Maastricht University

Improvement by movement : a study on therapeutic aspects and outcome assessment in ankylosing spondylitis

Citation for published version (APA):

van Tubergen, A. M. (2002). *Improvement by movement : a study on therapeutic aspects and outcome assessment in ankylosing spondylitis*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. https://doi.org/10.26481/dis.20020927at

Document status and date:
Published: 01/01/2002

DOI:
10.26481/dis.20020927at

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 20 Apr. 2024

Improvement by movement

A study on therapeutic aspects and outcome assessment in ankylosing spondylitis

© Astrid van Tubergen, Maastricht 2002

ISBN 90 5278 353 5

Cover design: Henk Droog | Dvada
Layout: Tiny Wouters
Production: Datawyse | Universitaire Pers Maastricht

The studies described in chapters 4 and 5 were financially supported by Land Salzburg, Gasteiner Tal Tourismusgesellschaft, Kurzentrum Thermentempel, and Gasteiner Heilstollen from Austria, and Zorgvoorzieningen Nederland NV, IZA Zorgverzekeringen, Dick van Tol Assurantiën BV, and Yakult BV from The Netherlands. Financial support for printing of this thesis by AMGEN BV, AstraZeneca, Boehringer Ingelheim BV, Dutch Arthritis Association, Janssen-Cilag BV, MSD, and Pharmacia BV was gratefully acknowledged.

Improvement by movement

A study on therapeutic aspects and outcome assessment in ankylosing spondylitis

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van Prof. dr. A.C. Nieuwenhuijzen Kruseman,
Rector Magnificus,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 27 september 2002 om 14.00 uur

door

Astrid Margaretha van Tubergen
Geboren op 1 december 1973 te Amsterdam

UM
UNIVERSITAIRE PERS MAASTRICHT

Promotores

Prof. dr. J.M.J.P. van der Linden
Prof. dr. D.M.F.M. van der Heijde

Co-promotor

Dr. R.B.M. Landewé

Beoordelingscommissie

Prof. dr. M.H. Prins (voorzitter)
Prof. dr. B.A.C. Dijkmans (Vrije Universiteit Amsterdam)
Prof. dr. W.J.A. van den Heuvel
Prof. dr. H.F.P. Hillen
Prof. dr. J.L. Severens

Contents

Chapter 1	Introduction	7
PART I	Therapeutic aspects of ankylosing spondylitis and other spondylarthropathies	15
Chapter 2	Spondylarthropathies: options for combination therapy	17
Chapter 3	A brief history of spa therapy	39
Chapter 4	Combined spa-exercise therapy is effective in patients with ankylosing spondylitis: a randomized controlled trial	49
Chapter 5	Cost-effectiveness of combined spa-exercise therapy in ankylosing spondylitis: a randomized controlled trial	67
PART II	Outcome assessment in ankylosing spondylitis	85
Chapter 6	Comparison of statistically-derived ASAS improvement criteria for ankylosing spondylitis with clinically relevant improvement according to an expert panel	87
Chapter 7	The use of a numerical rating scale as an answer modality in ankylosing spondylitis specific questionnaires	103
Chapter 8	Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis	119
Chapter 9	Assessment of disability with the WHODAS II in patients with ankylosing spondylitis	137
Chapter 10	Summary and general discussion	151
Chapter 11	Samenvatting in het Nederlands	163
	Dankwoord	169
	Curriculum Vitae	173

Chapter 1

Introduction

Introduction

Ankylosing spondylitis (AS) is a chronic inflammatory disease, in which predominantly the sacroiliac joints and the spine are affected.[1] Sacroiliitis is the hallmark of the disease. The prevalence of AS has long been reported to be approximately 0.1% of the European population, with a male to female ratio of 3:1.[1,2] Using magnetic resonance imaging techniques, studies from Berlin now estimate a prevalence for AS of 0.86% among the white population (almost comparable to the prevalence of rheumatoid arthritis).[3] In The Netherlands, at least 16,000 people suffer from the disease.

AS is the key disorder of the spondylarthropathies (SpA), a family of rheumatic diseases classified together because of common features. These features include the absence of rheumatoid factor, the presence of mono- or oligoarthritis predominantly of the lower limbs, and inflammatory back pain (spondylitis, sacroiliitis).[1] The SpA are strongly associated with the class-I molecule HLA B27, but other genetic factors, including HLA B60, also contribute in developing AS and other SpA.[4-6] Organ-specific manifestations such as acute anterior uveitis, inflammatory bowel disease, psoriasis, enthesitis, carditis, conjunctivitis, and genital inflammation are characteristic features for this group of diseases. Patients may have more than one symptom at the same time or at different moments during life. A positive family history is an important hallmark of SpA and reflects its genetic background. The risk for HLA B27 positive first-degree relatives of HLA B27 positive patients with AS to develop the disease is 10-20 times higher than for HLA B27 positive persons in the general population.[1]

In the first part of this thesis (chapters 2-5) various therapeutic aspects of the SpA, and AS in particular, are discussed. In the second part of this thesis (chapters 6-9) a number of aspects of outcome assessment in AS are discussed in detail.

Part I: Therapeutic aspects of ankylosing spondylitis and other spondylarthropathies

In chapter 2, an overview on drug treatment for SpA is provided, with special attention to evidence for efficacy of combination of treatments. Currently, no curative treatment is available for SpA, although the recently introduced so-called biologicals (anti-TNF-α therapy) may be viewed as very promising in this respect.[7-10] Treatment in SpA in general is mainly aimed at alleviation of symptoms of inflammation (with non-steroidal anti-inflammatory drugs [NSAIDs]), maintaining mobility (with physical exercise therapy), and long-term suppression of disease activity (with disease modifying anti-rheumatic drugs).[11-14]

The standard treatment in active AS consists of NSAIDs and physical exercises.[1,12,15,16] Patients are advised to exercise daily at home and to follow weekly group physical therapy.[16] In addition, patients may take courses of spa therapy in a spa resort, in which exercises also play a central role.

