRP versus CRP in classifying patients with early axSpA according to ASAS criteria is assessed in **chapter 7**.

Assessment of disease activity

Several clinical measures are available to monitor disease activity in patients with axSpA. Traditionally, the most commonly used has been the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), but this test is fully patient-reported. In order to overcome this limitation, the Ankylosing Spondylitis Disease Activity Score has been developed [20-22], which integrates four patient-reported measures and one objective marker of disease activity (CRP or ESR). The ASDAS has been developed and validated using the conventional CRP. With the availability of the hsCRP, then, the question arises if the ASDAS can be used without an adaptation of the formula, and what should be done if the conventional CRP falls below the limit of detection. For the latter situation, it has been suggested that 50% of the threshold value should be used, but this suggestion has not been underscored by data. In **chapter 8**, using data from an early axSpA cohort, we determine the best way to calculate the ASDAS when the conventional CRP level is below the limit of detection. We also address the question of how to implement hsCRP test results in the ASDAS formula.

Relationship between clinical disease activity measures and MRI-SI

Published data evaluating the relationship between inflammatory lesions detected by MRI and clinical disease activity measures are scarce and mainly obtained from studies including patients with established disease and employing only cross-sectional analysis [23-28]. In **chapter 9**, we investigate the relationship between MRI-SI inflammatory lesions and clinical disease activity measures in patients with early axSpA using longitudinal data analyses.

DATABASES USED IN THIS THESIS

In order to achieve the main objectives of this thesis, data from different clinical trials and cohorts were analyzed. These cohorts were considered most suitable for answering the different research questions in this thesis. A summary of these trials and cohorts is provided here:

Part I: In **chapter 3**, we have used data from 238 patients included in the Norwegian arm of the European Research on Incapacitating Diseases and Social Support (EURIDISS) cohort [29]. EURIDISS is a multinational cohort of patients with early RA initiated in 1996 and that had a mean follow-up of 10 years. Among others, radiographic damage and several measures of disability were evaluated. For **chapter 4 and 5**, the databases from 15 different randomised controlled trials designed for the approval of biologic treatments in patients with RA were analysed. Databases were comprehensive, had on average 517 patients per trial available for analysis, and split up separate joint scores for erosion and JSN according to the SHS method by patient and by reader.

Part II: For **chapter 6** the baseline data of the ESPeranza programme was used. This program is a Spanish prospective multicenter national health program aiming to facilitate early diagnosis of patients with SpA [30,31]. Starting in 2008, 775 patients (age 18-45 years) with short symptom duration (3-24 months) were referred from primary care physicians and other specialists to 25 rheumatology units when they had specific SpA features. In the ESPeranza programme, all the clinical, laboratory and imaging data used to make the clinical diagnosis were collected.

Finally, baseline data (for **chapter 7 and 8**) and 2-year follow-up data (for **chapter 9**) from the Devenir des Spondylarthopathies Indifférenciées Récentes [32] (DESIR) cohort were analysed. DESIR is an ongoing 10-year follow-up prospective French cohort that has started enrollment in 2007 and includes 708 patients from 25 centers (age 18-50 years old) with inflammatory back pain of recent onset (3-36 months), and a likelihood of axSpA of at least 50%, based on a physician's assessment [32]. All patients of DESIR used in this thesis had reached at least 2 years of follow-up, with visits every 6 months.

Lastly, **chapter 10** includes a summary of this thesis as well as the discussion for the results observed in the studies performed. In **chapter 11**, the summary of this thesis is also presented in Dutch.

REFERENCES

1. Bombardier C, Barbieri M, Parthan A, et al. The relationship between joint damage and functional disability in rheumatoid arthritis: a systematic review. Ann Rheum Dis 2012;71:836-44.
2. Smolen JS, Aletaha D. Monitoring rheumatoid arthritis. Curr Opin Rheumatol 2011;23:252-8.
3. Scott DL, Pugner K, Kaarela K, et al. The links between joint damage and disability in rheumatoid arthritis. Rheumatology (Oxford) 2000;39:122-32.
4. 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;73:1455-61.
5. Machado P, Landewe R, Braun J, et al. A stratified model for health outcomes in ankylosing spondylitis. Ann Rheum Dis 2011;70:1758-64.
6. Smolen JS, Aletaha D, Bijlsma JW, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis 2010;69:631-7.
7. Smolen JS, Landewe R, Breedveld FC, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. Ann Rheum Dis 2014;73:492-509.
8. van der Heijde D. Radiographic imaging: the 'gold standard' for assessment of disease progression in rheumatoid arthritis. Rheumatology (Oxford) 2000;39 Suppl 1:9-16.
9. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.
10. Ødegard S, Landewe R, van der Heijde D, et al. Association of early radiographic damage with impaired physical function in rheumatoid arthritis: a ten-year, longitudinal observational study in 238 patients. Arthritis Rheum 2006;54:68-75.
11. Swinkels HL, Laan RF, van 't Hof MA, et al. Modified Sharp method: factors influencing reproducibility and variability. Semin Arthritis Rheum 2001;31:176-90.
12. van der Heijde D, Simon L, Smolen JS, et al. How to report radiographic data in randomized clinical trials in rheumatoid arthritis: guidelines from a roundtable discussion. Arthritis Rheum 2002;47:215-8.
13. Bruynesteyn K, Boers M, Kostense P, et al. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.
14. Rudwaleit M, van der Heijde D, Khan MA, et al. How to diagnose axial spondyloarthritis early. Ann Rheum Dis 2004;63:535-43.
15. Pedersen SJ, Weber U, Østergaard M. The diagnostic utility of MRI in spondyloarthritis. Best Pract Res Clin Rheumatol 2012;26:751-66.
16. 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;68:777-83.
17. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 2008;359:2195-207.
18. Rudwaleit M, Landewe 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;68:770-6.
19. Nielen MM, van Schaardenburg D, Reesink HW, et al. Increased levels of C-reactive protein in serum from blood donors before the onset of rheumatoid arthritis. Arthritis Rheum 2004;50:2423-7.
20. Lukas C, Landewe R, Sieper J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. Ann Rheum Dis 2009;68:18-24.
21. Machado P, Landewe R, van der Heijde D. Endorsement of definitions of disease activity states and improvement scores for the Ankylosing Spondylitis Disease Activity Score: results from OMERACT 10. J Rheumatol 2011;38:1502-6.
22. Machado P, Landewe R, Lie E, et al. Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. Ann Rheum Dis 2011;70:47-53.
23. Kiltz U, Baraliakos X, Karakostas P, et al. The degree of spinal inflammation is similar in patients with axial spondyloarthritis who report high or low levels of disease activity: a cohort study. Ann Rheum Dis 2012;71:1207-11.
24. Konca S, Keskin D, Ciliz D, et al. Spinal inflammation by magnetic resonance imaging in patients with ankylosing spondylitis: association with disease activity and outcome parameters. Rheumatol Int 2012;32:3765-70.
25. Machado P, Landewe R, Braun J, et al. MRI inflammation and its relation with measures of clinical disease activity and different treatment responses in patients with ankylosing spondylitis treated with a tumour necrosis factor inhibitor. Ann Rheum Dis 2012;71:2002-5.
26. Jee WH, McCauley TR, Lee SH, et al. Sacroiliitis in patients with ankylosing spondylitis: association of MR findings with disease activity. Magn Reson Imaging 2004;22:245-50.
27. van der Heijde D, Machado P, Braun J, et al. MRI inflammation at the vertebral unit only marginally predicts new syndesmophyte formation: a multilevel analysis in patients with ankylosing spondylitis. Ann Rheum Dis 2012;71:369-73.
28. Weiss A, Song IH, Haibel H, et al. Good correlation between changes in objective and subjective signs

of inflammation in patients with short- but not long duration of axial spondyloarthritis treated with tumor necrosis factor-blockers. Arthritis Res Ther 2014;16:R35.

29. Syversen SW, Gaarder PI, Goll GL, et al. High anti-cyclic citrullinated peptide levels and an algorithm of four variables predict radiographic progression in patients with rheumatoid arthritis: results from a 10-year longitudinal study. Ann Rheum Dis 2008;67:212-7.
30. Fernandez Carballido C, on behalf of ESPeranza group. Diagnosing early spondyloarthritis in Spain: the ESPeranza program. Reumatol Clin 2010;6 Suppl 1:6-10.
31. Muñoz-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;70:827-30.
32. Dougados M, d'Agostino MA, Benessiano J, et al. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine 2011;78:598-603.

2

Relation between disease activity indices and their individual components and radiographic progression in rheumatoid arthritis: a systematic literature review

V. Navarro-Compán, A.M. Gherghe, J.S. Smolen, D. Aletaha,
R. Landewé, D. van der Heijde

Rheumatology (Oxford) 2015;54(6):994-1007

ABSTRACT

Objective

The aim of this study was to investigate the relationship between different disease activity indices (DAIs) and their individual components and radiographic progression in patients with RA.

Methods

A systematic literature review until July 2013 was performed by two independent reviewers using the Medline and EMBASE databases. Longitudinal studies assessing the relation between DAIs and single instruments and radiographic progression were included. The results were grouped based on the means of measurement (baseline versus time-integrated) and analysis (univariable or multivariable).

Results

Fifty-seven studies from 1,232 hits were included. All published studies that assessed the relationship between any time-integrated DAI including joint count and radiographic progression reached a statistically significant association. Among the single instruments, only swollen joint count and ESR were associated with radiographic progression, while no significant association was found for tender joint count. Data with respect to CRP are conflicting. Data on patient´s global health, pain assessment and evaluator´s global assessment are limited and do not support a positive association with progression of joint damage.

Conclusions

Published data indicate that all DAIs that include swollen joints are related to radiographic progression while, of the individual components, only swollen joints and acute phase reactants are associated. Therefore composite DAIs are the optimal tool to monitor disease activity in patients with RA.

INTRODUCTION

RA is a chronic inflammatory disease that may cause severe disability, reduction in quality of life and premature mortality [1-3], usually demonstrated by conventional radiography [4]. In addition, disease activity has been shown to be closely related to joint damage. Therefore the current treatment target in RA is controlling disease activity and achieving remission [5].

The principal manifestations of disease activity (swelling, tenderness, pain, stiffness and elevated acute phase reactants) can be measured separately using instruments or scales or in different combinations through composite disease activity indices (DAIs) [6]. Today the use of DAIs is recommended over separate instruments, in both clinical trials and clinical practice, as part of treat-to-target strategies, since they are more reliable and responsive to change. Most of the DAIs include a patient-reported assessment, a physician-assessed instrument and sometimes an acute phase reactant. However, some consider patient-reported outcomes, either as single measures or aggregated in a composite score, more feasible to monitor disease activity in a clinical setting because a physical examination is not required [6,7]. The association between these separate measures of disease activity and radiographic progression is important in deciding which measure should be used to monitor disease activity in patients with RA. Although some studies have provided some evidence in respect of this, this point is still unclear. The objective of this study is to investigate the relationship between the different DAIs and their individual components and radiographic progression in patients with RA.

METHODS

Research question and search strategy

A systematic literature review was performed using Medline and EMBASE databases in collaboration with an experienced librarian. The research question was formulated according to the Population, Intervention, Comparison, Outcome and Study design method [8] in which each of the items was defined as follows: population: patients with RA; intervention: DAIs including the DAS, the 28-joint DAS (DAS28), the Simplified Disease Activity Index (SDAI), the Clinical Disease Activity Index (CDAI), the Rheumatoid Arthritis Disease Activity Index (RADAI) and the Routine Assessment of Patient Index Data (RAPID) as well as individual instruments or scales including tender joint count (TJC), swollen joint count (SJC), patient´s general health (GH) assessment, patient´s global assessment (PGA), patient´s visual analogue scale of pain (pain VAS), evaluator´s global assessment (EGA), CRP and ESR; outcome: radiographic progression measured by any scoring method assessing erosions and/or joint space narrowing on conventional radiographs of hands and/or feet; setting: longitudinal

studies with at least 12 months of follow-up.

The search did not have any limit for starting date and included published studies up to 31 of July 2013. No language restriction was applied. The search terms are provided in supplementary Table S1.

Selection of studies

Two reviewers (V.N.C and A.M.G) independently screened the titles and abstracts of the citations and selected articles for full-text review. If there was no consensus between the two reviewers about including a particular study, the article was retrieved and included for full-text review. Later, both readers decided whether or not to include a study for data extraction after reading the full text of the article. In the case of lack of agreement, consensus was sought by discussion. Inclusion criteria were prospective or retrospective cohort studies or randomized controlled trials (RCTs) including only patients with RA or reporting separate data for them, available data for at least 50% of the initial sample size, an evaluation of the relationship between any of the DAIs or instruments described under intervention in the search, and radiographic progression and assessment of radiographic progression on conventional radiographs after a minimum of 12 months. No limitation for the type of analysis performed in the studies was used. However, a statistical hierarchy based on the use of the disease activity measure as an independent variable in the analysis (time-integrated measures were considered more appropriate than only baseline measures) and the type of analysis (multivariable was considered more appropriate than univariable) was taken into account for the results of this systematic review and to elaborate conclusions. For studies testing multivariable analysis, a minimum of 50 patients was required to be included. The exclusion criterion was lack of sufficient data reported in order to assess the relationship between the DAIs or single instruments and radiographic progression.

Data extraction

Using a systematic extraction data form developed for this specific purpose, both reviewers independently extracted data for each study including sociodemographic and disease characteristics, radiographic assessment, degree of disease activity at baseline and during follow-up and statistically relevant data (estimate and significance) for the reported relationship between disease activity measures and radiographic progression. Further, they also evaluated the quality and potential biases of the studies using the guidelines for assessing quality in prognostic studies, assigning an overall quality score per study of between 0 and 6 points according to Hayden et al [9].

Data synthesis

Data were summarized in two groups: DAIs and separate instruments were classified based on the way of measuring (time-integrated or baseline) and the type of analysis (multivariable or univariable) used. The relationship between disease activity measures and the outcome was shown in most of the studies as odds ratio, coefficients (β) or positive likelihood ratio with a corresponding p-value or 95% CI.

RESULTS

Characteristics of the studies

A detailed flowchart with the results of the literature search is depicted in supplementary Figure S1. The search retrieved 1,232 hits and 57 studies were included for data extraction [10-66]. The principal characteristics and overall quality score of the studies are shown in supplementary Table S2. Forty of these studies were observational studies and 17 used data from RCTs. Sample size ranged from 13 and 1,433 patients. Supplementary Table S3 presents the demographic and disease characteristics as well as the treatment for patients included in each study. The mean age of patients included in the studies ranged from 37 to 64 years and the percentage of female patients was between 59% and 100%. Most patients included in the studies were RF positive (range 48-100%) and anti-CCP antibody positive (32-90%). Of all the studies reporting disease duration, most had a symptom duration of <2 years (range 2.7-158 months). Further, most of the studies included patients receiving a synthetic DMARD (sDMARD). Most of the studies (70%) were prospective cohorts and had an overall quality score of at least 4 points (out of a maximum of 6; range 3-6). Radiographic progression was mainly assessed using the modified Sharp-van der Heijde or Larsen scoring method, and the time period to evaluate progression was between 12 and 240 months (supplementary Table S4).

The degree of disease activity and functional status at baseline visit of patients included in the studies are presented in supplementary Table S5 (single instruments) and supplementary Table S6 (DAIs), and considering the mean values of the DAIs, all studies reporting disease activity included patients with active disease.

DAIs

Table 1 shows the results for the studies that investigated the relationship between any of the DAIs and radiographic progression.

Table 1: Predictors of radiographic progression: disease activity indices

Author (ID) [ref number]	Univariable		Multivariable		
	Effect size, OR (95% CI)	p-value	Effect size, OR (95% CI)	p-value	Adjusted for
			DAS		
			Baseline measure		
Courvoisier (31)	1.36	NS			
Dixey (37)	1.39 (1.05, 1.84)				
Jantti (50)	1.50	NS			
Visser (30)	1.20 (MTX)	NS			
	1.20 (comb)	NS			
Drossaers-Bakker (46)	Pro vs non-prog	< 0.05			
Liao (21)	Pro vs non-prog	NS			
Park (23)	Increasing vs flat progression patterns	NS			
Descalzo (3)			Exp coeff=1.13	<0.01	
			Time-integrated measure		
Klarenbeek (20)			Prob(prog) 0.27 (0.20, 0.33)		Age, sex, treatment, CCP, bas SHS
Welsing (39)			β= 5.40 (2.1, 8.6)		Age, sex, DAS, RF, bas SHS
De Punder (2)	β= 4.46 (2.70-6.22)		β= 4.70 (2.96, 6.45)		Age, sex, bas Ratingen
			DAS28		
			Baseline measure		
Benbouazza (16)	1.4	NS			
Sanmartí (34)	1.18	NS			
Park (8)	β= 0.107	NS			
Contreras-Yáñez (17)	Pro vs non-prog	0.06			
Fautrel (11)	Pro vs non-prog	NS			
Imagama (4)	Pro vs non-prog	NS			
Lindqvist (43)			β= 1.40	<0.001	Age, sex, RF, DAS28, bas SHS
Hetland (32)	β= 0.020	NS	β= -0.65	NS	Age, sex, smoking, CCP, bas SHS, MRI
			Time-integrated measure		
Aletaha (36)	β= 0.580	<0.001			
Nishina (7)	β= 0.469	<0.001			
Ichikawa (26)	r= 0.618	0.001			
Felson (19)	LR+(≤2.6)= 2.2	<0.01			
Van Tuyl (13)	LR+(≤1)= 2.6	<0.01			

Smolen (9)	Pro vs non-prog (MTX) (Biologic) (Comb)	 <0.001 <0.001 NS			
Klarenbeek (20)			Prob(prog) 0.26 (0.20, 0.33)		Age, sex, treatment, CCP, bas SHS
Bakker (15)			β= 1.10 (1.0, 1.1)		RF, bas SHS
Salaffi (24)			β= 0.136	-	Age, sex, dis duration, RF, CCP, bas damage
Machold (33)	β= 0.543	<0.001	β= -0.387	<0.001	SJC, CRP, RF,CCP
		SDAI Baseline measure			
Smolen (12)	r= 0.18 (MTX) r= 0.005 (Comb)	<0.05 NS			
		Time-integrated measure			
Aletaha (36)	β= 0.59	<0.001			
Nishina (7)	β= 0.422	<0.001			
Felson (19)	LR+(≤3.3)= 4.8	<0.001			
Van Tuyl (13)	LR+(≤3.3)= 3.1	<0.001			
Kita (5)	Pro vs non-prog	0.03			
Klarenbeek			Prob(prog) 0.24 (0.18, 0.31)		Age, sex, treatment, CCP, bas SHSH
		CDAI Baseline measure			
Smolen (12)	r= 0.17 (MTX) - (Comb)	0.06 NS			
		Time-integrated measure			
Aletaha (36)	β= 0.54	<0.001			
Felson (19)	LR+(≤2.8)= 6.4	<0.001			
Van Tuyl (13)	LR+(≤2.8)= 3.2	<0.001			
Klarenbeek (20)			Prob(prog) 0.25 (0.18, 0.31)		Age, sex, treatment, CCP, bas SHS
		RADAI Baseline measure			
Berglin (41)	β= 0.430	<0.01			
Contreras- Yáñez (17)	Pro vs non-prog	NS			

bas damage: radiographic damage at baseline; bas SHS: baseline modified Sharp-van der Heijde score; CDAI: Clinical Disease Activity Index; comb: combined therapy (DMARD and biologic); DAS28: 28-joint DAS; dis duration: disease duration; (ID): identification number in supplementary Table S2; LR+: positive likelihood ratio; NS: not statistically significant; OR: odds ratio; prog vs non-prog: difference between progressors versus non-progressors; RADAI: Rheumatoid Arthritis Disease Activity Index; SDAI: Simplified Disease Activity Index; SJC: swollen joint count.

DAS and DAS28

Eleven studies evaluated the relation between DAS and radiographic damage. Most of the studies (8 of 11) assessed the association between DAS at baseline and radiographic progression and used univariable testing. Six of the eight (75%) studies did not show a statistically significant association. However, the three studies assessing the relationship between time-integrated DAS and radiographic progression used multivariable testing and showed a significant relationship.

Eighteen studies reported results for the association between DAS28 and radiographic damage. Among the studies that investigated the relationship to baseline DAS28, only 1 of 6 (17%) studies employing univariable analyses and 1 of 2 using multivariable analyses reached statistical significance. Further, 10 studies (6 with univariable and 4 with multivariable testing) evaluated the relationship for the time-integrated DAS28 and they all observed a significant relationship with progression of radiographic damage.

SDAI and CDAI

Seven studies (six with univariable and one with multivariable analyses) reported results for the relationship between baseline (n=1) or time-integrated (n=6) SDAI and progression of joint damage and all of them showed statistically significant associations.

Five studies (four with univariable and one with multivariable analyses) investigated the relationship between baseline (n=1) or time-integrated (n=4) CDAI and radiographic progression and all except one (with a similar trend) showed a significant association.

RADAI and RAPID

Only two studies evaluated the effect of the RADAI at baseline as a predictor of radiographic damage, one of which showed a significant relationship. Moreover, no study investigating the relationship between RAPID and radiographic progression was found.

Individual instruments

TJC and SJC

The results for those studies evaluating the relationship between TJC and SJC and radiographic progression are shown in Table 2. A total of 25 studies provided TJC results. Most of these studies (16 of 25) assessed the relationship with TJC at baseline and used univariable analysis. Eleven of these 16 studies did not report a statistically significant association. Furthermore, 6 of 25 studies reported a relationship between the time-integrated TJC and radiological progression, but only a significant univariate relationship was found.

In addition, 33 studies investigated the relationship between SJC and progression of joint damage. Among these, the majority (19 of 33) evaluated only SJC at baseline and

applied univariable statistical testing. Twelve of these 19 studies did not show a significant association. On the other hand, most of the studies evaluating the relationship between time-integrated SJC or using multivariable analysis observed a significant relationship.

Patient-reported outcomes (GH, PGA and pain VAS) and EGA
A summary with the results for these individual instruments is presented in Table 3.

Concerning the association with GH, only two studies evaluating the predictive value of baseline GH for radiographic progression were found: one study with univariable analysis did not observe an association, while another study with multivariable analysis found a weak but statistically significant relationship. Four studies all using univariable testing reported results for PGA and the two of them that assessed time-integrated PGA reached statistical significance. Moreover, eight studies evaluated the association of radiographic progression with pain VAS. Most of them (6 of 8) only used the baseline measure and did not report a significant relationship. Among the two studies that employed time-integrated pain VAS, only the one with univariable testing reached statistical significance.

Seven studies investigated the association between EGA and radiographic progression. Most of the studies (4 of 5) using the baseline measure did not observe a significant relationship, while the two studies evaluating the association with time-integrated EGA found a significant relationship.

ESR and CRP
Tables 4 and 5 show the results for studies investigating the relationship between ESR and CRP and radiographic progression. A total of 37 studies reported results for ESR. Among the studies assessing the relationship with baseline ESR (26 of 37), a significant relationship was observed in 6 of the 14 studies using univariable analysis and in the majority (10 of 12) of the studies with multivariable analysis. Furthermore, most of the studies (9 of 11) evaluating the relationship with time-integrated ESR reached statistical significance.

A total of 35 studies reported results for CRP. Of those studies, 25 evaluated only the association for baseline CRP. The results were inconsistent since 8 of 17 studies with univariable analysis and 5 of 8 studies with multivariable testing observed a significant relationship between CRP and radiographic progression. Among the studies assessing the relationship between time-integrated CRP and progression of joint damage, 4 of 6 with univariable analysis and 2 of 4 with multivariable testing reached statistical significance, with no differences based on mean disease duration of patients included in the studies.

As a summary of the main results for this review, Table 6 shows the number of studies that assessed each specific DAI or individual instrument as well as the percentage of studies that reached statistical significance grouped by the method of measuring disease activity

Table 2: Predictors of radiographic progression: disease activity instruments: tender joint count and swollen joint count.

Author (ID)	Univariable		Multivariable		
	Effect size, OR (95% CI)	p-value	Effect size, OR (95% CI)	p-value	Adjusted for
			TJC		
			Baseline measure		
Combe (47)	0.70	NS			
Courvoisier (31)[a]	0.48	NS			
Dixey (37)[b]	1.47 (1.8, 2.0)				
Sanmartí (34)	0.99	NS			
Sanmartí (44)	1.08	NS			
Visser (30)	1.0 (MTX)	NS			
	0.93 (comb)	NS			
Guillemin (42)[a]	β= 0.050				
	(-0.05, 0.16)				
Hetland (32)	β= -0.080	NS			
Nishina (7)	β= 0.060	NS			
Vastesaeger (29)	correlation	NS			
Drossaers-Bakker (46)[a]	Prog vs non-prog	<0.001			
Fautrel (11)	Prog vs non-prog	NS			
Liao (21)	Prog vs non-prog	NS			
Park (23)	Prog vs non-prog	0.03			
Paulus (38)	Prog vs non-prog	NS			
Vittecoq(45)[a]	Prog vs non-prog	0.03			
Smolen (12)			(MTX)	NS	SJC, ESR, CRP
			(Comb)	NS	
Weinblant (25)			1.15	0.04	Dis duration, RF
Smolen (35)	r= 0.045 (MTX)	NS	β= -0.007 (MTX)	NS	Age, sex, SJC, CRP, ESR, RF, bas SHS
	r= 0.020 (comb)	NS	β= -0.002 (comb)	NS	
			Time-integrated measure		
Machold (33)	β= 0.454	<0.001			
van Leeuwen (55)	β= 0.469	<0.001			
Ichikawa (26)	r= 0.275	<0.05			
Coste (52)[a]			β= -0.062	NS	Age, dis duration
Van den Broek (10)[c]			1.40	NS	Age, sex, BMI, treatment, DAS, ESR, RF,CCP
Graudal (49)	1.70	NS	0.70	NS	ESR, CRP
			SJC		
			Baseline measure		
Benbouazza (16)	1.10	0.03			
Combe (47)	0.80	NS			
Courvoisier (31)	1.05	NS			
Dixey (37)[b]	1.47 (1.8, 2.0)				

Lillegraven (6)	1.80	NS			
	1.5 (sDMARD)	NS			
	2.4 (bDMARD/comb)	0.08			
Sanmartí (34)	0.90	0.08			
Sanmartí (44)	0.91	NS			
Visser (30)	1.01 (MTX)	NS			
	1.0 (comb)	NS			
de Rooy (18)	β= 0.990	<0.001			
Hetland (32)	β= -0.040	NS			
Nishina (7)	β= 0.002	NS			
Vastesaeger (29)	SJC correlated with progression	No data			
Aletaha (1)	Rate of prog (MTX)	<0.01			
	Rate of prog (Comb)	NS			
Drossaers-Bakker (46)	Prog vs non-prog	<0.001			
Liao (21)	Prog vs non-prog	NS			
Lindqvist (43)	Prog vs non-prog	NS			
Park (23)	Prog vs non-prog	NS			
Paulus (38)	Prog vs non-prog	0.002			
Jantii (50)	SA= -0.030	NS			
Weinblant (25)			1.27	0.02	Dis duration, RF
Fautrel (11)	Prog vs non-prog	NS	1.25	0.06	CRP, CCP, bas erosion
Smolen (12)	r= 0.22 (MTX)	0.02	(MTX)	NS	TJC, ESR, CRP
	r= 0.007 (Comb)	NS	(Comb)	NS	
Berglin (41)	β= 0.395	<0.05	β= 0.685	0.04	Age, sex, therapeutic response, HAQ, CRP, SE, bas Larsen
Smolen (35)	β= 0.160 (MTX)	0.008	β= 0.039 (MTX)	0.01	Age, sex, TJC, ESR, CRP, RF, bas SHS
	β= -0.013 (comb)	NS	β= 0.002 (comb)	NS	
			Time-integrated measure		
Lukas (27)[d]	2.13	<0.001			
van Leeuwen (55)	β= 0.144	0.01			
Aletaha (14)	Prog vs non-prog	<0.001			
Coste (52)			β= -0.038	NS	Age, dis duration
Van den Broek (10)[c]			2.0 (1.1, 3.6)		Age, sex, BMI, treatment, DAS, ESR, RF,CCP
Graudal (49)	5.90	0.01	0.31	NS	ESR, TJC
Markatseli (22)	2.55	0.007	2.25	0.01	Dis duration, ESR, CRP, RF, CCP, bas Larsen
Ichikawa (26)	β= 0.523	<0.001	β= 3.521	<0.001	CRP, bas SHS
Machold (33)	β= 0.584	<0.001	β= 0.264	0.01	DAS28, CRP, RF, CCP

[a]Ritchie index. [b]OR for TJC and SJC together. [c]Tenderness without swelling vs no tenderness and no swelling (for TJC) and swelling without tenderness vs no swelling and no tenderness (for SJC). [d]Persistence/worsening of swelling joints versus improvement or no change. bas erosion: presence/number of erosions at baseline; bas SHS: baseline modified Sharp-van der Heijde score; comb: combined therapy (DMARD and biologic); DAS28: 28-joint DAS; dis duration: disease duration; (ID): identification number in supplementary Table S2; NS: not statistically significant; OR: odds ratio; prog vs non-prog: difference between progressors versus non-progressors; SA: Somer´s d asymmetrical association; SJC: swollen joint count; TJC: tender joint count.

Table 3: Predictors of radiographic progression: disease activity instruments: GH, PGA, EGA, pain VAS.

Author (ID)	Univariable		Multivariable		
	Effect size, OR (95% CI)	p-value	Effect size, OR (95% CI)	p-value	Adjusted for
			GH		
			Baseline measure		
Park (23)	Prog vs non-prog	NS			
Guillemin (42)[a]	0.05 (-0-06, 0.16)		β= 0.01	0.02	Age, sex, dis duration, extra-articular manifestations, treatment, ,Karnofsky, HAQ, country, ESR, RF, bas SHS
			PGA		
			Baseline measure		
Hetland (32)	β= 0.000	NS			
Nishina (7)	β= 0.085	NS			
Time-integrated measure					
Ichikawa (26)	β= 0.554	0.001			
Machold (33)	β= 0.443	<0.001			
			EGA		
			Baseline measure		
Guillemin (42)[b]	β= -0.270	<0.001			
Hetland (32)	β= -0,010	NS			
Jansen (48)	β= 0.257	NS			
Nishina (7)	β= 0.117	NS			
Park (23)	Prog vs non-prog	NS			
			Time-integrated measure		
Ichikawa (26)	β= 0.524	<0.001			
Coste (52)			β= -0.020 (-0.06, 0.02)		Age, dis duration
			Pain VAS		
			Baseline measure		
Hetland (32)	β= 0.010	NS			
Contreras-Yáñez (17)	Prog vs non-prog	NS			
Park (23)	Prog vs non-prog	NS			
Sanmartí (34)	1.01	0.09			
Combe (47)	1.80	0.08	2.41	NS	RF, SE, bas rad damage
Sanmartí (44)	1.03	0.02	1.02 (1.02, 1.09)		Dis duration, HAQ, bas Larsen
			Time-integrated measure		
Ichikawa (26)	β= 0.531	<0.001			
Coste (52)			β= -0.030	NS	Age, dis duration

[a](0-3 mm). [b]Karnofsky. bas SHS: baseline modified Sharp-van der Heijde score; dis duration: disease duration; EGA: evaluator´s global assessment for disease activity (0-100 mm); GH: patient´s global health assessment (0-100 mm); (ID): identification number in supplementary Table S2; NS: not statistically significant; OR: odds ratio; PGA: patient´s global assessment for disease activity (0-100 mm); prog vs non-prog: difference between progressors versus non-progressors; VAS pain: patient´s visual analogue scale for pain (0-100 mm).

parameters and the analysis they employed.

Finally, only six studies (four with univariable and two with multivariable analysis) stratified their results based on treatment and differentiated between patients receiving an sDMARD and patients receiving a biologic DMARD in monotherapy or in combination with an sDMARD. Visser et al did not find differences between the two groups for the DAS, TJC, SJC, ESR and CRP at baseline [39]. Smolen et al did not observe differences for the cumulative DAS28 either [18]. The study of Lillegraven et al did not find differences for baseline SJC and CRP [15]. Lastly, three studies (two with multivariable testing) found differences between the two treatment groups for the SJC and ESR/CRP at the study entry and one of them also showed this difference for the SDAI [10,21,44]. The relationship between these instruments and radiographic progression was only significant in the group of patients receiving sDMARD.

DISCUSSION

The findings of this review suggest that DAIs are more consistently and closely related to radiographic progression than their separate components. Furthermore, the number of studies assessing the different DAIs varies and therefore it is difficult to decide which of the DAIs performs better in this sense. While the association with DAS, and especially with the DAS28, has frequently been investigated, only a few studies have evaluated the relationship between the SDAI and CDAI and radiographic progression, with similar results for both (SDAI and CDAI) indices. At this time there is insufficient evidence for the relationship between the RADAI and radiographic progression, and the RAPID has not yet been evaluated in this regard.

Among the single-item instruments, only those reflecting inflammation -SJC and ESR- have been shown with sufficient evidence to be associated with radiographic progression. Most of the studies have not observed an association with patient-reported outcomes such as pain VAS or with other subjective measures such as TJC. Published data for GH, PGA and EGA are limited and do not support their use as unique tools related to progression of joint damage. Furthermore, data for CRP are less consistent and therefore ESR seems to be a better measure related to radiographic progression in RA. This observation was independent of disease duration.

Today, the claim is that biological therapy may disconnect the relationship between disease activity and joint damage [67]. This means that despite a lack of improvement in disease activity parameters, patients receiving this therapy still show an inhibition of radiographic progression. Furthermore, a recent meta-analysis has suggested that this effect may not be limited to biological therapy and could also be induced by conventional therapies [68]. In this review, only six studies assessed separately the relationship between disease activity

Table 4: Predictors of radiographic progression: disease activity instruments: ESR.

Author (ID)	Univariable		Multivariable		
	Effect size, OR (95% CI)	p-value	Effect size, OR (95% CI)	p-value	Adjusted for
			ESR Baseline measure		
Benbouazza (16)	1.0	NS			
Dixey (37)	2.16 (1.63, 2.87)				
Sanmartí (34)	1.02	0.03			
Sanmartí (44)	1.0	NS			
de Rooy (18)	β= 1.010	<0.001			
Hetland (32)	β= 0.040	NS			
Park (8)	β= 0.121	0.049			
Contreras-Yáñez (17)	Prog vs non-prog	NS			
Drossaers-Bakker (46)	Prog vs non-prog	NS			
Fautrel (11)	Prog vs non-prog	NS			
Park (23)	Prog vs non-prog	NS			
Paulus (38)	Prog vs non-prog	<0.001			
Vastesaeger (29)	ESR correlated with prog	No data			
Jantii (50)	SA= 0.08 (-0.07, 0.23)				
Nishina (7)			β= 0.220	NS	Sex, treatment, RF, CCP
Smolen (12)			(MTX) (Comb)	NS NS	TJC, SJC, CRP
Weinblant (25)			1.21	0.03	Dis duration, RF
Combe (47)	2.8	<0.001	3.44 (1.39, 8.50)		RF, SE, bas rad damage
Courvoisier (31)	1.47	NS	3.20 (1.17, 8.78)		Age, RF, anti-CCP, MMP3, anti-perinuclear anti-Keratin Ab, bas rad damage
Visser (30)	1.02 (MTX) 1.03 (comb)	<0.05 <0.05	0.80 (mod ESR) 2.73 (1.40, 5.33) (high ESR)	NS	Age, sex, BMI, dis duration, treatment, SJC, DAS, HAQ, RF, CCP, bas erosion
Lindqvist (43)	Prog vs non-prog	0.02	1.08	<0.001	RF, SE

Mamehara (28)	Prog vs non-prog	NS	0.31	NS	Treatment, CRP, RF, CCP, MMP3
Fex (54)	Prog vs non-prog	<0.001	β= 0.275	0.03	Age, sex, dis duration, SJC, TJC, RF, genetic factors, Hb, bas rad damage
Guillemin (42)	β= 0.320	<0.001	β= 0.020	<0.001	Age, sex, dis duration, extra-articular manifestations, treatment, ,Karnofsky, HAQ, country, ESR, CRP, bas SHS
Smolen (35)	β= 0.268 (MTX) β= 0.045 (comb)	<0.001 NS	β= 0.018 (MTX) β= 0.004 (comb)	0.003 NS	Age, sex, TJC, SJC, CRP,RF, bas SHS
Uhlig (51)	Prog vs non-prog	0.01	β= 0.350	<0.001	Sex, grip strength, pain, RF, bas rad damage
			Time-integrated measure		
Ichikawa (26)	β= 0.380	0.01			
Machold (33)	β= 0.346	0.01			
van Leeuwen (55)	β= 0.536	<0.001			
van Leeuwen (56)	β= 0.507	<0.001			
Wick (40)	β= 0.88	0.01			
van Leeuwen (53)	Using ESR gave same results that CRP				
Coste			β= 0.090	NS	Age, dis duration
Stockman			Prog vs non-prog	<0.05	Sex
van den Broek			1.01 (1.01, 1.02)		Age, sex, BMI, treatment, TJC, DAS, RF,CCP
Graudal	11.4	<0.001	6.5	0.01	TJC, SJC
Markatseli	1.85	0.03	1.34	NS	Dis duration, ESR, CRP, RF, CCP, bas Larsen

bas erosion: presence/number of erosion at baseline; bas SHS: baseline modified Sharp-van der Heijde score; comb: combined therapy (DMARD and biologic); dis duration: disease duration; ; Hb: haemoglobin; (ID): identification number in supplementary Table S2; MMP3: metalloproteinase 3; NS: not statistically significant; OR: odds ratio; prog vs non-prog: difference between progressors versus non-progressors; SA: Somer´s d asymmetrical association; SJC: swollen joint count; TJC: tender joint count.

Table 5: Predictors of radiographic progression: disease activity instruments: CRP.