In chapter 3, a brief general history of spa medicine is presented. Spa therapy has a longstanding history as a treatment modality for many rheumatic disorders, and today it gains popularity from a number of specialties in many countries. Nowadays, spa therapy is no longer confined to bathing and drinking of thermal water, but is usually offered in combination with other treatments, such as active physical exercise therapy, massages, or mud packs. In the past decade, only a few randomized controlled trials on the efficacy of spa therapy have been conducted in a number of rheumatic diseases, but not in AS.[17-22] In chapter 4, a randomized controlled trial on the efficacy of spa therapy in AS is described. This randomized controlled trial was designed to assess the efficacy of spa therapy in combination with exercise therapy, in addition to standard treatment with anti-inflammatory drugs and weekly group physical therapy, in patients with AS. Important outcome measures were functional ability, global well-being, and inflammatory signs of AS. Because spa therapy is usually considered expensive, cost-effectiveness and cost-utility analyses of spa-exercise therapy in AS were also performed alongside the clinical trial. The results of this full economic evaluation are presented in chapter 5.

Part II: Outcome assessment in ankylosing spondylitis

Starting in Amsterdam in 1995, the ASsessment in Ankylosing Spondylitis (ASAS) Working Group is an international group of rheumatologists, epidemiologists, patients with AS, and representatives of the pharmaceutic industry from more than 20 countries, who share their expertise in the field of AS. In 1998, the ASAS Working Group has selected "core sets" of outcome measures to be used in different kinds of trials in AS, in order to create uniformity and to promote comparability among these studies.[23] Recently, a small group of members from the ASAS Working Group has developed a preliminary definition of short term improvement in AS, based on the core set of clinical trials with NSAIDs.[24] Improvement criteria are helpful in defining at an individual level whether patients have responded to therapy. The composition and validation of the ASAS improvement criteria were performed by using a statistical approach. It was therefore questioned whether these statistically-derived improvement criteria reflect the opinions of clinicians and patients with AS. In chapter 6, the ASAS improvement criteria are evaluated for their clinical relevance as assessed by the opinion of an expert panel in the field of AS, consisting of members of the ASAS Working Group and patients with AS, by means of a 3-round consensus building Delphi-exercise.

Many outcome measures of the core set in AS research consist of questionnaires that should be valid in all their aspects. For this purpose the "OMERACT Filter" has been proposed, consisting of three domains: truth

(validity), discrimination (reproducibility and responsiveness) and feasibility.[25] In chapter 7, the feasibility and discriminative properties of three widely used AS-specific questionnaires on disease activity and functioning are described. One of the criteria for feasibility is the appropriateness of the answer scales used in questionnaires. Because some patients may experience difficulties with application of visual analog scales (VAS) or Likert-formatted scales (which are used in these AS-specific questionnaires), the numerical rating scale is slightly preferred in the literature.[26-34] In this study the different questionnaires are tested on both their original answer scale (VAS or Likert) and a numerical rating scale format.

Fatigue has been acknowledged to be a very important symptom in AS.[35-38] A specific instrument for measuring fatigue in AS was, however, not yet selected by the ASAS Working Group. In chapter 8, the value of a single-item questionnaire and a multidimensional questionnaire for measuring fatigue in AS is studied. In addition, factors associated with fatigue in AS are identified and the impact of fatigue on quality of life is investigated.

Several questionnaires have been developed to measure functional impairment in AS. Disability in AS is often equated with functional impairment and measured with the same questionnaires. However, disability encompasses more than loss of function alone, and includes also impairment at a social and personal level.[39,40] To measure disability in general, the WHO has developed a generic instrument that is conceptually compatible with the recently revised International Classification of Functioning and Disability (ICF): the WHO Disability Assessment Schedule II (WHODAS II).[41] In chapter 9, the value of the WHODAS II for measuring disability in AS is studied, and factors that are associated with short-term change on the WHODAS II are being identified, as are factors at baseline that can predict scores on the WHODAS II after 5 years.

Chapter 10 consists of a summary and general discussion on the findings of this thesis. A summary of this thesis in Dutch is provided in chapter 11.

References

1. van der Linden S, van der Heijde D. Ankylosing spondylitis. Clinical features. Rheum Dis Clin North Am 1998;24:663-76.
2. van der Linden SM, Valkenburg HA, de Jongh BM, Cats A. The risk of developing ankylosing spondylitis in HLA-B27 positive individuals. A comparison of relatives of spondylitis patients with the general population. Arthritis Rheum 1984;27:241-9.
3. Braun J, Bollow M, Remlinger G, et al. Prevalence of spondylarthropathies in HLA-B27 positive and negative blood donors. Arthritis Rheum 1998;41:58-67.
4. Brown M, Wordsworth P. Predisposing factors to spondyloarthropathies. Curr Opin Rheumatol 1997;9:308-14.
5. Khare SD, Luthra HS, David CS. HLA-B27 and other predisposing factors in spondyloarthropathies. Curr Opin Rheumatol 1998;10:282-91.
6. Wordsworth P. Genes in the spondyloarthropathies. Rheum Dis Clin North Am 1998;24:845-63.
7. Breban M, Gombert B, Amor B, Dougados M. Efficacy of thalidomide in the treatment of refractory ankylosing spondylitis. Arthritis Rheum 1999;42:580-1.
8. Brandt J, Haibel H, Cornely D, et al. Braun J. Successful treatment of active ankylosing spondylitis with the anti-tumor necrosis factor alpha monoclonal antibody infliximab. Arthritis Rheum 2000;43:1346-52.
9. Van den Bosch F, Kruithof E, Baeten D, De Keyser F, Mielants H, Veys EM. Effects of a loading dose regimen of three infusions of chimeric monoclonal antibody to tumour necrosis factor alpha (infliximab) in spondyloarthropathy: an open pilot study. Ann Rheum Dis 2000;59:428-33.
10. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. Lancet 2000;356:385-90.
11. Creemers MC, van Riel PL, Franssen MJ, van de Putte LB, Gribnau FW. Second-line treatment in seronegative spondylarthropathies. Semin Arthritis Rheum 1994;24:71-81.
12. Olivieri I, Cantini F, Salvarani C. Diagnostic and classification criteria, clinical and functional assessment, and therapeutic advances for spondyloarthropathies. Curr Opin Rheumatol 1997;9:284-90.
13. Dougados M, Revel M, Khan MA. Spondylarthropathy treatment: progress in medical treatment, physical therapy and rehabilitation. Baillieres Clin Rheumatol 1998;12:717-36.
14. Koehler L, Kuipers JG, Zeidler H. Managing seronegative spondarthritides. Rheumatology (Oxford) 2000;39:360-8.
15. Toussirot E, Wendling D. Current guidelines for the drug treatment of ankylosing spondylitis. Drugs 1998;56:225-40.
16. Hidding A, van der Linden S, Boers M, et al. Is group physical therapy superior to individualized therapy in ankylosing spondylitis? A randomized controlled trial. Arthritis Care Res 1993;6:117-25.

17. Sukenik S, Neumann L, Buskila D, Kleiner Baumgarten A, Zimlichman S, Horowitz J. Dead Sea bath salts for the treatment of rheumatoid arthritis. Clin Exp Rheumatol 1990;8:353-7.
18. Sukenik S, Buskila D, Neumann L, Kleiner Baumgarten A, Zimlichman S, Horowitz J. Sulphur bath and mud pack treatment for rheumatoid arthritis at the Dead Sea area. Ann Rheum Dis 1990;49:99-102.
19. Elkayam O, Wigler I, Tishler M, et al. Effect of spa therapy in Tiberias on patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 1991;18:1799-803.
20. Sukenik S, Neumann L, Flusser D, Kleiner Baumgarten A, Buskila D. Balneotherapy for rheumatoid arthritis at the Dead Sea. Isr J Med Sci 1995;31:210-4.
21. Wigler I, Elkayam O, Paran D, Yaron M. Spa therapy for gonarthrosis: a prospective study. Rheumatol Int 1995;15:65-8.
22. Nguyen M, Revel M, Dougados M. Prolonged effects of 3 week therapy in a spa resort on lumbar spine, knee and hip osteoarthritis: follow-up after 6 months. A randomized controlled trial. Br J Rheumatol 1997;36:77-81.
23. van der Heijde D, Calin A, Dougados M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. J Rheumatol 1999;26:951-4.
24. Anderson JJ, Baron G, van der Heijde D, Felson DT, Dougados M. Ankylosing spondylitis assessment group preliminary definition of short-term improvement in ankylosing spondylitis. Arthritis Rheum 2001;44:1876-86.
25. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology [editorial]. J Rheumatol 1998;25:198-9.
26. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. Ann Rheum Dis 1978;37:378-81.
27. Kremer E, Atkinson JH, Ignelzi RJ. Measurement of pain: patient preference does not confound pain measurement. Pain 1981;10:241-8.
28. Dixon JS, Bird HA. Reproducibility along a 10 cm vertical visual analogue scale. Ann Rheum Dis 1981;40:87-9.
29. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986;27:117-26.
30. Bird HA, Dixon JS. The measurement of pain. Baillieres Clin Rheumatol 1987;1:71-89.
31. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. J Chronic Dis 1987;40:1129-33.
32. Ferraz MB, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022-4.
33. Bolton JE, Wilkinson RC. Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. J Manipulative Physiol Ther 1998;21:1-7.
34. Eyres S, van der Heijde D, Dougados M, Tennant A. The Visual Analogue Scale: Deception on a 'VASt' scale? [Submitted].
35. Calin A, Edmunds L, Kennedy LG. Fatigue in ankylosing spondylitis-why is it ignored? J Rheumatol 1993;20:991-5.

36. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.
37. Jones SD, Koh WH, Steiner A, Garrett SL, Calin A. Fatigue in ankylosing spondylitis: its prevalence and relationship to disease activity, sleep, and other factors. J Rheumatol 1996;23:487-90.
38. Ward MM. Health-related quality of life in ankylosing spondylitis: a survey of 175 patients. Arthritis Care Res 1999;12:247-55.
39. Gray DB, Hendershot GE. The ICIDH-2: developments for a new era of outcomes research. Arch Phys Med Rehabil 2000;81:S10-4.
40. Chatterji S, Ustun B, Bickenbach JE. What is disability after all? Disabil Rehabil 1999;21:396-8.
41. Ustun TB, Chatterji S, Rehm J, et al. World Health Organization Disability Assessment Schedule II (WHODAS II): development, psychometric testing, and applications [Submitted].