Author (ID)	Univariable		Multivariable		
	Effect size, OR (95% CI)	p-value	Effect size, OR (95% CI)	p-value	Adjusted for
			CRP Baseline measure		
Benbouazza (16)	1.0	NS			
Combe (47)	2.90	<0.001			
Courvoisier (31)	1.32	NS			
Lillegraven (6)	1.20	NS			
	1.40 (sDMARD)	NS			
	0.9 (bDMARD)	NS			
Lindqvist (43)	1.01	NS			
Sanmartí (34)	1.07	NS			
de Rooy (18)	β= 1.010	<0.001			
Hetland (32)	β= 0.010	NS			
Sanmartí (43)	β= 0.345	<0.05			
Vastesaeger (29)	CRP correlated with prog	No data			
Aletaha (1)	Rate of prog (MTX)	0.03			
	Rate of prog (comb)	NS			
Kita (5)	Prog vs non-prog	NS			
Liao (21)	Prog vs non-prog	NS			
Park (23)	Prog vs non-prog	0.005			
Paulus (38)	Prog vs non-prog	<0.001			
Uhlig (51)	Prog vs non-prog	0.08			
Vittecoq (45)	Prog vs non-prog	0.03			

Berglin (41)			β= 0.182	0.07	Age, sex, SJC, HAQ, therapeutic response, SE, bas Larsen
Weinblant (25)			1.29	0.01	Dis duration, RF
Visser (30)	1.01 (MTX) 1.02 (comb)	<0.05 <0.05	1.54 (mod CRP) 4.76 (2.32, 9.73) (high CRP)	NS	Age, sex, BMI, dis duration, treatment, SJC, DAS, HAQ, RF, CCP, bas erosion
Mamehara (28)	Prog vs non-prog	NS	9.37	NS	treatment, ESR, RF, CCP, MMP3
Contreras-Yáñez (17)	Prog vs non-prog	0.02	HR= 1.20	0.01	DAS28, RF, CCP
Smolen (34)	β= 0.242 (MTX) β= 0.027 (comb)	<0.001 NS	β= 0.048 (MTX) β= 0.033 (comb)	NS NS	Age, sex, TJC, SJC, ESR, RF, bas SHS
Fautrel (11)	Prog vs non-prog	NS	β= 0.830	NS	SJC, CCP, bas erosion
Smolen (12)	r= 0.15 (MTX) r= 0.08 (comb)	0.11 NS	(MTX) (comb)	<0.01 NS	TJC, SJC, ESR
Time-integrated measure					
Aletaha (35)	β= 0.280	0.03			
Nishina (7)	β= 0.221	NS			
van Leeuwen (52)	β= 0.656	<0.001			
van Leeuwen (54)	β= 0.656	<0.001			
van Leeuwen (55)	β= 0.638	<0.001			
Aletaha (14)	Prog vs non-prog	NS			
Coste (52)			β= 0.040	NS	Age, dis duration
Ichikawa (26)	β= 0.573	<0.001	β= 4.837	0.002	VAS pain, SJC, bas SHS
Machold (33)	β= 0.360	0.007	β= 0.187	0.048	SJC, DAS28, RF, CCP
Markatseli (22)	3.71	<0.001	2.67	NS	Dis duration, SJC, ESR, RF, CCP, bas Larsen

bas erosion: presence/number of erosion at baseline; bas SHS: baseline modified Sharp-van der Heijde score; comb: combined therapy (DMARD and biologic); DAS28: 28-joint DAS; dis duration: disease duration; Hb: haemoglobin; (ID): identification number in supplementary Table S2; MMP3: metalloproteinase 3; NS: not statistically significant; OR: odds ratio; prog vs non-prog: difference between progressors versus non-progressors; SA: Somer´s d asymmetrical association; SJC: swollen joint count; TJC: tender joint count.

Table 6: Summary of studies evaluating the relationship between disease activity indices and their individual components and radiographic progression.

	Number of studies	Baseline measure		Time-integrated measure	
		Univariable studies, n (% sig)	Multivariable studies, n (% sig)	Univariable studies, n (% sig)	Multivariable studies, n (% sig)
Disease activity index					
DAS	11	7 (29)	1 (100)	-	3 (100)
DAS28	18	6 (17)	2 (50)	6 (100)	4 (≥75)
SDAI	7	1 (100)	-	5 (100)	1 (100)
CDAI	5	1 (0)	-	3 (100)	1(100)
RADAI	2	2 (50)	-	-	-
RAPID	0	-	-	-	-
Instrument or scale					
TJC	25	16 (31)	3 (33)	3 (100)	3 (0)
SJC	33	19 (37)	5 (≥60)	3 (100)	6 (67)
GH	2	1 (0)	1 (100)	-	-
PGA	4	2 (0)	-	2 (100)	-
VAS pain	8	4 (25)	2 (50)	1 (100)	1 (0)
EGA	7	5 (20)	-	1 (100)	1 (100)
ESR	37	14 (43)	12 (75)	6 (100)	5 (60)
CRP	35	17 (47)	8 (≥50)	6 (67)	4 (50)

Data show the total number of studies and the percentage of studies that reached statistically significance (% sig) based on the type of measure and analysis employed. CDAI: Clinical Disease Activity Index; DAS28: 28-joint DAS; EGA: evaluator´s global assessment; GH: patient´s general health assessment; PGA: patient´s global assessment; RADAI: Rheumatoid Arthritis Disease Activity Index; RAPID: Routine Assessment of Patient Index Data; SDAI: Simplified Disease Activity Index; SJC: swollen joint count; TJC: tender joint count, VAS pain: patient´s visual analogue scale of pain.

variables and radiographic progression in both biological and conventional treatment. The two studies using multivariable analysis assessed baseline measures of disease activity and support the theory that only biological therapy disconnects this relationship [21,42]. However, the remaining four studies have applied univariable analyses and only one of them showed a different relationship based on therapy [10,15,18,39]. Most of the studies in this systematic review included patients receiving sDMARDs. Therefore more data are required to clarify whether the relationship found between all the DAIs or instruments and radiographic progression may be different in patients receiving biological therapy.

Further methodological limitations of this review should be considered. While a hierarchy was considered to classify the published evidence and to elaborate the conclusions, the results of our study are simply based on whether or not the relationship between baseline or time-integrated disease activity measures and radiographic progression was statistically significant. Other potentially relevant factors such as the sample size or the quality of the studies were not taken into account. In any case, it is not only a difference in sample size

between the studies that explains the main results: in general, studies with a significant relationship did not necessarily have a greater number of patients than studies without a significant relationship. In addition, the estimate for each of the studies is shown in the tables of this review, but it was not considered appropriate to make direct comparisons, as different types of estimate have been applied. Another limitation could be publication bias. However, publication bias usually occurs if negative findings are not published. So if present, this would most likely confirm the absence of a relationship for the measures that do not show a relationship in the published literature.

Also, different scoring methods were used to assess radiographic progression, resulting in a different definition of outcome across studies. Finally, the majority of studies included patients with early disease and therefore it is not possible to investigate whether the relationship between DAI and radiographic progression is the same in patients with early and established RA.

In summary, published evidence indicates that DAIs are related to radiographic progression and thus have greater validity regarding a key construct of RA, i.e. that disease activity is related to structural damage. Among the single instruments, only measures reflecting inflammation, such as SJC and ESR, and not patient-reported measures, are robustly associated with radiographic progression. Based on these results, we recommend the use of one of the DAIs that assess at least the number of swollen joints to monitor disease activity in patients with RA.

REFERENCES

1. Nordgren B, Friden C, Demmelmaier I, et al. Long-term health-enhancing physical activity in rheumatoid arthritis - the PARA 2010 study. BMC Public Health 2012;12:397.
2. Radner H, Smolen JS, Aletaha D. Comorbidity affects all domains of physical function and quality of life in patients with rheumatoid arthritis. Rheumatology (Oxford) 2011;50:381-8.
3. Bombardier C, Barbieri M, Parthan A, et al. The relationship between joint damage and functional disability in rheumatoid arthritis: a systematic review. Ann Rheum Dis 2012;71:836-44.
4. van der Heijde D. Radiographic imaging: the 'gold standard' for assessment of disease progression in rheumatoid arthritis. Rheumatology (Oxford) 2000;39:9-16.
5. Smolen JS, Aletaha D, Bijlsma JW, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis 2010;69:631-7.
6. Smolen JS, Aletaha D. Monitoring rheumatoid arthritis. Curr Opin Rheumatol 2011;23:252-8.
7. Anderson JK, Zimmerman L, Caplan L, et al. Measures of rheumatoid arthritis disease activity: Patient (PtGA) and Provider (PrGA) Global Assessment of Disease Activity, Disease Activity Score (DAS) and Disease Activity Score with 28-Joint Counts (DAS28), Simplified Disease Activity Index (SDAI), Clinical Disease Activity Index (CDAI), Patient Activity Score (PAS) and Patient Activity Score-II (PASII), Routine Assessment of Patient Index Data (RAPID), Rheumatoid Arthritis Disease Activity Index (RADAI) and Rheumatoid Arthritis Disease Activity Index-5 (RADAI-5), Chronic Arthritis Systemic Index (CASI), Patient-Based Disease Activity Score With ESR (PDAS1) and Patient-Based Disease Activity Score without ESR (PDAS2), and Mean Overall Index for Rheumatoid Arthritis (MOI-RA). Arthritis Care Res (Hoboken) 2011;63:14-36.
8. AHRQ Effective Health Care Program Stakeholder Guide. AHRQ Publication No. 1: EHC069-EF July 201.
9. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med 2006;144:427-37.
10. Aletaha D, Alasti F, Smolen JS. Rituximab dissociates the tight link between disease activity and joint damage in rheumatoid arthritis patients. Ann Rheum Dis 2013;72:7-12.
11. de Punder YM, Hendrikx J, den Broeder AA, et al. Should we redefine treatment targets in rheumatoid arthritis? Low disease activity is sufficiently strict for patients who are anticitrullinated protein antibody-negative. J Rheumatol 2013;40:1268-74.
12. Descalzo MA, Garcia VV, González-Alvaro I, et al. Tackling missing radiographic progression data: multiple imputation technique compared with inverse probability weights and complete case analysis. Rheumatology (Oxford) 2013;52:331-6.
13. Imagama T, Tanaka H, Tokushige A, et al. Knee joint destruction driven by residual local symptoms after anti-tumor necrosis factor therapy in rheumatoid arthritis. Clin Rheumatol 2013;32:823-8.
14. Kita J, Tamai M, Arima K, et al. Significant improvement in MRI-proven bone edema is associated with protection from structural damage in very early RA patients managed using the tight control approach. Mod Rheumatol 2013;23:254-9.
15. Lillegraven S, Paynter N, Prince FH, et al. Performance of matrix-based risk models for rapid radiographic progression in a cohort of patients with established rheumatoid arthritis. Arthritis Care Res (Hoboken) 2013;65:526-33.
16. Nishina N, Kaneko Y, Kameda H, et al. Reduction of plasma IL-6 but not TNF-α by methotrexate in patients with early rheumatoid arthritis: a potential biomarker for radiographic progression. Clin Rheumatol 2013;32:1661-6.8.
17. Park YJ, Yoo SA, Choi S, et al. Association of polymorphisms modulating low-density lipoprotein cholesterol with susceptibility, severity, and progression of rheumatoid arthritis. J Rheumatol 2013;40:798-808.
18. Smolen JS, van der Heijde D, Keystone EC, et al. Association of joint space narrowing with impairment of physical function and work ability in patients with early rheumatoid arthritis: protection beyond disease control by adalimumab plus methotrexate. Ann Rheum Dis 2013;72:1156-62.
19. van den Broek M, Dirven L, Kroon HM, et al. Early local swelling and tenderness are associated with large-joint damage after 8 years of treatment to target in patients with recent-onset rheumatoid arthritis. J Rheumatol 2013;40:624-9.
20. Fautrel B, Granger B, Combe B, et al. Matrix to predict rapid radiographic progression of early rheumatoid arthritis patients from the community treated with methotrexate or leflunomide: results from the ESPOIR cohort. Arthritis Res Ther 2012;14:249.
21. Smolen JS, Avila JC, Aletaha D. Tocilizumab inhibits progression of joint damage in rheumatoid arthritis irrespective of its anti-inflammatory effects: disassociation of the link between inflammation and destruction. Ann Rheum Dis 2012;71:687-93.
22. van Tuyl LH, Britsemmer K, Wells GA, et al. Remission in early rheumatoid arthritis defined by 28 joint counts: limited consequences of residual disease activity in the forefeet on outcome. Ann Rheum Dis 2012;71:33-7.
23. Aletaha D, Alasti F, Smolen JS. Rheumatoid arthritis near remission: clinical rather than laboratory inflammation is associated with radiographic progression. Ann Rheum Dis 2011;70:1975-80.

24. Bakker MF, Jacobs JW, Kruize AA, et al. Misclassification of disease activity when assessing individual patients with early rheumatoid arthritis using disease activity indices that do not include joints of feet. Ann Rheum Dis 2012;71:830-5.

25. Benbouazza K, Benchekroun B, Rkain H, et al. Profile and course of early rheumatoid arthritis in Morocco: a two-year follow-up study. BMC Musculoskelet Disord 2011;12:266.

26. Contreras-Yáñez I, Rull-Gabayet M, Vázquez-Lamadrid J, et al. Radiographic outcome in Hispanic early rheumatoid arthritis patients treated with conventional disease modifying anti-rheumatic drugs. Eur J Radiol 2011;79:52-7.

27. de Rooy DP, van der Linden MP, Knevel R, et al. Predicting arthritis outcomes--what can be learned from the Leiden Early Arthritis Clinic? Rheumatology (Oxford) 2011;50:93-100.

28. Felson DT, Smolen JS, Wells G, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum 2011;63:573-86.

29. Klarenbeek NB, Koevoets R, van der Heijde D, et al. Association with joint damage and physical functioning of nine composite indices and the 2011 ACR/EULAR remission criteria in rheumatoid arthritis. Ann Rheum Dis 2011;70:1815-21.

30. Liao KP, Weinblatt ME, Cui J, et al. Clinical predictors of erosion-free status in rheumatoid arthritis: a prospective cohort study. Rheumatology (Oxford) 2011;50:1473-9.

31. Markatseli TE, Voulgari PV, Alamanos Y, et al. Prognostic factors of radiological damage in rheumatoid arthritis: a 10-year retrospective study. J Rheumatol 2011;38:44-52.

32. Park GS, Wong WK, Elashoff DA, et al. Patterns of radiographic outcomes in early, seropositive rheumatoid arthritis: a baseline analysis. Contemp Clin Trials 2011;32:160-8.

33. Salaffi F, Carotti M, Ciapetti A, et al. Relationship between time-integrated disease activity estimated by DAS28-CRP and radiographic progression of anatomical damage in patients with early rheumatoid arthritis. BMC Musculoskelet Disord 2011;12:120.

34. Weinblatt ME, Keystone EC, Cohen MD, et al. Factors associated with radiographic progression in patients with rheumatoid arthritis who were treated with methotrexate. J Rheumatol 2011;38:242-6.

35. Lukas C, van der Heijde D, Fatenajad S, et al. Repair of erosions occurs almost exclusively in damaged joints without swelling. Ann Rheum Dis 2010;69:851-5.

36. Ichikawa Y, Saito T, Yamanaka H, et al. Clinical activity after 12 weeks of treatment with non-biologics in early rheumatoid arthritis may predict articular destruction 2 years later. J Rheumatol 2010;37:723-9.

37. Mamehara A, Sugimoto T, Sugiyama D, et al. Serum matrix metalloproteinase-3 as predictor of joint destruction in rheumatoid arthritis, treated with non-biological disease modifying anti-rheumatic drugs. Kobe J Med Sci 2010;56:98-107.

38. Vastesaeger N, Xu S, Aletaha D, et al. A pilot risk model for the prediction of rapid radiographic progression in rheumatoid arthritis. Rheumatology (Oxford) 2009;48:1114-21.

39. Visser K, Goekoop-Ruiterman YP, de Vries-Bouwstra JK, et al. A matrix risk model for the prediction of rapid radiographic progression in patients with rheumatoid arthritis receiving different dynamic treatment strategies: post hoc analyses from the BeSt study. Ann Rheum Dis 2010;69:1333-7.

40. Courvoisier N, Dougados M, Cantagrel A, et al. Prognostic factors of 10-year radiographic outcome in early rheumatoid arthritis: a prospective study. Arthritis Res Ther 2008;10:106.

41. Hetland ML, Ejbjerg B, Hørslev-Petersen K, et al. MRI bone oedema is the strongest predictor of subsequent radiographic progression in early rheumatoid arthritis. Results from a 2-year randomised controlled trial (CIMESTRA). Ann Rheum Dis 2009;68:384-90.

42. Machold KP, Stamm TA, Nell VP, et al. Very recent onset rheumatoid arthritis: clinical and serological patient characteristics associated with radiographic progression over the first years of disease. Rheumatology (Oxford) 2007;46:342-9.

43. Sanmartí R, Gómez-Centeno A, Ercilla G, et al. Prognostic factors of radiographic progression in early rheumatoid arthritis: a two year prospective study after a structured therapeutic strategy using DMARDs and very low doses of glucocorticoids. Clin Rheumatol 2007;26:1111-8.

44. Smolen JS, van der Heijde D, St Clair EW, et al. Predictors of joint damage in patients with early rheumatoid arthritis treated with high-dose methotrexate with or without concomitant infliximab: results from the ASPIRE trial. Arthritis Rheum 2006;54:702-10.

45. Aletaha D, Nell VP, Stamm T, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. Arthritis Res Ther 2005;7:796-806.

46. Dixey J, Solymossy C, Young A; Early RA Study. Is it possible to predict radiological damage in early rheumatoid arthritis (RA)? A report on the occurrence, progression, and prognostic factors of radiological erosions over the first 3 years in 866 patients from the Early RA Study (ERAS). J Rheumatol Suppl 2004;69:48-54.

47. Paulus HE, Di Primeo D, Sharp JT, et al. Patient retention and hand-wrist radiograph progression of rheumatoid arthritis during a 3-year prospective study that prohibited disease modifying antirheumatic drugs. J Rheumatol 2004;31:470-81.

48. Welsing PM, Landewé R, van Riel PL, et al. The relationship between disease activity and radiologic progression in patients with rheumatoid arthritis: a longitudinal analysis. Arthritis Rheum 2004;50:2082-93.

49. Wick MC, Anderwald C, Weiss RJ, et al. Radiological

progression of joint damage in a longitudinal cohort of early DMARD-treated rheumatoid arthritis patients followed for 10 years. Scand J Rheumatol 2004;33:162-6.

50. Berglin E, Lorentzon R, Nordmark L, et al. Predictors of radiological progression and changes in hand bone density in early rheumatoid arthritis. Rheumatology (Oxford) 2003;42:268-75.

51. Guillemin F, Gerard N, van Leeuwen M, et al. Prognostic Factors for Joint Destruction in Rheumatoid Arthritis: A Prospective Longitudinal Study of 318 Patients. J Rheumatol 2003;30:2585-9.

52. Lindqvist E, Jonsson K, Saxne T, et al. Course of radiographic damage over 10 years in a cohort with early rheumatoid arthritis. Ann Rheum Dis 2003;62:611-6.

53. Sanmarti R, Gomez A, Ercilla G, et al. Radiological progression in early rheumatoid arthritis after DMARDS: a one-year follow-up study in a clinical setting. Rheumatology (Oxford) 2003;42:1044-9.

54. Vittecoq O, Pouplin S, Krzanowska K, et al. Rheumatoid factor is the strongest predictor of radiological progression of rheumatoid arthritis in a three-year prospective study in community-recruited patients. Rheumatology 2003;42:939-46.

55. Drossaers-Bakker KW, Zwinderman AH, Vliet Vlieland TP, et al. Long-term outcome in rheumatoid arthritis: a simple algorithm of baseline parameters can predict radiographic damage, disability, and disease course at 12-year follow-up. Arthritis Rheum 2002;47:383-90.

56. Combe B, Dougados M, Goupille P, et al. Prognostic factors for radiographic damage in early rheumatoid arthritis: a multiparameter prospective study. Arthritis Rheum 2001;44:1736-43.

57. Jansen LM, van der Horst-Bruinsma IE, van Schaardenburg D, et al. Predictors of radiographic joint damage in patients with early rheumatoid arthritis. Ann Rheum Dis 2001;60:924-7.

58. Graudal N, Tarp U, Jurik AG, et al. Inflammatory patterns in rheumatoid arthritis estimated by the number of swollen and tender joints, the erythrocyte sedimentation rate, and hemoglobin: longterm course and association to radiographic progression. J Rheumatol 2000;27:47-57.

59. Jäntti JK, Kaarela K, Luukkainen RK, et al. Prediction of 20-year outcome at onset of seropositive rheumatoid arthritis. Clin Exp Rheumatol 2000;18:387-90.

60. Uhlig T, Smedstad LM, Vaglum P, et al. The course of rheumatoid arthritis and predictors of psychological, physical and radiographic outcome after 5 years of follow-up. Rheumatology (Oxford) 2000;39:732-41.

61. Coste J, Spira A, Clerc D, et al. Prediction of articular destruction in rheumatoid arthritis: disease activity markers revisited. J Rheumatol 1997;24:28-34.

62. van Leeuwen MA, van Rijswijk MH, Sluiter WJ, et al. Individual relationship between progression of radiological damage and the acute phase response in early rheumatoid arthritis. Towards development of a decision support system. J Rheumatol 1997;24:20-7.

63. Fex E, Jonsson K, Johnson U, et al. Development of radiographic damage during the first 5-6 yr of rheumatoid arthritis. A prospective follow-up study of a Swedish cohort. . Br J Rheumatol 1996;35:1106-15.

64. van Leeuwen MA, van der Heijde D, van Rijswijk MH, et al. Interrelationship of outcome measures and process variables in early rheumatoid arthritis. A comparison of radiologic damage, physical disability, joint counts, and acute phase reactants. J Rheumatol 1994;21:425-9.

65. van Leeuwen MA, van Rijswijk MH, van der Heijde D, et al. The acute-phase response in relation to radiographic progression in early rheumatoid arthritis: a prospective study during the first three years of the disease. Br J Rheumatol 1993;32:9-13.

66. Stockman A, Emery P, Doyle T, et al. Relationship of progression of radiographic changes in hands and wrists, clinical features and HLA-DR antigens in rheumatoid arthritis. J Rheumatol 1991;18:1001-7.

67. Landewe R, Keystone E, Smolen JS, et al. Disconnect between disease activity and joint space narrowing for patients with early RA treated with adalimumab plus methotrexate but not methotrexate alone: case for anti-TNF cartilage protection. J Rheumatol 2011;38:1156.

68. Boers M, van Tuyl L, van den Broek M, et. Meta-analysis suggests that intensive non-biological combination therapy with step-down prednisolone (COBRA strategy) may also 'disconnect' disease activity and damage in rheumatoid arthritis. Ann Rheum Dis 2013;72(3):406-9.

SUPPLEMENTARY MATERIAL

Supplementary Table S1: Terms used for each item of the PICOS during the search strategy.

Population	"Arthritis, Rheumatoid"[Majr] OR "Rheumatoid Arthritis"[TIAB]
Intervention	"Severity of Illness Index"[Mesh] "disease activity index"[all fields] OR "disease activity indices"[all fields] OR "disease activity score"[all fields] OR "disease activity scores"[all fields] OR "disease activity scoring"[all fields] OR "disease activity measurement"[all fields] OR "disease activity measures"[all fields] OR "disease activity instruments"[all fields] "disease activity scales"[all fields] OR "disease activity assessment"[all fields] OR "disease activity questionnaire"[all fields] OR "DAS"[all fields] OR "DAS28"[all fields] OR "SDAI"[all fields] OR "CDAI"[all fields] OR "RAPID"[all fields] OR "routine assessment of patient index data"[all fields] OR "RADAI"[all fields] OR “erythrocyte sedimentation rate”[all fields] OR “ESR”[all fields] OR “C-reactive protein”[all fields] OR “C reactive protein”[all fields] OR “CRP”[all fields] OR “patient assessment”[all fields] OR “patient global assessment”[all fields] OR “physician global assessment”[all fields] OR “tender joint”[all fields], “tender joints”[all fields] OR “TJC”[all fields] OR “swollen joint”[all fields] OR “swollen joints”[all fields] OR “SJC”[all fields]
Outcome	(("sharp"[tiab] OR "heijde"[tiab] OR "van der heijde"[tiab] OR "larsen"[tiab] OR "genant"[tiab]) AND ("score"[tiab] OR "scoring"[tiab] OR "scores"[tiab])) OR (("damage"[tiab] AND "radiograph*"[tiab]) OR "radiologic*"[tiab] OR "progression"[tiab] OR "erosions"[tiab] OR "joint space"[tiab]

Supplementary Table S2: General characteristics of the studies.

ID	Author	Year	Study design	Initial sample	Final sample	Quality (Hayden) (0-6)
1	Aletaha	2013	Post-hoc RCT	392	392	5
2	de Punder	2013	prospective cohort	267	267	5
3	Descalzo	2013	prospective cohort	96	39	3
4	Imagama	2013	prospective cohort	32	32	4
5	Kita	2013	prospective cohort	13	13	5
6	Lillegraven	2013	prospective cohort	478	478	5
7	Nishina	2013	prospective cohort	62	62	6
8	Park	2013	prospective cohort	302	302	5
9	Smolen	2013	post-hoc RCT	799	638	3
10	van den Broek	2013	post-hoc RCT	290	290	6
11	Fautrel	2012	prospective cohort	370	370	6
12	Smolen	2012	Post-hoc RCT	273	273	6
13	van Tuyl	2012	prospective cohort	421	421	5
14	Aletaha	2011	Pool data of 6 trials	2377	2377	5
15	Bakker	2011	RCT	265	200	4
16	Benbouazza	2011	retrospective cohort	51	45	6
17	Contreras-Yáñez	2011	prospective cohort	82	82	5
18	de Rooy	2011	prospective cohort	676	676	6
19	Felson	2011	Pool data of 3 trials	192	192	5
20	Klarenbeek	2011	post-hoc RCT	508	508	6
21	Liao	2011	prospective cohort	271	271	5
22	Markatseli	2011	retrospective	144	144	5
23	Park	2011	propective cohort	41	41	5
24	Salaffi	2011	retrospective cohort	59	48	5
25	Weinblant	2011	post-hoc RCT	169	169	6
26	Ichikawa	2010	RCT	55	55	5
27	Lukas	2010	post-hoc RCT	495		5
28	Mamehara	2010	retrospective study	58	58	6
29	Vastesaeger	2009	post-hoc RCT	1049	-	4
30	Visser	2009	post-hoc RCT	508	465	5
31	Courvoisier	2008	prospective cohort	191	112	5
32	Hetland	2008	post-hoc RCT	160	130	6
33	Machold	2007	prospective cohort	138	55	3
34	Sanmartí	2007	prospective cohort	115	105	6
35	Smolen	2006	post-hoc RCT	1049	1004	6
36	Aletaha	2005	prospective cohort	106	56	4
37	Dixey	2004	prospective cohort	866	866	4
38	Paulus	2004	prospective RCT	1433	824	3
39	Welsing	2004	prospective cohort/post-hoc RCT	185/152	57/32	5
40	Wick	2004	prospective cohort	54	54	5
41	Berglin	2003	prospective cohort	43	43	6
42	Guillemin	2003	prodpective cohort	516	318	5
43	Lindqvist	2003	prospective cohort	183	181	6
44	Sanmartí	2003	prospective cohort	65	60	5
45	Vittecoq	2003	prospective cohort	102	91	5
46	Drossaers-Bakker	2002	prospective cohort	135	112	5
47	Combe	2001	prospective cohort	191	177	6
48	Jansen	2001	prospective cohort	130	114	5
49	Graudal	2000	prospective cohort	112	65	4
50	Jantti	2000	prospective cohort	121	66	5
51	Uhlig	2000	prospective cohort	238	182	4
52	Coste	1997	prospective cohort	72	60	5
53	van Leeuwen	1997	prospective cohort	149	149	4
54	Fex	1996	prospective cohort	113	84	5
55	van Leeuwen	1994	prospecive cohort	162	149	4
56	van Leeuwen	1993	prospective cohort	110	110	4
57	Stockman	1991	retrospective cohort	112	112	4

Supplementary Table S3: Demographic and disease characteristics of patients included.

Author	Age (SD or range)	Female (%)	Disease duration, months (SD or range)	RF+ (%)	Anti-CCP+ (%)	Treatment[a]
Aletaha	48 (13)		12 (14)	88		Mixed population
De Punder	57 (13)	64		69	69	sDMARD
Descalzo	53 (17)	68	5.1 (3)	56		
Imagama	59 (11)	94	158 (133)			sDMARD+Biologic
Kita	60 (10)	71	3.0 (2.2)			Mixed population
Lillegraven	59 (50-66)	84	144 (48-276)	66	70	Mixed population
Nishina	56 (15)	79	6.3 (8)	74	74	sDMARD
Park	54 (12)	79	72 (36-144)	69	75	Mixed population
Smolen	53 (13)	74	8.8 (10)	86		Mixed population
Van den Broek	52 (12)	67				Mixed population
Fautrel	49 (11)	73	3.8 (3.8)	55	50	sDMARD
Smolen	50 (12)		104 (98)	82		Mixed population
van Tuyl	55 (13)	69	7.2 (7.2)	56	57	sDMARD
Aletaha	52 (13)	73	26.8 (39)	74		Mixed population
Bakker	53 (14)	66		68		Mixed population
Benbouazza	47 (11)	89	6.00 (4)	63	57	sDMARD
Contreras-Yáñez	39 (16-78)	87	4.9 (1-12)	76	72	sDMARD
de Rooy	56 (16)	68	6.1 (5)	58	32	sDMARD
Felson						sDMARD
Klarenbeek						Mixed population
Liao	48 (14)	80	47.4 (37)	48	51	Mixed population
Markatseli	53 (13)	76	20.0 (20)	50	56	sDMARD
Park	54 (46-59)	80	5.7 (3-10)			sDMARD
Salaffi	56 (11)	73	9.5 (1.9)			Mixed population
Weinblant	53 (13)	79	59.4 (18-121)	76		sDMARD+Biologic
Ichikawa	51 (12)	78	9.2 (5)	91	90	sDMARD
Lukas						Mixed population
Mamehara	55 (16)	81	54 (48)	64	64	sDMARD
Vastesaeger	51 (41-60)	71	7.2 (5-13)			Mixed population
Visser	54 (14)	68	5.5	64	61	Mixed population
Courvoisier	50 (13)	80	3.9 (3)	79	58	Mixed population
Hetland	53 (44-63)	65	3.3 (2.4-4.9)	67	61	sDMARD
Machold	51 (15)	75				sDMARD
Samartí	55 (15)	81	10.0 (7)	74	70	sDMARD
Smolen	51 (18-76)					Mixed population
Aletaha	51 (16)	75	2.7 (3)	78		sDMARD
Dixey		66	6.0 (4-11)	64		sDMARD
Paulus	53 (11)	71	42.0 (23)	67		NSAID
Welsing	54 (14) / 50(13)	64/59	4.8 (4-7) / 3.6(2-7)	77/73		sDMARD/ sDMARD
Wick	49 (9)	75		66		sDMARD
Berglin	51 (14)	70	7.5 (3)	91		sDMARD
Guillemin	52 (12)	70	24.0 (14)	77		
Lindqvist	51 (12)	63	12	71		sDMARD
Sanmartí	52 (16)	78	9.5 (7)	78	56	sDMARD
Vittecoq	49 (12)	66	26.4 (14)	51		sDMARD
Drossaers-Bakker	37 (8)	100	12.0 (0-60)	71		sDMARD
Combe	51 (15)	73	3.3 (3)	81		sDMARD
Jansen	64 (21-86)	68	3.0 (0-24)	51		sDMARD
Graudal		72				sDMARD
Jantti				100		sDMARD
Uhlig	51 (13)	74	26.4 (12)	69		sDMARD
Coste	56 (13)	72	72.0 (108-348)	81		sDMARD
van Leeuwen	53 (16-77)	63	7.0 (1-12)	81		sDMARD
Fex	53 (13)	66	11.4 (7)	63		sDMARD
van Leeuwen	53 (16-77)	63	6.5 (2-13)	81		
van Leeuwen	51 (16-77)	63	5.95 (1-12)	85		sDMARD
Stockman	55 (1)	65	128.4 (11)	87		sDMARD

SD: standard deviation; RF: rheumatoid factor; Anti-CCP: anticyclic citrullinated peptide antibodies; [a]Mixed population: some of the patients with synthetic DMARD and the others with biological (monotherapy or combination) therapy.

Supplementary Table S4: Radiographic assessment in the studies.

Author	Scoring method	Readers	Blind	Trained	ICC	Cut-off	% progressors	Progression (SD or range)	Time to outcome (months)
Aletaha	TGSS	2	no			0.25		0.5 (1.5)	12
De Punder	R score	1	no		0.85		63	3 (0-16)	36
Descalzo	SHS	1	no	yes					24
Imagama	Larsen	2				1	23		24
Kita	Genant	1				1			12
Lillegraven	SHS[a]	1	yes	yes	0.93	5	11		24
Nishina	SHS	2	yes			1.3	50	0.5 (2.9)	12
Park	SHS	2	yes	no	0.91	4			24
Smolen	SHS	2	yes						24
Van den Broek	Larsen	1				1	64		96
Fautrel	SHS	1	yes	yes	0.99	5	34	1.6 (5.5)	12
Smolen	TGSS	2	no			0.25		0.6 (1.4)	12
van Tuyl	SHS	1				1	43	2.7	12
Aletaha	SHS					0,5	44	1.43 (6.5)	12
Bakker	SHS	2	no						60
Benbouazza	SHS	2	no	yes	0.89	5.4	33	3 (1-5)	24
Contreras-Yáñez	erosions	2	no				34		36
de Rooy	SHS	1	no	yes	0.91			34.8 (2.1)	60
Felson	SHS					1			24
Klarenbeek	SHS	2	no			3			60
Liao	SHS[a]	2	yes		0.93				24
Markatseli	Larsen	1	no			10	51	2.1 (1.3)[c]	120
Park	SHS	2		yes				12.6 (6.6-21.4)	
Salaffi	SHS	2	yes	yes		9.5	54	1.7 (2.5-6.5)[c]	36
Weinblant	SHS	2	yes	yes		2.5	31		12
Ichikawa	SHS[a]	3						24.2 (26.4)	24
Lukas	SHS	2	yes			1			12
Mamehara	SHS	2	no	yes		1	55		12
Vastesaeger	SHS	2	yes		0.88	5	16	2.1	12
Visser	SHS	2	yes		0.93	5			12
Courvoisier	SHS	1	yes		0.94	5	70	29.6	120
Hetland	SHS	1	no				30	3.4 (0-43)	24
Machold	Larsen	2	no	yes	0.86[b]				36
Samartí	Larsen	1	no		0.77[b]	2	32	4.9 (6.6)	24
Smolen	SHS					0.1	50	1.53 (7)	12
Aletaha	Larsen	2	no	yes	0.86				36
Dixey	Larsen		yes				38		36
Paulus	Genant, SHS[a]	3	yes	yes	0.89[b]		57	8.7 (13.6)	36
Welsing	SHS / SHS	1/2	no / no		0.88 / 0.91	1.4 / 8.7		9.5 / 7.7	108 / 72
Wick	Larsen				0.94	2			120
Berglin	Larsen	1				2		19.5	24
Guillemin	SHS[a]	1	no	yes	0.98		14	6 (0.5-18)	30
Lindqvist	Larsen	1	no		0.96	11	47	43	120
Sanmartí	Larsen	1	no	yes		2	37	3.7 (6.5)	12
Vittecoq	SHS	2	yes	yes		14	25	13.7	36
Drossaers-Bakker	SHS	1	yes						144
Combe	SHS	2	no		0.85	3.4	42	6.1 (6.2)	36
Jansen	SHS	1	yes	yes			9	3 (-7-77)	12
Graudal	Larsen								
Jantti	Larsen								240
Uhlig	SHS[a]	1	no			30			60
Coste	SHS	1	yes		0.97			3 (2-9)	24
van Leeuwen	SHS	2	no		0.9		83	22 (6-47)	36
Fex	Larsen	1		yes	0.94		44	28 (9-53)	60
van Leeuwen	SHS							22	36
van Leeuwen	SHS		no				86	34.8 (0-203)	36
Stockman	SHS[a]	3	yes		0.84				24

SHS: sharp-van der Heijde scoing method; [a]SHS assessed in hands; ICC: intraclass correlation coefficient; [b]Kappa value; [c]mean yearly progression; SD: standard deviation.

Supplementary Table S5: Disease activity instruments at baseline.

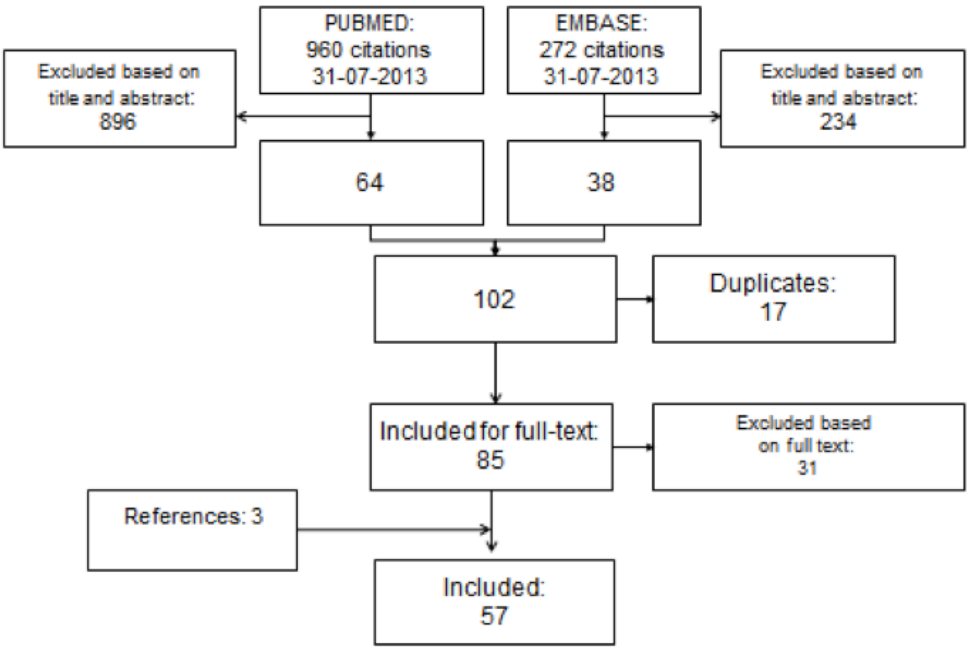
Author	TJC	SJC	VAS pain	GH	PGA	EGA	ESR	CRP
Aletaha	18 (7.2)	14 (6.4)	64 (23)		82 (22)	69 (17)	60 (25)	30 (27)
De Punder	13	14 (10-23)		47 (25)			30 (9-56)	
Descalzo								
Imagama								
Kita								12 (11)
Lillegraven		6 (2-13)						
Nishina	4	4 (2-9)			39 (12-65)	34 (25-51)	33 (20-68)	6 (1-14)
Park								
Smolen	31 (14)	21 (11)						39 (40)
Van den Broek								
Fautrel	8.7 (6.9)	7.9 (5.4)					33 (25)	25 (38)
Smolen	14.9 (6.5)	11.8 (5.3)	53 (22)		61 (23)	61 (16)	45 (24)	22 (23)
van Tuyl	7 (6)	8 (5)	50 (20)		50 (20)	40 (20)	32 (21)	30 (40)
Aletaha	15.2 (6.9)	12.8 (6.1)			61 (22)	61 (19)	43 (26)	31 (37)
Bakker	9.1 (4.8)	9.9 (4.7)	52 (27)	53 (23)			41 (29)	
Benbouazza	20.4 (10.1)	9.6 (6.5)	66 (18)		34 (16)	62 (17)	57 (26)	27 (22)
Contreras-Yáñez			69 (7-10)				24 (2-105)	12 (0-147)
de Rooy		9.5 (7.4)					40 (27)	30 (35)
Felson								
Klarenbeek								
Liao	7.4 (7.9)	6.3 (7)						32
Markatseli							51 (28)	26 (27)
Park	20 (12-32)	17 (11-27)	58 (35-75)	53 (43-73)		48 (31-65)	40 (17-57)	15 (9-42)
Salaffi								
Weinblant	31.5	20.5	67.5		7	70 (60-80)	40 (15-72)	23 (4-69)
Ichikawa	14.4 (8.8)	10 (6.1)	66 (24)		67 (24)	66 (18)	69 (32)	41 (38)
Lukas[a]								
Mamehara							39 (26)	10 (17)
Vastesaeger	41	19 (14-26)					40 (23-61)	14
Visser	12.7	13,3					36	21
Courvoisier	17.3 (8.5)	8.9 (5.8)	55 (23)				38 (27)	29 (40)
Hetland								
Machold	11 (3-17)	8 (5-12)			54 (34-69)	44 (32-61)	56 (26-80)	23 (12-55)
Samartí	10.1 (5.9)	8,3 (4.1)	51 (21)		58 (15)	56 (14)	40 (25)	2.8 (2.9)
Smolen								
Aletaha	8 (3-16)	7 (4-13)	50 (32-66)		51 (33-66)	44 (31-48)	49 (24-70)	51 (19-170)
Dixey								
Paulus	29.19	21,7			78	76	49	24
Welsing								
Wick							40 (44)	
Berglin	4 (1-11)	6 (3-12)						
Lindqvist							28 (14-52)	
Guillemin	10.3 (8.2)			51 (25)		76 (11)	26 (21)	
Sanmartí	10.2 (5.5)	8 (4.1)	51 (22)				45 (28)	29 (31)
Vittecoq	11.4 (9.7)							14 (17)
Drossaers-Bakker	9	3.5					27	
Combe	21 (10)	9 (5.9)	58 (22)				40 (29)	34 (43)
Jansen	9 (0-30)	12 (1-33)					40 (24)	34 (42)
Graudal	6.4 (6.4)	7.2 (1-26)					39 (3-127)	
Jantti								
Uhlig	9.4 (6)						26 (20)	
Coste	10 (13)		41 (33)			26 (34)	26 (29)	10 (18)
van Leeuwen								20 (1-260)
Fex	8 (4-12)	6 (3-12)		1.5 (11.8)[b]			28 (12-48)	
van Leeuwen[a]								
van Leeuwen[a]								
Stockman							32 (2)	

[a]Manuscript referring to the first published article with the same population but not reporting specific data regarding these variables; [b](0-3 mm); TJC: tender joint count; SJC: swollen joint count; VAS pain: patient´s visual analogue scale for pain (0-100 mm); GH: patient´s global health assessment (0-100 mm); PGA: patient´s global assessment for disease activity (0-100 mm); EGA: evaluator´s global assessment for disease activity (0-100 mm); ESR: erythrocyte sedimentation rate (ESR); CRP: C-reactive protein.