Part I

Therapeutic aspects of ankylosing spondylitis and other spondylarthropathies

Chapter 2

Spondylarthropathies: options for combination therapy

AM van Tubergen, RBM Landewé, Sj van der Linden

Springer Seminars in Immunopathology 2001; 23:147-63

Introduction

The spondylarthropathies (SpA) are a family of rheumatic diseases classified together because of common features, including the absence of rheumatoid factor, the presence of mono- or oligoarthritis predominantly of the lower limbs, and inflammatory back pain (spondylitis, sacroiliitis). They are strongly associated with the class-I HLA-antigen B27. Organ-specific manifestations such as acute anterior uveitis, inflammatory bowel disease, psoriasis, enthesitis (inflammation of the tendon-bone junction), carditis, conjunctivitis, and genital inflammation are characteristic features for this group of diseases. Patients may have more than one symptom at the same time or at different moments during life. A positive family history is an important hallmark of SpA, and reflects its genetic background.

Two criteria sets for classifying the SpA have been proposed by Amor et al.[1] and the European Spondylarthropathy Study Group (ESSG)[2] respectively. The ESSG criteria were developed to encompass not only the major disorders belonging to the group of SpA (for which specific classification criteria already existed), but also to include the so-called undifferentiated SpA, syndromes that share certain features of major SpA disorders, but that did not fulfill the available classification criteria for these disorders.[2]

The family of SpA includes ankylosing spondylitis (AS), psoriatic arthritis (PsA), reactive arthritis, arthritis associated with inflammatory bowel disease (Crohn's disease and ulcerative colitis), one subform of juvenile chronic arthritis, as well as undifferentiated spondylarthropathy. Of some disorders -SAPHO (synovitis, acne, pustulosis, hyperostosis, and osteomyelitis) syndrome, Behçet's disease, and Whipple's disease- the association with the family of SpA is still somewhat controversial.

Currently, no curative treatment is available for SpA. Treatment is aimed at alleviation of symptoms of inflammation (with non-steroidal anti-inflammatory drugs [NSAIDs]), maintaining mobility (with physical therapy), and long-term suppression of disease activity (with so-called disease modifying anti-rheumatic drugs [DMARDs]). Only very recently, the armamentarium of drugs available for disease suppression has been expanded with "biologicals", products of modern biotechnology that specifically inhibit mediators of inflammation.

Here we review drug treatment for SpA. Unfortunately, only a limited number of studies explicitly report combinations of treatment. Unlike diseases such as rheumatoid arthritis (RA), in which combination therapy refers to combinations of DMARDs, in SpA it is difficult to state what combination therapy means, since many combinations can be made. For instance, patients with SpA who use second-line drugs will usually continue their NSAIDs, and patients with AS are recommended to do special physical exercises in addition to drug treatment. Although usually not identified as such in the literature, the treatment of SpA is often a combination of several strategies.

A second limitation we encountered in preparing our review was the absence of clear response criteria for clinical trials in SpA in general, and in subgroups of SpA. Comparison of results is therefore difficult. It is only very recently that core sets for clinical record keeping, physical therapy, symptom-modifying anti-rheumatic drugs and disease controlling anti-rheumatic therapy in AS have been selected to facilitate comparison across studies.[3]

We focus here on the treatment of SpA in general. However, much clinical research has been performed in more homogeneous patient groups, comprising patients with for instance only AS or PsA. Although SpA are classified together for their common features, subgroups may respond differently to similar treatments. We will therefore also discuss the main therapeutic options for (combination) therapy in AS and PsA. We have made our choice on the basis of the prevalence of the underlying disorders and the availability of (controlled) clinical trials.

General treatment of the SpA

NSAIDs

In the treatment of SpA the NSAIDs play a central role. NSAIDs have a rapid effect on inflammatory manifestations. In particular, patients with involvement of the axial skeleton respond well to NSAID treatment. Amor et al. even included the presence of such a response as one of the criteria for SpA.[1]

A major disadvantage of the NSAIDs are the commonly occurring gastrointestinal adverse events. The prevalence of gastric and duodenal ulceration in patients with RA and osteoarthritis regularly using NSAIDs for less than 1 year was reported to be 14%, which increased to 26% in patients taking NSAIDs for more than 1 year up to 15 years.[4] It is unknown whether the prevalence of NSAID-gastropathy is similar in SpA. Singh and Rosen Ramey described differences among NSAIDs with respect to toxicity.[5] They found ibuprofen and aspirin to be the least toxic NSAID, whereas fenoprofen, indomethacin and ketoprofen were the most toxic. However, the new generation NSAIDs (cyclo-oxygenase (COX)-2 selective drugs) that are presumed to have less gastrointestinal toxicity were not yet included in this study.

NSAIDs inhibit COX, of which two isoforms exist. COX-1 is the constitutive form and has physiological functions; COX-2 is induced in inflammation.[6] The anti-inflammatory effects of NSAIDs are attributed to inhibition of COX-2, whereas inhibition of COX-1 may result in adverse events. The expression of COX-2, but not the expression of COX-1, is increased not only in synovial tissue of patients with inflammatory arthritis, but also, and preferably, in synovial tissue of patients with AS.[7] These findings may have implications for the use of COX-2 selective NSAIDs in patients with AS, who might benefit more from these drugs compared with patients with other rheumatic disorders. These new COX-2 selective drugs have been recently introduced. A reduction in the number of gastrointestinal side effects compared with established non-selective NSAIDs was

found in osteoarthritis patients, the level now being similar or equivalent to placebo.[6] As expected, COX-2 selective NSAIDs were not more effective than standard NSAIDs in these patients. We are now awaiting the results of COX-2 selective studies in SpA.

Local steroids

In case of a contraindication for the use of NSAIDs or an inadequate disease control with these drugs, several other treatment options are available. In acute peripheral joint inflammation local steroid injections may reduce inflammatory symptoms in SpA. Three studies reported the efficacy of corticosteroid injections in the sacroiliac joint, two of which were double-blind randomized controlled studies.[8-10] In all studies a significant reduction of pain up to 80% was observed. The use of NSAIDs decreased in 50% of the patients, and some even discontinued. Braun et al. observed a clear improvement of MRI assessed sacroiliitis and inflammatory back pain in 25 of 30 patients treated, subjectively lasting for a mean of 8.9 months (0.2-15 months).[9]

DMARDs

Second-line drugs or DMARDs may be considered for patients inadequately responding to NSAIDs, or with persisting peripheral joint involvement. Most experience with DMARDs in SpA is with sulfasalazine (SSZ). SSZ is a slow-acting drug, the effects of which become apparent after several weeks or months of intake. The dosage is gradually increased, and SSZ is often supplied in combination with an appropriate NSAID.

The efficacy and tolerability of SSZ in the entire group of SpA was investigated by Dougados et al.[11] In a randomized double-blind placebo-controlled trial patients with SpA refractory to treatment with NSAIDs received SSZ or placebo. Patients suffered from AS (n=134), PsA (n=136), or reactive arthritis (n=81). The withdrawal rates due to side effects were 5% in the placebo group, and 16% in the SSZ group. After 6 months, in the intention-to-treat analysis, a statistically significant difference was found with respect to the patient's overall assessment of disease activity (60% of the patients in the SSZ group and 44% of the placebo patients improved by at least 1 point on a 5-point scale). There was a strong tendency in favor of SSZ with respect to improvement in pain and physician's overall assessment. In the completer analysis, a significant difference for these three outcome variables was found in favor of the SSZ group. Subgroup analysis showed that most impressive improvements were seen in patients with PsA, especially in those with polyarticular involvement.

Clegg et al. performed a re-analysis of three previously conducted randomized double-blind controlled trials of SSZ versus placebo in 264 AS, 221 PsA, and 134 reactive arthritis patients.[12] They focused on the efficacy of SSZ in peripheral versus axial manifestations of SpA, and found that in particular patients with peripheral joint involvement benefited from SSZ treatment.

There is still controversy as to which is the active moiety of SSZ: the sulfonamide sulfapyridine (SP) or the 5-aminosalicylic acid (5-ASA) moiety. Taggart et al. showed in a comparative randomized study of SSZ versus SP or 5-ASA that SP appeared to be the active moiety.[13] Patients receiving SP had significantly better outcome compared with those treated with 5-ASA. However, the results of the SSZ group showed a trend towards more improvement compared with SP alone. More recently, two open studies in patients with active SpA showed positive results for a group of patients treated with 5-ASA, suggesting that 5-ASA might be the active moiety.[14,15]

In conclusion, there seems to be some place for SSZ as DMARD in patients with SpA, preferably in those with polyarticular peripheral arthritis.

Combination DMARD therapy

A comparison of different combination strategies in SpA has been reported by Çalgüneri et al.[16] They studied in 138 patients with SpA whether a combination of two or three DMARDs given concurrently was superior to one DMARD. Patients with active SpA and peripheral joint involvement were divided into three groups: group 1 (n=48) SSZ 1-2 g/day; group 2 (n=45) SSZ 1-2 g/day in combination with methotrexate (MTX) 7.5-15 mg/week; group 3 (n=45) SSZ 1-2 g/day in combination with MTX 7.5-15 mg/week and hydroxychloroquine 200 mg/day. After 2 years significant improvements compared with baseline were found with all three regimens, but the improvements were greater with combination therapies. Group 3 showed most improvement. Statistically significantly more improvement was found in group 3 versus 1 or 2 with respect to erythrocyte sedimentation rate (ESR), morning stiffness, total joint count, spinal mobility, patients' and physician's global assessments, and analgesics consumption. This preliminary study emphasizes that combination DMARD treatment may certainly be a valuable option in patients with SpA. More studies and studies including subgroups of patients with SpA, performed on the basis of proper response criteria to increase statistical power, are needed to establish the role of DMARD-combination therapy in SpA.

Biologicals

Although a number of patients experience adequate disease control with an NSAID, with or without corticosteroid injections or DMARD therapy, a group of patients remains refractory to these treatments. The latest developments in the treatment of SpA comprise the administration of the tumor necrosis factor-alpha (TNF-α) inhibiting drugs, including thalidomide, as well as the soluble TNF-α receptors (etanercept), and chimerised monoclonal antibodies against human TNF-α (infliximab). TNF-α is a pro-inflammatory cytokine mainly produced by macrophages and monocytes, and plays a major role in inflammation and immune responses in general.[17] In patients with AS, high levels of TNF-α mRNA were found in the sacroiliac joints, implying that TNF-α may be important in inflammatory sacroiliac joint disease.[18] Based on these findings inhibition of TNF-α may be an interesting therapeutic option. So far, only a limited number

of studies with anti-TNF-α therapy in SpA in general have been published, but remarkable improvements of symptoms were observed.[19,20]

El Hassani et al.[19] assessed the efficacy of thalidomide, a drug that inhibits production of TNF-α, in 11 patients with SpA (AS n=7, PsA n=1, undifferentiated SpA n=3) refractory to standard treatment. Patients received thalidomide 100-300 mg/day for 6 months. Three patients discontinued because of side effects. Clinical parameters improved in five patients, and remained stable in three; biological parameters of inflammation improved in six patients.

Van den Bosch et al. assessed the efficacy of infliximab in an open pilot study of 21 patients with active SpA (AS n=10, PsA n=9, undifferentiated spondylarthropathy n=2) refractory to established treatments.[20] The patients received three infusions of 5 mg/kg infliximab at weeks 0, 2, and 6. A rapid and statistically significant improvement in subjective and objective variables was observed from day 3 onwards, and persisted for another 12 weeks. After 12 weeks both patients' and physician's global assessment improved with 82%, and pain with 86%. The mean ESR decreased from 44 mm/h to 4 mm/h, and mean CRP decreased from 4.6 mg/dL to 0.6 mg/dL. Morning stiffness reduced from 90 minutes to 5 minutes on average. In patients with peripheral joint involvement, the number of tender and swollen joints significantly decreased; patients with axial involvement showed significant improvements in functioning and disease activity. Furthermore, and interestingly, a decrease in psoriatic skin involvement was observed in patients with PsA. Only minor side effects were reported, which did not cause withdrawal from the study.