Supplementary Table S6: Disease activity indices and health assessment questionnaire at baseline.

Author	DAS	DAS28	SDAI	CDAI	RADAI	HAQ
Aletaha		7.1 (1.0)	49.4 (15)	46.4 (14)		1.83 (0.7)
De Punder	4.0 (1.3)					
Descalzo		5.6 (1.2)				1.3 (0.7)
Kita			22.5 (11)			
Imagama		4.9 (1.1)				
Lillegraven		5.1 (1.2)	28 (13)	25 (12)		1.2 (0.7)
Nishina		4.4 (3.6-5.6)	16.2 (9-28)		6 (0.8-9.9)	0.63 (0.3-1.2)
Park		4.2 (2.9-5.4)				
Smolen		6.3 (0.9)				1.51 (0.7)
Van den Broek	4.3 (0.9)					1.3 (0.6)
Fautrel		5.4 (1.2)				1.03 (0.7)
Smolen van Tuyl		6.4 (0.9)	40.9 (13)	38.8 (12)		1.5 (0.6)
Aletaha		6.4 (1.1)	43.4 (15)	40.3 (14)		1.47 (0.7)
Bakker		5.5 (1.1)				
Benbouazza		6.9 (1.3)				2.2 (0.6)
Contreras-Yáñez		6.4 (3.1-8.6)				1.6 (0-3)
de Rooy						
Felson						
Klarenbeek	4.4 (0.9)					1.4 (0.9)
Liao		3.8 (1.6)				1.65 (1.5)
Markatseli		5.8 (1.2)				
Park	4.5 (3.8-5.8)					1.3 (0.8-1.8)
Salaffi						
Weinblant						1.95 (1.3-2.4)
Ichikawa		4.8 (0.9)				0.76 (0.4)
Lukas[a]						
Mamehara						
Vastesaeger						1.5 (1.0-1.9)
Visser	4.4 (0.8)					1.4 (0.8)
Courvoisier	4.0 (0.7)					1.29 (0.7)
Hetland		5.6 (4.7-6.1)				0.9 (0.4-1.5)
Machold						0.88 (0.1-1.8)
Samartí		5.7 (0.9)				1 (0.6)
Smolen						
Aletaha						0.75 (0-1.5)
Dixey						
Paulus						
Welsing	3.6 (1.0) / 4.8(1.1)	4.8 (1.0) / 6.1(1.1)				0.7 (1.4) / 1.4(0.7)
Wick						
Berglin		4.6 (4.1-7.0)				0.69 (0.3-1.0)
Guillemin						0.99 (0.7)
Lindqvist						0.8 (0.5-1.2)
Sanmartí		5.8 (0.8)				1 (0.5)
Vittecoq						0.8 (0.7)
Drossaers-Bakker	2.8					0.75 (0-2.9)
Combe	4.1 (0.8)					1.3 (0.7)
Jansen	5.4 (1.2)					1 (0.8)
Graudal						
Jantii						
Uhlig						0.96 (0.7)
Coste						1.2 (1.5)
van Leeuwen						
Fex						0.8 (0.4-1.2)
van Leeuwen[a]						
van Leeuwen[a]						
Stockman						

[a]Manuscript referring to the first published article with the same population but not reporting specific data regarding these variables. DAS: Disease Activity Score; DAS28: Disease Activity Score using 28 joints count; SDAI: Simplified Disease Activity Index; CDAI: Clinical Disease Activity Index; RADAI: Rheumatoid Arthritis Disease Activity Index; HAQ: Health Assessment Questionnaire.



Supplementary Figure S1: Flow chart of the results of the literature search.

3

Relationship between types of radiographic damage and disability in patients with rheumatoid arthritis in the EURIDISS cohort: a longitudinal study

V. Navarro-Compán, R. Landewé, S.A. Provan, S. Ødegård, T. Uhlig, T.K. Kvien, A.P. Keszei, S. Ramiro, D. van der Heijde

Rheumatology (Oxford) 2015;54:83-90

ABSTRACT

Objective

The aim of this study was to assess if any of the different types of radiographic damage [true joint space narrowing (JSN), (sub)luxation, and erosions] are preferentially related to disability in patients with RA.

Methods

Longitudinal data from 167 RA patients from the European Research on Incapacitating Diseases and Social Support study over 10 years were analyzed to investigate the relationship between the three types of radiographic damage and disability [grip strength, HAQ and the dexterity scale in the Arthritis Impact Measurement Scales (AIMS)]. A longitudinal analysis including separate models per type of damage and joint group and combined models including all information was conducted.

Results

All types of damage were inversely related to grip strength in the analysis of separate models, but only true JSN independently remained statistically significant in the combined analysis [β: -0.087 (95% CI -0.151, -0.022)]. Neither JSN, (sub)luxation nor erosions were associated with HAQ score, while erosions were associated with AIMS dexterity only in the analysis of separate models. After stratifying for hand joint group, erosions at MCP joints [β: -0.288 (95% CI -0.556, -0.019)] and true JSN at wrist [β: -0.132 (95% CI -0.234, -0.030)] were significantly related to grip strength. Erosions at PIP [β: 0.017 (95% CI 0.005, 0.028)] and MCP joints [β: 0.114 (95% CI 0.010, 0.217)] was the only type of damage associated with HAQ and AIMS dexterity, respectively.

Conclusion

All types of radiographically visible joint damage interfere with important aspects of physical functions. True JSN is most closely related to hand function.

INTRODUCTION

RA may cause severe disability, impaired quality of life and premature mortality if not treated properly [1,2]. Usually the degree of disability and impairment in patients with RA is determined in part by joint destruction [3]. There are three main anatomical structures in the joint that can be affected: (i) cartilage, (ii) ligaments or tendons and (iii) bone. Structural damage of the joint can be visualized by simple radiographs, but the lesions observed depend on the affected anatomical structure. Cartilage damage presents as joint space narrowing (JSN) on X-rays, damage to tendons or ligaments may present as (sub)luxation and bone damage as erosions.

A recommended method to quantify radiographic changes is the modified Sharp-van der Heijde method (SHS) [4,5]. This method assesses the presence of erosions and JSN separately, but also accounts for the presence of (sub)luxation, included in the JSN score but not otherwise indicated.

In agreement with other studies [6], we previously reported a longitudinal relationship between the SHS score and disability [7]. What is still a matter of debate is whether the three types of structural damage contribute independently and similarly to the degree of disability or whether one or two of them mainly explain functional impairment in RA patients. This may be -or may become- relevant in order to select the best treatment option for patients, as targeted drugs may increasingly be used for specific kinds of structural damage (e.g. denosumab) [8].

Very few studies have previously assessed the effects of erosions and JSN on disability separately, and the results suggest that JSN rather than erosions explain disability [9,10]. Nevertheless, these studies are based on limited information. First, the extent to which the JSN score captures cartilage damage or (sub)luxation in these studies remains uncertain since a differentiation between both was not made. Second, cartilage damage, and therefore true JSN preferentially occurs in some joints (e.g. the wrist), which may influence the relationship between type of damage and disability (e.g. the wrist might be more important for function than proximal interphalangeal (PIP) joints) [11]. Finally, most of the studies assessed relative short follow-up periods and used only general measurements of disability such as HAQ score to evaluate functional status, but did not evaluate more proximal measures of function such as grip strength and dexterity [12,13].

Based on this, we hypothesized that JSN is more strongly related than other types of damage to disability and that the wrist is the localization in the hand more strongly associated with disability in patients with RA. The primary objective of this study was to assess whether any of the different types of radiographic damage [JSN, (sub)luxation and erosions] is preferentially related to disability in patients with RA. Disability was assessed here in three

different ways: grip strength, HAQ and dexterity component of the original Arthritis Impact Measurement Scales (AIMS). As secondary objective, the influence of radiographic damage in different hand joint groups (PIP, metacarpophalangeal (MCP) and wrist joints) on disability was investigated.

METHODS

Study design and population

These analyses were based on data collected during 10-year follow-up of the Norwegian arm of the European Research on Incapacitating Diseases and Social Support (EURIDISS) study. Details of this study have been previously described [7,14]. In short, the EURIDISS study included 238 patients with an age at baseline between 20 and 70 years, a diagnosis of RA according to the 1987 ACR criteria [15] and a disease duration ≤4 years. Patients with other incapacitating diseases or with a functional stage IV according to Steinbrocker's classification were excluded. For this study, data from patients with radiographs available for both hands for at least two visits of the study were selected for the analysis. The study was approved by the Norwegian Regional Committee for Research Ethics.

Data collection and clinical measures

In the EURIDISS study, socio-demographic and disease data were collected at baseline and during the follow-up assessments at 1, 2, 5 and 10 years. These data included age, gender, disease duration since diagnosis and duration since first symptoms, Ritchie's Articular Index, presence of subcutaneous nodules, and of any other extra-articular manifestations, ESR, RF and patient's overall assessment of health on a 100 mm visual analogue scale.

Disability measurements

For the current analysis, the main outcome was the change in grip strength. Grip strength was assessed in both hands at every visit using a hand-held dynamometer (JAMAR Technologies, Hatfield, PA, USA) displaying grip force of 0-90 kg. The best performance of 2 attempts in each hand was used to calculate the average grip strength in both hands. Other outcomes for disability included the HAQ (range 0-3) [16] and the AIMS dexterity scale (range 0-5) [17]. Unlike grip strength, increasing scores in HAQ and AIMS dexterity scale represent worse functioning. The AIMS dexterity scale addresses an individual's motor skills with regards to fingers, hands, and arms. The dexterity scale of the AIMS consists of five items listed in Guttman's scale order, which assumes that subjects failing an item in the list with a certain level of difficulty also tend to fail all items with perceived higher levels of difficulty in the group.

Radiographs

Postero-anterior radiographs of both hands were performed at every time point. All radiographs (hands only) were read by one trained reader who was aware of the chronological time order of the radiographs. For this study, another trained reader who was also aware of the time order evaluated all radiographs with an initial individual joint score of 3 or 4 for the JSN component of the SHS in order to differentiate true JSN from (sub)luxation. The categories were 0 for no (sub)luxation, 1 for subluxation and 2 for luxation. The overall range for the different types of damage were 0-160 for erosions, 0-120 for true JSN and 0-20 for (sub)luxation. Further, the radiographic sum scores were calculated per joint group: PIPs, MCPs and wrist, taking into account that the presence of (sub)luxation was only assessed at the level of MCP joints but not at the level of PIP joints and the wrist, as (sub)luxation is typically not present in these locations.

Statistical analysis

Data for continuous variables are presented as mean (SD) for approximately normally distributed variables, or as median (range) if appropriate. For categorical variables, results are shown as percentage and relative frequencies. The percentage of missing joint scores per time point increased over time but was small at all time points, ranging from 0.07% to 0.81%. However, to make sure that SHS scores did not decrease due to the missing values, last observation carried forward was used to impute these few joint scores.

The three types of joint radiographic damage were analysed as continuous, dichotomous (presence of at least one joint with the specific type of damage) and count (number of joints with that damage) variables. The HAQ score was analysed as a continuous and as a dichotomous (based on cut-off of ≤0.5 units) variable [18]. Grip strength was analysed as a continuous variable and AIMS dexterity as a dichotomous variable, using the median value of all patients as the cut-off point.

Longitudinal analysis was applied to investigate the relationship between the three types of radiographic joint damage and functional outcomes over a follow-up period of 10 years. In longitudinal analysis, the model adjusts for within-patient correlation [19]. For this study, generalized estimating equations (GEE) models were built in two consecutive steps. First, functional outcomes were modelled by each type of damage –erosion or JSN or (sub)luxation- separately (analysis of separate models). Second, functional outcomes were modelled by the three types of damage –erosion, JSN and (sub)luxation- in the same model (analysis of combined models). Possible confounders in all these models included age, Ritchie articular index and ESR as continuous variables and gender and treatment with DMARDs as categorical variables.

For the GEE models, the linear scale response (with β as a parameter estimate) was chosen

for continuous outcomes and the binary logistic response was chosen for the dichotomous outcomes. For all analyses, two working correlation matrix structures (exchangeable and autoregressive) were tested for fitting data best [using quasilikelihood under the Independence model Criterion (QIC)]. Interactions of types of damage with hand side (left and right) and gender were tested in the separate models. All data were analysed using SPSS version 20.0 (SPSS, Chicago, IL, USA).

RESULTS

Baseline characteristics

A total of 177 RA patients with available radiographic data were included in this study. From these, 167 patients had radiographic assessment at two or more study visits. These patients were very similar to those in the entire EURIDISS cohort (Table 1) and reflect a usual cohort of patients with RA as seen in the 1990s. With a mean (SD) HAQ score of 0.90 (0.62), physical function was moderately impaired. The mean (SD) values for grip strength [15.6 (7.6) in females and 33.8 (12.8) kg in males] were lower than the reported reference values for grip strength in the healthy population (27.7 for females and 47.2 kg for males) [20].

Table 1: Baseline characteristics for patients included in the Norwegian EURIDISS cohort and for the subgroup of patients included in this study.

	EURIDISS patients (n=238)	EURIDISS patients with radiographs (n=177)
Age, years	51.9 (13.0)	50.5 (12.8)
Female, n (%)	175 (73.5)	129 (72.9)
Disease duration, years	2.3 (1.5)	2.2 (1.2)
RF, n (%)	174 (73.1)	131 (74.0)
Ritchie index	9.8 (6.0)	9.5 (6.1)
VAS pain (0-100)	33 (25)	32 (24)
ESR, mm/h	25.9 (20)	24.5 (18.8)
CRP, mg/L	12.6 (16.9)	12.4 (17.4)
HAQ score	0.93 (0.64)	0.90 (0.62)
Grip strength, kg	19.2 (11.6)	20.5 (12.3)
Left hand	17.7 (11.7)	19.1 (12.4)
Right hand	20.6 (12.3)	21.9 (12.9)
AIMS dexterity scale	1.6 (1.4)	1.6 (1.5)
≤ 1, n (%)	143 (62)	115 (65)
≥ 2, n (%)	86 (38)	62 (35)
Treatment, n (%)		
Glucocorticoids	65 (27.3)	44 (24.9)
DMARDs	124 (52.1)	93 (52.5)

Data are given as mean (SD) unless otherwise specified. VAS: visual analogue scale; AIMS: Arthritis Impact Measurement Scales.

Core radiographic data are presented in Table 2. The median (range) SHS score at baseline was 2 (0-61) and had increased 20 units after 10 years. The majority of patients had at least one type of damage present at baseline. The most frequent type of damage observed was the combination of true JSN and erosion, followed by the sole presence of true JSN. Importantly, the sole presence of (sub)luxation was not seen in any of the patients during the study, and only 8% of all patients had this lesion in combination with any type of damage at the end of the study.

Table 2: Change over time in disability measurements and radiographic damage.

	Baseline (n=177)	1 year (n=176)	2 year (n=170)	5 year (n=155)[a]	10 year (n=142)
HAQ score	0.90 (0.62)	0.87 (0.63)	0.86 (0.66)	0.88 (0.62)	0.92 (0.70)
Grip strength (kg)	20.5 (12.3)	20.6 (12.6)	20.8 (12.9)	20.0 (12.2)	19.9 (11.7)
Females	15.6 (7.6)	15.6 (7.7)	15.4 (7.7)	15.9 (7.9)	15.7 (7.8)
Males	33.8 (12.8)	33.8 (13.7)	35.6 (12.8)	31.2 (14.8)	33.1 (13.2)
AIMS dexterity scale, n (%)					
≤ 1	115 (65)	121 (69)	109 (64)	78 (51)	81 (57)
≥ 2	62 (35)	55 (31)	61 (36)	77 (49)	61 (43)
SHS (median, range)	2 (0-61)	4 (0-66)	6 (0-80)	14 (0-117)	22 (0-150)
Type of damage, n (%)					
None	68 (45)	52 (34)	46 (29)	29 (21)	22 (16)
Only Ero	7 (5)	11 (7)	14 (9)	7 (5)	8 (6)
Only true JSN	26 (17)	22 (15)	18 (11)	16 (11)	10 (7)
Only Sub	-	-	-	-	-
Ero + true JSN	51 (34)	68 (44)	76 (49)	82 (59)	84 (61)
Ero + Sub	-	-	-	-	-
True JSN + Sub	-	-	1 (1)	1 (1)	-
Ero + true JSN + Sub	-	-	1 (3)	4 (3)	14 (8)

Data are given as mean (SD) unless otherwise specified. AIMS: Arthritis Impact Measurement Scales; SHS: Sharp-van der Heijde score; ero: erosion; JSN: joint space narrowing; sub: subluxation or luxation. [a] Data for AIMS dexterity scale available in 128 patients.

The association of radiographic damage and grip strength

Interactions between each type of damage and the hand side on grip strength were not statistically significant, but a significant interaction was found for gender and true JSN ($p=0.03$), and for gender and erosion ($p<0.01$), but not for gender and (sub)luxation ($p=0.8$), with male patients showing the strongest association. After adjusting for confounders, the analysis of separate models showed a negative association between all the 3 types of damage and grip strength for both genders (Table 3), but the effect of erosion and true JSN on grip strength was almost twice as strong in men as in women. However, in the combined analysis including the three types of damage, only the association for true JSN remained independently statistically significant [β= -0.087 (95% CI -0.151, -0.022)] (Table 4) with the loss of an independent contribution of erosions and/or (sub)luxation.

Table 3: Analysis of separate models for the relationship between type of radiographic damage and outcomes and between the damage in different hand joint groups and outcomes in the EURIDISS study over 10 years.

	HAQ score		Grip strength				Dexterity	
			Females		Males			
	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Type of damage								
Erosion	0.003	-0.002 to 0.008	-0.110*	-0.167 to -0.052	-0.230*	-0.366 to -0.093	0.027*	0.004 to 0.049
JSN	0.002	-0.001 to 0.004	-0.082*	-0.120 to -0.044	-0.152*	-0.259 to -0.046	0.009	-0.003 to 0.021
Subluxation	0.039	-0.014 to 0.092	-0.948*	-1.349 to -0.546	-0.597	-0.709 to 1.904	0.143	-0.288 to 0.575
Joint group								
Erosion								
PIP	0.012*	0.003 to 0.022	-0.121	-0.338 to 0.095	-0.650*	-0.996 to 0.003	0.094*	0.035 to 0.153
MCP	0.003	-0.008 to 0.013	-0.323*	-0.458 to -0.189	-0.400*	-0.710 to -0.090	0.073*	0.019 to 0.127
Wrist	0.005	-0.005 to 0.015	-0.118*	-0.220 to -0.015	-0.549*	-1.095 to -0.002	0.014	-0.025 to 0.053
JSN								
PIP	0.005	-0.005 to 0.014	-0.231*	-0.384 to -0.077	-0.248	-0.597 to 0.101	0.040	-0.001to 0.082
MCP	-0.001	-0.007 to 0.006	-0.171*	-0.262 to -0.080	-0.212	-0.523 to -0.099	0.021	-0.009 to 0.052
Wrist	0.004	-0.001 to 0.009	-0.127*	-0.190 to -0.064	-0.376*	-0.574 to -0.178	0.010	-0.013 to 0.032
Subluxation								
MCP	0.039	-0.014 to 0.092	-0.948*	-1.349 to -0.546	-0.597	-0.709 to 1.904	0.143	-0.288 to 0.575

The table shows one model for each type of damage and each region in the left column of the table. Analysis adjusted for age, gender (for HAQ and dexterity scale), Ritchie index, ESR and treatment. *$p < 0.05$. JSN: joint space narrowing; subluxation: subluxation or luxation; AIMS: Arthritis Impact Measurement Scales.

Next, we investigated the association between the types of damage and grip strength in each joint group (PIPs, MCPs and wrist) separately. In the combined analysis, erosions in MCPs [β= -0.288 (95% CI -0.556, -0.019)] and JSN in the wrist [β= -0.132 (95% CI -0.234, -0.030)] (Table 4) contributed significantly and independently to explaining variation in grip strength.

The association of radiographic damage and HAQ score

None of the separate types of damage was significantly related to the HAQ score (Table 3), but in the analysis per joint group a significant relationship between PIP erosions and HAQ score was found [β= 0.017 (95% CI 0.005, 0.028)] (Table 4).

The association of radiographic damage and the AIMS dexterity scale

A marginally significant association between erosions and the AIMS dexterity scale was found in the separate analysis [β= 0.027 (95% CI 0.004, 0.049)] (Table 3), but this association disappeared in the combined analysis (Table 4). When hand joint groups were analysed separately, a significant association between MCP erosion and AIMS dexterity scale was observed [β= 0.114 (95% CI 0.010, 0.217)].

Table 4: Analysis of combined models for the relationship between the type of radiographic damage and outcomes and between the damage in different hand joint groups and outcomes in the EURIDISS study over 10 years.

	HAQ score		Grip strength		Dexterity	
	β	95% CI	β	95% CI	β	95% CI
Type of damage						
Erosion	0.003	-0.004 to 0.011	-0.027	-0.141 to 0.088	0.039	-0.002 to 0.081
JSN	0.000	-0.005 to 0.004	-0.087*	-0.151 to -0.022	-0.010	-0.035 to 0.015
Subluxation	0.029	-0.024 to 0.082	0.037	-0.871 to 0.796	0.006	-0.428 to 0.440
Joint group						
Erosion						
PIP	0.017*	0.005 to 0.028	0.025	-0.291 to 0.341	0.066	-0.011 to 0.144
MCP	-0.004	-0.020 to 0.013	-0.288*	-0.556 to -0.019	0.114*	0.010 to 0.217
Wrist	0.000	-0.012 to 0.013	0.050	-0.095 to 0.194	-0.004	-0.055 to 0.047
JSN						
PIP	-0.001	-0.013 to 0.011	-0.071	-0.327 to 0.185	0.000	-0.055 to 0.054
MCP	-0007	-0.018 to 0.003	0.056	-0.111 to 0.222	-0.047	-0.107 to 0.013
Wrist	0.005	-0.002 to 0.012	-0.132*	-0.234 to -0.030	-0.007	-0.044 to 0.029
Subluxation						
MCP	0.043	-0.013 to 0.098	0.196	-0.765 to 1.157	-0.067	-0.555 to 0.422

The table shows one model including all types of damage and one model including all types of damage per joint group. Analysis adjusted for age, gender, Ritchie index, ESR and treatment. *p < 0.05. JSN: joint space narrowing; subluxation: subluxation or luxation; AIMS: Arthritis Impact Measurement Scales.

DISCUSSION

This study evaluated the contribution of different kinds of radiographic joint damage in explaining several types of functional outcome in patients with RA. A worsening in every type of radiographic damage was related to a decrease in grip strength. True JSN, likely reflecting cartilage loss, showed the strongest relationship with grip strength. Overall, radiographic damage was associated with HAQ score, but this association was stronger for the sum of the total SHS than for its constituent parts.

Our results suggest that true JSN is the most important type of damage related to grip strength and that (dominant) hand side does not influence this relationship, which is in agreement with previous studies [20]. Based on the estimate, an increment of 11 JSN units will lead to a decrease of 1 kg in grip strength over a period of 10 years.

In a recent study, HAQ score appeared to be mainly dependent on JSN [21]. Several reasons may explain this discrepancy with our results. First, the feet, which are involved in part of the activities assessed by the HAQ, were not radiographically evaluated in our study. Another possible explanation is the different sample sizes of the studies, as the one from Aletaha et al included a wider range of abnormalities and therefore could have been more apt to find any type of association.

It may also simply be true that the relationship between radiographic score and HAQ score is only pertinent for the combined SHS score, as shown by us in a previous analysis [7], rather than for its components. This finding supports the use of the total SHS rather than its components that do not assess sufficient joints or only reflect one type of (qualitative) damage. As such, the total SHS score performs like an appropriate index: the combined effect is stronger than the summed effects of its components. Second, a longitudinal study in a different cohort also has reported that neither overall JSN nor erosions had a significant effect on HAQ [22]. Third, a recent analysis from a randomised controlled trial reported no association between erosions and JSN with HAQ at baseline [23]. Finally, Aletaha et al´s study methodology, while unprecedented and elegant, was not beyond argumentation as we have pointed out previously [24].

Furthermore, the results on dexterity suggest that erosions are more important than true JSN for skilled movements. However, we believe the dexterity measurement itself may have limitations that might have affected the results [17]. The AIMS dexterity scale does not follow an entirely logical order. For example, this scale considers writing with a pen or pencil, which requires coordinated motor actions rather than force, as much more difficult than opening a jar of food, which mainly requires force. For an RA patient with active disease and significant joint destruction, (lack of) force or strength may be a limiting factor of greater importance than coordination.

Regarding the results of the analyses per joint group, identification of the wrist as the predominant site related to grip strength is not very surprising because JSN is frequently observed in the wrist and therefore disproportionally contributes to the total JSN SHS score [24]. In addition, the absence of association observed between JSN and HAQ is consistent with the data published in the study of Koevoets et al [22]. Moreover, this study also found a relationship between erosion and HAQ, but the predominant site was different, as the PIP joints were the predominant site in our study and the wrist was predominant in the study of Koevoets et al.

Our study has important strengths, such as the long follow-up period and the additional assessment of more proximal measures of disability (grip strength). Although most of the RA studies report the HAQ as a unique instrument of disability [18], the HAQ score does not optimally reflect what patients can do with their hands in their daily life [12,13]. Additionally, the results of this study for all the analyses using different types of variables for radiographic damage and for HAQ were consistent. Very few patients had been exposed to biological DMARDs, which is important due to the uncoupling of inflammation and bone destruction in patients exposed to these agents [23]. The 10-year follow-up data have therefore also been useful in the identification of genetic and soluble biomarkers for prediction of radiographic damage and for prediction of cardiovascular outcomes [25-28]. Furthermore, the main results of the study were robust against using different types of correlation structures in the GEE models, which adds to the credibility of the data.

Finally, this study also has limitations. Obviously the most important limitation of this study is the lack of radiographic evaluation of the feet, which could have influenced the association with the HAQ score in both directions. Additionally, the fact that radiographic images and data were not available for one-quarter of patients included in the EURIDISS study may have resulted in bias by completion. However, no differences at baseline characteristics were observed between patients with radiographic data available compared to the entire cohort. Furthermore, the population of this study included patients in an early stage of the disease with non-biologic DMARD therapy in most cases and therefore with greater probability of having radiographic damage and progression. A similar relationship might not be found in patients treated with biologic DMARDs. It should also be noted that only one reader scored all images. For observational studies like this, one experienced reader is considered appropriate. However, this reader could have over- or underscored the radiographic damage. It is difficult to estimate how this limitation could have influenced the direction of the relationship for all types of lesions.

Additionally, (sub)luxation in this study was rare. Only a few patients showed (sub)luxation and we may have missed a potentially relevant association between this type of damage and disability measurements. Patients with a follow-up duration of more than 10 years treated in the pre-biologic era had ample time to develop (sub)luxation, which they did not. It is to

be expected that with increasing attention for early treatment and treat-to-target principles, (sub)luxation may disappear from the radar [29].

Finally, some methodological limitations need to be considered. First, patients in the EURIDISS cohort were included according to specified inclusion criteria, but the patients included in the current analyses can be considered as a convenience sample based on loss to follow-up as in all longitudinal cohort studies. Thus a formal sample size calculation was not performed a priori. As a consequence, these data might not be appropriate to detect very subtle associations between joint damage and function. On the other hand, we used sensitive analytical methodology to find subtle relationships, and the flip-side of this type of post-hoc analysis obviously is that you may occasionally find spurious associations that cannot be confirmed. In this respect, it is important to state that our findings are in agreement with, and therefore supportive of, previous findings by others [22]. Second, we have applied last observation carried forward, which in general is not the most appropriate method of imputation in radiographic datasets. However, in this particular analytical situation we had very few missing joints, so we did not find it necessary to apply other imputation methods (such as multiple imputation).

In summary, all different types of radiographically visible joint damage interfere with important functions of daily living in patients with RA. True JSN, especially of the wrist, contributes more to hand function than erosion and (sub)luxation, while all three types of radiographic damage contribute similarly to overall disability.

REFERENCES

1. Nordgren B, Friden C, Demmelmaier I, et al. Long-term health-enhancing physical activity in rheumatoid arthritis--the PARA 2010 study. BMC Public Health 2012;12:397.
2. Radner H, Smolen JS, Aletaha D. Comorbidity affects all domains of physical function and quality of life in patients with rheumatoid arthritis. Rheumatology (Oxford) 2011;50:381-8.
3. Bombardier C, Barbieri M, Parthan A, et al. The relationship between joint damage and functional disability in rheumatoid arthritis: a systematic review. Ann Rheum Dis 2012;71:836-44.
4. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.
5. Guideline on clinical investigation of medicinal products other than NSAIDs for treatment of rheumatoid arthritis. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/12/WC500119785.pdf.)
6. Scott DL, Pugner K, Kaarela K, et al. The links between joint damage and disability in rheumatoid arthritis. Rheumatology (Oxford) 2000;39:122-32.
7. Ødegard S, Landewe R, van der Heijde D, et al. Association of early radiographic damage with impaired physical function in rheumatoid arthritis: a ten-year, longitudinal observational study in 238 patients. Arthritis Rheum 2006;54:68-75.
8. Cohen SB, Dore RK, Lane NE, et al. Denosumab treatment effects on structural damage, bone mineral density, and bone turnover in rheumatoid arthritis: a twelve-month, multicenter, randomized, double-blind, placebo-controlled, phase II clinical trial. Arthritis Rheum 2008;58:1299-309.
9. Lillegraven S, van der Heijde D, Uhlig T, et al. What is the clinical relevance of erosions and joint space narrowing in RA? Nat Rev Rheumatol 2012;8:117-20.
10. van der Heijde D. Erosions versus joint space narrowing in rheumatoid arthritis: what do we know? Ann Rheum Dis 2011;70 Suppl 1:i116-8.
11. Kuper HH, van Leeuwen MA, van Riel PL, et al. Radiographic damage in large joints in early rheumatoid arthritis: relationship with radiographic damage in hands and feet, disease activity, and physical disability. Br J Rheumatol 1997;36:855-60.
12. Dellhag B, Bjelle A. A five-year followup of hand function and activities of daily living in rheumatoid arthritis patients. Arthritis Care Res 1999;12:33-41.
13. Sheehy C, Gaffney K, Mukhtyar C. Standardized grip strength as an outcome measure in early rheumatoid arthritis. Scand J Rheumatol 2013;42:289-93.
14. Syversen SW, Goll GL, van der Heijde D, et al. Prediction of radiographic progression in rheumatoid arthritis and the role of antibodies against mutated citrullinated vimentin: results from a 10-year prospective study. Ann Rheum Dis 2010;69:345-51.
15. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
16. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.
17. Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis. The arthritis impact measurement scales. Arthritis Rheum 1980;23:146-52.
18. Maska L, Anderson J, Michaud K. Measures of functional status and quality of life in rheumatoid arthritis: Health Assessment Questionnaire Disability Index (HAQ), Modified Health Assessment Questionnaire (MHAQ), Multidimensional Health Assessment Questionnaire (MDHAQ), Health Assessment Questionnaire II (HAQ-II), Improved Health Assessment Questionnaire (Improved HAQ), and Rheumatoid Arthritis Quality of Life (RAQoL). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S4-13.
19. Twisk JW. Applied longitudinal data analysis for epidemiology: a practical guide: Cambridge University Press; 2013.
20. Peters MJ, van Nes SI, Vanhoutte EK, et al. Revised normative values for grip strength with the Jamar dynamometer. J Peripher Nerv Syst 2011;16:47-50.
21. Aletaha D, Funovits J, Smolen JS. Physical disability in rheumatoid arthritis is associated with cartilage damage rather than bone destruction. Ann Rheum Dis 2011;70:733-9.
22. Koevoets R, Dirven L, Klarenbeek NB, et al. 'Insights in the relationship of joint space narrowing versus erosive joint damage and physical functioning of patients with RA'. Ann Rheum Dis 2013;72:870-4.
23. Smolen JS, van der Heijde D, Keystone EC, et al. Association of joint space narrowing with impairment of physical function and work ability in patients with early rheumatoid arthritis: protection beyond disease control by adalimumab plus methotrexate. Ann Rheum Dis 2013;72:1156-62.
24. Landewe R, van der Heijde D. Joint space narrowing, cartilage and physical function: are we deceived by measurements and distributions? Ann Rheum Dis 2011;70:717-8.
25. Lie BA, Viken MK, Ødegard S, et al. Associations between the PTPN22 1858C->T polymorphism and radiographic joint destruction in patients with rheumatoid arthritis: results from a 10-year longitudinal study. Ann Rheum Dis 2007;66:1604-9.
26. Syversen SW, Gaarder PI, Goll GL, et al. High anti-cyclic citrullinated peptide levels and an algorithm of four variables predict radiographic progression in patients with rheumatoid arthritis: results from a 10-year longitudinal study. Ann Rheum Dis 2008;67:212-7.

27. Syversen SW, Goll GL, van der Heijde D, et al. Cartilage and bone biomarkers in rheumatoid arthritis: prediction of 10-year radiographic progression. J Rheumatol 2009;36:266-72.

28. Provan S, Angel K, Semb AG, et al. NT-proBNP predicts mortality in patients with rheumatoid arthritis: results from 10-year follow-up of the EURIDISS study. Ann Rheum Dis 2010;69:1946-50.

29. Rahman MU, Buchanan J, Doyle MK, et al. Changes in patient characteristics in anti-tumour necrosis factor clinical trials for rheumatoid arthritis: results of an analysis of the literature over the past 16 years. Ann Rheum Dis 2011;70:1631-40.

4

Rate of adjudication of radiological progression in rheumatoid arthritis randomized controlled trials depending on preset limits of agreement: a pooled analysis from 15 randomized trials

V. Navarro-Compán, R. Landewé, H.A. Ahmad, C.G. Miller, D. Xu, R. Wolterbeek, D. van der Heijde

Rheumatology (Oxford) 2013;52:1404-7

ABSTRACT

Objective

The aim of this study is to provide data on the adjudication rate for a predetermined threshold of difference in change score between two readers in randomized controlled trials (RCTs).

Methods

Fifteen datasets from RCTs in RA were scored by 13 experienced readers as pairs according to the modified Sharp-van der Heijde method. The theoretical adjudication rates for thresholds of between 3 and 20 units were calculated. We investigated the influence of the number of time points within the same session, the length of the interval and disease duration on the adjudication rates.

Results

A total of 21,295 time points from 7,643 patients from 15 databases were included in the analysis. The adjudication rate was inversely related to the threshold. Higher adjudication rates were observed with a higher number of time points, longer time intervals, and in early versus established RA. The adjudication rates ranged from 0% to 22% depending on the scenario.

Conclusions

With trained and experienced readers, the adjudication rate in RA RCTs is low even with very conservative adjudication thresholds.

INTRODUCTION

Radiographically defined joint damage is an important outcome in RA, as it is the consequence of joint inflammation and is related to disability [1]. Prevention of structural damage is included as one of the claims in the US Food and Drug Administration (FDA) [2] and the European Medicines Agency (EMA) Guidelines on clinical investigation of new drugs for treatment of RA [3].

The gold standard imaging modality to assess the degree of structural damage attributable to RA is conventional radiography [4]. Several methodologies have been developed to quantify the radiographic progression [5], but the modified Sharp methods are most frequently used for registration purposes of new drugs. The Sharp-van der Heijde (SHS) method [6] is the recommended method by the EMA [3].

The quality of a scoring method is defined partially by its precision, which is the ability of the measurement to be consistently reproduced [7]. Moreover, the degree of baseline and progression of radiographic damage in patients participating in RA randomized controlled trials (RCTs) is decreasing over time [8]. Precision may be adversely affected by inter-observer variability. Common measures to restrain measurement error include the use of two readers and averaging their scores. If a difference in change score between readers exceeds a certain threshold, adjudication using a third reader or rereading films could be used to further increase precision.

The choice of threshold for adjudication is arbitrary and has historically been set at an inter-reader difference of 7-15 SHS units as compared with baseline. However, there are no published reports regarding the impact of a particular threshold nor which threshold is most optimal for a particular clinical trial. Regulatory authorities have expressed concerns if 20% or more of the cases are the result of adjudication for a given clinical trial. To our knowledge, no previous study has analyzed data to determine the number of cases that would result in adjudication at a predefined difference in change score between readers.

The principal aim of this study is to provide data on how a selected threshold for adjudication results in a specific adjudication rate between two trained readers using the SHS method. The secondary aim is to investigate if additional factors such as the number of time points within the same reading session, the length of the interval between baseline and follow-up visits or disease duration of patients included in the studies influence the adjudication rate.

METHODS

Data were extracted from 15 databases from RCTs for the approval of biologic treatments in patients with RA. We selected these studies because they had been used for product registration purposes using similar methodologies and evaluated according to the SHS method and detailed data were available. Thirteen experienced readers who received the same training had scored all digitized films as pairs on a 21 CFR Part 11 compliant read system deployed by BioClinica, Inc. The readers were blinded to patient identification, treatment, and order of the time points.

Available information for each RCT included the following variables: study duration, number and time interval of films, mean disease duration of the patients and erosion and joint space narrowing raw score per joint according to the SHS method.

The total SHS was calculated per visit and per reader, to obtain the change score per reader between baseline and all follow-up visits. Then, considering all treatment groups as one per trial, the adjudication rate per study was calculated for thresholds between 3 and 20 units for the difference in SHS change score between two readers from baseline to all follow-up visits.

We further investigated whether the number of time points within the same reading session, the length of the interval between baseline and follow-up visits and the disease duration of patients included in the studies influenced the adjudication rate. A cut-off of 3 years for mean disease duration was used to differentiate between RCTs in early and established RA studies.

Statistical analysis

For descriptive purposes, characteristics of the separate RCTs are presented by mean ± standard deviation (SD), while data across studies are shown as medians [interquartile range (IQR)/range]. The datasets were analyzed to yield the theoretical number of cases, represented as a percentage, that would be adjudicated to a third reader based on a predefined threshold for difference of change score between two readers as compared with baseline. For this purpose, we applied a marginal model using generalized estimating equations having identified study as cluster level and patients clustered within studies. Statistical analysis was done using SPSS software version 20.0.