Ankylosing spondylitis

AS is a key disorder of the family of SpA, because of its rather high prevalence (0.1%), and because AS potentially includes all symptoms of SpA. In AS, it is predominantly the axial skeleton which is affected. Sacroiliitis is the hallmark of the disease. AS may cause serious functional impairment as the disease progresses. Treatment in AS is aimed at reducing inflammatory symptoms and preventing or at least minimizing deformity and disability. The standard therapy in active AS remains drug therapy with NSAIDs, which reduce inflammatory symptoms. In patients not adequately responding to NSAIDs, DMARDs may be added to the treatment.

Ward and Kuzis surveyed 226 patients with AS, participating in a longitudinal outcome study, and 123 American rheumatologists about their judgments of the effectiveness of different treatments for AS.[21] Of these patients, 83% reported using NSAIDs, 27% analgesics, 17% second-line medication, and 10% no medication. Twenty-five percent used indomethacin, 4% phenylbutazone, 12% SSZ, 5% MTX, and 12% received physical exercise therapy. This is in contrast to the opinion of the rheumatologists: 81% rated indomethacin and 90% phenylbutazone "very" or "extremely effective" in treating active AS.

Indomethacin was recommended by 82% of rheumatologists for patients with active AS; 64% recommended SSZ, and 20% MTX for patients refractory to NSAIDs.

Physical therapy

Physical exercises are important in the treatment of AS to maintain and improve mobility, fitness and strength, and to prevent deformity.[22] Exercises may be done at home or at a physiotherapy practice, or in group settings supervised by a physiotherapist. Hidding et al. showed in a randomized controlled study that patients attending weekly group physical therapy had more improvement in global health, thoraco-lumbar mobility, and fitness than patients continuing unsupervised therapy at home.[22]

Spa therapy is also frequently performed by patients with AS. Recently, the efficacy of a 3-week course of spa therapy in combination with exercises and followed by weekly group physical therapy was compared with weekly group physical therapy alone in a randomized controlled trial.[23] Significantly more improvement in the primary outcomes measures, i.e. functioning, patients' global well-being, and inflammatory symptoms, were observed in the spa-exercise groups compared with the physical therapy group. Beneficial effects lasted for at least 40 weeks.

Sulfasalazine

SSZ has been used since mid 1980s in the treatment of AS. In 1990, a meta-analysis of five randomized controlled studies of SSZ versus placebo in patients with AS showed that SSZ was a well-tolerated and effective drug.[24] The pooled estimate of improvement in favor of SSZ was 28% in duration of morning stiffness, 31% in severity of morning stiffness, 27% in severity of pain, 7% in general well-being, 9% in ESR, and 12% in IgA, all over placebo. Adverse events were more frequently observed in the SSZ group than in the placebo group (odds ratio 1.57), but this difference was not statistically significant. However, studies published after this meta-analysis showed less convincing evidence for the efficacy of SSZ in AS, especially in patients with axial involvement and longstanding disease duration.[25,26]

Kirwan et al. studied 89 patients with established AS, most of them with symptoms confined to the axial skeleton, allocated to either SSZ or placebo treatment (it was unclear whether the allocation was random).[25] After 3 years, the results for the SSZ group were not better than those for the placebo group, although a trend favoring SSZ was seen with respect to spinal mobility, stiffness, and sleep quality. The only statistically significant difference included a lower number of episodes of peripheral joint symptoms in the SSZ group (0.298 episodes/year) compared with the placebo group (0.392 episodes/year). Although the results in this study may have suffered from insufficient statistical power, the lack of significant differences emphasizes that the effects of SSZ in AS are certainly not dramatic.

Clegg et al. also compared SSZ with placebo in patients with longstanding AS (n=264).[26] Response rates (based on non-validated criteria such as patients' and physician's assessment, duration of morning stiffness, and severity of back pain) were 38% for the SSZ group and 36% for the placebo group. Only the ESR had significantly more decreased in the SSZ group compared with controls. However, the subgroup of patients with peripheral joint involvement showed significantly better response rates (56%) compared with placebo (30%).

Therefore, SSZ may have some value in the treatment of patients with AS, but should probably be prescribed preferentially to those with short disease duration and peripheral joint involvement.

Methotrexate

Patients failing both on NSAIDs and SSZ may benefit from MTX.[27,28] In an open study Creemers et al. treated 11 patients with AS (3-26 years disease duration and refractory to NSAIDs and SSZ) with MTX 7.5-15 mg/week.[27] Nine patients were evaluable after 24 weeks and improvements of more than 15% were observed for pain in 6 of 9 patients, general well-being in 6 of 9, morning stiffness in 5 of 9, functioning in 5 of 9, spinal mobility (Schober index) in 6 of 9, and peripheral joint involvement in 5 of 6, although individual patients responded differently. Side effects were mild and reversible.

Biasi et al. treated 17 patients with short disease duration (4.8 years) with MTX 7.5-10 mg/week in an open study.[28] After 3 years, 16 patients were evaluable. Night pain had completely disappeared, general well-being improved with 94%, Schober index improved from 2 cm to 5 cm (59%), finger-floor distance decreased from 33 cm to 7 cm (78%), and both ESR and CRP significantly decreased. The number of swollen joints declined from 5 to 2, the number of tender joints from 8 to 3. Fifteen patients showed a significant clinical amelioration. Nine patients had discontinued NSAIDs.

Contrary to these positive results, Ostendorf et al. did not observe significant clinical and laboratory improvements in 10 patients with AS refractory to NSAIDs and SSZ (disease duration 11 years) after 1 year of treatment with MTX 7.5-15 mg/week.[29] In the same study, however, two other groups of 10 patients each, with RA and PsA respectively, responded significantly to treatment with MTX.

To our knowledge, no controlled studies have been published, so that the real value of MTX therapy in AS can not be established, yet.

Other therapeutic options

Only a few studies have tested the efficacy of systemic corticosteroids in AS.[30,31] Mintz et al. studied the effects of pulse therapy with 1 g of methylprednisolone 1, 2, 3, or 4 times on consecutive days in five patients with severe AS.[30] The duration of improvements in spinal mobility, morning stiffness, and pain after 1 or 2 pulses was short, but after 3 or 4 pulses improvements lasted for 1 year on average. A comparative study of 1 g versus 375 mg methylprednisolone did not show statistically significant differences between the two groups, although there

was a trend towards more and longer lasting improvements in the high-dose group.[31] Unfortunately, no other controlled studies have been conducted.

It has been suggested that bisphosphonates may suppress the generation of pro-inflammatory cytokines and they have, therefore, been studied by Maksymowych et al. in 16 patients with AS (disease duration 3-32 years).[32] Eight patients received intravenous pamidronate 30 mg each month for 3 months, followed by an additional 3 months of 60 mg (group 1), and eight patients received 60 mg monthly for 3 months only (group 2). After 6 months, improvements of 37% in disease activity, 21% in functioning, and 36% in metrology index were found for group 1; and of 31%, 18%, and 31% respectively for group 2. However, no randomized controlled trials have yet been published.

D-penicillamine in the treatment of AS has been investigated in a single small randomized controlled trial.[33] After 6 months, none of the patients felt improved, and no significant changes were observed. Although the study was small, the results do not support a role for penicillamine in the treatment of AS.

Biologicals

Anti-TNF-α therapy in AS is now under investigation. Breban et al. reported impressive clinical and laboratory improvements in two patients refractory to NSAIDs and DMARDs after treatment with thalidomide.[34] Recently, Brandt et al. conducted an open trial on the efficacy of treatment with three infusions of infliximab in 11 patients with AS.[35] One patient had withdrawn after the first infusion, because of a generalized rash. At weeks 2 and 4 improvements of more than 50% in disease activity, functioning, and pain were observed in 9 of 10 patients. The positive effects persisted until 12 weeks in 8 of 10 patients. Quality of life significantly improved in six of the nine dimensions of the Short Form-36 at 4 weeks. The use of NSAIDs decreased with 50%, and five patients completely stopped taking NSAIDs. Randomized controlled studies are now in progress.

Psoriatric arthritis

PsA is an inflammatory arthritis associated with psoriasis. More than 1% of the population suffers from psoriatic skin lesions. Arthritis may complicate psoriasis in approximately 10% of the patients, although the skin lesions may be absent. This figure makes PsA one of the more common disorders of the SpA-group. PsA is a heterogenic disease and as such has an unpredictable disease course. Different treatment strategies may be necessary. Predictive factors for disease severity include significant inflammation at the first visit and carriership for HLA B27, B39, and/or DQw3, "protective" factors are a low ESR at presentation and carriership for HLA DR7.[36,37]

The majority of patients with PsA can be managed with NSAIDs alone, if necessary combined with local corticosteroid injections. In patients not

adequately responding to NSAIDs and in severe progressively destructive PsA, second-line drugs can be added. MTX, cyclosporine A (CsA), and SSZ are currently the most frequently used DMARDs in PsA. A systematic review on drug therapy in PsA concluded that only the efficacy of SSZ and parental high-dose MTX had been sufficiently proven.[38] Azathioprine, etretinate, oral low-dose MTX, and perhaps colchicine were suggested to be effective, but further clinical trials are necessary to establish efficacy. No studies on combination therapy were reviewed.

Chang performed a survey among 109 American rheumatologists determining their perceptions on the effectiveness of DMARDs and treatment choices for peripheral PsA; 59% of the rheumatologists responded.[39] MTX was ranked as the most effective drug by 92% of the rheumatologists. To their opinion, the next most effective drugs were SSZ (28%), gold (23%), steroids (21%), and CsA (15%). In "mild" PsA, SSZ was the drug of first choice in 55%, and MTX second choice (37%). In "moderate" and "severe" PsA, MTX was the drug of first choice in 77% and 92%, respectively. Thirty-six percent used combinations of MTX with another DMARD (not specified).

Sulfasalazine

Several randomized controlled trials have studied the efficacy of SSZ in PsA.[40-44] Three small studies, with less than 40 patients in each, and in which the patient groups were not well defined, all suggested beneficial effects of SSZ.[40-42] More recently, two large multicenter trials confirmed earlier observations.[43,44]

Combe et al. compared the efficacy of SSZ 2.0 g/day (n=53) versus placebo (n=64) in patients with PsA (disease duration of arthritis 8 years).[43] Pain significantly decreased in the SSZ group with 41%, compared with 24% in the placebo group. Morning stiffness, joint pain/tenderness index, and ESR showed a trend favoring SSZ.