RESULTS

Initially, 23,672 time points and 4,010,390 joint scores from 8,435 patients were included in datasets. From all RCTs, 2 studies had two time points, 10 studies had three time points and 3 studies had four time points analysed. A total of 10,577 (0.26%) joint scores from 131 (1.6%) patients were missing, primarily due to surgery or joint replacement. The difference in change score between the two readers was missing for 792 (9.4%) patients because at least the score of one time point for one of the readers was not available. The median (range) number- and percentage of missing patients for all studies was 48 (0-161) patients and 8.6% (0-19.1), respectively. Finally, a total of 21,295 time points from 7,643 patients were included in this analysis. The median (range) sample size of the studies was 517 (103-901) patients, and the number of time points within one reading session was two for 1,172 patients, three for 5,296 patients and four for 1,175 patients. Median (IQR) baseline radiological damage and progression scores across all studies were 32 (18-48) and 1.05 (0.5-2.0), respectively. Detailed characteristics of studies are shown in supplementary Table 1.

The adjudication rate was inversely related with the threshold for difference in change score from baseline between two readers (Figure 1), and this relationship was stronger as the thresholds decreased. The number of time points within the same reading session influenced the adjudication rate as follows: the higher the number of time points, the higher the adjudication rate for a given threshold.

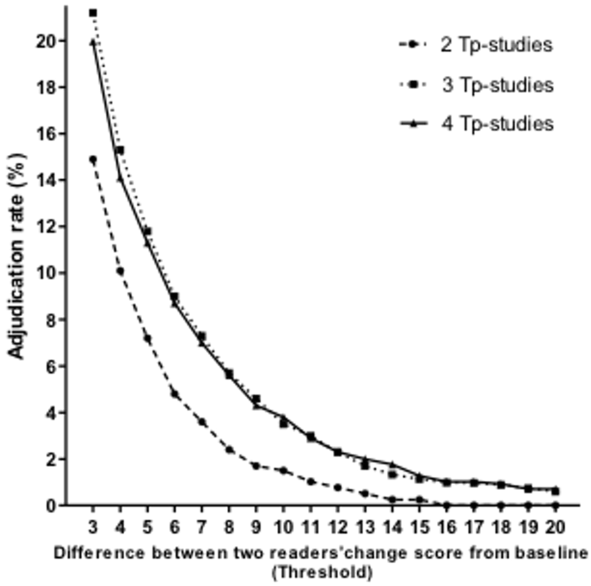


Figure 1: Adjudication rate in relation to the threshold based on the number of time points (Tp) of the studies.

The length of the interval between baseline and time points also influenced the adjudication rate, especially when the allowed difference in change score from baseline between two readers was strict. The longer the time interval, the higher the adjudication rate, but it remained below 22% for all intervals and thresholds.

Four RCTs were classified as early RA while 11 studies included patients with established RA. For the same threshold, the adjudication rate was higher for the early RA studies versus the established ones, especially for low thresholds (Figure 2). In all cases, the adjudication rate remained below 20%, even with a very conservative threshold selection.

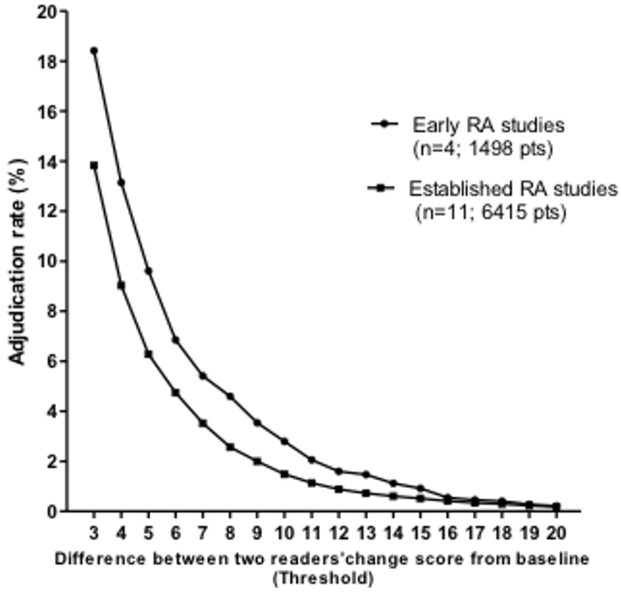


Figure 2: Adjudication rate at first follow-up time point based on disease duration.

DISCUSSION

This study shows the number of cases that would result in adjudication for a predetermined threshold of difference in change score between two readers in RA RCTs using the SHS score. It also provides rational data to select the optimal threshold when designing the imaging plan for a prospective trial. As expected, the percentage of adjudicated cases was inversely related to the threshold, and in most cases it remained below 20%, which has been mentioned by regulatory authorities as a reason for concern, even with a very low threshold.

The number of time points within the same reading session and the length of the interval between baseline and follow-up time points increased the adjudication rate; this was especially true when looking at low thresholds. The selection based on 3 years (in contrast to 1 or 2 years) disease duration might have influenced this result, as well as the relatively low number of trials in this group (n=4). Moreover, the percentage of adjudicated cases was higher in early RA RCTs compared to the established ones. Although the subsequent course of radiological progression in RA is highly variable [9], usually radiological damage increases over time and patients with a longer disease duration usually have higher SHS scores [10]. However, the adjudication rate also remained below 20% in both groups.

One limitation of this study is that imputed scores were ignored, which may have influenced the adjudication rates. But only less than 0.3% of joints were missing, and we do not feel that ignoring this low number of missing values has influenced the difference in change score between readers. Additionally, all the readers were experienced and had received the same training and all used the SHS scoring method. It is unknown if these results could be extrapolated to other readers using different training or scoring systems. Moreover, all images were scored with unknown time order, so these results cannot be applied when images are scored with known time order. However, the SHS method is one of the methods that is accepted by agencies for the approval of a new treatment in patients with RA, being the setting that we explored in this study [11]. Another possible limitation is that we selected a cut-off period of 3 years to define early RA, while a cut-off of 1 or 2 years is often used [12]. The reason was to include a representative number of patients in both groups.

In conclusion, the results of this study give guidance as to the adjudication rate that can be expected given a predetermined threshold for the difference in change score between two readers in RA RCTs using the SHS score. When using trained and experienced readers, the adjudication rate in RCTs with patients with RA is quite low, even with very conservative thresholds less than 20%, which adds to the validity of the SHS method as a precise and reliable outcome measure in RA.

REFERENCES

1. Scott DL, Pugner K, Kaarela K, et al. The links between joint damage and disability in rheumatoid arthritis. Rheumatology (Oxford) 2000;39:122-32.
2. US Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices and biological products for the treatment of rheumatoid arthritis. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071579.pdf.)
3. Guideline on clinical investigation of medicinal products other than NSAIDs for treatment of rheumatoid arthritis. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/12/WC500119785.pdf.)
4. van der Heijde D. Radiographic imaging: the 'gold standard' for assessment of disease progression in rheumatoid arthritis. Rheumatology (Oxford) 2000;39 Suppl 1:9-16.
5. van der Heijde D. Plain X-rays in rheumatoid arthritis: overview of scoring methods, their reliability and applicability. Baillieres Clin Rheumatol 1996;10:435-53.
6. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.
7. Swinkels HL, Laan RF, van 't Hof MA, et al. Modified Sharp method: factors influencing reproducibility and variability. Semin Arthritis Rheum 2001;31:176-90.
8. Rahman MU, Buchanan J, Doyle MK, et al. Changes in patient characteristics in anti-tumour necrosis factor clinical trials for rheumatoid arthritis: results of an analysis of the literature over the past 16 years. Ann Rheum Dis 2011;70:1631-40.
9. Plant MJ, Jones PW, Saklatvala J, et al. Patterns of radiological progression in early rheumatoid arthritis: results of an 8 year prospective study. J Rheumatol 1998;25:417-26.
10. Lukas C, van der Heijde D, Fatenajad S, et al. Repair of erosions occurs almost exclusively in damaged joints without swelling. Ann Rheum Dis 2010;69:851-5.
11. van der Heijde D, Boonen A, Boers M, et al. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. Rheumatology (Oxford) 1999;38:1213-20.
12. van der Horst-Bruinsma IE, Speyer I, Visser H, et al. Diagnosis and course of early-onset arthritis: results of a special early arthritis clinic compared to routine patient care. Br J Rheumatol 1998;37:1084-8.

SUPPLEMENTARY MATERIAL

Supplementary Table 1: Study Characteristics.

Study	Visits (n)	Initial patients (n)	Missed patients (n)	Included patients (n)	Mean disease duration (years)	Baseline SHS (mean ± SD)	SHS progression (mean ± SD)
1	2	546	3	543	6	47.7 ± 55.0	0.91 ± 3.03
2	3	244	0	244	6	53.4 ± 53.1	2.00 ± 4.91
3	3	1008	87	921	7	31.5 ± 50.2	0.53 ± 3.59
4	4	509	0	509	7	29.1 ± 48.0	0.71 ± 5.12
5	3	839	103	736	12	63.6 ± 71.2	1.76 ± 5.35
6	4	299	4	295	12	55.9 ± 60.5	2.03 ± 7.00
7	3	607	48	559	1	13.2 ± 26.8	1.05 ± 5.79
8	4	371	0	371	1	13.2 ± 28.1	1.29 ± 7.05
9	3	123	20	103	2	17.7 ± 20.1	5.71 ± 9.30
10	3	636	77	559	6	18.7 ± 32.8	0.70 ± 4.03
11	3	636	119	517	6	14.4 ± 27.0	0.40 ± 4.15
12	3	444	85	359	8	29.7 ± 43.6	1.06 ± 4.88
13	3	994	161	833	6	37.8 ± 46.4	0.35 ± 3.39
14	2	629	0	629	6	43.8 ± 54.2	0.53 ± 6.83
15	3	550	85	465	3	43.6 ± 43.3	4.48 ± 10.27

5

Measurement error in the assessment of radiographic progression in rheumatoid arthritis clinical trials: the smallest detectable change revisited

V. Navarro-Compán, D. van der Heijde, H.A. Ahmad, C.G. Miller,
R. Wolterbeek, R. Landewé

Ann Rheum Dis 2014;73:1067-70

ABSTRACT

Objectives

To evaluate if the mean smallest detectable change (SDC) of multiple time intervals using the Bland & Altman (B&A) levels of agreement (LoA) method is an appropriate surrogate for the generalizability analysis method for estimating the overall SDC of radiological progression in rheumatoid arthritis (RA) trials. Secondly, to compare the SDC based on 95% LoA with the SDC based on 80% LoA, and to investigate the association between SDC and baseline damage and progression.

Methods

Fifteen datasets from randomized controlled trials in RA were scored by 13 experienced readers as pairs according to the modified Sharp-van der Heijde method. The SDC using the 95% and 80% LoA and the generalizability methods were calculated.

Results

21,295 radiographic time points from 7,643 patients were included. The mean (range) SDC for the LoA and the generalizability methods was 3.1 (2.3-4.3) and 3.2 (2.3-4.6) units, respectively. The mean ± SD difference between the two methods was -0.13 ± 0.28. The mean SDC including all intervals (n=31) was 3.0 ± 0.7 for 95% LoA and 2.0 ± 0.4 for 80% LoA. No relationship was observed between baseline damage and the SDC, whereas the SDC increased with increasing radiological progression.

Conclusions

The mean of the interval SDCs obtained by the simple LoA method is a valid surrogate for the SDC obtained by complex generalizability methods. The SDC depends on the level of radiographic progression rather than on the level of absolute damage. In addition, the use of an SDC based on 80% rather than on 95% LoA is proposed.

INTRODUCTION

Prevention of structural damage has been included as one of the claims in the US Food and Drug Administration [1] and the European Medicines Agency (EMA) Guidelines on clinical investigation of new drugs for treatment of rheumatoid arthritis (RA) [2]. The gold standard imaging technique for assessing the degree of structural damage is conventional radiography [3], and the Sharp-van der Heijde (SHS) method recommended by the EMA for assessing radiological progression [2].

Reliability of the scoring method is essential to be able to detect differences in radiological progression between treatment arms, in order to assess the efficacy of therapeutic interventions in randomized controlled trials (RCTs). Reliability can be reported in relative terms using statistics such as the intraclass correlation coefficients; however, descriptions of other statistics such as the smallest detectable difference (SDD) and the smallest detectable change (SDC) is recommended because they provide an estimation of absolute rather than relative reliability, and they may give clinical guidance for assessing real changes at the individual patient level [4]. While the SDD has been recommended as one of the measures in the guidelines on reporting radiographic data of RCTs in RA, the SDC is nowadays recognized as the preferred measure for absolute agreement [5,6]. The SDD is appropriate to determine if progression in patient A is different from progression in patient B. In order to determine if progression in an individual patient is beyond measurement error, however, the SDC is the most appropriate statistic [5,6].

At least two analytical methods for estimating the SDC are available: one 'simple' method is based on the standard deviation (SD) of the difference between change scores obtained by two readers resulting in 95% levels of agreement (LoA) (also referred to as the Bland & Altman (B&A) method); the other method is more complex and based on generalizability analysis (the analysis of variance (ANOVA) method).

Two arguments challenge the methodology of obtaining SDC cut-off levels as appropriate surrogates for inter-reader reliability:

1) The simple LoA method is only applicable if two scores (twice scored by the same observer or two observers) are obtained. In the case of multiple time points or multiple readers (complex databases), which is common in RA trials, only a generalizability analysis is appropriate. However, estimating the SDC in complex databases requires more statistical expertise and is more laborious, and a simpler method is warranted.

2) The SDC is calculated using 95% LoA, basically assuming that 95% of the inter-reader differences of paired observations in a scenario with two readers is captured within the area delineated by the upper and lower 95% LoA, and not more than 5% of the differences

are more extreme. It can be argued that this requirement is rather strict. For example, it has been shown that the SDD (SDC multiplied by SQRT2) is a conservative estimate, as rheumatologists have rated progression at or below this level as clinically significant [7]. Further, there is no scientific basis for choosing a 95% limit over a less strict limit, and one may argue that the use of 80% LoA is not only sufficiently strict to select a cut-off to determine if a patient shows progression beyond measurement error, but is also closer to reality in terms of what clinicians consider relevant.

The principal aim of this study is twofold: first, to evaluate if the mean SDC of multiple time intervals in complex databases using the 'simple' LoA method per interval is an appropriate surrogate for the generalizability analysis for estimating the overall SDC of radiological progression; second, to compare the SDC based on 95% LoA with the SDC based on 80% LoA, and to investigate the association between baseline radiological damage/radiological progression and the magnitude of the SDC.

METHODS

Data were extracted from 15 databases of RCTs testing biological treatments in patients with RA. All these trials were performed according to good clinical practice and all studies received ethical approval. We selected these studies because they had been used for registration purposes using similar methodologies and all scored by members of our group according to the SHS method [8]. Thirteen experienced readers, who had all received the same training, scored all digitized films in pairs on a 21 CFR Part 11 compliant read system deployed by BioClinica, Code of Federal Regulations. The readers were blinded to patient identification, treatment and chronology of the time points. Initially, the total SHS for all patients was calculated per visit and per reader for all visits. Next, the SDC was calculated using the simple LoA [9] and a generalizability analysis as follows.

LoA (B&A method)

First, the change score per reader was calculated on a per time-point basis [baseline-first follow-up, first follow-up-second follow-up and second follow-up-third follow-up (if applicable)], and subsequently the difference in change scores between the two readers was calculated. Second, the SD of that difference was calculated. The SDC for all intervals of each trial was estimated using the formula $(\pm 1.96*SD)/(\sqrt{2}*\sqrt{k})$ for 95% LoA and $(\pm 1.28*SD)/(\sqrt{2}*\sqrt{k})$ for 80% LoA, in which k represents the number of readers within the same reading session (equals 2 in this study). Finally, we estimated a mean 95% LoA SDC per study by calculating the average of the 95% LoA SDCs of all intervals of the study $(SDC^{1st\ interval} + SDC^{2nd\ interval} + SDC^{3rd\ interval} \ldots\ldots + SDC^{n\ interval})/n$.

Generalizability analysis (ANOVA method)

For the generalizability analysis, we performed an ANOVA as proposed by Bruynesteyn et al [6]. Random variation in change scores (the residual error) per trial was determined, taking into account all the time points from the same trial, using a full-factorial univariate linear model, as detailed in the statistical analysis below. The standard error of the mean (SEM) was calculated by taking the square root of this residual error and the SDC for all intervals of each trial was estimated using the formula (±1.96*SEM)/√k for 95% LoA and (±1.28*SEM)/√k for 80% LoA, where k represents the number of readers (equals 2 in this study).

To compare the B&A method with the ANOVA method, we excluded studies with only two time points (n=2) from this analysis.

Statistical analysis

For descriptive purpose, the values including the characteristics of all RCTs are presented as median (interquartile range -IQR-). All treatment arms were considered as one per trial. The variance components (including residual error) were estimated by three-way ANOVA, with change score between two time points per reader as the dependent variable, patient and reader as random factors, and time interval as fixed factor and all possible interactions (patient*reader, patient*time interval and patient*reader*time interval) were also included in the ANOVA to obtain the residual error components. Statistical analysis was performed using SPSS software version 18.0.

RESULTS

A total of 21,295 time points from 7,643 patients were included in the analysis. From all RCTs, two studies had two time points, 10 studies had three time points and three studies had four time points. The median (range) sample size of the studies was 517 (103-921) patients, and the number of time points within one reading session was two for 1,172 patients, three for 5,296 patients and four for 1,175 patients. The median (IQR) disease duration of patients included in the studies was 6 (3-7) years, and the median (IQR) baseline radiological damage and progression in SHS to last follow-up across all studies was 32 (18-48) and 1.1 (0.5-2.0), respectively.

Since the principal aim of this study was to propose a surrogate for the ANOVA method for calculating the SDC when more than two time points are scored within the same reading session, we evaluated the agreement between the two different methods (LoA method and ANOVA method) employing a B&A plot; this means plotting the difference between the methods against their mean as shown in Figure 1. The mean (range) SDC over the included studies based on the 95% LoA and ANOVA methods was 3.1 (2.3-4.3) and 3.2 (2.3-4.6) units,

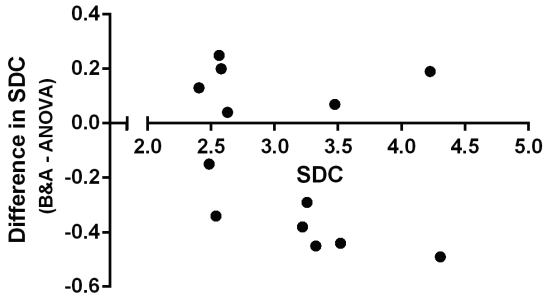


Figure 1: Difference in smallest detectable change (SDC) between the Bland & Altman (B&A) levels of agreement (LoA) method (mean of all intervals) and the analysis of variance (ANOVA) method (taking into account all intervals), for trials with two or more intervals.

respectively. The mean ± SD difference between the two methods was -0.13 ± 0.28, range (-0.48, 0.25) units. The mean of the SDC for all studies was somewhat higher for the ANOVA method (not statistically significant). No particular trend was observed, and therefore the difference between the two methods did not tend to get larger (or smaller) as the average discrepancy increased. The variability was also consistent along the range of observations (homoscedasticity of the scatter). Moreover, median values for the difference between the 95% LoA and ANOVA methods were higher in studies with less radiographic damage at baseline and less radiographic progression compared with studies with more radiographic damage and progression (-0.22 vs 0.07 and -0.22 vs 0.04, respectively), but no differences in the range was observed (supplementary Table S1).

Second, we compared the SDC based on the 95% LoA with the SDC based on the 80% LoA using the LoA method. Figure 2 shows the SDC values for all intervals and studies based on both LoAs. The median (range) difference between the 95% and 80% LoA SDCs was 1.1 (0.8-1.6) for the first interval, 0.9 (0.7-1.5) for the second interval and 1.3 (1.1-1.4) for the third interval. The mean ± SD SDC including all the SDCs calculated for all intermediate intervals (n=31) was 3.0 ± 0.7 for 95% LoA and 2.0 ± 0.4 for 80% LoA.

Finally, we also investigated if there was an association between baseline radiological damage and radiographic progression to last follow-up with the SDC. We did not observe any relationship between the degree of damage at baseline (SHS) and the SDC (r^2= 0.01, p=0.8) (Figure 3a), while an association between radiological progression and the SDC was obvious (r^2= 0.64, p<0.001) (Figure 3b), indicating that the SDC is higher in trials with more progression, although this relationship is strongly influenced by two trials with the highest progression rate.

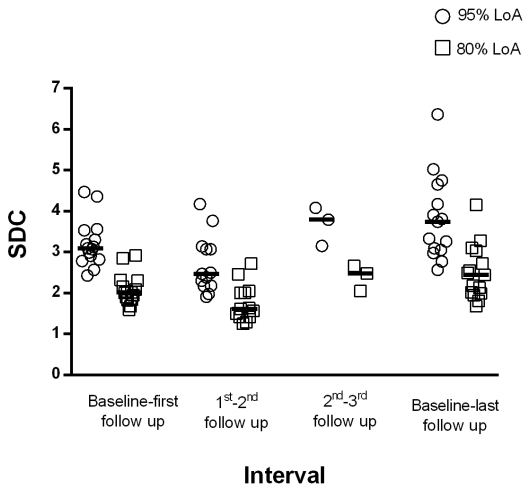


Figure 2: Smallest detectable change (SDC) for 95% and 80% levels of agreement (LoA) based on the LoA method.

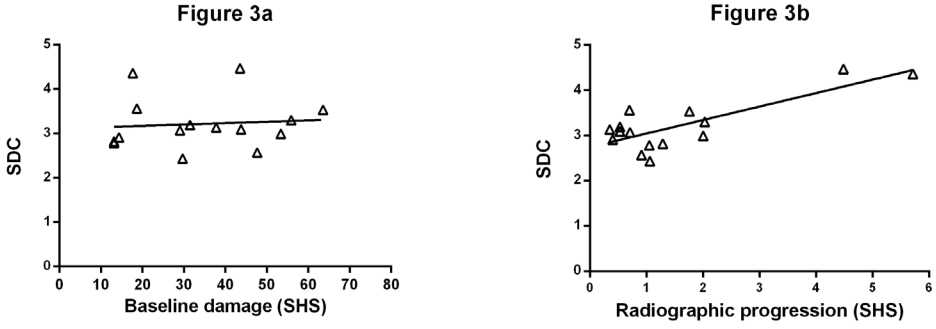


Figure 3: Association between baseline radiographic damage and radiographic progression with smallest detectable change (SDC) for 95% level of agreement. Regression lines summarize the association between radiographic damage (Figure 3a) and radiographic progression to last follow-up (Figure 3b) with the SDC. SHS: Sharp-van der Heijde score.

DISCUSSION

The results of this analysis suggest that, in complex databases with multiple time points and time intervals, the mean of the interval SDCs obtained by the simple LoA method is a valid surrogate for the 'umbrella SDC' obtained by complex methods based on the generalizability theory. Further, we have found arguments that the SDC is dependent on the level of radiographic progression in a trial rather than on the level of absolute damage. In addition, we here propose to consider the use of an SDC based on 80% rather than 95% LoA, for reasons explained below.

The maximum discrepancy in SDCs of 0.48 units when calculated by the two methods, and the systematic difference of only 0.13 units is negligible, in the light of the minimal clinically important difference (MCID) of 3-4.5 units for radiographic progression [10]. On the other hand, we should take into account the fact that this MCID cut-off was selected based on results of a study performed when biological agents were just entering clinical use and therefore when tolerance for progressive joint damage was less strict. However, we consider it very unlikely that an updated MCID would even approach 0.48 units.

Further, there was consistency in variability, and no particular trend was observed when the two methods were compared, which adds to the validity of the mean LoA method. Obviously, the mean LoA method has important advantages in that it is simpler, less time consuming, and more familiar to researchers.

With respect to the proposal to base SDCs on the 80% LoA, it can be argued that there is no solid scientific basis to choose a 95% instead of a lower LoA. The 95% cut-off level has its basis in distribution theory, where it is a boundary for including 95% of observations of a distribution with standard normal ('bell-shaped') properties (the mean ± 2 SDs), and was therefore probably chosen because it resembles the 95% confidence interval (CI) used in statistical hypothesis testing. Conceptually, though, CI and LoA are not related: whereas 95% CI statistically tests the null hypothesis that the mean difference in change scores obtained by two readers is zero, the 95% LoA quantifies the boundaries that include 95% of all paired observations and has nothing in common with hypothesis testing [11]. The justification for choosing boundaries other than 95% as LoA depends on the relevance of avoiding potential misclassifications. In radiographic analysis, the SDC concept is used to determine whether a patient is a 'true progressor' (i.e. progression beyond reasonable measurement error) or not (i.e. progression still compatible with measurement error, and therefore classified as zero progression). If an 80% LoA is accepted as the basis for the SDC, and the SDC is accepted as the level that distinguishes 'progression beyond measurement error' from 'progression still compatible with measurement error', more patients will be accepted as 'true progressors'. Obviously, there will also be some more 'progressors' for whom progression is due to measurement error, but this misclassification will affect both arms of an RCT in an unbiased

manner. Given the context of the RCT, in which a treatment is tested against a comparator for its potential to *avoid* radiographic progression, and the current mean progression scores observed in such trials, it is unlikely that a cut-off level based on an 80% SDC will spuriously influence the trial results. In fact, a trial with higher percentages of patients with progression per trial arm may provide increased conservatism, which is advantageous from the perspective of internal validity of a trial. In the light of the well-recognized phenomenon of deflating radiographic progression rates over time in clinical trials [12], increased rates of 'progressors' per trial arm using more lenient SDC cut-offs is advantageous for the statistical power of a trial. Increased misclassification is unlikely to be relevant here, since one may expect that these misclassifications will be evenly distributed among trial arms. However, the ultimate effect will depend on the analyses and the degree of misclassification.

It is therefore proposed to use 80% LoA-based SDCs instead of 95% LoA-based SDCs, so that measurement error is substantially lower than the change in radiographic damage that rheumatologists consider clinically relevant: approximately 3 units [10].

Another observation of note was that the degree of joint damage at baseline did not influence the SDC, whereas the level of radiographic progression did have a slight influence on the SDC, that is, the SDC tended to increase with increasing radiographic progression rates. The first observation is somewhat unexpected, since readers usually recognize unaffected joints relatively easily and in general achieve a high level of agreement, while they have to make far more decisions in case of multiple affected joints with different states of joint involvement. An explanation could be that the studies included in this analysis covered a relatively small range of potential involvement at baseline (10 to 65 units). SDCs may therefore still be relatively low if baseline joint damage is low to moderate, but increase if baseline damage exceeds 65 units. Moreover, trials with even lower baseline damage were not included and therefore baseline damage below 10 could not be tested. This study does not provide resolution for this. The second observation of increasing SDCs with increasing progression rates is in compliance with what has been found in detailed analyses. In a recent analysis of the TEMPO trial [13], with four independent reads of the same patient, a very high level of agreement was reached for the great majority of individual joints that showed zero progression in SHS. In contrast, agreement on a per joint basis was poor in those joints that were scored as 'progressive' by at least one of the four readers. This lack of agreement is lost when total per-patient scores are calculated, as is standard practice or in evaluation in RCTs, explaining increasing SDCs with increased progression rates. A limitation of this observation is that this positive correlation was largely determined by two trials.

A limitation of our study is that the maximum number of time points within the same reading session in RCTs included was four, so we do not know if these results would be applicable if five or more time points are present. However, not many clinical trials include more than four time points in one read campaign, and the question is therefore rather theoretical.

Importantly, all images were scored in unknown chronological order by experienced readers, and these results cannot be extrapolated to reads with known time order and to reads by inexperienced readers or those that have not been trained similarly. Although a recent study suggests that chronological reading is more precise than random reading [14], regulatory agencies still require radiographs to be scored randomly. Random scoring is therefore still considered the reference setting [15]. Moreover, although not tested, it may be assumed that the issues addressed in this paper are equally applicable to studies scored in chronological order, as the topics under investigation in this manuscript are not directly influenced by the (un)blinding of the time order.

In conclusion, for reasons of convenience, we propose to report the mean of all interval SDCs as an appropriate surrogate for the ANOVA-based SDC in trials with multiple time points. In addition, we consider an SDC based on an 80% LoA to be an acceptable alternative to an SDC based on a 95% LoA. For the SHS method, based on these large datasets involving many different readers, we propose a cut-off level of 3.0 units for a 95% LoA SDC and of 2.0 units for an 80% LoA SDC as the threshold for deciding if the RA of an individual patient shows radiographic progression.

REFERENCES

1. US Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices and biological products for the treatment of rheumatoid arthritis. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071579.pdf.)
2. Guideline on clinical investigation of medicinal products other than NSAIDs for treatment of rheumatoid arthritis. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/12/WC500119785.pdf.)
3. van der Heijde DM. Radiographic imaging: the 'gold standard' for assessment of disease progression in rheumatoid arthritis. Rheumatology (Oxford) 2000;39 Suppl 1:9-16.
4. de Vet HC, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033-9.
5. van der Heijde D, Simon L, Smolen J, et al. How to report radiographic data in randomized clinical trials in rheumatoid arthritis: guidelines from a roundtable discussion. Arthritis Rheum 2002;47:215-8.
6. Bruynesteyn K, Boers M, Kostense P, et al. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.
7. Bruynesteyn K, van der Heijde D, Boers M, et al. Minimal clinically important difference in radiological progression of joint damage over 1 year in rheumatoid arthritis: preliminary results of a validation study with clinical experts. J Rheumatol 2001;28:904-10.
8. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.
9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
10. Bruynesteyn K, van der Linden S, Landewe R, et al. Progression of rheumatoid arthritis on plain radiographs judged differently by expert radiologists and rheumatologists. J Rheumatol 2004;31:1088-94.
11. Landewe RB, van der Heijde D. Principles of assessment from a clinical perspective. Best Pract Res Clin Rheumatol 2003;17:365-79.
12. Rahman MU, Buchanan J, Doyle MK, et al. Changes in patient characteristics in anti-tumour necrosis factor clinical trials for rheumatoid arthritis: results of an analysis of the literature over the past 16 years. Ann Rheum Dis 2011;70:1631-40.
13. Lukas C, van der Heijde D, Fatenajad S, et al. Repair of erosions occurs almost exclusively in damaged joints without swelling. Ann Rheum Dis 2010;69:851-5.
14. van Tuyl LH, van der Heijde D, Knol DL, et al. Chronological reading of radiographs in rheumatoid arthritis increases efficiency and does not lead to bias. Ann Rheum Dis 2014;73:391-5.
15. van der Heijde D, Boonen A, Boers M, et al. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. Rheumatology (Oxford) 1999;38:1213-20.

Supplementary Table S1: Characteristics of studies and difference in smallest detectable change between the Bland & Altman levels of agreement method (average of all intervals) and the analysis of variance (ANOVA) method (taking into account all intervals).

Study ID	Mean disease duration (years)	Study duration (weeks)	Baseline SHS (mean ± SD)	SHS progression to last follow up (mean ± SD)	Bland&Altman method	ANOVA method	Difference between two methods
2 Time points							
1	6	24	47.7 ± 55.0	0.91 ± 3.03	-	2.57	-
14	6	24	43.8 ± 54.2	0.53 ± 6.83	-	3.09	-
3 Time points							
2	6	24	53.4 ± 53.1	2.00 ± 4.91	2.69	2.44	0.25
3	7	48	31.5 ± 50.2	0.53 ± 3.59	2.68	2.48	0.20
5	12	48	63.6 ± 71.2	1.76 ± 5.35	3.30	3.74	-0.44
7	1	52	13.2 ± 26.8	1.05 ± 5.79	2.47	2.34	0.13
9	2	52	17.7 ± 20.1	5.71 ± 9.30	4.06	4.55	-0.48
10	6	52	18.7 ± 32.8	0.70 ± 4.03	3.03	3.41	-0.38
11	6	104	14.4 ± 27.0	0.40 ± 4.15	2.41	2.56	-0.15
12	8	104	29.7 ± 43.6	1.06 ± 4.88	2.37	2.71	-0.35
13	6	52	37.8 ± 46.4	0.35 ± 3.39	3.10	3.55	-0.45
15	3	52	43.6 ± 43.3	4.48 ± 10.27	4.32	4.13	0.19
4 Time points							
4	7	96	29.1 ± 48.0	0.71 ± 5.12	3.11	3.40	-0.29
6	12	96	55.9 ± 60.5	2.03 ± 7.00	3.51	3.44	0.07
8	1	108	13.2 ± 28.1	1.29 ± 7.05	2.65	2.61	0.04

5

6

Spondyloarthritis features forecasting the presence of HLA-B27 or sacroiliitis on magnetic resonance imaging in patients with suspected axial spondyloarthritis: results from a cross-sectional study in the ESPeranza cohort

V. Navarro-Compán, E. de Miguel, D. van der Heijde, R. Landewé, R. Almódovar, C. Montilla, E. Beltrán, P. Zarco

Arthritis Res Ther (accepted for publication)

ABSTRACT

Introduction

Chronic back pain (CBP) is frequently the presenting symptom in patients with suspected axial spondyloarthritis (axSpA). Presence of sacroiliitis on magnetic resonance imaging (MRI) or HLA-B27 adds to diagnostic certainty. However, these costly tests cannot be applied in all patients with CBP. This study aims to investigate which SpA features increase the likelihood of a positive HLA-B27 or positive MRI of the sacroiliac-joints (MRI-SI) in patients with suspected axSpA.

Methods

Data from 665 patients with CBP within the ESPeranza Programme were analysed. Diagnostic utility measures (LR+, LR-) for a positive MRI-SI or HLA-B27 were calculated for various definitions of inflammatory back pain (IBP), their separate items and for other SpA features.

Results

Pretest probabilities of a positive result were 41% for MRI-SI and 40% for HLA-B27. For a positive MRI-SI result, the most useful IBP characteristic was alternating buttock pain (LR+=2.6). Among the IBP-criteria, fulfillment of the ASAS criteria (LR+=2.1) was most contributory. Interestingly, the addition of alternating buttock pain to the Calin/ASAS-IBP criteria (LR+=6.0 and 5.5, respectively) or the addition of awakening at second half of night to the Calin-IBP criteria (LR+=5.5) increased the pre-test probability of MRI-sacroiliitis from 41% to 79-80%. Dactylitis (LR+=4.1) and inflammatory bowel disease (IBD) (LR+=6.4) increased this probability to 73% and 81%, respectively. To forecast HLA-B27 positivity, awakening at the second half of the night, fulfillment of the ASAS-IBP definition and uveitis were the most useful, but only marginally predictive (LR+=1.3, 1,6 and 2.6, respectively).

Conclusions

If patients with suspected axial SpA have either 1) IBP according to Calin/ASAS definition plus alternating buttock pain, or 2) IBP according to Calin definition plus awakening at night, or 3) dactylitis or 4) IBD, the probability of finding a positive MRI-SI increases significantly.

INTRODUCTION

Axial spondyloarthritis (axSpA) has a major impact on physical function and quality of life [1]. Nevertheless, despite these important consequences patients with axSpA have been traditionally diagnosed after several years of symptoms [2]. In this sense, magnetic resonance imaging of sacroiliac joints (MRI-SI) has become important in the last decade, especially in the early stages of the disease. Nowadays, performing imaging, and testing human leucocyte antigen B27 (HLA-B27) are among the most important diagnostic procedures in patients with (suspicion of) axSpA. Accordingly, imaging and HLA-B27 results are also the entry criterion for classifying patients with chronic back pain (CBP) as axSpA based on the Assessment of SpondyloArthritis international Society (ASAS) classification criteria [3].

Furthermore, in patients with suspected axSpA the starting point of the disease is usually the presence of CBP. However, CBP is one of the most prevalent symptoms in the general population and therefore it is essential to be able to select which patients with CBP have highest chance of being diagnosed as axSpA. For this purpose, several referral strategies including different manifestations at the beginning of the disease have been developed for primary physicians [4-6], but in clinical practice most patients are referred from primary care to rheumatologists because of inflammatory back pain (IBP) [4]. However, despite several IBP definitions have been published [7-9], there is only poor agreement about the presence of IBP between primary physicians and rheumatologists. In addition to this, the original algorithm for diagnosing axSpA has been recently modified excluding IBP as an obligatory entry criterion. This modification has been applied based on the finding that up to 30% of axSpA patients do not have IBP and therefore its inclusion as obligatory entry criterion results in too many misdiagnoses [10]. According to the modified algorithm, complementary examinations (HLA-B27 and MRI) need to be considered in those patients without sacroiliitis on conventional x-rays or with less than 4 SpA features, which are usually most of the referred patients. But in clinical practice, due to efficiency reasons these tests must be usually restricted to patients with higher probability of a positive result.

Based on clinical features, it would be very helpful for rheumatologists to identify which referred patients have highest likelihood of ultimately being diagnosed as axSpA. Such SpA features that can be obtained by history taking or simple physical examination, could potentially contribute to an efficient 'test-sequence' to be applied in patients presenting with CBP and to optimise the use of supplementary tests in these patients. Based on this, this study aims to investigate which SpA features may increase the pre-test probability of a positive test result of HLA-B27 or MRI-SI in patients with suspected axSpA, which is a step forward in making a diagnosis of axSpA.

METHODS

Study design and population

This study was performed within the context of the ESPeranza Programme. The details from this initiative have been previously reported [11-13]. In summary, the ESPeranza Programme is a Spanish prospective multicenter national health program aiming to facilitate early diagnosis of patients with spondyloarthritis. The program was designed in compliance with the Helsinki Declaration and approved by the Ethical Committee of Research Unit of Hospital Reina Sofía in Córdoba. During three years since April 2008, primary physicians and other specialists were asked by rheumatologists to refer patients meeting the following criteria: 1) age from 18 to 45 years; 2) symptom duration between 3 and 24 months and 3) fulfilling one of the following three symptoms: IBP, asymmetrical arthritis or spinal/joint pain plus the presence of at least one of the following features: psoriasis, inflammatory bowel disease (IBD), anterior uveitis, radiographic sacroiliitis, HLA-B27 positivity, or a family history of spondylitis. All patients signed the informed consent. In total, 25 centers across the country participated in the program and 775 patients met the inclusion criteria. For the current study, only those patients referred with axial symptoms (n=665) were selected.

Data collection and clinical measures

In the ESPeranza programme, socio-demographic and disease data were collected including age, sex, symptom duration, family history of SpA or related diseases, peripheral manifestations (enthesitis, dactylitis and/or arthritis) and extra-articular manifestations (psoriasis, uveitis, IBD). In addition, all typical characteristics of IBP were separately collected including: morning stiffness, improvement of back pain with exercise and not with rest, alternating buttock pain, insidious onset and awakening at second half of night. Moreover, the SpA feature related with CBP 'good response to nonsteroidal anti-inflammatory drugs (NSAIDs)' was collected too. Based on this information, the fulfilment of the different existing IBP criteria was established as follows: Calin (at least 4 out of the following 5: age <40 years, insidious onset, duration >3 months, morning stiffness, improvement with exercise), Berlin (at least 2 out of the following 4: morning stiffness, improve with exercise and not with rest, awakening at second half of night, alternating buttock pain) and ASAS (at least 4 out of the following 5: age <40 years, insidious onset, improvement with exercise, no improvement with rest, night pain).

Clinical assessments were also performed including: disease activity (Bath Ankylosing Spondylitis Disease Activity Index, BASDAI), function (Bath Ankylosing Spondylitis Functional Index, BASFI), laboratory-tests (C-Reactive Protein, CRP). The presence of HLA-B27 was tested in the local lab of each center according to the standard procedure. Regarding imaging, the evaluation of conventional radiographs of the cervical- and lumbar -spine and

of the pelvis was part of the protocol. MRI-SI was not included in the protocol as mandatory. Nevertheless, all rheumatologists were asked to perform MRI-SI if possible. However, due to existing differences between centers on MRI accessibility including budget limitations and waiting lists for MRI, MRI-SI has not been performed all patients. MRI-SIs were read locally by one reader at each hospital, who evaluated the presence or absence of sacroiliitis according to the ASAS definition [14].

Statistical analysis

For descriptive purpose, results for continuous variables are presented as mean and standard deviation (SD) and for categorical variables are shown as percentage and relative frequencies. The diagnostic utility of each IBP characteristic, good response to NSAIDs, IBP definitions (Calin, Berlin, ASAS) and other SpA features for a positive MRI-SI or HLA-B27 was calculated. Diagnostic utility was assessed based on sensitivity, specificity, positive and negative predictive values (PPV and NPV) and especially positive and negative likelihood ratios (LR+; LR-). Importantly, to define a SpA feature as useful, a cut-off value of 4.0 was used for LR+. The reason for the selection of this cut-off is that it has been associated with a moderate increase (approximately 25%) of the likelihood of disease [15]. Finally, post-test probability was calculated based on the positive and negative likelihood ratios using the following conversion formula: Odds = Probability / (1 – Probability). According to Bayes' law (post-test odds = pre-test odds* LR, in which LR- is used if a feature is absent and the LR+ if a feature is present) the likelihood of finding a positive MRI/HLA-B27 was estimated in patients with this particular feature present or absent.

RESULTS

Baseline characteristics

A total of 665 patients with axial symptoms were referred and included in ESPeranza. Fifty five percent were male, mean (SD) age was 33.2 (7.1) years and mean (SD) symptoms duration was 12.6 (6.6) months. Most of them (n=653; 98.2%) had also data collected for HLA-B27 testing and were included in the analysis to evaluate the association between IBP characteristics and the presence of HLA-B27. Approximately half of the patients (n=326; 49%) had MRI-SI performed and could be included in the analysis to investigate the association of IBP characteristics and a positive MRI-SI. Demographic and disease characteristics were similar for patients in in the whole cohort compared with patients in whom MRI-SI had been performed (Table 1). In total, 270 (41%) patients were HLA-B27 positive and 130 (40%) patients had a positive MRI-SI result ('pre-test probabilities').

Table 1: Baseline characteristics for all patients with axial symptoms included in ESPeranza (left column) and for the subgroup of patients with MRI-SI assessed (right column).

Characteristic	Patients with axial symptoms referred to ESPeranza N (%)=665	Patients with axial symptoms and MRI-SI available N (%)= 326
Age (years)	33.2 ± 7.1	32.8 ± 7.0
Male	363 (54.6)	175 (53.7)
Symptoms duration (months)	11.9 ± 6.6	12.6 ± 6.4
Morning stiffness	393 (59.1)	195 (59.8)
Improve with exercise, not with rest	208 (31.3)	97 (29.8)
Alternating buttock pain	197 (29.6)	98 (30.1)
Insidious onset	430 (64.7)	256 (78.5)
Awakening at second half of night	315 (47.4)	169 (51.8)
Response to NSAIDs	407 (61.2)	200 (61.3)
Enthesitis	107 (16.1)	45 (13.8)
Psoriasis	82 (12.3)	35 (10.7)
Dactylitis	26 (3.9)	14 (4.3)
IBD	26 (3.9)	10 (3.1)
Uveitis	34 (5.1)	16 (4.9)
Arthritis	95 (14.3)	37 (11.3)
Family history	170 (25.6)	89 (27.3)
HLA-B27	270 (40.6)	144 (44.2)
Elevated CRP (mg/L)	177 (26.6)	89 (27.3)
BASDAI	4.0 ± 2.3	3.9 ± 2.3
BASFI	2.5 ± 2.4	2.3 ± 2.4
ASAS criteria for axial SpA	291 (43.8)	167 (51.2)
Imaging arm	194 (29.2)	132 (40.5)
MRI-SI positive	85 (12.8)	85 (26.1)
mNY positive*	109 (16.4)	47 (14.4)
Clinical arm only	97 (14.6)	35 (10.7)

MRI-SI: magnetic resonance imaging of sacroiliac joints; NSAIDs: nonsteroidal anti-inflammatory drugs; IBD: Inflammatory bowel disease; HLA-B27: human leucocyte antigen B27; CRP: C-reactive protein; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASFI: Bath Ankylosing Spondylitis Functional Index; ASAS: Assessment of SpondyloArthritis international Society; mNY: modified New York criteria for ankylosing spondylitis.
*Patients with both (MRI-SI and mNY positive) are included.

IBP characteristics forecasting the presence of sacroiliitis on MRI or HLA-B27

Table 2 shows the results for the diagnostic utility of these characteristics. While alternating buttock pain was most contributory in predicting a positive MRI-SI result (LR+= 2.6; LR-= 0.6), awakening at second half of night was the most contributory in predicting a positive HLA-B27 result (LR+= 1.3; LR-= 0.8). But none of them separately had a very strong contribution.

IBP definitions forecasting the presence of sacroiliitis on MRI or HLA-B27

The three multi-item IBP definitions performed more or less similarly in terms of LR+. To forecast a positive MRI-SI, the ASAS criteria performed slightly better but still had very low

Table 2: Values for diagnostic utility measures for each of the inflammatory back pain characteristics in relation to the presence of sacroiliitis on MRI and a positive HLA-B27 testing.

	Sacroiliitis on MRI						HLA-B27					
	Sen	Spe	LR+	LR-	PPV	NPV	Sen	Spe	LR+	LR-	PPV	NPV
Morning stiffness >30 minutes	68.5	45.9	1.3	0.7	45.6	68.7	63.3	43.6	1.1	0.8	44.2	62.8
Improve with exercise, not with rest	31.5	71.4	1.1	0.1	42.3	61.1	33.7	70.2	1.1	0.9	44.4	60.0
Alternating buttock pain	47.7	81.6	2.6	0.6	63.3	70.2	31.9	71.3	1.1	0.9	43.9	59.7
Insidious onset	87.7	27.6	1.2	0.4	44.5	77.1	68.1	37.1	1.1	0.9	43.3	62.3
Awakening at second half of night	64.6	56.6	1.5	0.6	49.7	70.7	55.2	57.4	1.3	0.8	47.8	64.5
Response to NSAIDs	69.2	43.9	1.2	0.7	45.0	68.3	67.4	43.3	1.2	0.8	45.6	65.4

MRI: magnetic resonance imaging; HLA-B27: human leucocyte antigen B27; Sen: sensitivity; Spe: specificity; LR+: positive likelihood ratio; LR-: negative likelihood ratio; PPV: positive predictive value; NPV: negative predictive value; NSAIDs: nonsteroidal anti-inflammatory drugs.

LRs (LR+= 2.1; LR-= 0.7) (Table 3). Based on these results, it was decided to assess whether or not the addition of other IBP characteristics (or good response to NSAIDs) not being part of the IBP definitions could increase their diagnostic utility. This means that patients must fulfil the specific IBP definition plus the additional characteristic. The results for this analysis are depicted in Table 3. Of all possible combinations, the ones that performed better were the combination of the Calin/ASAS-definition and alternating buttock pain (LR+= 6.0 and 5.5, respectively; LR-= 0.7 for both combinations) and the combination of the Calin-definition and awakening at second half of night (LR+= 5.5; LR-= 0.8). These combinations had a post-test probability of sacroiliitis on MRI of 79-80%% if the combined criterion was present, and 32-35%, if the combined criterion was absent.

Similar to sacroiliitis, the three IBP definitions performed similarly well to predict a positive HLA-B27 result (Table 3). The ASAS criteria were best performing but still with a very poor LR+ (1.6) and LR- (0.9). On the other hand, and in contrast to the results for sacroiliitis, the addition of other IBP characteristics (or good response to NSAIDs) to the existing IBP definitions only increased their diagnostic utility slightly.

Other SpA features forecasting the presence of sacroiliitis on MRI or HLA-B27

The utility of SpA features other than IBP in helping to anticipate a positive (or a negative) MRI-SI or a positive (or negative) HLA-B27 test result was evaluated too. These SpA features included peripheral and extra-articular manifestations, family history of SpA or related diseases, elevated CRP and HLA-B27 (only for MRI-SI). While dactylitis and IBD appeared to be the most useful SpA features to forecast a positive MRI-SI (LR+= 4.1 and 6.4, respectively), uveitis was the best for a positive HLA-B27, but had a LR+ of only 2.6 (Table 3). The presence or history of dactylitis or IBD resulted in a post-test probability for a positive MRI of 73% and 81%, respectively compared with the pre-test probability of 40%. The post-test probability went from 40% to 35% if both, dactylitis or IBD were confirmed absent.

Table 3: Values for diagnostic utility measures for each of the inflammatory back pain definitions and other SpA features in relation to the presence of sacroiliitis on MRI and a positive HLA-B27 testing.

	Sacroiliitis on MRI						HLA-B27					
	Sen	Spe	LR+	LR-	PPV	NPV	Sen	Spe	LR+	LR-	PPV	NPV
	IBP Definition											
Calin criteria	51.5	70.9	1.8	0.7	54.0	68.8	36.3	74.2	1.4	0.9	49.7	62.3
Berlin criteria	72.3	50.5	1.5	0.6	49.2	73.3	64.1	49.3	1.3	0.7	47.1	66.1
ASAS criteria	47.7	77.0	2.1	0.7	57.9	68.9	31.5	80.7	1.6	0.9	53.5	62.3
	IBP Definition plus other IBP characteristics											
Calin + alt. buttock	30.8	94.9	6.0	0.7	80.0	67.4	17.4	86.4	1.3	1.0	51.0	56.3
Calin + NSAIDs	36.2	80.1	1.8	0.8	54.7	65.4	28.5	75.7	1.2	0.9	48.8	56.5
Calin + night	28.0	94.9	5.5	0.8	77.8	67.4	16.2	87.9	1.3	1.0	52.3	56.3
Calin + alt. buttock + night	26.9	94.9	5.3	0.8	77.8	66.2	16.0	88.1	1.3	1.0	52.3	56.3
Calin + alt. buttock + NSAIDs	20.0	96.9	6.5	0.8	81.3	64.6	10.4	91.0	1.2	1.0	48.4	55.5
Calin + night + NSAIDs	29.2	87.2	2.3	0.8	60.3	65.0	21.5	88.3	1.8	0.9	56.3	61.5
Berlin + insidious	64.6	62.2	1.7	0.6	53.1	72.6	50.7	53.1	1.1	0.9	46.8	57.0
Berlin + NSAIDs	53.8	68.4	1.7	0.7	53.0	69.1	48.9	66.1	1.4	0.8	50.4	64.7
Berlin + insidious + NSAIDs	47.7	76.0	2.0	0.7	56.9	68.7	34.0	67.2	1.0	1.0	45.8	55.6
ASAS + alt. buttock	30.8	94.4	5.5	0.7	78.4	94.4	18.1	86.4	1.3	0.9	52.0	56.5
ASAS + NSAIDs	36.9	86.2	2.7	0.7	64.0	67.3	24.4	82.0	1.4	0.9	56.4	61.9
ASAS + stiffness	39.2	81.1	2.1	0.7	58.0	66.8	31.3	76.3	1.3	0.9	51.7	57.7
ASAS + alt. buttock + NSAIDs	21.5	96.4	6.0	0.8	80.0	64.9	10.7	94.3	1.9	0.9	56.9	94.3
ASAS + stiffness + NSAIDs	28.5	88.8	2.5	0.8	62.7	65.2	19.4	83.1	1.1	1.0	48.3	55.9
ASAS + stiffness + alt. buttock	26.9	94.9	5.3	0.8	77.8	66.2	16.0	88.1	1.3	1.0	52.3	56.3
	Other SpA features											
Enthesitis	11.5	84.7	0.8	1.0	33.3	59.1	17.8	85.1	1.2	0.9	85.1	45.7
Peripheral arthritis	13.8	90.3	1.4	0.9	48.6	61.2	16.7	87.7	1.4	0.9	87.7	48.9
Dactylitis	23.1	94.4	4.1	0.8	21.4	59.3	4.8	96.9	1.6	09	96.9	52.0
Uveitis	6.2	95.9	1.5	0.9	50.0	60.6	8.1	96.9	2.6	0.9	96.9	64.7
Psoriasis	10.0	88.8	0.9	1.0	37.1	59.8	7.0	83.8	0.4	1.1	83.8	23.5
Inflammatory bowel disease	23.1	96.4	6.4	0.8	30.0	59.8	1.5	94.5	0.3	1.0	94.5	16.0
Family history of SpA	27.7	73.0	1.0	0.9	40.4	60.3	33.7	80.2	1.7	0.8	80.2	54.5
Elevated CRP	40.8	77.5	1.8	0.8	55.1	66.0	38.5	77.1	1.7	0.8	77.1	54.6
HLA-B27	59.7	65.1	1.7	0.6	53.5	70.6	-	-	-	-	-	-

MRI: magnetic resonance imaging; HLA-B27: human leucocyte antigen B27; Sen: sensitivity; Spe: specificity; LR+: positive likelihood ratio; LR-: negative likelihood ratio; PPV: positive predictive value; NPV: negative predictive value; ASAS: Assessment of SpondyloArthritis international Society; NSAIDs: nonsteroidal anti-inflammatory drugs; SpA: Spondyloarthritis; CRP: C-reactive protein.

DISCUSSION

This study evaluates the utility of IBP characteristics, separately and combined, as well as other SpA features to anticipate the presence of sacroiliitis on MRI or a positive result in HLA-B27 testing in patients with suspected axSpA.

In contrast to most prediction analyses, we have adopted a Bayesian approach based on likelihood ratios, rather than a 'frequentist approach' with odds ratios focusing on positive test results. The Bayesian approach is far closer to clinical reality since it directly visualizes the consequences of a positive test result (here: IBP criteria present) and a negative test result (here: IBP criteria absent) on the likelihood of the outcome (here: positive MRI-SI result) in a cohort with a given prevalence of the outcome. In contrast, the effects of odds ratios obtained with regular regression analysis – either significant or not – on the outcome of interest are entirely dependent on the prevalence of the outcome, and the clinical value is difficult to disentangle (and often disappointingly low).

To predict the presence of sacroiliitis on MRI, we have found that alternating buttock pain and the ASAS-definition for IBP are the most useful characteristic and criterion, respectively, but neither of them performed sufficiently well when tested alone. Interestingly, the addition of the criterion 'alternating buttock pain' to the Calin or ASAS definition increased the positive likelihood ratio for MRI-SI importantly compared to the Calin or ASAS definition alone (LR+ 6.0 vs 1.8 respectively for the Calin definition and 5.5 vs 2.1 respectively for the ASAS definition). When the Berlin algorithm had been designed, the criterion 'alternating buttock pain' also increased the specificity for axSpA but at the cost of a very low sensitivity (because alternating buttock pain was a rare finding) [16]. When the ASAS definition for IBP was later developed, it therefore did not contribute independently to the presence of IBP according to experts´ opinion and therefore was not included in the proposed definition [7]. In the current study, we decided to re-examine the performance of the criterion 'alternating buttock pain' because of the high cost associated with MRI-SI: we were searching for a criterion that could help us in deciding which patients should undergo MRI-SI. Moreover, the addition of the IBP characteristic 'awaking at second half of night' to the Calin criteria also increased significantly the likelihood ratio of having a positive MRI-SI compared to the performance of the Calin criteria alone (LR+ 5.5 vs 1.8, respectively). Furthermore, on top of IBP, other SpA features have been identified for predicting the presence of sacroiliitis on MRI in patients with suspected axSpA. The most useful features we found are dactylitis and IBD.

In addition, in order to recommend a tool as diagnostically relevant in the context of a Bayesian approach, both positive and negative LR should be explored [15]. It is obvious that single features fall short in this regard. But if a patient with CBP has any of the four following characteristics: 1. IBP according to Calin or ASAS definition plus alternating buttock pain;

2. IBP according to the Calin definition plus awakening at second half at night; 3. dactylitis or 4. IBD, the probability of having a positive MRI-SI is very high (73-81%). So in case of limited resources, the presence of any of these four characteristics may help to increase the diagnostic efficiency of the MRI-SI.

Forecasting a positive HLA-B27 test result is more difficult: Neither the individual IBP characteristics, nor the existing- or other possible combinations were found useful to predict a positive HLA-B27 test. Among other SpA features, uveitis seems to be best related to a positive HLA-B27 but it is not sufficiently useful.

To our knowledge, this is the first study investigating this topic. This study has a few strengths: First, many patients with suspected axSpA from all over Spain were included. Second, the mean symptom duration in these patients was very short compared to most diagnostic utility studies and this is an advantage in the referral of patients to early axSpA clinics. And most importantly, patients included in this analysis came from a national health programme representing a typical scenario of common clinical practice that involved primary physicians, rheumatologists and other specialists who usually take care of patients with suspected axSpA. Relying on the conclusions of a local reader (radiologist or rheumatologist) rather than on a report of a central reading committee (as in clinical trials) should also be interpreted in this context, but this is exactly the clinical situation that we wanted to address.

Limitations should be considered too: The most important limitation is the possibility of confounding by indication. For the analysis of the association between IBP characteristics or other SpA features and the presence of sacroiliitis on MRI only data from those patients in whom a MRI-SI was ordered by the rheumatologist were included. Since there is always a reason to do so, patient-selection is very likely, which may have implications for the external validity of this study. On the other hand, patients included in this analysis (because they had MRI ordered) had similar characteristics to the overall group of patients. Second, inter-reader variation by 25 different local readers that have assessed all MRIs may have influenced this study too. In this sense, it can be argued that the results of this study are applicable to clinical practice where usually many different readers are involved. Third, all patients included in this study came from the same country and the same cohort, which may limit the extrapolation of the results to other countries. However, in a sub-analysis on Spanish patients participating in the international study evaluating several referral strategies to diagnose axSpA, the national results were very similar to those observed in the global population [17]. Fourth, It should be pointed out that the diagnostic performance is dependent on the prior probability of the outcome of interest: while the diagnostic performance in terms of likelihood ratios is considered to be rather stable across cohorts, the application of these likelihood ratios in cohorts with other prior probabilities are very different. A final limitation is obviously that we failed in framing a group of patients in whom the likelihood of a positive MRI and/or a positive HLA-B27 was very low, since the negative likelihood ratios of all features tested fell short as the lowest LR- was 0.6.

In summary, if a patient has IBP according to the Calin or ASAS definition plus alternating buttock pain or IBP according to the Calin criteria plus awakening at second half of night or dactylitis or IBD, the probability of a positive MRI-SI increases from 40 to approximately 75%. So in case of limited resources, the presence of any of these four characteristics may improve the efficiency of ordering MRI in patients with suspected axSpA. None of the IBP definitions nor any of the other SpA features seem to be useful to forecast a positive HLA-B27 test.

REFERENCES

1. Landewe R, Dougados M, Mielants H, et al. Physical function in ankylosing spondylitis is independently determined by both disease activity and radiographic damage of the spine. Ann Rheum Dis 2009;68:863-7.
2. Feldtkeller E, Bruckel J, Khan MA. Scientific contributions of ankylosing spondylitis patient advocacy groups. Curr Opin Rheumatol 2000;12:239-47.
3. 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;68:777-83.
4. Sieper J, Srinivasan S, Zamani O, et al. Comparison of two referral strategies for diagnosis of axial spondyloarthritis: the Recognising and Diagnosing Ankylosing Spondylitis Reliably (RADAR) study. Ann Rheum Dis 2013;72:1621-7.
5. Braun A, Saracbasi E, Grifka J, et al. Identifying patients with axial spondyloarthritis in primary care: how useful are items indicative of inflammatory back pain? Ann Rheum Dis 2011;70:1782-7.
6. Poddubnyy D, Vahldiek J, Spiller I, et al. Evaluation of 2 screening strategies for early identification of patients with axial spondyloarthritis in primary care. J Rheumatol 2011;38:2452-60.
7. 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;68:784-8.
8. Calin A, Porta J, Fries JF, et al. Clinical history as a screening test for ankylosing spondylitis. JAMA 1977;237:2613-4.
9. 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;54:569-78.
10. van den Berg R, de Hooge M, Rudwaleit M, et al. ASAS modification of the Berlin algorithm for diagnosing axial spondyloarthritis: results from the SPondyloArthritis Caught Early (SPACE)-cohort and from the Assessment of SpondyloArthritis international Society (ASAS)-cohort. Ann Rheum Dis 2013;72:1646-53.
11. Fernandez Carballido C, on behalf of ESPeranza group. Diagnosing early spondyloarthritis in Spain: the ESPeranza program. Reumatol Clin 2010;6 Suppl 1:6-10.
12. Muñoz-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;70:827-30.
13. Villaverde V, Carmona L, Lopez Robledillo JC, et al. Motivations and objections to implement a spondyloarthritis integrated care pathway. A qualitative study with primary care physicians. Reumatol Clin 2013;9:85-9.
14. Rudwaleit M, Jurik AG, Hermann KG, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. Ann Rheum Dis 2009;68:1520-7.
15. McGee S. Simplifying likelihood ratios. J Gen Intern Med 2002;17:646-9.
16. Rudwaleit M, van der Heijde D, Khan MA, et al. How to diagnose axial spondyloarthritis early. Ann Rheum Dis 2004;63:535-43.
17. Juanola X, Fernandez-Sueiro JL, Torre-Alonso JC, et al. Comparison of 2 referral strategies for the diagnosis of axial spondyloarthritis in Spain. The RADAR study. Reumatol Clin 2013;9:348-52.

6

7

Value of high-sensitivity C-reactive protein for classification of early axial spondyloarthritis: results from the DESIR cohort

V. Navarro-Compán, D. van der Heijde, B. Combe,
C. Cosson, F.A. van Gaalen

Ann Rheum Dis 2013;72:785-6

The average delay in axial spondyloarthritis (axSpA) diagnosis after symptom onset is one of the longest among inflammatory rheumatic diseases [1]. New tools, such as magnetic resonance imaging [2], have been developed to reduce this delay. Elevated C-reactive protein (CRP) has been incorporated as one of the features for Assessment of SpondyloArthritis international Society (ASAS) SpA criteria [3], and in the Berlin diagnostic algorithm [4]. However, CRP levels are elevated in only a minority of early SpA patients [5]. More sensitive tests, so-called high-sensitivity CRP (hsCRP), have been developed and can detect lower concentrations of CRP compared with traditional methods [6]. HsCRP levels are increased in other rheumatic chronic inflammatory diseases [7], and show a better correlation with disease activity parameters compared with routine CRP in patients with axSpA [8]. Therefore, hsCRP could be more sensitive than traditional CRP in diagnosing axSpA. The aim of this study was to assess the contribution of hsCRP versus CRP to classification of early axSpA using the ASAS criteria.

Baseline data from 648 patients with inflammatory back pain (IBP) duration of more than 3 months, but less than 3 years, from the Devenir des Spondylarthopathies Indifferenciees Recentes cohort was used. Database was locked on 12 December 2011. Design, inclusion criteria, CRP measurements using conventional methods, and imaging have previously been described [9]. One biological resources centre (Paris Bichat, Joëlle Benessiano) was in charge of centralizing and managing biological data collection. Serum levels hsCRP were measured by particle-enhanced immunoturbidimetry (Roche Diagnostics, Switzerland). The cut-off values selected to define positive hsCRP and CRP were ≥2 mg/l and ≥5 mg/l [10], respectively.

Using ASAS axSpA criteria, 444 (69%) patients were classified as axSpA, and 203 (31%) patients as no-SpA. Socio-demographic and IBP characteristics and imaging and lab results are displayed in Table 1. Serum levels of CRP were higher in axSpA versus no-SpA (p=0.02).

Table 1: Characteristics of patients with early inflammatory back pain from the DESIR cohort.

	ASAS axSpA criteria yes n=444 (69%)	ASAS axSpA criteria no n=203 (31%)
Age (years)	32.4 ± 8.6	34.8 ± 8.3
Male (%)	223 (50)	76 (37)
Caucasian (%)	400 (90)	159 (88)
Back pain duration (years)	1.0 ± 0.9	0.9 ± 8.4
HLA-B27 positive (%)	368 (83)	8 (4)
Sacroiliitis on MRI (%)	231 (52)	0 (0)
Sacroiliitis on X-Ray (%)	107 (24)	0 (0)
CRP (mg/l)	9.9 ± 14.6	6.8 ± 14

ASAS: Assessment of SpondyloArthritis international Society; CRP: C-reactive protein; DESIR: Devenir des Spondylarthopathies Indifferenciees Recentes.

In the subgroup of patients with negative CRP, mean serum levels of hsCRP were also higher in axSpA patients compared with no-SpA patients (1.7 mg/l vs 1.5 mg/l, p=0.03). Moreover, after dichotomizing hsCRP, more patients within the axSpA group had positive hsCRP (33.1%) versus the no-SpA group (27.6%), although this difference was not statistically significant (p=0.06) (Table 2). Substituting the ASAS classification criteria for the probability of SpA according to the treating physician [9], also did not indicate increased levels of hsCRP in patients with a high probability of SpA (data not shown).

Table 2: Serum levels of hs-CRP in patients with normal CRP values (<5 mg/l).

hsCRP (mg/l)	ASAS axSpA criteria yes (n=260)	ASAS axSpA criteria no (n=152)
<2 (%)	174 (66.9)	110 (72.4)
≥2 - <3 (%)	41 (15.8)	30 (19.7)
≥3 - <4 (%)	34 (13.1)	11 (7.2)
≥4 - <5 (%)	7 (2.7)	0 (0)
≥5 (%)	4 (1.5)	1 (0.7)

ASAS: Assessment of SpondyloArthritis international Society; CRP: C-reactive protein.

Finally, we investigated how many extra patients from the no-SpA group (n=203) would be classified as axSpA substituting the traditional CRP by hsCRP. No patients without SpA had sacroiliitis on imaging, so none of them could meet the imaging arm of the ASAS axSpA criteria. In the clinical arm (HLA-B27 arm) of the criteria, only 4 (2%) extra patients had two axSpA features instead of only one feature (IBP), substituting hsCRP by CRP (195 vs 191 patients), but none of them was HLA-B27 positive. Consequently, none of the no-SpA patients met ASAS axSpA criteria applying this modification.

In conclusion, in patients with a normal CRP, hsCRP is increased in axSpA patients compared with patients without SpA. However, hsCRP measurement in patients with IBP did not add any extra value for classifying axSpA patients. Future studies including patients with chronic back pain (without inflammatory characteristics) are required to confirm these results.

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;23:61-6.
2. Rudwaleit M, Landewe 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;68:770-6.
3. 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;68:777-83.
4. Rudwaleit M, van der Heijde D, Khan MA, et al. How to diagnose axial spondyloarthritis early. Ann Rheum Dis 2004;63:535-43.
5. Hermann J, Giessauf H, Schaffler G, et al. Early spondyloarthritis: usefulness of clinical screening. Rheumatology (Oxford) 2009;48:812-6.
6. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 2008;359:2195-207.
7. Nielen MM, van Schaardenburg D, Reesink HW, et al. Increased levels of C-reactive protein in serum from blood donors before the onset of rheumatoid arthritis. Arthritis Rheum 2004;50:2423-7.
8. Poddubnyy DA, Rudwaleit M, Listing J, et al. Comparison of a high sensitivity and standard C reactive protein measurement in patients with ankylosing spondylitis and non-radiographic axial spondyloarthritis. Ann Rheum Dis 2010;69:1338-41.
9. Dougados M, d'Agostino MA, Benessiano J, et al. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine 2011;78:598-603.
10. Whicher JT. BCR/IFCC reference material for plasma proteins (CRM 470). Community Bureau of Reference. International Federation of Clinical Chemistry. Clin Biochem 1998;31:459-65.

8

Calculating the ankylosing spondylitis disease activity score if the conventional C-reactive protein level is below the limit of detection or if high-sensitivity C-reactive protein is used: an analysis in the DESIR cohort

P. Machado, V. Navarro-Compán, R. Landewé, F.A. van Gaalen, C. Roux, D. van der Heijde

Arthritis Rheumatol 2015;67(2):408-13

ABSTRACT

Objectives

The Ankylosing Spondylitis Disease Activity Score (ASDAS) is a composite measure of disease activity in axial spondyloarthritis. The aims of this study were to determine the most appropriate method for calculating the ASDAS using C-reactive protein (CRP) level when the conventional CRP level is below the limit of detection, to determine how low CRP values obtained by high-sensitivity CRP (hsCRP) measurement influence ASDAS-CRP results, and to test agreement between different ASDAS formulae.

Methods

Patients with axial spondyloarthritis who had a conventional CRP level below the limit of detection (5 mg/liter) were selected (n=257). The ASDAS-conventional CRP with 11 different imputations for the conventional CRP value (range 0-5 mg/liter, at 0.5 mg/liter intervals) was calculated. The ASDAS-hsCRP and ASDAS using the erythrocyte sedimentation rate (ESR) were also calculated. Agreement between ASDAS formulae was tested.

Results

The ASDAS-hsCRP showed better agreement with the ASDAS-CRP calculated using the conventional CRP imputation values of 1.5 and 2.0 mg/liter and with the ASDAS-ESR than with other imputed formulae. Disagreement occurred mainly in lower disease activity states (inactive/moderate disease activity). When the CRP value was <2 mg/liter, the resulting ASDAS-CRP scores may have been inappropriately low.

Conclusion

When the conventional CRP level is below the limit of detection or when the hsCRP level is <2 mg/liter, the constant value of 2 mg/liter should be used to calculate the ASDAS-CRP score. There is good agreement between the ASDAS-hsCRP and ASDAS-ESR; however, formulae are not interchangeable.

INTRODUCTION

The Ankylosing Spondylitis Disease Activity Score (ASDAS) is a composite index to assess disease activity in axial spondyloarthritis (SpA) [1-3]. It combines five single disease activity variables in such a manner that it optimally conveys information, resulting in one single score with better validity, enhanced discriminative capacity, and improved ability to detect change as compared to separate variables [1-5]. ASDAS cut-off values have been developed to define disease activity states and response criteria [2].

The ASDAS has been endorsed by the Assessment of SpondyloArthritis international Society (ASAS) and by the Outcome Measures in Rheumatology study group and validated in various populations worldwide [5-10]. The ASAS membership has selected the ASDAS using the C-reactive protein (CRP) levels as the preferred version and the ASDAS using the erythrocyte sedimentation rate (ESR) as an alternative [1-3]. The same validated cut-off values apply to both the ASDAS-CRP and the ASDAS-ESR [2].

The development and validation of the ASDAS was based on conventional CRP values. It has been suggested that when the conventional CRP is below the limit of detection and high-sensitivity CRP (hsCRP) is not available, 50% of the threshold value should be used to calculate the ASDAS-CRP [2]. However, this recommendation is not based on data-driven testing and the effect of using the hsCRP has not been determined. Further testing is required.

The aims of this study were to determine the best way to calculate the ASDAS when the conventional CRP is below the limit of detection, to study the influence of low CRP values obtained by hsCRP in the ASDAS-CRP, and to test agreement between different ASDAS formulae.

PATIENT AND METHODS

Patients

Baseline data from the Devenir des Spondylarthropathies Indifférenciées Récentes (DESIR) cohort was used. Details of the DESIR cohort have been previously described [11]. Briefly, DESIR is a French multicenter, prospective study of patients with early (<3 years´ duration) inflammatory back pain (IBP) suggestive of SpA. A total of 708 patients were included in the DESIR cohort at baseline. For the present study, we selected all patients who fulfilled the ASAS classification criteria for axial SpA [12] and who had a conventional CRP value below the limit of detection as well as the results of hsCRP testing; we used data from baseline

assessments only. We used the dataset locked on 12 December 2011.

ASDAS calculation

ASDAS-CRP and ASDAS-ESR scores were calculated based on 5 variables: acute-phase reactant levels (either CRP or ESR) and 4 patient-reported variables [1,2], namely back pain (question 2 on the Bath Ankylosing Spondylitis Disease Activity Index [BASDAI]) [13], duration of morning stiffness (question 6 on the BASDAI), peripheral pain/swelling (question 3 on the BASDAI), and patient global assessment of disease activity. All the patient-reported variables were scored on a scale of 0-10. ASDAS scores were also categorised according to previously published cut-off values for disease activity: an ASDAS score of <1.3 = inactive disease, ≥1.3-<2.1 = moderate activity, ≥2.1-3.5 = high activity, and >3.5 = very high disease activity [2]. Disease activity was quantified using the following equations:

ASDAS-CRP = (0.12*back pain) + (0.06*duration of morning stiffness) + (0.11*patient global) + (0.07*peripheral pain/swelling) + (0.58*ln[CRP +1])

or

ASDAS-ESR = (0.08*back pain) + (0.07*duration of morning stiffness) + (0.11*patient global) + (0.09*peripheral pain/swelling) + (0.29*√ESR)

The limit of detection by the conventional CRP assay was 5 mg/liter. The ASDAS-conventional CRP with 11 different imputations (from 0 mg/liter [ASDAS-CRP(0)] to 5 mg/liter [ASDAS-CRP(5)], at 0.5 mg/liter intervals) to replace the undetermined conventional CRP value was calculated. High-sensitivity CRP was measured by particle-enhanced immunoturbidimetry on a Cobas Integra 800 or Modular Analytics P800 device according to the instructions of the manufacturer (Roche Diagnostics). (Measurement was performed at Paris Bichat, a biologic resource center, by Dr. Joëlle Benessiano).

To gain insight into how low CRP values influence the total ASDAS-CRP score, we plotted CRP values against the CRP term 0.58*ln(CRP+1) from the ASDAS-CRP formula and displayed the ASDAS-CRP scores that were calculated using multiple CRP values (from 0 to 5 mg/liter) and different fixed values (from 0 to 5 units) for the 4 other variables included in the ASDAS-CRP formula (back pain, duration of morning stiffness, peripheral pain/swelling, and patient global assessment of disease activity).

Statistical analysis

The two-way mixed single-measures (absolute agreement) intraclass correlation coefficient (ICC) was used to assess agreement between the ASDAS-hsCRP and other ASDAS formulae (ASDAS-conventional CRP with different imputation strategies and ASDAS-ESR). The ICC can have values between 0 (no agreement) and 1 (perfect agreement).

Scatterplots were created to provide an additional view of the deviation of ASDAS-conventional CRP and ASDAS-ESR from ASDAS-hsCRP. Mean differences (and 95% confidence intervals) between ASDAS-hsCRP and other ASDAS formulae were also calculated.

Agreement between ASDAS-determined disease activity states was assessed using the kappa statistic. The kappa statistic represents the actual agreement beyond chance as a proportion of the potential agreement beyond chance. Since disease activity states are ordered categories, we used the weighted kappa value. The kappa statistic can have values between 0 (agreement equivalent to chance) and 1 (perfect agreement). The strenght of agreement was determined as follows: kappa values of <0.20 indicate poor agreement, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 good, 0.81-1.00 very good. SPSS version 22 and MedCalc version 13.1 were used in the statistical analyses.

RESULTS

Patient characteristics

A total of 260 patients fulfilled the inclusion criteria. Three patients had missing ASDAS results; therefore, data from 257 patients were available. Demographic and clinical characteristics of the study population are shown in supplementary Table S1.

Agreement between ASDAS formulae

Table 1 shows the level of agreement between the results obtained using the different ASDAS formulae, both in terms of continuous variable (scores on the ASDAS) and in terms of categorical variable (the disease activity state as determined by the score).

Quantitatively, the best agreement between the ASDAS-hsCRP and ASDAS-conventional CRP scores occurred with the imputed CRP values 1.0, 1.5, 2.0 and 2.5 mg/liter (ICC=0.94, 0.95, 0.94 and 0.92, respectively, representing very good agreement). Agreement between ASDAS-hsCRP and ASDAS-ESR was also very good (ICC=0.91) (Table 1). As shown in the scatterplots presented in Figure 1, use of conventional CRP imputation values ≤1.0 mg/liter systematically resulted in lower scores of the ASDAS-conventional CRP as compared to the ASDAS-hsCRP, while conventional CRP imputation values ≥2.5 mg/liter systematically resulted in higher scores on the ASDAS-conventional CRP compared to the ASDAS-hsCRP.

Qualitatively, the best agreement between ASDAS-hsCRP and ASDAS-conventional CRP disease activity states occurred with the conventional CRP imputation values of 1.5 and 2 mg/liter (weighted kappa=0.75 and 0.76, respectively, representing good agreement) (Table 1). Agreement between ASDAS-hsCRP and ASDAS-ESR disease activity states was

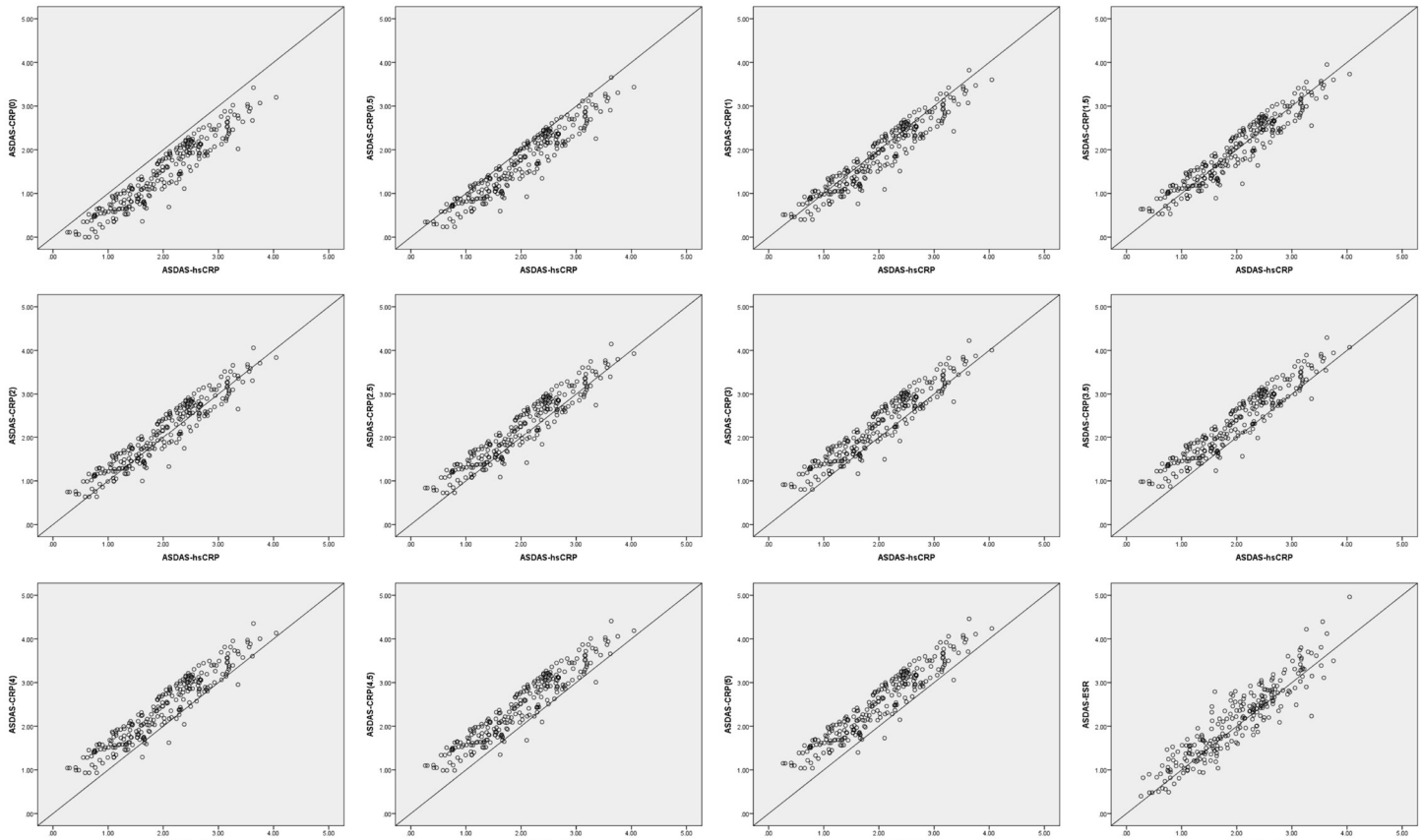


Figure 1: Scatterplots showing the relationship between the Ankylosing Spondylitis Activity Score (ASDAS) using high-sensitivity C-reactive protein (hsCRP) measurement and other ASDAS formulae (ASDAS-conventional CRP [ASDAS-CRP] with multiple imputation of CRP values [ranging from 0 to 5 mg/liter, at 0.5 mg/liter intervals] and ASDAS using the erythrocyte sedimentation rate [ESR]). The line indicates exact agreement between the ASDAS formulae. Data on 257 patients were used for all analyses except for analysis on the relationship between ASDAS-hsCRP and ASDAS-ESR, where data on 246 patients were used.

Table 1: Agreement between results obtained using the ASDAS-hsCRP and results obtained using other ASDAS formulae (ASDAS-conventional CRP with multiple imputation and ASDAS-ESR)*.

	ASDAS-hsCRP vs. ASDAS calculated using other formulae (continuous variable)		ASDAS-hsCRP vs. ASDAS disease activity states calculated using other formulae (categorical variable)
	ICC (95% CI)	Mean (95% CI) difference in ASDAS score	Weighted kappa (95% CI)
ASDAS-CRP(0)	0.78 (-0.06 to 0.94)	-0.52 (-1.02 to -0.03)	0.51 (0.44 to 0.57)
ASDAS-CRP(0.5)	0.89 (0.33 to 0.96)	-0.29 (-0.79 to 0.21)	0.73 (0.67 to 0.79)
ASDAS-CRP(1)	0.94 (0.89 to 0.96)	-0.12 (-0.62 to 0.38)	0.73 (0.67 to 0.79)
ASDAS-CRP(1.5)	0.95 (0.93 to 0.96)	0.01 (-0.49 to 0.51)	0.75 (0.69 to 0.81)
ASDAS-CRP(2)	0.94 (0.90 to 0.96)	0.11 (-0.38 to 0.61)	0.76 (0.70 to 0.81)
ASDAS-CRP(2.5)	0.92 (0.70 to 0.96)	0.20 (-0.29 to 0.70)	0.71 (0.65 to 0.77)
ASDAS-CRP(3)	0.89 (0.37 to 0.96)	0.28 (-0.22 to 0.78)	0.66 (0.60 to 0.73)
ASDAS-CRP(3.5)	0.86 (0.11 to 0.96)	0.35 (-0.15 to 0.85)	0.64 (0.58 to 0.70)
ASDAS-CRP(4)	0.83 (0.00 to 0.95)	0.41 (-0.09 to 0.91)	0.61 (0.54 to 0.67)
ASDAS-CRP(4.5)	0.81 (-0.04 to 0.94)	0.47 (-0.03 to 0.96)	0.59 (0.53 to 0.65)
ASDAS-CRP(5)	0.78 (-0.06 to 0.94)	0.52 (0.02 to 1.01)	0.50 (0.44 to 0.57)
ASDAS-ESR	0.91 (0.85 to 0.94)	0.13 (-0.52 to 0.79)	0.69 (0.63 to 0.76)

*The Ankylosing Spondylitis Disease Activity Score (ASDAS) using the conventional C-reactive protein (CRP) level with 11 different imputations [ASDAS-CRP(0) to ASDAS-CRP(5), representing CRP values from 0 to 5 mg/liter, at 0.5 mg/liter intervals] and the ASDAS using the erythrocyte sedimentation rate (ESR) were calculated. Two hundred fifty-seven patients were used in all analyses except for the analyses of ASDAS-ESR (n= 246). hsCRP: high-sensitivity CRP; ICC: intraclass correlation coefficient; 95% CI: 95% confidence interval.

also good (weighted kappa=0.69). Disease activity states according to ASDAS-CRP(1.5) and ASDAS-CRP(2) had 78.2% and 78.1% agreement with ASDAS-hsCRP disease activity states, respectively. This percentage decreased to 53.3-75.6% when other CRP values were imputed. Disagreement was evident in lower disease activity states, namely shifts between inactive disease and moderate disease activity (supplementary Table 2).

Effect of low CRP values on ASDAS-CRP scores

The values corresponding to y=0.58*ln(CRP+1), the CRP term from the ASDAS-CRP formula, according to CRP values between 0 and 5 mg/liter, were calculated. The function approximates y=0 asymptotically. For higher values, the relationship between CRP and 0.58*ln(CRP+1) is roughly linear. However, for lower values, small differences in the CRP value represent larger steps in the term 0.58*ln(CRP+1) because the steepness of the curve increases in this area, which may result in inappropriately low ASDAS scores. This implies that it may be better not to use very low CRP values when calculating the ASDAS-CRP. The decision about the optimal CRP threshold value can be made by examining hypothetical case scenarios. A graphic representation of the results of this analysis, illustrating that this threshold should be between 1.5 and 2.5 mg/liter, is presented in supplementary Figure 1.

Table 2: ASDAS-conventional CRP results for different CRP values and different fixed values for all the other four variables used in the calculation of the ASDAS-CRP formula.

	ASDAS-CRP									
CRP (mg/liter)	All other variables=0	All other variables=1	All other variables=1.5	All other variables=2	All other variables=2.5	All other variables=3	All other variables=3.5	All other variables=4	All other variables=4.5	All other variables=5
0	0.0	0.4	0.5	0.7	0.9	1.1	1.3	1.4	1.6	1.8
0.5	0.2	0.6	0.8	1.0	1.1	1.3	1.5	1.7	1.9	2.0
1	0.4	0.8	0.9	1.1	1.3	1.5	1.7	1.8	2.0	2.2
1.5	0.5	0.9	1.1	1.3	1.4	1.6	1.8	2.0	2.2	2.3
2	0.6	1.0	1.2	1.4	1.5	1.7	1.9	2.1	2.3	2.4
2.5	0.7	1.1	1.3	1.4	1.6	1.8	2.0	2.2	2.3	2.5
3	0.8	1.2	1.3	1.5	1.7	1.9	2.1	2.2	2.4	2.6
3.5	0.9	1.2	1.4	1.6	1.8	2.0	2.1	2.3	2.5	2.7
4	0.9	1.3	1.5	1.7	1.8	2.0	2.2	2.4	2.6	2.7
4.5	1.0	1.3	1.5	1.7	1.9	2.1	2.2	2.4	2.6	2.8
5	1.0	1.4	1.6	1.8	1.9	2.1	2.3	2.5	2.7	2.8

*The Ankylosing Spondylitis Disease Activity Score (ASDAS) using the conventional C-reactive protein (CRP) level was calculated using multiple CRP values (ranging from 0 to 5 mg/liter, at 0.5 mg/liter intervals) and multiple fixed values (from 0 to 5 units, at 0.5-unit intervals) for the other 4 variables used in the calculation of the ASDAS-CRP (back pain, duration of morning stiffness, peripheral pain/swelling, and patient global assessment of disease activity). ASDAS scores were categorized as follows: <1.3 = inactive disease (lightly shaded), ≥1.3-<2.1 = moderate disease activity (shaded), ≥2.1-3.5 = high activity (darkly shaded), and >3.5 = very high disease activity.

Table 2 is a matrix showing ASDAS-CRP scores for hypothetical scenarios in which different CRP values and different fixed values for the other 4 items used in the ASDAS-CRP formula were imputed. The 1.5, 2.0 and 2.5 mg/liter imputation strategies perform well with very subtle differences. However, looking at individual cases is particularly informative. If all the other variables are equal to 4, disease activity is rated as moderate when a CRP constant value of 1.5 is used (ASDAS 2.0) but high when a CRP constant value of 2 is used (ASDAS 2.1). Clinically, the latter scenario makes more sense. Further, if all the other variables are equal to 1.5, disease activity is rated as moderate when a constant value of 2.5 is used (ASDAS 1.3) but inactive when a CRP constant value of 2 is used (ASDAS=1.2). Again, clinically the latter scenario makes more sense. These two examples favour the use of the constant value of 2 mg/liter rather than 1.5 or 2.5 mg/liter as the ideal imputation strategy for very low CRP levels.

DISCUSSION

The availability of conventional CRP and hsCRP determinations in the DESIR cohort allowed us to perform this analysis in a large population of patients with early IBP who fulfilled the ASAS classification criteria for axial SpA. Our study shows that when the conventional CRP value is below the limit of detection, the value of 2 mg/liter should be used to calculate the ASDAS-CRP. Furthermore, when the hsCRP value is below 2 mg/liter, the constant value of 2 mg/liter should also be used to calculate the ASDAS-CRP.

We have shown that for very low hsCRP values, small differences represent larger steps in the CRP term of the ASDAS formula and therefore larger steps in the total ASDAS-CRP score. The final choice of the best imputation value was made by looking at a matrix of clinical scenarios (Table 2) according to different imputation strategies. Differences between the imputation of the 1.5, 2.0 and 2.5 mg/liter CRP values were small, but the analysis of individual cases regarding the repercussion of these different imputation strategies in ASDAS disease activity states allowed us to conclude that the best option was not to use hsCRP values below 2 mg/liter.

Disagreement between the ASDAS-hsCRP and other ASDAS formulae was mainly evident among lower disease activity states (inactive/moderate disease activity), a shift that has fewer therapeutic implications than the shift between moderate and high/very high disease activity. This is particularly important given recent evidence that the ASDAS cut-off for high disease activity (ASDAS ≥2.1) is likely to be the most appropriate ASDAS cut-off value for use in the selection of patients for tumor necrosis factor blocker treatment [14,15]. Further evidence supports the replacement of the commonly used BASDAI selection cut-off of 4 units (on a 0-10 scale) by the ASDAS high disease activity cut-off [16]. There was also a high

level of agreement between the ASDAS-hsCRP and ASDAS-ESR. However, it is important to highlight that formulae are not interchangeable.

One of the limitations of our study is the fact that this is a selected population with early disease. Therefore results might not be generalizable to the entire spectrum of axial SpA patients, in particular to patients with advanced disease/ankylosing spondylitis. However, a lack of generalizability is unlikely given the fact that CRP is more frequently elevated in ankylosing spondylitis than in non-radiographic axial SpA, so the need to substitute conventional CRP values below the limit of detection or very low hsCRP values will occur more often in early disease than in late disease [17].

The ASDAS is increasingly being used as a measure of disease activity in clinical practice, clinical trials and observational studies [16]. This study contributes to further standardization of the ASDAS and to a more homogeneous and reproducible application of this new index.

REFERENCES

1. Lukas C, Landewe R, Sieper J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. Ann Rheum Dis 2009;68:18-24.
2. Machado P, Landewe R, Lie E, et al. Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. Ann Rheum Dis 2011;70:47-53.
3. van der Heijde D, Lie E, Kvien TK, et al. ASDAS, a highly discriminatory ASAS-endorsed disease activity score in patients with ankylosing spondylitis. Ann Rheum Dis 2009;68:1811-8.
4. Machado P, Landewe R, Braun J, et al. MRI inflammation and its relation with measures of clinical disease activity and different treatment responses in patients with ankylosing spondylitis treated with a tumour necrosis factor inhibitor. Ann Rheum Dis 2012;71:2002-5.
5. Pedersen SJ, Sorensen IJ, Garnero P, et al. ASDAS, BASDAI and different treatment responses and their relation to biomarkers of inflammation, cartilage and bone turnover in patients with axial spondyloarthritis treated with TNFalpha inhibitors. Ann Rheum Dis 2011;70:1375-81.
6. Machado P, Landewe R, van der Heijde D. Endorsement of definitions of disease activity states and improvement scores for the Ankylosing Spondylitis Disease Activity Score: results from OMERACT 10. J Rheumatol 2011;38:1502-6.
7. Arends S, Brouwer E, van der Veer E, et al. Baseline predictors of response and discontinuation of tumor necrosis factor-alpha blocking therapy in ankylosing spondylitis: a prospective longitudinal observational cohort study. Arthritis Res Ther 2011;13:R94.
8. Fernandez-Espartero C, de Miguel E, Loza E, et al. Validity of the Ankylosing Spondylitis Disease Activity Score (ASDAS) in patients with early spondyloarthritis from the ESPeranza programme. Ann Rheum Dis 2014;73:1350-5.
9. Popescu C, Trandafir M, Badica A, et al. Ankylosing spondylitis functional and activity indices in clinical practice. J Med Life 2014;7:78-83.
10. Xu M, Lin Z, Deng X, et al. The Ankylosing Spondylitis Disease Activity Score is a highly discriminatory measure of disease activity and efficacy following tumour necrosis factor-alpha inhibitor therapies in ankylosing spondylitis and undifferentiated spondyloarthropathies in China. Rheumatology 2011;50:1466-72.
11. Dougados M, d'Agostino MA, Benessiano J, et al. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine 2011;78:598-603.
12. Rudwaleit M, Jurik AG, Hermann KG, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. Ann Rheum Dis 2009;68:1520-7.
13. 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;21:2286-91.
14. Fagerli KM, Lie E, van der Heijde D, et al. Selecting patients with ankylosing spondylitis for TNF inhibitor therapy: comparison of ASDAS and BASDAI eligibility criteria. Rheumatology (Oxford) 2012;51:1479-83.
15. Vastesaeger N, Cruyssen BV, Mulero J, et al. ASDAS high disease activity versus BASDAI elevation in patients with ankylosing spondylitis as selection criterion for anti-TNF therapy. Reumatol Clin 2014;10:204-9.
16. Machado P, Landewe R. Spondyloarthritis: Is it time to replace BASDAI with ASDAS? Nat Rev Rheumatol 2013;9:388-90.
17. Poddubnyy DA, Rudwaleit M, Listing J, et al. Comparison of a high sensitivity and standard C reactive protein measurement in patients with ankylosing spondylitis and non-radiographic axial spondyloarthritis. Ann Rheum Dis 2010;69:1338-41.

SUPPLEMENTARY MATERIAL

Supplementary table 1: Summary of the baseline clinical and demographic characteristics of the study population (n=257)*.

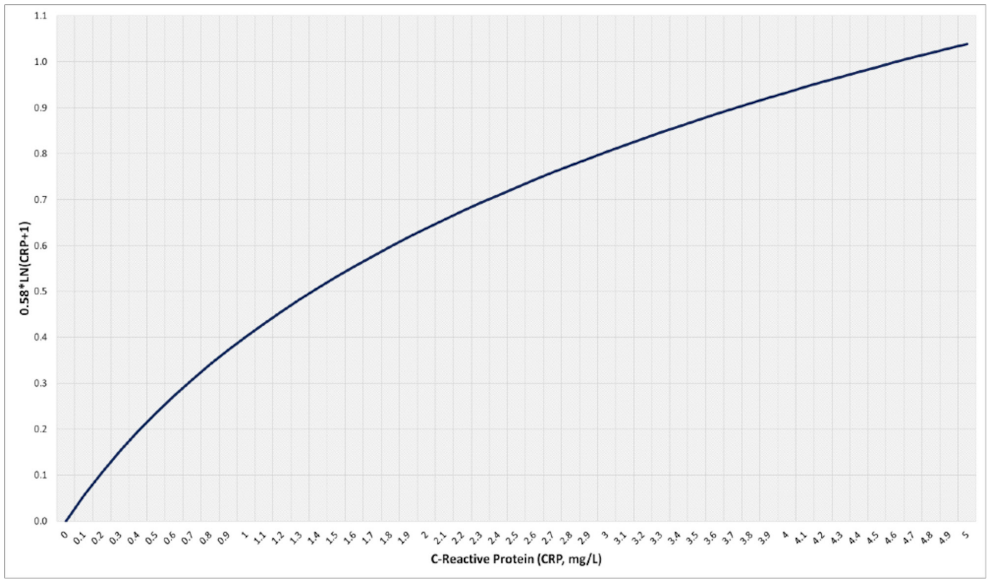
Male, no (%)	121 (47.1)
Caucasian, no (%)	234 (91.1)
Age, years	33.2 (8.8)
HLA-B27 positive, no (%)	191 (89.7)
ASDAS-hsCRP	2.0 (0.8)
ASDAS-ESR[a]	2.2 (0.9)
hsCRP, mg/liter	1.7 (1.4)
ESR†, mmHg	8.2 (6.9)
BASDAI (0-10 scale)	4.0 (2.1)
Patient global assessment (0-10 scale)	4.6 (2.7)
Physician global assessment (0-10 scale)	3.9 (2.2)
BASMI (0-10 scale)	2.2 (0.9)
BASFI (0-10 scale)	2.6 (2.2)

*Except were indicated otherwise, values are the mean (standard deviation). [a]ESR was not available in 4.3% (11/257) of the patients. ASDAS: Ankylosing Spondylitis Disease Activity Score; BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASMI: Bath Ankylosing Spondylitis Metrology Index; BASFI: Bath Ankylosing Spondylitis Functional Index; CRP: C-reactive protein; hsCRP: high sensitivity CRP; ESR: erythrocyte sedimentation rate.

Supplementary table 2: Percentage and causes of disagreement in ASDAS disease activity states using ASDAS-hsCRP and other ASDAS formulae (ASDAS-conventional CRP with multiple imputation strategies* and ASDAS-ESR).

ASDAS formulae	ASDAS-hsCRP ASDAS disease activity states, percentage and causes of disagreement
ASDAS-CRP(0)	Disagreement: 46.7% MDA → ID: 21.0%; HDA → ID: 2%; HDA → MDA: 20.6%; VHDA → HAD: 3.1%
ASDAS-CRP(0.5)	Disagreement: 25.0% ID → MDA: 0.4%; MDA → ID: 12.5%; MDA → HAD: 0.4%; HDA → MDA: 8.6%; HAD → ID: 0.4%; VHDA → HAD: 2.7%
ASDAS-CRP(1)	Disagreement: 24.4% ID → MDA: 3.1%; MDA → ID: 10.1%; MDA → HAD: 2.3%; HDA → MDA: 6.2%; HAD → ID: 0.4%; VHDA → HAD: 2.3%
ASDAS-CRP(1.5)	Disagreement: 21.9% ID → MDA: 5.1%; MDA → ID: 5.4%; MDA → HAD: 4.7%; HDA → MDA: 4.7%; HAD → ID: 0.4%; HDA → VHDA: 0.4%; VHDA → HAD: 1.2%
ASDAS-CRP(2)	Disagreement: 21.8% ID → MDA: 6.6%; MDA → ID: 2.7%; MDA → HAD: 6.6%; HDA → MDA: 4.3%; HDA → VHDA: 1.2%; VHDA → HAD: 0.4%
ASDAS-CRP(2.5)	Disagreement: 25.3% ID → MDA: 10.9%; MDA → ID: 1.6%; MDA → HAD: 8.2%; HDA → MDA: 2.3%; HDA → VHDA: 1.9%; VHDA → HAD: 0.4%
ASDAS-CRP(3)	Disagreement: 29.1% ID → MDA: 13.2%; MDA → ID: 0.4%; MDA → HAD: 10.1%; HDA → MDA: 1.9%; HDA → VHDA: 3.1%; VHDA → HAD: 0.4%
ASDAS-CRP(3.5)	Disagreement: 31.6% ID → MDA: 14.8%; MDA → ID: 0.4%; MDA → HAD: 11.3%; HDA → MDA: 0.8%; HDA → VHDA: 4.3%
ASDAS-CRP(4)	Disagreement: 34.3% ID → MDA: 15.6%; MDA → ID: 0.4%; MDA → HAD: 13.2%; HDA → MDA: 0.8%; HDA → VHDA: 4.3%
ASDAS-CRP(4.5)	Disagreement: 35.8% ID → MDA: 17.5%; MDA → HAD: 10.1%; HDA → MDA: 2%; HDA → VHDA: 6.2%
ASDAS-CRP(5)	Disagreement: 43.6% ID → MDA: 17.9%; MDA → HAD: 18.7%; HDA → MDA: 0.4%; HDA → VHDA: 6.6%
ASDAS-ESR	Disagreement: 28.1% ID → MDA: 7.7%; MDA → ID: 3.3%; MDA → HAD: 8.5%; HDA → MDA: 3.3%; HDA → VHDA: 4.1%; VHDA → HAD: 1.2%

*ASDAS-CRP(0) to ASDAS-CRP(5) represents the ASDAS-CRP results with 11 imputation strategies for the conventional CRP, from 0 to 5 mg/liter, at 0.5 mg/liter intervals. ASDAS: Ankylosing Spondylitis Disease Activity Score; CRP: C-reactive protein; ESR: erythrocyte sedimentation rate; hsCRP: high sensitivity CRP; ID: inactive disease; MDA: moderate disease activity; HAD: high disease activity; VHDA: very high disease activity. Data on 257 patients were used for all analyses except for the ASDAS-ESR, where data on 246 patients were used.



Supplementary figure 1: Graphic displaying the results of the C-reactive protein (CRP) component of the ASDAS-CRP formula (0.58*ln(CRP+1)) according to the CRP value, from 0 to 5 mg/liter, at 0.1 mg/liter intervals.

9

Disease activity is longitudinally related to sacroiliac inflammation on MRI in male patients with axial spondyloarthritis: 2-year of the DESIR cohort

V. Navarro-Compán, S. Ramiro, R. Landewé, M. Dougados, C. Miceli-Richard, P. Richette, D. van der Heijde

Ann Rheum Dis (accepted for publication)

ABSTRACT

Objectives

To investigate the longitudinal relationship between inflammatory lesions in sacroiliac joints on magnetic resonance imaging (MRI-SI) and clinical disease activity measures (DA) in patients with axial spondyloarthritis (axSpA).

Methods

Two-year follow-up data from 167 patients (50% males, mean (SD) age 33 (9) years) fulfilling ASAS axSpA criteria in the DESIR cohort with MRI-SI at baseline, 1- and 2-years were analysed. The relationship between MRI-SI (as dependent variable) and DA (ASDAS, BASDAI, patient´s global disease activity, night pain, CRP and ESR, as independent variables) was investigated using two types of generalized estimating equations (GEE) models: model of absolute-scores and model of change-scores.

Results

In the model of absolute-scores**,** the relationship between DA and MRI-SI was different for males and females: In males, but not in females, a statistically significant relationship with MRI-SI was found for all DA except BASDAI. In the model of changes, only ASDAS [beta (95%CI): 2.79 (0.85-4.73)] and pain at night [0.97 (0.04-1.90)] were significantly associated in males while again in females no significant relationship was found. ASDAS fitted the data best.

Conclusions

In male, but not in female patients with axSpA, clinical disease activity, especially if measured by ASDAS, is longitudinally associated with MRI-SI inflammatory lesions.

INTRODUCTION

The diagnostic value of magnetic resonance imaging of sacroiliac joints (MRI-SI) in axial spondyloarthritis (axSpA) is nowadays clearly established [1]. For reasons of understanding the disease axSpA better, it is important to know whether or not clinical signs and symptoms are related to the degree of inflammation that is occurring in the sacroiliac joints, which can be observed on MRI. Data evaluating this are scarce. Furthermore, most of them stem from clinical trials including patients with active and mostly established disease but almost no data are available for patients in early stages of the disease or with low disease activity [2-5]. Additionally, all these studies only provided cross-sectional correlations but in none of them a longitudinal analysis was conducted. Based on this, the current study aims to investigate the longitudinal relationship between MRI-SI inflammatory lesions and clinical disease activity measures (DA) in patients with early axSpA; i.e. whether or not changes on MRI-SI inflammation are related to changes on DA over time.

METHODS

Study population

Two-year follow-up data from the DESIR (DEvenir des Spondylarthopathies Indifférenciées Récentes) cohort were analysed. Design and inclusion criteria of this cohort have been reported [6]. In summary, 708 patients with inflammatory back pain (IBP) (>3 months but <3 years), age 18-50 years old, and probability of SpA >50% based on the physician's assessment, were recruited. Database used for this analysis was locked on June 30th 2014.

For this study, patients fulfilling ASAS criteria for axSpA at baseline with MRI-SI and DA available for at least two consecutive visits were included.

Clinical disease activity measures

DA were collected every six months and included: Ankylosing Spondylitis Disease Activity Score with C-reactive protein (ASDAS_CRP), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), patient´s visual analogue scale (VAS) of night pain and global assessment of disease activity, CRP and erythrocyte sedimentation rate (ESR). If serum CRP was below the limit of detection, a value of 2 was used for the ASDAS calculation [7].

Magnetic resonance imaging of sacroiliac joints (MRI-SI)

MRI-SI (T1 and short tau inversion recovery –STIR- sequences) was performed at baseline. Additionally, MRI-SI was repeated at 12 and 24 months but due to budget restrictions,

only in 9 out of the 25 participating centers. Two trained readers, blinded to clinical data and time-point, independently scored all MRI-SI images for the presence of inflammatory lesions according to the Spondyloarthritis Research Consortium of Canada (SPARCC) scoring method (range 0-72) [8]. The mean of the two readers was used for this analysis. The intraclass-correlation coefficient was 0.94 for absolute baseline scores and 0.93 for change scores (baseline-12 months). The smallest detectable difference was 6.1 (absolute baseline scores) and the smallest detectable change was 3.6 (mean of two change-intervals).

Where dichotomous MRI-data (positive vs. negative MRI) were required, a cut off level of 2 SPARCC units was assumed to appropriately reflect the distinction between an Assessment of SpondyloArthritis international Society (ASAS) positive and negative MRI [9].

Statistical analysis

The longitudinal relationship between MRI-SI (SPARCC) and DA was analyzed using generalized estimating equations (GEE) models, with MRI-SI (SPARCC) as dependent variable and the separate clinical measures as independent variable. Two types of models were run: 1) a standard model, for which the absolute scores of SPARCC and DA were used; 2) a model of changes, for which the change scores between two consecutive measurements of both SPARCC and DA were employed. While the standard model includes a pooled analysis of longitudinal (within-subject) and cross-sectional (between-subjects) relationships, the model of changes 'removes' the cross-sectional part of the analysis enabling a real longitudinal interpretation [10].

Interactions with age, gender, HLA-B27 and symptom duration were tested. If relevant interactions were found, analyses were stratified. If not, variables were entered as covariates. Model fit was estimated with the Quasi-likelihood under the Independence model Criterion (QIC): the lower the QIC, the better the data fit the model.

Additional sensitivity analyses were performed to examine the potential influence of the following characteristics on the association between DA and MRI-SI: i) reader-variation, ii) variation in MRI-SI status at baseline (positive versus negative according to ASAS definition); and iii) anti-tumor necrosis factor –TNF- alpha therapy during the study. Data were analysed using SPSS 20.0.

RESULTS

Baseline characteristics

A total of 486 patients fulfilled ASAS criteria for axSpA in the DESIR cohort. Out of these patients, 167 patients met the inclusion criteria for this study. Overall, baseline data of

Table 1: Characteristics of patients with axial spondyloarthritis included in this study and for all patients in DESIR cohort.

	Patients includedin this study			DESIR cohort		
	Total	Males	Females	Total	Males	Females
	n=167	n=83 (50%)	n=84 (50%)	(n=486)	n=244 (50%)	n=242 (50%)
Age (years)	33.0 ± 9.0	31.4 ± 8.4	34.6 ± 9.4	33.0 ± 8.6	31.8 ± 8.4	34.1 ± 8.7
Male	83 (50)			238 (50)		
Fulfilling ASAS criteria						
Imaging arm only	32 (19)	13 (16)	19 (23)	81 (17)	34 (14)	47 (19)
Clinical arm only	59 (35)	23 (28)	36 (43)	201 (41)	78 (32)	123 (51)
Both arms	76 (46)	47 (57)	29 (34)	204 (42)	132 (54)	72 (30)
Back pain duration (mo)	17.8 ± 10.8	16.2 ± 9.5	19.4 ± 11.8	18.0 ± 10.6	17.6 ± 10.4	18.5 ± 10.9
HLA-B27 positive	136 (81)	71 (86)	65 (74)	406 (83)	210 (86)	196 (81)
ASDAS (range)	2.6 (0.6-5.5)	2.5 (0.7-5.5)	2.6 (0.6-5.2)	2.6 (0-5-6.1)	2.5 (0-5-6.1)	2.6 (0.6-5.2)
BASDAI (0-10) (range)	4.3 (0.2-9.8)	4.1 (0.2-7.9)	4.5 (0.2-9.8)	4.3 (0-9.8)	3.9 (0-9.7)	4.6 (0.2-9.8)
BASFI (0-10)	2.9 ± 2.3	2.6 ± 2.1	3.1 ± 2.5	3.0 ± 2.3	2.7 ± 2.1	3.3 ± 2.3
Night pain (0-10) (range)	4.3 (0-10)	3.9 (0-10)	4.8 (0-10)	4.5 (0-10)	4.1 (0-10)	4.9 (0-10)
Pt's global (0-10) (range)	4.8 (0-10)	4.6 (0-10)	5.0 (0-10)	4.9 (0-10)	4.5 (0-10)	5.3 (0-10)
ESR (mm/h) (range)	15.5 (1-76))	14.7 (1-73)	16.3 (2-76)	14.9 (1-124)	14.1 (1-76)	15.8 (1-124)
CRP (mg/L) (range)	9.1 (0.1-88)	11.4 (0.3-88)	7.0 (0.1-67)	9.0 (0.1-91)	11.1 (0.3-91)	6.8 (0.1-74)
Elevated CRP	48 (30)	26 (33)	22 (26)	162 (35)	89 (39)	73 (31)
ASQoL (0-18)	8.9 ± 5.1	7.7 ± 5.0	10.0 ± 4.9	9.0 ± 5.0	7.9. ± 5.0	10.2 ± 4.8
Treatment (baseline)						
NSAID	162 (97)	79 (95)	83 (99)	460 (95)	232 (95)	228 (94)
Anti-TNFα	0	0	0	0	0	0
Treatment (at 24 mo)						
NSAID	112 (67)	52 (63)	60 (71)	278 (57)	129 (53)	149 (62)
Anti-TNFα	58 (35)	30 (36)	27 (32)	130 (27)	62 (25)	67 (28)
SPARCC (0-72) (range)	5.6 (0-42)	8.8 (0-42)	2.4 (0-23)	5.1 (0-55)	7.7 (0-55)	2.5 (0-26)
Positive MRI				NA	NA	NA
MRIpos/MRIpos**	49 (29)	30 (36.5)	19 (23)			
MRIpos/MRIneg	23 (14)	17 (21)	6 (7)			
MRIneg/MRIpos	11 (7)	5 (6)	6 (7)			
MRIneg/MRIneg	82 (50)	30 (36.5)	52 (63)			

Unless otherwise is specified, table shows mean ± standard deviation or absolute number (percentage) for all patients included in this analysis and the entire cohort. Furthermore, the results are also stratified by gender. Mo (months); HLA-B27: human leucocyte antigen B27; ASDAS: Ankylosing Spondylitis Disease Activity Index, BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; BASFI: Bath Ankylosing Spondylitis Functional Index; pt´s global: patient´s global assessment for disease activity; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein; ASQoL: Ankylosing Spondylitis Quality of Life; NSAIDs: nonsteroidal anti-inflammatory drugs; Anti-TNF: Anti TNF: Anti-tumor necrosis factor alpha therapy; MRI-SIJ: sacroiliac joint magnetic resonance imaging; ASAS: Assessment of SpondyloArthritis international Society; SPARCC: Spondyloarthritis Research Consortium of Canada scoring system. ** baseline/12 months.

included patients were similar to the entire cohort (Table 1).

The distribution of SPARCC scores, stratified for males and females, is shown in supplementary Figure S1. The results show that both the frequency (55 vs. 26%) and the amplitude (42 vs. 23 units) of positive MRI scores are higher in males than in females.

Relationship between clinical disease activity measures and MRI-SI

Standard model (absolute scores)

A significant interaction was found between gender and several DA (ASDAS, BASDAI, pain at night and patient´s global assessment) with regard to the relationship with MRI-SI. No other relevant interactions were found. In males, a significant relationship between all DA, except BASDAI, and MRI was found in the standard models (Table 2), with the lowest QIC values for the models with ASDAS and acute phase reactants (APRs). In contrast, in females none of the clinical measures were significantly related to MRI-SI inflammation.

Table 2: Longitudinal relationship (standard model) between disease activity measures and inflammation degree (SPARCC) on MRI-SI stratified for gender in six separate models.

	Males			Females		
	Beta	95% CI	QIC	Beta	95% CI	QIC
ASDAS	2.408	1.127 to 3.690	18,228	0.301	-0.530 to 1.133	3,700
BASDAI (0-10)	0.309	-0.330 to 0.947	20,558	-0.165	-0.473 to 0.143	3,588
Night pain (0-10)	0.497	0.040 to 0.954	20,525	0.014	-0.146 to 0.175	3,689
Pt's gb disease (0-10)	0.451	0.062 to 0.841	20,060	-0.028	-0.189 to 0.132	3,676
ESR (mm/h)	0.181	0.009 to 0.354	13,571	0.066	-0.007 to 0.139	2,279
CRP (mg/L)	0.149	0.028 to 0.270	18,128	0.047	-0.059 to 0.154	3,685

Six separate models were built for both genders, each one of them including one of the disease activity parameters. ASDAS: Ankylosing Spondylitis Disease Activity Index, BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; pt´s gb disease: patient´s global assessment for disease activity; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein. All models were adjusted for age, symptom duration and HLA-B27. 95% CI: 95% Confidence interval; QIC: Quasi likelihood under Independence Model Criterion.

Model of changes (change scores)

In males, changes in ASDAS, BASDAI and night pain between two consecutive visits were significantly associated with changes in SPARCC over the same time interval (Table 3, no longitudinal data for ESR available). The model with the ASDAS had the lowest QIC. Consistent with the models with absolute scores, in females the change in SPARCC was not significantly related with the change in any of the clinical measures.

All analyses were repeated for both readers separately and showed identical results for

both types of models. Analyses restricted to patients i) with a positive MRI at baseline (supplementary Table S1 and S2); ii) with an ASAS classification according to the 'imaging arm' (supplementary Table S3 and S4); and iii) who did not use anti-TNF therapy (not shown) yielded similar results.

Table 3: Longitudinal relationship (model of changes) between disease activity measures and inflammation degree (SPARCC) on MRI-SI stratified for gender in five separate models.

	Males			Females		
	Beta	95% CI	QIC	Beta	95% CI	QIC
ASDAS	2.792	0.850 to 4.734	10,739	0.909	-0.112 to 1.931	1,462
BASDAI (0-10)	0.971	0.040 to 1.903	12,099	0.244	-0.118 to 0.606	1,534
Night pain (0-10)	0.852	0.227 to 1.476	11,982	0.040	-0.140 to 0.220	1,578
Pt's gb disease (0-10)	0.574	-0.042 to 1.191	12,272	0.108	-0.066 to 0.282	1,579
CRP (mg/L)	0.119	-0.004 to 0.242	11,834	0.085	-0.039 to 0.208	1,531

Five separate models were built for both genders, each one of them including one of the disease activity parameters. ASDAS: Ankylosing Spondylitis Disease Activity Index, BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; pt´s gb disease: patient´s global assessment for disease activity; CRP: C-reactive protein. All models were adjusted for age, symptom duration and HLA-B27. 95% CI: 95% Confidence interval; QIC: Quasi likelihood under Independence Model Criterion.

DISCUSSION

This study shows that DA is longitudinally associated with MRI-SI inflammatory lesions over time, with ASDAS as the measure most closely related to MRI: an increase of one unit in ASDAS coincides with an increase of 2.8 units in SPARCC score. Further, this association is specific for males.

Among the investigated measures, ASDAS and pain at night were significantly related to MRI-SI inflammatory lesions in both types of models. Thus fulfilling a requirement for a longitudinal interpretation.

Unlike ASDAS, BASDAI was not significantly associated with MRI-SI in the standard model but were associated in the model of changes. Most likely this means that BASDAI, which is fully patient-reported, is not only reflecting inflammation (such as in SI-joints) but also patient-specific factors independent of inflammation, while ASDAS is likely a better reflection of inflammation than BASDAI, as has been shown by previous cross-sectional data too [4,11-14].

These longitudinal observations couple typical imaging findings such as bone marrow edema in the SI-joints to clinical signs and symptoms (disease activity) in patients with axSpA. But the question of whether or not to use of MRI-SI in monitoring patients axSpA has not been

solved yet and requires an in depth analysis of MRI-SI as a biomarker [15]. Several studies have partially assessed this but no study has evaluated both of them in the same cohort [11,16-18].

The gender-difference found in this study is remarkable and has not been addressed previously in detail. A recent study of ours has shown that the association between ASDAS and radiographic progression is stronger in males than in females [19]. In that AS-study it could not be entirely excluded that a low female patient number has precluded a robust conclusion. But in DESIR gender distribution is equal, and gender disparity should be considered as a true phenomenon. The question is what the gender disparity here means: Many will claim that in females a positive MRI is infrequent and of lower intensity. In this regard, we cannot entirely exclude that a true relationship does exist in females but that we have not been able to demonstrate it because of statistical reasons. Others argue that many females classified as axial SpA do not truly have axSpA, and that MRI in females often is false positive. While misclassification in females cannot be completely ruled out here, it cannot be an explanation for the gender disparity: even in the imaging positives, the association between disease activity and MRI-SI scores is gender-dependent. Differences in anti-TNF treatment between males and females can also not explain the gender disparity: analyses in those that did not use anti-TNF treatment (70% of patients) yielded similar results.

So the most likely explanation for the gender-disparity is therefore that axSpA has a different expression in males than in females, as has been suggested previously [20]. While in male patients clinical signs and symptoms coincide with MRI-positivity and with subsequent structural damage [19], in female patients symptoms attributed to SpA (and measured by patient-reported outcomes) occur independently of MRI-inflammation and subsequent structural damage. The reasons for such gender-related uncoupling are unknown and deserve further research.

This study has limitations. First, symptom duration was short while clinical measures and MRI-SI correlate better in patients with long disease duration [12]. Second, MRI-spine was not analyzed here.

Pertinent strengths of this study are the longitudinal analysis, the inclusion of patients with axSpA (and not only those fulfilling the modified New York criteria), and the analytical rigor: results obtained in all patients were similar when re-analyzed in patients that were imaging positive, and when separate MRI-reader scores in stead of average reader scores, were used.

In conclusion, we can state that in male, but not in female patients with axSpA, clinical disease activity measured by ASDAS is longitudinally associated with MRI-SI inflammatory lesions.

REFERENCES

1. Pedersen SJ, Weber U, Østergaard M. The diagnostic utility of MRI in spondyloarthritis. Best Pract Res Clin Rheumatol 2012;26:751-66.
2. van der Heijde D, Sieper J, Maksymowych WP, et al. Spinal inflammation in the absence of sacroiliac joint inflammation on magnetic resonance imaging in patients with active nonradiographic axial spondyloarthritis. Arthritis Rheumatol 2014;66:667-73.
3. Kiltz U, Baraliakos X, Karakostas P, et al. The degree of spinal inflammation is similar in patients with axial spondyloarthritis who report high or low levels of disease activity: a cohort study. Ann Rheum Dis 2012;71:1207-11.
4. Machado P, Landewe R, Braun J, et al. MRI inflammation and its relation with measures of clinical disease activity and different treatment responses in patients with ankylosing spondylitis treated with a tumour necrosis factor inhibitor. Ann Rheum Dis 2012;71:2002-5.
5. Jee WH, McCauley TR, Lee SH, et al. Sacroiliitis in patients with ankylosing spondylitis: association of MR findings with disease activity. Magn Reson Imaging 2004;22:245-50.
6. Dougados M, d'Agostino MA, Benessiano J, et al. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine 2011;78:598-603.
7. Machado P, Navarro-Compan V, Landewe R, et al. How to calculate the ASDAS if the conventional CRP is below the limit of detection or if using high sensitivity CRP? - An analysis in the DESIR cohort. Arthritis Rheumatol 2015;67:408-13.
8. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Rheum 2005;53:502-9.
9. Rudwaleit M, Jurik AG, Hermann KG, et al. Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. Ann Rheum Dis 2009;68:1520-7.
10. Twisk JW. Applied longitudinal data analysis for epidemiology: a practical guide: Cambridge University Press; 2013.
11. Baraliakos X, Heldmann F, Callhoff J, et al. Which spinal lesions are associated with new bone formation in patients with ankylosing spondylitis treated with anti-TNF agents? A long-term observational study using MRI and conventional radiography. Ann Rheum Dis 2014;73:1819-25.
12. Weiss A, Song IH, Haibel H, et al. Good correlation between changes in objective and subjective signs of inflammation in patients with short- but not long duration of axial spondyloarthritis treated with tumor necrosis factor-blockers. Arthritis Res Ther 2014;16:R35.
13. Pedersen SJ, Sorensen IJ, Hermann KG, et al. Responsiveness of the Ankylosing Spondylitis Disease Activity Score (ASDAS) and clinical and MRI measures of disease activity in a 1-year follow-up study of patients with axial spondyloarthritis treated with tumour necrosis factor alpha inhibitors. Ann Rheum Dis 2010;69:1065-71.
14. Konca S, Keskin D, Ciliz D, et al. Spinal inflammation by magnetic resonance imaging in patients with ankylosing spondylitis: association with disease activity and outcome parameters. Rheumatol Int 2012;32:3765-70.
15. Maksymowych WP, Landewe R, Tak PP, et al. Reappraisal of OMERACT 8 draft validation criteria for a soluble biomarker reflecting structural damage endpoints in rheumatoid arthritis, psoriatic arthritis, and spondyloarthritis: the OMERACT 9 v2 criteria. J Rheumatol 2009;36:1785-91.
16. Maksymowych WP, Chiowchanwisawakit P, Clare T, et al. Inflammatory lesions of the spine on magnetic resonance imaging predict the development of new syndesmophytes in ankylosing spondylitis: evidence of a relationship between inflammation and new bone formation. Arthritis Rheum 2009;60:93-102.
17. Chiowchanwisawakit P, Lambert RG, Conner-Spady B, et al. Focal fat lesions at vertebral corners on magnetic resonance imaging predict the development of new syndesmophytes in ankylosing spondylitis. Arthritis Rheum 2011;63:2215-25.
18. van der Heijde D, Machado P, Braun J, et al. MRI inflammation at the vertebral unit only marginally predicts new syndesmophyte formation: a multilevel analysis in patients with ankylosing spondylitis. Ann Rheum Dis 2012;71:369-73.
19. 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;73:1455-61.
20. van Onna M, Jurik AG, van der Heijde D, et al. HLA-B27 and gender independently determine the likelihood of a positive MRI of the sacroiliac joints in patients with early inflammatory back pain: a 2-year MRI follow-up study. Ann Rheum Dis 2011;70:1981-5.

SUPPLEMENTARY MATERIAL

Supplementary Table S1: Longitudinal relationship (standard model) between disease activity measures and inflammation degree (SPARCC) on MRI-SI stratified for gender in six separate models, in patients with a positive MRI-SI at baseline.

	Males (n=43)			Females (n=23)		
	Beta	95% CI	QIC	Beta	95% CI	QIC
ASDAS	4.265	2.404 to 6.127	11,009	0.600	-1.047 to 2.246	1,960
BASDAI (0-10)	0.833	-0.113 to 1.778	13,117	-0.259	-0.910 to 0.393	1,910
Night pain (0-10)	0.715	0.101 to 1.329	13,164	0.158	-0.248 to 0.565	1,841
Pt's gb disease (0-10)	0.721	0.103 to 1.339	12,579	0.005	-0.415 to 0.425	1,960
ESR (mm/h)	0.240	0.007 to 0.473	8,009	0.113	0.046 to 0.179	873
CRP (mg/L)	0.185	0.025 to 0.345	11,737	0.105	-0.174 to 0.384	1,954

Six separate models were built for both genders, each one of them including one of the disease activity parameters. ASDAS: Ankylosing Spondylitis Disease Activity Index, BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; pt´s gb disease: patient´s global assessment for disease activity; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein. All models were adjusted for age, symptom duration and HLA-B27. 95% CI: 95% Confidence interval; QIC: Quasi likelihood under Independence Model Criterion.

Supplementary Table S2: Longitudinal relationship (model of changes) between disease activity measures and inflammation degree (SPARCC) on MRI-SI stratified for gender in five separate models, in patients with a positive MRI-SI at baseline.

	Males (n=43)			Females (n=23)		
	Beta	95% CI	QIC	Beta	95% CI	QIC
ASDAS	4.440	1.683 to 7.197	7,840	1.781	-0.755 to 4.317	1,194
BASDAI (0-10)	1.627	0.408 to 2.845	9,416	0.447	-0.661 to 1.554	1,268
Night pain (0-10)	1.299	0.603 to 1.996	9,596	0.134	-0.319 to 0.588	1,322
Pt's gb disease (0-10)	0.987	0.015 to 1.960	10,020	0.155	-0.378 to 0.688	1,329
CRP (mg/L)	0.186	-0.012 to 0.384	9,452	0.257	-0.097 to 0.611	1,212

Five separate models were built for both genders, each one of them including one of the disease activity parameters. ASDAS: Ankylosing Spondylitis Disease Activity Index, BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; pt´s gb disease: patient´s global assessment for disease activity; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein. All models were adjusted for age, symptom duration and HLA-B27. 95% CI: 95% Confidence interval; QIC: Quasi likelihood under Independence Model Criterion.

Supplementary Table S3: Longitudinal relationship (standard model) between disease activity measures and inflammation degree (SPARCC) on MRI-SI stratified for gender in six separate models, in patients with an ASAS classification according to the imaging arm.

	Males (n=60)			Females (n=48)		
	Beta	95% CI	QIC	Beta	95% CI	QIC
ASDAS	3.068	1.539 to 4.597	15,081	0.660	-0.533 to 1.852	2,958
BASDAI (0-10)	0.549	-0.269 to 1.367	17,476	-0.085	-0.546 to 0.376	2,959
Night pain (0-10)	0.726	-0.192 to 1.259	17,368	0.091	-0.140 to 0.323	2,948
Pt's gb disease (0-10)	0.748	0.264 to 1.233	16,748	0.016	-0.222 to 0.253	2,989
ESR (mm/h)	0.201	-0.014 to 0.417	11,356	0.098	0.004 to 0.193	1,677
CRP (mg/L)	0.166	0.032 to 0.300	15,334	0.111	-0.114 to 0.335	2,933

Six separate models were built for both genders, each one of them including one of the disease activity parameters. ASAS: Assessment of SpondyloArthritis international Society, ASDAS: Ankylosing Spondylitis Disease Activity Index, BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; pt´s gb disease: patient´s global assessment for disease activity; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein. All models were adjusted for age, symptom duration and HLA-B27. 95% CI: 95% Confidence interval; QIC: Quasi likelihood under Independence Model Criterion.

Supplementary Table S4: Longitudinal relationship (model of changes) between disease activity measures and inflammation degree (SPARCC) on MRI-SI stratified for gender in five separate models, in patients with an ASAS classification according to the imaging arm.

	Males (n=60)			Females (n=48)		
	Beta	95% CI	QIC	Beta	95% CI	QIC
ASDAS	3.265	1.092 to 5.438	9,840	1.446	-0.016 to 2.908	1,291
BASDAI (0-10)	1.168	0.124 to 2.212	7819	0.380	-0.159 to 0.920	1344
Night pain (0-10)	1.141	0.498 to 1.783	11,040	0.090	-0.142 to 0.322	1,432
Pt's gb disease (0-10)	0.899	0.187 to 1.611	11,416	0.167	-0.096 to 0.429	1,430
CRP (mg/L)	0.134	-0.018 to 0.286	11,123	0.193	-0.060 to 0.446	1,332

Five separate models were built for both genders, each one of them including one of the disease activity parameters. ASAS: Assessment of SpondyloArthritis international Society, ASDAS: Ankylosing Spondylitis Disease Activity Index, BASDAI: Bath Ankylosing Spondylitis Disease Activity Index; pt´s gb disease: patient´s global assessment for disease activity; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein. All models were adjusted for age, symptom duration and HLA-B27. 95% CI: 95% Confidence interval; QIC: Quasi likelihood under Independence Model Criterion.

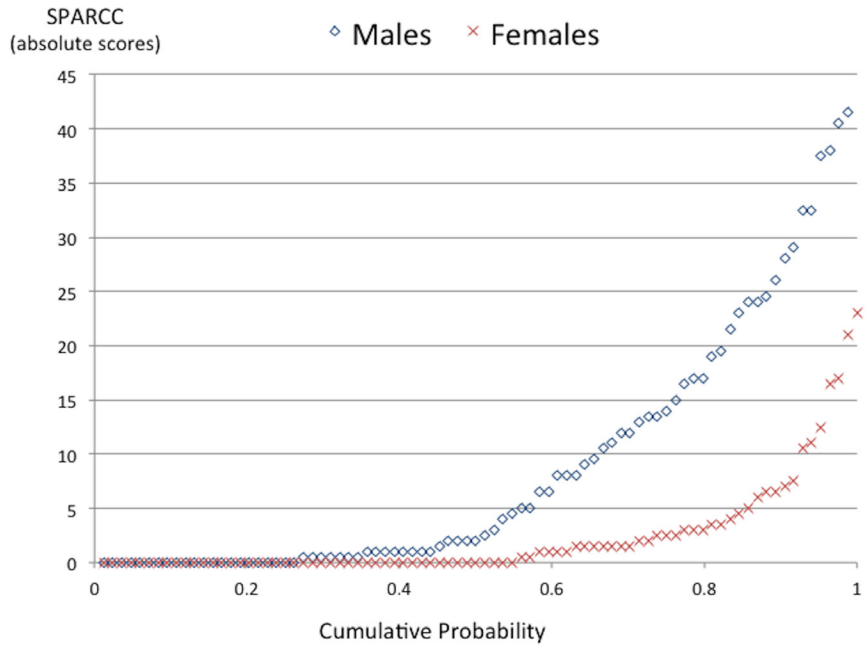


Figure S1A: Individual absolute SPARCC scores at baseline.

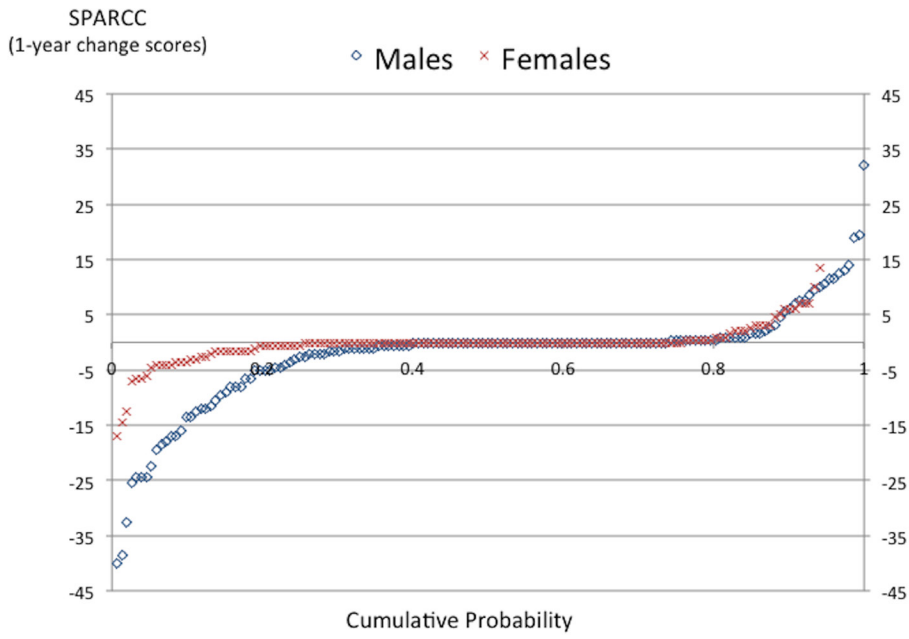


Figure S1B: Individual 1-year change SPARCC scores.

Supplementary Figure S1: Individual absolute (figure S1A) and 1-year change (figure S1B) SPARCC scores for males and females.

10

Summary and Conclusions

SUMMARY AND CONCLUSIONS

The studies of this thesis cover outstanding aspects of research methodology in the assessment of inflammation and damage in patients with rheumatoid arthritis (RA) and axial spondyloarthritis (axSpA). The studies pertaining to part I focus on RA and may help to better understand the relationship between disease activity, radiographic damage and disability in patients with RA. More specifically, they evaluate which disease activity measure is best associated with radiographic damage and which of the structural lesions detectable on radiographs is most associated with disability. In addition, these studies also address methodological issues related to the optimal assessment of radiographic progression in clinical trials. In part II of this thesis, the studies have axSpA as the main topic. In detail, they evaluate the yield of tests reflecting inflammation in general. The purpose of these tests is to help classifying patients or to monitor disease activity. Among others, the studies described in this thesis provide data for a better usage of these tests in clinical practice.

In this chapter, we present a summary of the main findings of these studies. Thereafter, we will come to a synthesis of these findings with a focus on principles of assessment and their relevance to determining outcome in chronic inflammatory rheumatic diseases. Simultaneously, we will shape potential future research questions in the field of assessing outcome.

RHEUMATOID ARTHRITIS

Relationship between disease activity measures and radiographic damage

At the start of the studies described in this thesis, disease activity indices (DAIs) were considered better than single variable instruments, including patient-reported outcomes (PROs), for monitoring disease activity in patients with RA [1]. This prioritization was largely based on expert opinion but was supported by limited data only. We have performed a systematic literature review to explore the existing knowledge about the relationship between the DAIs and their individual components on the one hand, and radiographic progression on the other hand. We considered this type of information essential to decide which tool should be employed to monitor disease activity (**chapter 2**) [2]. In total, 57 studies were included in this review and it was shown that all DAIs that include a joint count are related to radiographic progression. In addition, among the single variable instruments, only measures reflecting inflammation, such as swollen joint count (SJC) and erythrocyte sedimentation rate (ESR), were related to radiographic progression. Importantly, PROs did not show such an association. Accordingly, we have therefore recommended the use of one of the DAIs including a SJC to monitor disease activity in patients with RA in clinical practice.

Relationship between types of radiographic damage and disability

Radiographic damage assessed in patients with RA usually combines three types of radiological lesions reflecting different kinds of structural damage: 'true' joint space narrowing (JSN), erosion and subluxation or complete luxation -(sub)luxation- [3]. In previous studies the level of radiographic damage was shown to be associated with disability in patients with RA, but the contribution of each subtype of lesions to several aspects of disability was not known [4,5].

In the study described in **chapter 3** we have performed a longitudinal analysis on the 10-year follow-up data from patients included in the Norwegian arm of the European Research on Incapacitating Diseases and Social Support (EURIDISS) cohort [6]. The topic of interest was the relationship between each subtype of lesion and disability. All 3 subtypes observed on radiographs (true JSN, erosion and (sub)luxation) were related to grip strength but only JSN, especially in the wrist, appeared to be independently associated: we found that an increment of 11 true JSN units in both hands will on average lead to a decrease of 1 kg in mean grip strength of both sides over a period of 10 years. Most reported studies have established a clear relationship between absolute radiographic damage (i.e. the sum of the three type of lesions scores) and overall disability, measured by the health assessment questionnaire (HAQ)-score. In our study, none of the three subtypes of lesions was independently associated with HAQ-score. Based on this, we have concluded that true JSN in the hands may contribute more to explaining variation in hand function than erosion and (sub)luxation do, while at the patient level all 3 types of radiographic damage in an aggregated fashion (as total score) contribute to explaining variation in overall disability.

Methodological aspects of assessment of radiographic damage

Because of the association with (changes in) disease activity and functional disability, inhibition (or slowing) of radiographic progression has been established by regulatory authorities as one of the claims that could be granted for new treatments in RA [7]. In order to appropriately investigate such a claim, a number of typical methodological issues were outstanding.

Adjudication

The first issue pertains to 'adjudication'. Usually, in clinical trials two readers independently provide scores to sets of images that they have to judge with unknown chronology. If an important discrepancy between the scores of these two readers occurs, a third reader (named the adjudicator) is asked to provide an adjudication score that, together with the closest score of the two initial readers, is used to obtain the 'mean change score'. This process, which serves to constrain measurement error in the trial, is methodologically legitimate but requires a certain threshold for the difference between the two primary readers before adjudication is started. This threshold is arbitrary, but regulatory authorities have expressed concerns if -for a given clinical trial- 20% or more of the final mean change scores of patients

are resolved by adjudication. Often, using the Sharp-van der Heijde (SHS) method, differences between reader-scores of 7-15 units were operationalized as adjudication threshold, but this choice was completely arbitrary. Being aware of this methodological issue, we aimed to provide data regarding the proportions of patients to be adjudicated given a predetermined threshold for the difference in change score between two initial readers in RA trials (**chapter 4**). We have analysed datasets of 15 recent randomised controlled trials (RCTs) with 2-4 time points per trial in which radiographic progression had been assessed by 13 readers acting in pairs. As expected, the adjudication rate was inversely related to the threshold for the difference between the two readers. But the rate was rather low (always below 22%) even if very conservative thresholds of ≥3 units were applied. In addition, we found that particular features could influence the adjudication rates: the adjudication rate increased by increasing number of time points per trial, by a longer time gap between visits and by shorter disease duration.

Smallest detectable change

The second unresolved methodological issue pertains to the question how to decide if radiographic progression in an individual patient who has participated in a clinical trial is 'true' (beyond measurement error) or not. There is consensus about the preferred metric here: the smallest detectable change (SDC) [8,9]. But there is no consensus about how to determine this SDC if RCTs with more than two time points are analysed. At the beginning of this study, two analytical methods for estimating an SDC were available: One 'simple' method based on Bland & Altman (B&A) analysis; and a more complex method based on generalizability theory and involving analysis of variance (ANOVA) [9]. The simple B&A-method suffices for two time points trials, but for complex datasets with several time points such as in most recent RCTs only the method based on generalizability theory is available. In **chapter 5**, using same data employed in the analysis of the study in **chapter 4**, we propose a simple extension of the B&A-method: we have investigated if the mean of all interval-SDCs obtained by the simple B&A method is an appropriate surrogate for the 'ANOVA-method' for estimating the overall SDC of radiological progression in complex databases. For this purpose, we have evaluated the agreement between the two different methods. The mean (standard deviation) difference observed between the two methods was only −0.13 (0.28), range (−0.48, 0.25) units. If we consider the minimal clinically important difference for radiographic progression (3 units of SHS method) [10], this difference of 0.13 units is negligible. Accordingly, we propose to report the average of all interval SDCs as an appropriate surrogate for the ANOVA-based SDC in complex databases.

We have addressed another methodological issue of B&A-based SDC calculations: The B&A method purports to determine an upper and lower level of agreement (LoA), which are boundaries enclosing an area within which it cannot be precluded that a measured change (here: progression score) in an individual patient is due to measurement error rather than to true change. The upper and lower LoAs are statistically derived and enclose 95% of all

observed data. The choice for 95% threshold is arbitrary and not well justified in literature. The 95% LoA resemble 95%-confidence intervals but are fundamentally different [11]. We thought that, because of this fundamental difference, 95% LoA are unnecessarily conservative in the context of radiographic progression scores, where they importantly limit the sensitivity to detect subtle changes. We have compared SDCs based on the 95% LoA with the SDCs based on the 80% LoA (**chapter 4**). The mean (standard deviation) SDC was 3.0 (0.7) when based on 95% LoA and 2.0 (0.4) when based on 80% LoA. If we choose 80% LoA, more patients would classify as 'true progressors' in both groups, which is of (modest) statistical benefit (more statistical power). But there is a more clear advantage related to this: a more liberal SDC will lead to more patients being classified as 'progressor'. The aim of an RCT is to investigate if a new treatment *reduces* the number of 'progressors'. In order to reliably demonstrate this, it is important that there is a sufficient proportion of patients with 'progression beyond measurement error' in the control group, and a 80% LoA SDC helps achieving this without unacceptably jeopardizing trial-results.

AXIAL SPONDYLOARTHRITIS

Issues of diagnosis

HLA-B27 and MRI-SI

Unlike many other diseases in rheumatology, axSpA is a disease in which supplementary tests and procedures play an important role to make a diagnosis or to measure disease activity. These supplementary procedures include relatively simple laboratory tests such as tests to quantify the level of acute phase reactants (APRs) and a blood test to determine human leukocyte antigen B27 (HLA-B27) carriership, but also relatively complicated imaging procedures such as magnetic resonance imaging of the sacroiliac joints (MRI-SI) that quantifies bone marrow edema [12]. In part II of this thesis, several aspects of these supplementary procedures for different purposes were further evaluated.

First, cost-, availability- and feasibility issues make that HLA-B27 testing and especially MRI-SI cannot be performed in all patients presenting with chronic back pain (CBP), as this is a very prevalent symptom in clinical practice and still the most common presenting symptom in patients with axSpA [13].

In **chapter 6**, we have addressed the principles of sequential testing by investigating if particular SpA [14] features point to a higher likelihood of a subsequent positive MRI-SI or a positive HLA-B27 test in patients with CBP referred to rheumatologists participating in the ESPeranza Programme [15]. Such SpA features, that can be obtained by history taking or simple physical examination, could potentially contribute to an efficient 'test-sequence' to

be applied in patients presenting with CBP. The prevalence of a positive MRI-SI or a positive HLA-B27 test in patients in ESPeranza was 41% and 40%, respectively. Unfortunately, we did not find any of the SpA features to increase the likelihood of a subsequent positive HLA-B27 test. But interestingly, we found that the presence of the symptom 'awakening at second half of night' together with inflammatory back pain (IBP) according to the Calin definition or the presence of the rather uncommon symptom 'alternating buttock pain' together with IBP according to the Assessment of SpondyloArthritis international Society (ASAS)/Calin criteria increased the likelihood on a subsequent positive MRI-SI from 40 to 79-80%. In addition, we found that the findings 'dactylitis' and 'presence of inflammatory bowel disease' increased this likelihood from 41 to 73% and 81%, respectively. Based on these results, and in case of limited resources, the presence of any of these four characteristics may be valuable in helping rheumatologists to improve the efficiency of ordering MRI-SI in patients with suspected axSpA.

High sensitivity CRP

Concerning how to diagnose patients with early axSpA, the new classification ASAS criteria developed in 2009 have shown to outperform the older criteria [16]. 'Elevated C-reactive protein –CRP-' is one of the SpA features in the ASAS criteria. At the time of development of these criteria, CRP was measured with the conventional CRP-methods. Recently, high sensitivity CRP (hsCRP)-methods, which are more sensitive than conventional CRP-methods, have been developed [17] and have replaced the conventional CRP methods in many hospitals. However, hsCRP-methods have not been validated for application in the ASAS criteria. We have evaluated in patients included in the Devenir des Spondylarthopathies Indifférenciées Récentes (DESIR) cohort [18] to what extent the replacement of the conventional-CRP test by the hsCRP test would (maybe inappropriately) increase the sensitivity of the ASAS criteria for classifying patients with axSpA (**chapter 7**). In the subgroup of DESIR-patients with normal conventional CRP, we have observed higher hsCRP mean serum levels in the patients with axSpA as compared to patients without axSpA. But importantly, when elevated conventional CRP was substituted by elevated hsCRP, not one single patient could additionally be classified as axSpA. This means that –for the purpose of classifying patients with axSpA– conventional CRP-methods can be replaced by hsCRP-methods without further consideration.

Issues of monitoring disease activity

High sensitivity CRP

Changing assessment-methods, such as the methods to measure CRP, may not only affect the performance of criteria for axSpA but also the instruments to measure disease activity of axSpA. The Ankylosing Spondylitis Disease Activity Score (ASDAS), an index that was recently developed, integrates several PROs as well as an APR reflecting inflammation. At the development of the ASDAS, only conventional CRP (but not hsCRP) was used. Arbitrarily,

and awaiting further data driven guidance, it was suggested to use 50% of the threshold value of CRP in the ASDAS formula if conventional CRP was below the detection limit. This suggestion was not supported by data. In the study described in **chapter 8**, we have first selected patients from DESIR with a conventional CRP below the limit of detection, then have calculated the ASDAS using hsCRP values and finally compared these ASDAS values with the values of ASDAS obtained by imputing CRP with 11 different artificial threshold-values (range 0-5 mg/L, at 0.5 mg/L intervals). The best agreement with truly measured hsCRP (the external standard) was found for imputed values of 1.5 and 2 mg/L, with only a minimal discrepancy between them. Based on the performance in clinically relevant disease activity states a value of 2 mg/L was finally proposed as the best option for imputation: if the conventionally measured CRP level yields a result reported as 'below the limit of detection' or if the hsCRP is reported as 'below 2 mg/L', a value of 2 mg/L should be imputed for CRP in the ASDAS formula. In addition, we also took the opportunity to evaluate if the ASDAS employing ESR (instead of CRP) was in agreement with the ASDAS using hsCRP. Our findings confirmed that the agreement between the two formulae is good.

MRI-SI

While the value of MRI-SI in diagnosing patients with axSpA is undisputed, the added value of MRI-SI in monitoring disease activity in patients with axSpA was (and still is) a matter of debate. In this regard, first it was important to investigate if signs and symptoms, quantified by simple clinical or laboratory measures, are longitudinally associated with the presence of inflammatory lesions detected by MRI-SI. Data regarding this topic were scarce and had mainly been obtained from studies including patients with established disease and employing only cross-sectional analysis [19,20]. In **chapter 9**, we have investigated and observed for the first time the existence of a longitudinal relationship between clinical disease activity measures and inflammatory lesions detected by MRI-SI, assessed according to the Spondyloarthritis Research Consortium of Canada (SPARCC) score [14]. Among all measures, the ASDAS was best associated to MRI-SI: on average, an increase of one unit in ASDAS is associated with an increase of 2.8 units in SPARCC score. Interestingly, we also have shown for the first time that the relationship between clinical disease activity measures and MRI-SI inflammatory lesions is different in males and females: there was a marked association between ASDAS and MRI-SI in males, while this association was absent in females. Translating these data into clinical terms, we suggest that monitoring the ASDAS in male patients with axSpA provides useful 'subjective' information as well as 'objective information' that is congruent with the pathophysiological hallmark of the disease, being this inflammation of the sacroiliac joints (on MRI). In females with axSpA the situation is far less clear: the lack of association between the ASDAS and MRI-SI in females points again to the observation that axSpA has a different expression in males than females (see below).

IMPLICATIONS AND RELEVANCE OF THIS THESIS

The central theme of this thesis is *assessment*. Reading the chapters of this thesis it becomes clear that *assessment* influences many aspects of outcome in inflammatory rheumatic diseases. The basic construct underlying the research in this thesis is the well-known appreciation that inflammation in a chronic inflammatory rheumatic disease such as RA or axSpA leads to some kind of structural damage, which in turn contributes to functional impairment, which can be considered a long-term outcome.

It is important to realize that all these levels can and should be *assessed* so that diseases can better be diagnosed, prognostic factors for outcome can better be identified, pivotal associations can be constructed, long-term functional impairment can better be explained, interventions can be designed to improve long-term outcome, etc: better *assessment* will lead to a better understanding and a better management of our chronic inflammatory diseases.

The importance of *assessment* in understanding the different levels of the here defined outcome pyramid will be briefly discussed below one-by-one:

1. Assessment starts at diagnosis. Inflammatory rheumatic diseases are typically diagnoses by 'pattern recognition': the diagnostician combines his knowledge about the pattern of a disease (the 'Gestalt') with information obtained from patients presenting to him with complaints and symptoms. Rheumatology is a medical discipline in which the pattern of the disease is spelled out by classification criteria, not by an unambiguous external [21] standard! These criteria are the product of consensual deliberations among experts (including patients) who have integrated the best of their knowledge into classification algorithms, followed by validation studies. RA and axSpA are examples of diseases in which the pattern is defined by classification criteria. Many of these criteria should be quantified, *assessed*. This feature makes criteria susceptible to technical and methodological developments over time. An often overlooked consequence of this is that every change that takes place in the *assessment* of criteria must be validated to some extent in order to be able to oversee the consequences of such a change. The replacement of conventional CRP tests by hsCRP tests, as highlighted in this thesis, seems a logical move, but is in fact an example of a change with potentially bad consequences for the classification of axSpA. But obviously issues like this do not pertain solely to laboratory tests: MRI-SI is another example of a procedure with relevance for a diagnosis or a classification. Changing the content of a positive MRI-result, or the threshold, would imply a change in the *assessment* of the disease that may have important consequences for classification and diagnosis.

In this thesis we have also addressed a slightly different aspect of diagnostic *assessment*: we have explored sequential testing in axSpA; we have investigated if simple clinical

tests may increase the likelihood of a certain result obtained by another diagnostic test (MRI-SI or HLA-B27). Sequential testing is a technique that is implicitly applied by experienced clinicians who make a diagnosis. Essentially, sequential testing means that the ordering of a second (often more costly or incriminating) test is made on the basis of the result of the first (often simple) test. Sequential testing is clinical reasoning, which is the opposite of 'checkbox medicine' that implies ticking checkboxes in a set of classification criteria. Sequential testing is also congruent with Bayesian reasoning in that 'prior knowledge' is taken into consideration when deciding if additional testing in an individual patient should be carried out. We have given an example of a study here that makes transparent that sequential testing (in axSpA) may increase the yield of subsequent more delicate diagnostic procedures (e.g. MRI-SI). It is our conviction that Bayesian reasoning in Rheumatology will lead to more accurate diagnoses, is less expensive, and more satisfactory to patient and physician, than 'cookbook medicine' by protocol based on sets of classification criteria that nowadays and unfortunately is propagated in many countries as a means to control health care costs.

2. The next levels of our outcome construct pertain to the *assessment* of disease activity and structural damage. While we do not dispute anymore that in inflammatory rheumatic diseases inflammation (operationalized as disease activity) leads to damage, we may ignore too often that the strength of this relationship is subjected to methodological principles of *assessment* underlying disease activity and damage. In other words: if we change the way in which we measure disease activity or damage, or change the components of a disease activity measure or a damage measure, we may also change the strength of the association, and consequently the impact that this association may have on our perception of the disease. In this thesis we have investigated several aspects of *assessment* of disease activity and damage. Not only have we substantiated once again that the association between disease activity and damage is best served by choosing DAIs rather than by single components, but we have also worked out 'open ends' in the assessment of structural damage by addressing topics like 'adjudication' and 'smallest detectable change'. At first glance, these topics may seem trivial, until one realizes that –in the context of clinical trials– adjudication influences the precision of a scoring result, and the determination of the SDC directly influences the number of patients with progression of damage in a trial. In summary, these seemingly trivial aspects of *assessment* may each have its influence on the strength of the association between disease activity and structural damage, and are therefore relevant to mention. Along similar lines, one may argue that relatively subtle changes in the method to assess disease activity, such as the replacement of conventional CRP by hsCRP in the formula of the ASDAS may or may not have repercussions on the performance of the ASDAS, and consequently on the association of ASDAS and structural damage in axSpA [22].

3. The last level of our outcome pyramid pertains to disability. Physical function or the impairment thereof, has many components in itself: in RA impaired grip-strength implies impairment of the hand-function, but an increased HAQ-score far more refers to general

disability. Needless to say that if one aims at investigating the association between damage and physical function in RA, it is crucial to define how to *assess* damage and how to *assess* disability. Far too often, these associations have been investigated at a generic level by combining a total damage score (such as the total SHS in RA or the modified Stoke Ankylosing Spondylitis Score in axSpA) with a generic measure for disability (such as the HAQ-score in RA or the Bath Ankylosing Spondylitis Functional Index in axSpA). By exploring the example of RA, we have disentangled in this thesis the subcomponents of the total SHS (JSN, (sub)luxation and erosion), and demonstrated that the strength of the associations between damage and function is dependent on the choice of the damage *assessment* and the choice of the disability *assessment*. Future investigators should better realize that these associations are not 'a given' but are dependent on the context, and that choices regarding the *assessment* of damage and disability matter with regard to the interpretation of the results.

As described above, principles of *assessment* are of pivotal importance for the understanding of our chronic inflammatory diseases, and we have argued how seemingly trivial aspects of *assessment* may have a far-stretching impact on our perception of the diseases in all their facets in this thesis.

Assessment may also influence our pathogenic thinking about disease, and we have given an example of this in the last chapter of this thesis, where we have investigated the longitudinal relationship between disease activity assessed by the ASDAS and inflammation assessed by MRI-SI: not only have we formally established a longitudinal association between ASDAS and MRI-SI, but more importantly have we demonstrated that this association was fundamentally different in males as compared to females. This observation, which cannot be explained by methodological artefacts, points to the importance of gender in explaining phenotypical differences in axSpA, and fits neatly with previous observations in ankylosing spondylitis that disease activity (assessed by ASDAS) and syndesmophyte formation are associated in males but not in females [22]. Future researchers can take this example of dissociation of symptoms and signs of inflammation by gender as a starting point for further (translational and clinical) research.

IN CONCLUSION

In this thesis we have argued that methodological principles of assessment are as important as issues related to content of measurement instruments for measuring outcome in patients with inflammatory rheumatic conditions and for better understanding these diseases.

If improving the outcome of these diseases is a major goal of research in the field, sufficient attention for principles of assessment and proper implementation of them into clinical practice are of pivotal importance.

REFERENCES

1. Smolen JS, Aletaha D, Bijlsma JW, et al. Treating rheumatoid arthritis to target: recommendations of an international task force. Ann Rheum Dis 2010;69:631-7.
2. Scott DL, Pugner K, Kaarela K, et al. The links between joint damage and disability in rheumatoid arthritis. Rheumatology (Oxford) 2000;39:122-32.
3. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.
4. Bombardier C, Barbieri M, Parthan A, et al. The relationship between joint damage and functional disability in rheumatoid arthritis: a systematic review. Ann Rheum Dis 2012;71:836-44.
5. Ødegard S, Landewe R, van der Heijde D, et al. Association of early radiographic damage with impaired physical function in rheumatoid arthritis: a ten-year, longitudinal observational study in 238 patients. Arthritis Rheum 2006;54:68-75.
6. Syversen SW, Gaarder PI, Goll GL, et al. High anti-cyclic citrullinated peptide levels and an algorithm of four variables predict radiographic progression in patients with rheumatoid arthritis: results from a 10-year longitudinal study. Ann Rheum Dis 2008;67:212-7.
7. Smolen JS, Aletaha D. Monitoring rheumatoid arthritis. Curr Opin Rheumatol 2011;23:252-8.
8. van der Heijde D, Simon L, Smolen J, et al. How to report radiographic data in randomized clinical trials in rheumatoid arthritis: guidelines from a roundtable discussion. Arthritis Rheum 2002;47:215-8.
9. Bruynesteyn K, Boers M, Kostense P, et al. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. Ann Rheum Dis 2005;64:179-82.
10. Bruynesteyn K, van der Heijde D, Boers M, et al. Minimal clinically important difference in radiological progression of joint damage over 1 year in rheumatoid arthritis: preliminary results of a validation study with clinical experts. J Rheumatol 2001;28:904-10.
11. Landewe R, van der Heijde D. Principles of assessment from a clinical perspective. Best Pract Res Clin Rheumatol 2003;17:365-79.
12. Rudwaleit M, van der Heijde D, Khan MA, et al. How to diagnose axial spondyloarthritis early. Ann Rheum Dis 2004;63:535-43.
13. Sieper J, Srinivasan S, Zamani O, et al. Comparison of two referral strategies for diagnosis of axial spondyloarthritis: the Recognising and Diagnosing Ankylosing Spondylitis Reliably (RADAR) study. Ann Rheum Dis 2013;72:1621-7.
14. Maksymowych WP, Inman RD, Salonen D, et al. Spondyloarthritis Research Consortium of Canada magnetic resonance imaging index for assessment of spinal inflammation in ankylosing spondylitis. Arthritis Rheum 2005;53:502-9.
15. Muñoz-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;70:827-30.
16. 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;68:777-83.
17. Poddubnyy DA, Rudwaleit M, Listing J, et al. Comparison of a high sensitivity and standard C reactive protein measurement in patients with ankylosing spondylitis and non-radiographic axial spondyloarthritis. Ann Rheum Dis 2010;69:1338-41.
18. Dougados M, d'Agostino MA, Benessiano J, et al. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine 2011;78:598-603.
19. Jee WH, McCauley TR, Lee SH, et al. Sacroiliitis in patients with ankylosing spondylitis: association of MR findings with disease activity. Magn Reson Imaging 2004;22:245-50.
20. Kiltz U, Baraliakos X, Karakostas P, et al. The degree of spinal inflammation is similar in patients with axial spondyloarthritis who report high or low levels of disease activity: a cohort study. Ann Rheum Dis 2012;71:1207-11.
21. Smolen JS, van der Heijde D, Keystone EC, et al. Association of joint space narrowing with impairment of physical function and work ability in patients with early rheumatoid arthritis: protection beyond disease control by adalimumab plus methotrexate. Ann Rheum Dis 2013;72:1156-62.
22. 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;73:1455-61.

11

Samenvatting en conclusies

SAMENVATTING EN CONCLUSIES

Reumatoïde artritis (RA) en axiale spondyloartritis (axSpA) zijn chronische ziekten die gekenmerkt worden door ontstekingen. Dergelijke ontstekingen kunnen verschillende weefsels en organen in het lichaam aantasten. Maar meestal starten de ontstekingen in de gewrichten waar ze klachten als pijn en stijfheid veroorzaken. In RA ontstaan de ontstekingen meestal in de kleine gewrichten van de handen en voeten. Bij axSpA ontstaan de ontstekingen juist in de grote gewrichten van het bekken en het heiligbeen (de sacro-iliacale of SI gewrichten) en de rug.

In de loop van de tijd kunnen dergelijke ontstekingen leiden tot aantasting van kraakbeen en bot, maar ook van pezen en ligamenten in en rondom de gewrichten. De ontstekingen kunnen ook leiden tot het aanmaken van extra bot. Dergelijke schade kan uiteindelijk leiden tot vervorming van de gewrichten (anatomische veranderingen), welke op hun beurt kunnen leiden tot belemmeringen in het dagelijks leven waardoor de kwaliteit van leven vermindert. Dus de ontstekingen en de onomkeerbare consequenties kunnen -indien onbehandeld - grote invloed hebben op het dagelijks leven van patiënten met RA en axSpA.

Er is niet een enkel symptoom, kenmerk of aanvullende test op basis waarvan de diagnose RA of axSpA te stellen is. Daarom stelt de reumatoloog in de dagelijkse praktijk de diagnose vaak op basis van een patroon van meerdere symptomen en kenmerken, al dan niet in combinatie met aanvullende laboratoriumtesten of beeldvormende technieken. Voor wetenschappelijk onderzoek zijn verschillende kenmerken en testen gecombineerd in classificatiecriteria. Voor RA en voor axSpA bestaan verscheidene classificatiecriteria. De meest toegepaste classificatiecriteria voor RA zijn de American College of Rheumatology (ACR) en European League Against Rheumatism (EULAR) criteria, en voor axSpA zijn dat de Assessment of SpondyloArthritis international Society (ASAS) criteria.

Aangezien persisterende ontstekingen leiden tot onomkeerbare schade, is strikte controle van de ziekteactiviteit, en specifiek van de ontsteking, de belangrijkste korte termijn doelstelling in patiënten met RA en axSpA. Om dit te bereiken is het van essentieel belang dat de mate van ontstekingen wordt vastgelegd en gecontroleerd tijdens de bezoeken aan de reumatoloog. Op basis hiervan kan dan een intensievere behandeling ingezet worden als blijkt dat de ziekte actief is. Er zijn verschillende instrumenten ontwikkeld om ziekteactiviteit in RA en axSpA patiënten te meten. Deze instrumenten kunnen bestaan uit individuele testen zoals het aantal ontstoken gewrichten, het meten van pijn (gerapporteerd door de patiënt) of een laboratoriumtest. Maar er bestaan ook verschillende samengestelde indices waarin dergelijke instrumenten en/of testen worden gecombineerd.

Helaas lukt het slechts in een klein deel van de patiënten om de ziekteactiviteit volledig te controleren. Veel patiënten blijven dan ook tot op zekere hoogte last houden van

persisterende ontstekingen, met alle onomkeerbare consequenties - de verwoesting van de gewrichten, pezen en ligamenten - van dien. Dit wordt ook wel 'structurele schade' genoemd en kun je zien op röntgenfoto's. Bij RA worden de botten voornamelijk verwoest door het ontstaan van erosies; bij axSpA wordt juist voornamelijk nieuw bot gevormd. De mate van schade zoals te zien op röntgenfoto's (radiologische schade) is direct geassocieerd met belemmeringen in het dagelijks leven en ook met afgenomen kwaliteit van leven. Daarom is het belangrijk dat geneesmiddelen radiologische schade kunnen vertragen of zelfs voorkómen. Instanties die de effectiviteit van geneesmiddelen beoordelen vinden dit dan ook een belangrijke vereiste bij het goedkeuren van nieuwe therapieën voor RA. Ondanks dat er al een duidelijke relatie is aangetoond tussen ziekteactiviteit, radiologische schade en beperkingen in het dagelijks leven in patiënten met RA en axSpA hebben we bepaalde aspecten van deze relatie nog niet volledig weten te doorgronden.

Het feit dat andere gewrichten zijn aangedaan in RA dan in axSpA heeft niet alleen consequenties voor het stellen van een diagnose of het classificeren van de ziekte, maar ook voor het monitoren van de ziekteactiviteit en schade in de dagelijkse praktijk en in wetenschappelijk onderzoek. In patiënten met RA kunnen artsen, maar ook patiënten zelf, vrij gemakkelijk zwelling van de perifere gewrichten herkennen als eerste teken van ontsteking, wat de diagnose en het monitoren van de ziekte vergemakkelijkt. Bij patiënten met axSpA is het echter lastiger om ontsteking direct te observeren omdat de aangedane SI gewrichten en de gewrichten in de rug zich veel dieper in het lichaam bevinden. Aanvullende testen zijn daarom nodig. Dit kunnen testen in het bloed zijn zoalds de zogenaamde acute fase eiwitten (o.a. de bezinking). Maar ook beeldvormende technieken zoals magnetic resonance imaging (MRI) zijn dan ook relevant, met name in patiënten met axSpA. Hiermee is het mogelijk ontstekingen in deze diepgelegen gewrichten aan te tonen.

Dit proefschrift

In dit proefschrift worden een aantal studies beschreven over het beoordelen van ontsteking en schade in patiënten met RA en axSpA en worden beschreven in twee delen.

De studies behorende tot deel I van dit proefschrift richten zich op RA en helpen in het beter begrijpen van de relatie tussen ziekteactiviteit, radiologische schade en beperkingen in het dagelijks leven. De studies evalueren welke van de beschikbare instrumenten (individuele testen of samengestelde indices) geassocieerd zijn met radiologische schade. Daarnaast wordt ook onderzocht welke afwijkingen op röntgenfoto's geassocieerd zijn met beperkingen in het dagelijks leven. Daarnaast worden in deze studies methodologische kwesties die gerelateerd zijn aan het optimaal beoordelen van radiologische progressie in wetenschappelijk onderzoek onder de loop genomen.

Deel II van dit proefschrift richt zich op axSpA. De studies beschreven in dit deel evalueren het nut van aanvullende testen die ontsteking aan kunnen tonen (markers in het bloed,

MRI). Er wordt onderzocht wat hun rol is in het classificeren (diagnosticeren) van patiënten en het monitoren van de ziekteactiviteit. De studies beschreven in dit proefschrift geven inzicht hoe deze testen het best gebruikt kunnen worden in de dagelijkse praktijk.

DEEL I: REUMATOIDE ARTRITIS

De relatie tussen ziekteactiviteit en radiologische schade

Ten tijde van de start van de studies beschreven in dit proefschrift, ging men er van uit dat samengestelde indices beter waren dan individuele instrumenten om ziekteactiviteit in patiënten met RA te monitoren. Dit was grotendeels gebaseerd op 'expert opinie' en werd slechts ondersteund door beperkt wetenschappelijk bewijs. De relatie tussen individuele instrumenten – met name instrumenten die het oordeel van de patiënt over de ziekte meten zoals pijnschalen of de duur van ochtendstijfheid - en radiologische progressie werd namelijk nog niet compleet begrepen. De resultaten in **hoofdstuk 2** komen uit een systematische literatuurstudie die we hebben uitgevoerd, en beschrijven de bestaande kennis over de relatie tussen de samengestelde indices en hun individuele componenten met radiologische progressie. In totaal hebben we 57 studies over dit onderwerp samengevat in ons literatuuroverzicht. Uit dit overzicht is gebleken dat alle samengestelde indices waarin een telling van pijnlijke of gezwollen gewrichten is meegenomen, gerelateerd zijn aan radiologische progressie. Bovendien is gebleken dat alleen de individuele instrumenten die objectief de ontsteking meten, zoals het tellen van de gezwollen gewrichten en laboratoriumtesten, gerelateerd zijn aan radiologische progressie. Dit in tegenstelling tot subjectieve maten zoals het oordeel van de patiënt over de ziekteactiviteit die niet gerelateerd bleken aan radiologische progressie. Op basis van deze resultaten bevelen we dan ook aan om gebruik te maken van samengestelde indices waarin tenminste een telling van de gezwollen gewrichten is opgenomen om de ziekteactiviteit in patiënten met RA in de dagelijkse praktijk te monitoren.

Het beoordelen van radiologische schade

Om de relatie tussen ziekteactiviteit, radiologische schade en de beperkingen in het dagelijks leven goed te kunnen begrijpen is het belangrijk om de radiologische schade op röntgenfoto's in maat en getal uit te drukken. We onderscheiden drie type laesies: 1. Kraakbeenschade; 2. Schade aan gewrichten of banden; 3. Botschade. Kraakbeen bedekt de uiteinden van de botten in de gewrichten en schade hieraan is te herkennen aan een versmalling van de ruimte tussen de botuiteinden die een gewricht vormen (in het Engels: joint space narrowing (JSN)). Schade aan de gewrichten of banden kan resulteren in standsafwijkingen van botten waardoor een gewricht gedeeltelijk of helemaal 'uit de kom' raakt, wat ook wel een (sub) luxatie wordt genoemd. Schade aan het bot zelf is te herkennen als gaten in het oppervlak

van het bot, ook wel erosies genoemd.

In de loop van de tijd zijn er verscheidene methodes ontwikkeld om deze radiologische laesies te kwantificeren. De meest gebruikte methode is de Sharp-van der Heijde (SHS) methode met een score die gaat van 0 (helemaal normaal) tot 448 (alle gewrichten volledig kapot). Met de SHS methode worden 32 gewrichten in de handen en 12 gewrichten in de voeten beoordeeld op de aanwezigheid van erosies en JSN, waarbij de JSN score daadwerkelijke versmalling van de gewrichtsspleet maar ook de aanwezigheid van (sub) luxatie kan aanduiden.

De relatie tussen verschillende types radiologische schade en beperkingen in het dagelijks leven

Eerdere studies hebben al aangetoond dat de mate van radiologische schade geassocieerd is met beperkingen in het dagelijks leven van patiënten met RA, maar het was niet duidelijk welke type laesies aan de beperkingen bijdragen. In de studie beschreven in **hoofdstuk 3** hebben we geprobeerd hier meer inzicht in te krijgen door data van patiënten te bestuderen die geïncludeerd zijn in de Noorse arm van het Europese Onderzoek naar Invaliderende Ziekten en Maatschappelijke Ondersteuning (in het Engels: European Research on Incapacitating Diseases and Social Support (EURIDISS)). In deze patiënten zijn is de progressie van radiologische schade gemeten over een periode van 10 jaar. Tegelijkertijd zijn ook verschillende maten van beperkingen gemeten. Dit zijn de handfunctie (gemeten met knijpkracht) en het algemeen dagelijks functioneren (gemeten met de Health Assessment Questionnaire (HAQ) score, een vragenlijst door patiënten ingevuld).

De meeste studies die eerder zijn uitgevoerd vonden allemaal een duidelijke relatie tussen alle radiologische schade samen, oftewel de som van de drie type laesies, en het algemeen dagelijks functioneren zoals gemeten met de HAQ score. Maar uit onze studie is gebleken dat geen van de individuele type laesies apart geassocieerd is met de HAQ score. Wel is gebleken dat ieder type laesie geassocieerd is met knijpkracht. De sterkste relatie werd gevonden voor de 'echte' versmalling van de gewrichtsspleet (als maat voor kraakbeenschade). Hieruit hebben we geconcludeerd dat kraakbeenverlies in de handen waarschijnlijk het meeste effect hebben op de mate van beperking in handfunctie. Alle drie type laesies zijn belangrijk voor de mate van algemeen functioneren.

Methodologische aspecten van het beoordelen van radiologische schade

Registratie autoriteiten die nieuwe medicijnen goedkeuren vinden radiologische progressie een belangrijke uitkomstmaat, maar er bestaat nog enige onduidelijkheid over een aantal methodologische vraagstukken in de beoordeling en analyse van de radiologische progressie.

Adjudication

Het eerste vraagstuk heeft betrekking op 'adjudication'. In wetenschappelijke studies naar de effectiviteit van medicaties worden röntgenfoto's meestal door twee onafhankelijke lezers beoordeeld om te bepalen of er sprake is van radiologische progressie. Als er een belangrijk verschil tussen de scores van deze twee lezers wordt gevonden, dan wordt een derde lezer (de zogenaamde adjudicator; een soort scheidsrechter) gevraagd om de röntgenfoto's waarover discrepantie bestaat te beoordelen (de adjudicator score). Om de kwaliteit te waarborgen van de lezers heeft men als stelregel dat adjudication niet te vaak moet optreden; zo wordt meestal adjudication in maximaal 20% van de patiënten in een wetenschappelijke studie geaccepteerd. In **hoofdstuk 4** hebben we onderzocht hoe vaak adjudication daadwerkelijk wordt toegepast in 15 recente wetenschappelijke studies waarin radiologische progressie is beoordeeld, en we hebben gevonden dat er altijd in minder dan 21% van de patiënten adjudication nodig was. Dit kun je zien als een maat voor de betrouwbaarheid van deze scoringsmethode om radiologische progressie te beoordelen aan. Daarnaast hebben we gevonden dat bepaalde kenmerken van dergelijke wetenschappelijke studies de mate waarin adjudication nodig is beïnvloedt. Een voorbeeld is het aantal foto's dat er van patiënten aanwezig is.

Kleinst detecteerbare verandering, na correctie voor meetfout.

Elke meting die bij een patiënt gedaan wordt bestaat uit twee componenten: een 'juiste meting' en een 'meetfout'. Per definitie is onbekend hoe groot iedere component in een meting is. Hoe groter de meetfout, des te onbetrouwbaar het resultaat van de meting. Ondanks dat de precieze grootte van de meetfout niet bekend is per meting, kan de bijdrage van deze meetfout aan het gemiddelde van alle metingen wel worden bepaald. Om inzicht te krijgen in de grootte van de meetfout kan men alle foto's door twee lezers laten scoren. Vervolgens kun je met een statistische toets de grootte van de meetfout inschatten. Een voorbeeld van de bepaling van de meetfout van de gemiddelde radiologische progressie in een groep patiënten met RA is de kleinst detecteerbare verandering (in het Engels: smallest detectable change (SDC)). Meestal wordt hier een marge van 95% voor genomen: 95% van de verschillen tussen de lezers moet binnen deze bandbreedte vallen. Deze 95% SDC geeft een indicatie van de marge van onzekerheid: een gemeten verandering over de tijd kleiner dan de berekende 95% SDC kan beschouwd worden als meetfout, en alleen van veranderingen groter dan de 95% SDC beschouwt men als een 'daadwerkelijke verandering' is. Zo kun je het aantal patiënten met een verandering groter dan de 95% SDC bepalen. De 95% SDC is een bruikbaar concept bij het beoordelen van radiologische progressie in wetenschappelijke studies, maar het heeft twee theoretische tekortkomingen. De eerste is dat de 95% SDC met name geschikt is voor studies met twee meetmomenten, maar dat de berekening van de 95% SDC lastig is als er meer dan twee meetmomenten zijn. De tweede tekortkoming is dat er geen echte reden was om 95% betrouwbaarheid te eisen. En mogelijk geeft deze een 'te streng' afkappunt van de meetfout.

In **hoofdstuk 5** hebben we aangetoond dat voor wetenschappelijke studies met meer dan twee meetmomenten een simpele berekening van het gemiddelde van alle 95% SDC's van alle mogelijke intervallen (elk gebaseerd op twee meetmomenten) volstaat om een goed beeld te krijgen van de 95% SDC over de totale periode. In dit hoofdstuk hebben we ook een 80% SDC voorgesteld in plaats van een 95% SDC, waarbij we hebben laten zien dat een 80% SDC realistischer schattingen van de meetfout geeft en daarom informatief kan zijn in hedendaagse wetenschappelijke studies waarin radiologische progressie vaak (veel) kleiner is dan in oudere studies.

DEEL II: AXIALE SPONDYLOARTRITIS

Beoordelen van objectieve ontsteking voor de diagnose

HLA-B27 en MRI van de SI-gewrichten

Chronische rugpijn is vaak het eerste symptoom van axSpA, maar is ook een veelgehoorde klacht onder de algehele bevolking met een verscheidenheid aan onderliggende oorzaken. Daarom zijn aanvullende testen, gericht op het meten van ontsteking, belangrijk om de juiste diagnose te stellen. Aanvullende testen omvatten onder andere relatief simpele laboratoriumtesten zoals bloedtesten om de hoeveelheid acute fase eiwitten (tekenen van ontsteking in het bloed zoals de bezinking en C-reactieve proteïne (CRP)) te meten. Ook het bepalen van de aanwezigheid van een gen geassocieerd met axSpA (human leukocyte antigen B27 (HLA-B27)) is van belang in de diagnostiek. En tenslotte beeldvormende technieken zoals een MRI-scan van de SI-gewrichten (MRI-SI). Maar helaas zijn sommige van deze aanvullende testen duur en/of tijdrovend en kunnen daarom niet in alle patiënten met chronische rugklachten worden toegepast, omdat uiteindelijk slechts een klein deel van deze patiënten daadwerkelijk axSpA heeft. Het zou daarom goed zijn als reumatologen kunnen inschatten welke patiënten met chronische rugklachten meer kans op axSpA hebben. De testen kunnen dan vooral bij deze patiënten nuttig zijn als aanvullende test. In **hoofdstuk 6** hebben we onderzocht of, en zo ja, welke typische kenmerken van axSpA die verzameld kunnen worden door middel van simpele vragen aan de patiënt of via lichamelijk onderzoek, een grote waarschijnlijkheid geven op het vinden van een positieve MRI-SI (oftewel: de aanwezigheid van ontstekingslaesies) of een positieve test voor HLA-B27. We hebben deze studie uitgevoerd in het ESPeranza Programma, een Spaans nationaal prospectief multicenter gezondheidsprogramma met als doel een vroege diagnose van patiënten met SpA te faciliteren. We hebben data gebruikt van patiënten met chronische rugklachten die werden doorverwezen met de verdenking op SpA. Uit onze studie is gebleken geen enkel apart SpA kenmerk de waarschijnlijkheid op het vinden van een positieve HLA-B27 test deed toenemen. We hebben wel drie SpA kenmerken gevonden waarbij de MRI-SI vaker positief gevonden werd: 1. De aanwezigheid van een bepaald type rugpijn; 2. De aanwezigheid

van dactylitis (een zogenaamde worstvinger of worstteen); 3. De aanwezigheid van een inflammatoire darmziekte (de ziekte van Crohn of colitis ulcerosa). Deze drie kenmerken kunnen daarom mogelijk waardevol zijn voor reumatologen bij het bepalen welke patiënten een MRI-SI zouden moeten ondergaan en welke niet.

Zeer gevoelige meting van CRP

Het is aangetoond dat de ASAS criteria beter zijn dan de oudere criteria om axSpA in een vroeg stadium vast te stelln. Één van de SpA kenmerken opgenomen in de ASAS criteria is verhoogd CRP in het bloed, een marker van ontsteking. CRP kan bepaald worden via een conventionele methode maar ook via een nieuwere methode waarbij CRP met een hogere gevoeligheid wordt gemeten (in het Engels: high sensitivity CRP meting, kortweg hsCRP meting). Deze hsCRP meet lagere hoeveelheiden CRP in het bloed. Ten tijde van de ontwikkeling van de ASAS criteria werd voornamelijk de conventionele CRP meting gebruikt, maar tegenwoordig is in veel ziekenhuizen de conventionele methode vervangen door de hsCRP meting. In **hoofdstuk 7** hebben we onderzocht of, en zo ja, hoe de vervanging van de conventionele CRP methode door de hsCRP meting de ASAS criteria zou beïnvloeden. Dit hebben we onderzocht in het DEvenir des Spondylarthopathies Indifférenciées Récentes (DESIR) cohort. We hebben gezien dat niet één extra patiënt werd geclassificeerd als axSpA wanneer gebruik werd gemaakt van de hsCRP meting in plaats van de conventionele methode. Dit betekent dat de conventionele CRP methode prima vervangen kan worden door de hsCRP meting wanneer het gaat om het classificeren van patiënten met axSpA.

Beoordelen van ziekteactiviteit

Zeer gevoelige meting van CRP

Er bestaan verschillende klinische maten om ziekteactiviteit in patiënten met axSpA te monitoren. Van oudsher wordt de Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) het meest gebruikt. Deze vragenlijst wordt volledig door de patiënt ingevuld en bevat geen objectieve maat van ontsteking. Dit kan als een beperking worden gezien, en daarom is recent de Ankylosing Spondylitis Disease Activity Score (ASDAS) ontwikkeld. Dit is een samengestelde index waarin patiënt gerapporteerde symptomen en de CRP meting gecombineerd worden. Er bestaat ook een ASDAS met de bezinking als objectieve maat. De hsCRP meting kan zeer lage hoeveelheden CRP in het bloed detecteren terwijl de conventionele methode waarden onder een bepaalde drempel (meestal <5 mg/L) als 'onmeetbaar klein' afgeeft. Daarom hebben we in **hoofdstuk 8** gekeken welke waarde van CRP het beste kan worden ingevuld in de ASDAS formule wanneer de conventionele methode 'onmeetbaar klein' is of de hsCRP meting een zeer kleine waarde heeft aangegeven. Dit hebben we ook in het DESIR cohort onderzocht. Uit alle mogelijke CRP waardes die we hebben getest, bleek een kunstmatige waarde van 2 mg/L het beste te werken in de ASDAS berekening. Daarnaast hebben we ook gekeken of de ASDAS berekend met de

bezinking in plaats van CRP, ook overeenstemde met de ASDAS berekening waarin hsCRP is meegenomen. Onze resultaten laten zien dat er een goede overeenstemming is tussen beide ASDAS berekeningen.

Relatie tussen klinische ziekteactiviteit en MRI-SI

MRI-SI

De belangrijke rol van MRI-SI in het diagnosticeren van patiënten met axSpA in een vroeg stadium staat tegenwoordig buiten kijf. Echter, de mogelijke rol van MRI-SI in het monitoren van ziekteactiviteit over de tijd is nog steeds onderwerp van discussie. In **hoofdstuk 9** hebben we hier onderzoek naar gedaan en hebben we voor het eerst aangetoond dat een verandering in klinische maten zoals ASDAS en BASDAI tussen twee visites gepaard gaat met een verandering in ontsteking zoals te zien op MRI-SI over dezelfde periode. We hebben gevonden dat van de klinische maten, ASDAS het sterkst geassocieerd is met MRI-SI. Bovendien hebben we gevonden dat de relatie tussen klinische maten van ziekteactiviteit en ontsteking op MRI-SI verschillend is in mannen en vrouwen: bij mannen was er een duidelijke associatie tussen ASDAS en MRI-SI, maar deze was afwezig in vrouwen. Wanneer we deze bevindingen vertalen naar de kliniek, dan suggereren we dat het monitoren van ziekteactiviteit met behulp van de ASDAS in mannen ook een afspiegeling is van de ontstekingen op de MRI-SI, terwijl dat in vrouwen niet het geval is. Het gebrek aan een associatie tussen ASDAS en ontstekingen op MRI-SI in vrouwen, samen met resultaten uit andere studies, lijkt er op te wijzen dat axSpA zich in vrouwen anders uit dan in mannen.

CONCLUSIES

Er lagen twee breed gedragen gedachtes ten grondslag aan de studies in dit proefschrift. Ten eerste dat ontsteking in chronische reumatische ontstekingsziektes zoals RA en axSpA tot structurele schade leidt. Ten tweede dat deze structurele schade op de lange termijn leidt tot functieverlies en beperkingen in het dagelijks leven. Het centrale thema in dit proefschrift is *'meting'*. Wanneer men de hoofdstukken van dit proefschrift doorleest, wordt het duidelijk dat een degelijke *meting* belangrijk is voor veel aspecten van uitkomstmaten van reumatische ontstekingsziektes en ook om deze ziektes beter te begrijpen. Methodologische aspecten van *meting* zijn net zo belangrijk als de inhoud van deze *metingen*. Het verbeteren van uitkomstmaten voor patiënten met reumatische ontstekingsziektes is een belangrijk doel van onderzoek in dit vakgebied. Daarom is het van cruciaal belang dat er voldoende aandacht wordt geschonken aan de principes van *meting* en dat deze op de juiste manier worden geïmplementeerd in de dagelijkse praktijk. Een betere *meting* zal uiteindelijk leiden tot een beter begrip en tot betere behandeling van chronische reumatische ontstekingsziektes.

LIST OF PUBLICATIONS

1. Ariza-Ariza R, Navarro-Sarabia F, Hernández-Cruz B, Rodríguez-Arboleya L, **Navarro-Compán V**, Toyos J. Dose escalation of the anti-TNF-alpha agents in patients with rheumatoid arthritis. A systematic review. Rheumatology (Oxford) 2007;46(3):529-32.

2. Navarro-Sarabia F, Ruiz-Montesinos D, Hernandez B, **Navarro-Compán V**, Marsal S, Barcelo M, Perez-Pampín E, Gómez-Reino JJ. DAS-28-based EULAR response and HAQ improvement in rheumatoid arthritis patients switching between TNF antagonists. BMC Musculoskelet Disord 2009;10:91.

3. Navarro Sarabia F, **Navarro Compán V**. Efectos de certolizumab pegol en la discapacidad y la calidad de vida en la artritis reumatoide. Reumatol Clín 2009;9(4):12-6.

4. Garrido López BC, **Navarro Compán MV**, F Navarro Sarabia. Vaccines and chemo-prophylaxis in rhemautoid arthritis: is a vaccine calendar necessary? Reumatol Clin 2011;7:412-6.

5. **Navarro Compán V**, Navarro Sarabia F. Safety and efficacy of the newer biological therapeutics in the treatment of the rheumatoid arthritis. Clinical Medicine Insights: Therapeutics 2011;3:195–203.

6. **Navarro-Compán V**, Moreira V, Ariza-Ariza R, Hernández-Cruz B, Vargas-Lebrón C, Navarro-Sarabia F. Low doses of etanercept can be effective in ankylosing spondylitis patients who achieve remission of the disease. Clin Rheumatol 2011;30(7):993-6.

7. Abdel-Kader N, Cardiel MH, **Navarro Compán V**, Piedra Priego J, González A. Cushing's disease as a cause of severe osteoporosis: a clinical challenge. Reumatol Clin 2012;8(5):278-9.

8. Ariza-Ariza R, Hernandez-Cruz B, **Navarro-Compán V**, Leyva Pardo C, Juanola X, Navarro-Sarabia F. A comparison of telephone and paper self-completed questionnaires of main patient-related outcome measures in patients with ankylosing spondylitis and psoriatic arthritis. Rheumatol Int 2013;33(11):2731-6.

9. de Hooge M, van den Berg R, **Navarro-Compán V**, van Gaalen F, van der Heijde D, Huizinga T, Reijnierse M. Magnetic resonance imaging of the sacroiliac joints in the early detection of spondyloarthritis: no added value of gadolinium compared with short tau inversion recovery sequence. Rheumatology (Oxford) 2013;52(7):1220-4.

10. **Navarro-Compán V**, Landewé R, Ahmad HA, Miller CG, Xu D, Wolterbeek R, van der Heijde D. Rate of adjudication of radiological progression in rheumatoid arthritis randomized controlled trials depending on preset limits of agreement: a pooled analysis from 15 randomized trials. Rheumatology (Oxford) 2013;52(8):1404-7.

11. **Navarro-Compán V**, Melguizo-Madrid E, Hernandez-Cruz B, Santos-Rey K, Leyva-Prado C, González-Martín C, Navarro-Sarabia F, González-Rodríguez C. Interaction between oxidative stress and smoking is associated with an increased risk of rheumatoid arthritis: a case-control study. Rheumatology (Oxford) 2013;52(3):487-93.

12. **Navarro-Compán V**, van der Heijde D, Combe B, Cosson C, van Gaalen FA. Value of high-sensitivity C-reactive protein for classification of early axial spondyloarthritis: results from the DESIR cohort. Ann Rheum Dis 2013;72(5):785-6.

13. Bautista-Molano W, **Navarro-Compán V**, Landewé 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.

14. **Navarro-Compán V**, van der Heijde D, Ahmad HA, Miller CG, Wolterbeek R and Landewé R. Measurement error in the assessment of radiographic progression in rheumatoid arthritis clinical trials: the smallest detectable change revisited. Ann Rheum Dis 2014;73(6):1067-70.

15. Melguizo E, **Navarro V**, Hernández B, Santos K, Arrobas T, Domínguez C, Navarro F, González C. Diagnostic utility of oxidative damage markers for early rheumatoid arthritis in non-smokers and negative anti-CCP patient. An Sist Sanit Navar. 2014;37(1):109-15.

16. Machado P, **Navarro-Compán V**, Landewé R, van Gaalen FA, Roux C, van der Heijde D. Calculating the ankylosing spondylitis disease activity score if the conventional C-reactive protein level is below the limit of detection or if high-sensitivity C-reactive protein is used: an analysis in the DESIR cohort. Arthritis Rheumatol 2015;67(2):408-13.

17. Mandl P, **Navarro-Compán V**, Terslev L, Aegerter P, van der Heijde D, D'Agostino MA, Baraliakos X, Pedersen SJ, Jurik AG, Naredo E, Schueller-Weidekamm C, Weber U, Wick MC, Bakker PA, Filippucci E, Conaghan PG, Rudwaleit M, Schett G, Sieper J, Tarp S, Marzo-Ortega H, Østergaard M. EULAR recommendations for the use of imaging in the diagnosis and management of spondyloarthritis in clinical practice. Ann Rheum Dis 2015;74(7):1327-39.

18. **Navarro-Compán V**, Gherghe AM, Smolen JS, Aletaha D, Landewé R, van der Heijde D. Relationship between disease activity indices and their individual components and radiographic progression in RA: a systematic literature review. Rheumatology (Oxford) 2015;54(6):994-1007.

19. **Navarro-Compán V**, Landewé R, Provan SA, Ødegård S, Uhlig T, Kvien TK, Keszei AP, Ramiro S, van der Heijde D. Relationship between types of radiographic damage and disability in patients with rheumatoid arthritis in the EURIDISS cohort: a longitudinal study. Rheumatology (Oxford) 2015;54(1):83-90.

20. **Navarro-Compán V**, Smolen JS, Huizinga TW, Landewé R, Ferraccioli G, da Silva JA, Moots RJ, Kay J, van der Heijde D. Quality indicators in rheumatoid arthritis: results from the METEOR database. Rheumatology (Oxford) 2015;54(9):1630-9.

21. van den Berg R, de Hooge M, Bakker PA, van Gaalen F, **Navarro-Compán V**, Fagerli KM, Landewé R, van Oosterhout M, Ramonda R, Reijnierse M, van der Heijde D. Metric properties of the SPARCC score of the sacroiliac joints - data from baseline, 3-month, and 12-month follow up in the SPACE Cohort. J Rheumatol 2015;42(7):1186-93.

22. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, Kvien TK, **Navarro-Compán MV**, Oliver S, Schoels M, Scholte-Voshaar M, Stamm T, Stoffer M, Takeuchi T, Aletaha D, Andreu JL, Aringer M, Bergman M, Betteridge N, Bijlsma H, Burkhardt H, Cardiel M, Combe B, Durez P, Fonseca JE, Gibofsky A, Gomez-Reino JJ, Graninger W, Hannonen P, Haraoui B, Kouloumas M, Landewé R, Martin-Mola E, Nash P, Ostergaard M, Östör A, Richards P, Sokka-Isler T, Thorne C, Tzioufas AG, van Vollenhoven R, de Wit M, van der Heijde D. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. Ann Rheum Dis published 12 May 2015. doi:10.1136/annrheumdis-2015-207524. [Epub ahead of print].

23. Stoffer MA, Schoels MM, Smolen JS, letaha D, Breedveld FC, Burmester G, Bykerk V, Dougados M, Emery P, Haraoui B, Gomez-Reino J, Kvien TK, Nash P, **Navarro-Compán V**, Scholte-Voshaar M, van Vollenhoven R, van der Heijde D, Stamm TA. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search update. Ann Rheum Dis published 19 May 2015. doi:10.1136/annrheumdis-2015-207526. [Epub ahead of print].

24. Plasencia C, Kneepkens EL, Wolbink G, Krieckaert CLM, Turk S, **Navarro-Compán V**, L´Ami M, Nurmohamed M, van der Horst-Bruinsma I, Jurado T, Diego C, Bonilla G, Villalba A, Peiteado D, Nuño L, van der Kleij D, Rispens T, Martín-Mola E, Balsa A, Pascual-Salcedo D. Comparing tapering strategy to standard dosing regimen of TNF inhibitors in patients with spondyloarthritis in low disease activity in daily clinical practice. J Rheumatol published 15 Jul 2015. pii: jrheum.141128. [Epub ahead of print].

25. Ventura-Ríos L, **Navarro-Compán V**, Aliste M, Alva Linares M, Areny R, Audisio M, Bertoli AM, Cazenave T, Cerón C, Díaz ME, Gutiérrez M, Hernández C, Navarta DA, Pineda C, Py GE, Reginato AM, Rosa J, Saaibi DL, Sedano O, Solano C, Castillo-Gallego C, Falçao S, De Miguel E. Is Entheses Ultrasound Reliable? A Reading Latin American Exercise. Clin Rheumatol published 22 Jul 2015. [Epub ahead of print].

26. de Hooge M, van den Berg R, **Navarro-Compán V**, Reijnierse M, van Gaalen F, Fagerli K, Landewé R, van Oosterhout M, Ramonda R, Huizinga T, van der Heijde D. Patients with chronic back pain of short duration from the SPACE-cohort: which MRI structural lesions in the sacroiliac joints and inflammatory and/or structural lesions in the spine are most specific for axial spondyloarthritis? Ann Rheum Dis published 18 Jul 2015. doi: 10.1136/

annrheumdis-2015-207823. [Epub ahead of print].

27. **Navarro-Compán V**, Ramiro S, Landewé R, Dougados M, Miceli-Richard C, Richette P, van der Heijde D. Disease Activity is Longitudinally Related to Sacroiliac inflammation on MRI in Male patients with Axial Spondyloarthritis: 2-year of the DESIR cohort. Ann Rheum Dis. (Accepted for publication).

28. **Navarro-Compán V**, de Miguel E, van der Heijde D, Landewé R, Almódovar R, Montilla C, Beltrán E, Zarco P. Sponyloarthritis features forecasting the presence of HLA-B27 or sacroiliitis on magnetic resonance imaging in patients with suspected axial spondyloarthritis: results from a cross-sectional study in the ESPeranza cohort. Arthritis Res Ther. (Accepted for publication).

29. Río-Martínez P, **Navarro-Compán V**, Díaz-Miguel C, Almodóvar R, Mulero J, De Miguel E and ESPeranza group. Similarities and differences between patients fulfilling axial and peripheral ASAS criteria for spondyloarthritis: Results from the ESPeranza cohort. Semin Arthritis Rheum (Accepted for publication).

30. **Navarro-Compán V**, Plasencia-Rodríguez C, de Miguel E, Balsa A, Martín-Mola E, Seoane-Mato D and Cañete JD. Anti-TNF discontinuation and tapering strategies in patients with axial spondyloarthritis: A systematic literature review. (Submitted).

31. Fernández-Carballido C, **Navarro-Compán V**, Castillo-Gallego C, Castro-Villegas MC, Collantes-Estevez E and de Miguel E, on behalf of the ESPeranza Study Group. Disease activity is the major determinant of quality of life and physical function in patients with early axial spondylarthritis: results from the ESPeranza cohort. (Submitted).

32. Almodóvar R, **Navarro-Compán V**, Fernández-Carballido C, Azucena Hernández A, de Miguel E, Zarco P and ESPeranza Study Group. Differences between familial and sporadic early spondyloarthritis: results from the ESPeranza cohort. (Submitted).

33. Rubio Vargas R, Melguizo-Madrid E, González-Rodríguez C, Navarro-Sarabia F, Domínguez-Quesada C, Ariza-Ariza R and **Navarro-Compán V**. Association between serum dickkopf-1 levels and disease duration in axial spondyloarthritis. (Submitted).

34. Plasencia C, Wolbink G , Krieckaert CLM, Kneepkens EL, Turk S, Jurado T, Martínez-Feito A, **Navarro-Compán V**, Bonilla G, Villalba A, Peiteado D, Nuño L, Martín-Mola E, Nurmohamed MT, van der Kleij D, Rispens T, Pascual-Salcedo D, Balsa A. Comparing a tapering strategy to the standard dosing regimen of TNF inhibitors in patients with rheumatoid arthritis with low disease activity. (Submitted).

35. Gossec G, Portier A, Landewé R, Etcheto A, **Navarro-Compán V**, Kroon F, Heijde D, Dougados M. Preliminary definitions of “flare” in axial spondyloarthritis, based on pain, BASDAI and ASDAS-CRP: an ASAS initiative. (Submitted).

CURRICULUM VITAE

Victoria Navarro Compán was born on October 8th, 1979 in Sevilla, Spain. She studied at the school San José SS.CC in Sevilla. During high school, she spent a year at Washington High School in Pittsburg, USA. Thereafter, she started her studies in Medicine at the University of Cádiz and obtained her medical degree at the University of Sevilla in 2003. Later, she started her training in Rheumatology at the University Hospital Virgen Macarena in the same city. She registered as a rheumatologist in 2009. As part of her training, she did a rotation during four months at the Charité Universitätsmedizin in Berlin, Germany. During the next two years, she stayed at the University Hospital Macarena working as a clinician and got involved in research. At the end of 2011, she moved to the Netherlands and focused on research at the Leiden University Medical Center. At the same time, she took a Master in Epidemiology at the University of Maastricht, where she graduated 'cum laude' in 2013. After more than two years abroad, she returned to Spain, where she started working at the University Hospital La Paz in Madrid combining research with clinical duties.

ACKNOWLEDGEMENTS

This thesis was written while living in four different cities located in three different countries. It is simply impossible to thank all people who helped me one way or the other. So, I would like to start thanking those people that are not mentioned individually. To all of you, thank you!

Dear Prof. van der Heijde and Prof. Landewé, dear Désirée and Robert, I have no words to express my gratitude and how lucky I feel. I have got much more from you than I could have ever imagined. A deep thank you for everything you have taught me, for your patience, for the nice moments we shared and especially for being so generous. Fortunately, this is just the beginning of a long and beautiful supervision-relationship.

I am very thankful to the department of Rheumatology at the LUMC for all the help and care I received. Despite not speaking Dutch, you made me feel as "one of yours". Dear Prof. Huizinga, dear Tom, thanks for your warm welcome and for all facilities. Dear roommates (Annemarie, Jessica, Emilia, Rachel and Diederik) and successor roommates (Pauline, Iris, Féline, Miranda, Zineb), thank you for all the good moments and laughs we shared and for all your support, including helping me to handle any type of electronic device or to find out how daily things work in the Netherlands. Dear Joyce, Hughine, Nancy, Jozé and Cedric, thank you! Also, I like to show my gratitude to all doctors and trainees. Dr. van Gaalen, dear Floris, it was a pleasure working with you.

I would like to thank all the co-authors of the manuscripts too. Dear Prof. Smolen and Prof. Dougados, it has been a real privilege learning from you.

Dear Prof. de Miguel, Eugenio, thank you for all your support. You have opened the doors of the department of Rheumatology at University Hospital La Paz. To all my colleagues in Madrid, thank you for your really warm welcome. Prof. Martín Mola, dear Emilio, thank you for including me as part of this wonderful team. Prof. Balsa, dear Alejandro, thank you for all your help and persistence to facilitate my work. Also, I would like to thank IdiPaz for your support, and my new roommates (María and Dani) for such a nice ambiance to work.

Obviously, this thesis is also the product of who I was and what I learnt before I went to Leiden: Dear Prof. Sieper, dear Jochen, thank you for the first opportunity to perform research abroad. Dear colleagues from Sevilla at University Hospital Virgen Macarena (Cármenes, Javier, Loli, Silvia, Manuel, Lola, Belén, Virginia), a deep thank you to all of you.

While working abroad, I was lucky to not only meet many new friends (ITC-friends, EMEUNET-friends, Pedro, César, Anna, Fran) but also to keep in close contact with my 'lifelong friends' from school (Feria 2050) and university (Carolina, Rodrigo, Carla, Juanjo). To all of you, thank you for being part of my life. Dear Manouk and my dear paranymphs, Sofia and Rosaline,

meeting with you is among the best experiences while working on this thesis. The three of you were unconditionally present any time I needed help. Manouk, we shared so many good moments inside and outside of the hospital… it was a real pleasure! Rosaline, thank you for being so patient to help me. I enjoyed a lot while scoring (and talking) next to you. Sofia, I will always be thankful for hosting me in Maastricht: the lunches with you were the basis for our future friendship.

To my brothers Kiko, Víctor and Paco and to my entire family (uncles, cousins, parents in law, and sisters and brother in law), thank you for being always there. All your support was essential to achieve this goal.

Dear mum, mamá, your strength, optimism and unlimited generosity are an example to follow. Thank you for all your love! Dear Prof. Navarro Sarabia, dear Federico, dear mentor, dear "Jefe", papá, you have been my inspiration for my professional career and the model to pursue. You encourage me every day to expand my knowledge. I have no words to thank you for everything I have learnt from you.

Finally, my dear Ignacio, while you did not actually write one single word of this thesis I know that you are at the basis of every single word. You are able to help me in such a manner that no one ever could. As partners in life, we had already shared many things before going abroad. But this period was really special: The birth of our Little Victoria was the best thing that could have ever happened to us. To both of you, THANK YOU!