Clegg et al. also compared the efficacy of SSZ 2.0 g/day (n=109) with placebo (n=112) in patients with PsA resistant to NSAIDs (disease duration 12 years).[44] After 36 weeks, a treatment response (based on non-validated criteria such as improvements in patients' and physician's global assessment, joint pain/tenderness, and joint swelling) of 58% was found in the SSZ group versus 45% in the placebo group ($P=0.05$). Only the difference in patients' global assessment of the treatment response reached statistical significance, and a trend favoring SSZ over placebo with respect to physician's global assessment was found. The difference between SSZ and placebo reached a maximum at 36 weeks, suggesting that the full effects of SSZ may take some time to become apparent. In all studies SSZ was well tolerated. Side effects were often reported, and led to withdrawal in 13-30% for the SSZ group, and 5-23% for the placebo group. Gastrointestinal side effects were most often reported.

Methotrexate

Although MTX is currently considered as the preferred second-line drug treatment in PsA, no large randomized double-blind placebo controlled studies

could be found in the English literature. Two small randomized controlled trials on the efficacy of respectively high-dose MTX (n=21)[45] and low-dose MTX (n=37)[46] in PsA have been published. Both studies showed significant improvements in skin involvement. Black et al. observed significant improvement in joint involvement and ESR.[45] In the study by Wilkens et al., only the between-group difference with respect to physician's assessment of disease severity reached statistical significance.[46] There was, however, a trend favoring MTX with respect to morning stiffness and patients' assessment of disease severity. Both studies might have suffered from insufficient statistical power. It seems likely that a well-powered clinical trial may confirm the efficacy of MTX in PsA.

Several open studies and one case-controlled study of MTX in patients with PsA with both peripheral and axial joint involvement were performed after the above-mentioned randomized controlled trials.[47-50] The numbers of patients studied ranged from 28 to 54 (disease duration approximately 10 years). The patients received MTX 7.5-15 mg/week and were followed for 12 months up to 2.8 years. Withdrawals due to side effects were found in up to 21% of the patients. Different outcome measures were used, but all showed a significant reduction in the number of swollen and tender joints. Improvements in the Ritchie index (joint tenderness count) up to 83%, a reduction in pain of up to 66%, and significant decreases in ESR and CRP were observed. Skin involvement improved in all studies. Nail involvement dramatically improved in the study by Espinoza et al.,[48] but did not change in the study by Pigatto et al.[49]

Abu Shakra et al. described the results of a case-control study on clinical and radiological outcome of 38 patients with PsA treated with MTX for 24 months, compared with 38 matched controls who had never taken MTX in the past.[50] Clinical response was defined as "improvement" if there was a reduction in the number of actively inflamed joints of >40%, and "poor" if there was a reduction of <20% or any increase in joint count. After 24 months similar treatment responses were found between the MTX group and matched controls, of which half were identified as being "improved". The radiographs showed an increase in damage score of 63% in the MTX group, and 47% in the matched controls (not significant).

Taking into account that most rheumatologists world-wide prefer MTX as a drug of first choice in patients with severe and less severe PsA, the evidence to endorse such a policy is extremely sparse. It is to be recommended that in PsA –before exploring the potential of MTX as a member in DMARD combination therapy– first of all the place of MTX as monotherapy should be established by sufficiently large, well-performed randomized controlled trials.

Cyclosporine A

CsA is also frequently prescribed as a DMARD in PsA, but mainly uncontrolled studies have been conducted.[51-57] In these open studies the results of 6-55 patients with PsA treated with CsA 1.5-7 mg/kg/day for 2 months to 2 years were described. Side effects that led to discontinuation were found in up to 15% of the patients. Significant improvements of 46-57% in number of painful joints, 49-72% in the Ritchie index, 30-81% in number of swollen joints, 35-61% in

pain, and 37-95% in morning stiffness were found. Also clinically relevant reductions in skin involvement were observed.

Salvarani et al. performed a multicenter randomized open trial comparing a group with CsA 3 mg/kg/day, a group with SSZ 2 g/day, and a group with symptomatic treatment only (NSAIDs, analgesics, and/or prednisone <5 mg/day), in 99 patients with active PsA.[58] The primary end-point was the 6-month change in pain. The CsA-group showed significantly more improvement in pain and skin involvement compared with both the SSZ group and the symptomatic treatment group. No significant clinical differences between SSZ and symptomatic therapy were found. ESR significantly decreased in the SSZ group versus symptomatic therapy; CRP significantly decreased in the CsA group. Using the American College of Rheumatology (ACR) criteria for defining improvement in patients with RA, the number of patients meeting an ACR70 response (almost similar to clinical remission) was significantly higher in the CsA group compared with SSZ (P=0.05). These results seem promising, but one should keep in mind that the ACR criteria have been developed for patients with active RA who are enrolled in clinical trials, so that a 70% response in PsA is not comparable with such a response in RA.

Spadaro et al. compared in a randomized controlled trial the efficacy of CsA 3-5 mg/kg/day (n=17) with MTX 7.5-15 mg/week (n=18) in patients with PsA with active peripheral joint involvement.[59] After 1 year, significant improvements in all clinical parameters were observed in both groups. No significant differences between the two treatment groups were found, but more patients from the CsA group had withdrawn because of side effects. This again, probably, is a study that suffered from insufficient statistical power, which does not allow a clear distinction between the two drugs with respect to efficacy.

Discontinuation of treatment is frequently observed in patients on CsA. Using a life-table analysis Spadaro et al. compared the efficacy and toxicity of several DMARDs prescribed in PsA.[60] The study showed that patients with PsA had a significantly lower probability for continuing CsA over time compared with MTX. The most common cause for discontinuation of CsA was hypertension (29%). MTX showed a significantly lower rate of withdrawal for inefficacy, compared with gold and CsA. Since patients taking CsA compared with MTX are less likely to continue long-term treatment, CsA seems not to be the drug of first choice in PsA.

Combination therapy

In two small and uncontrolled studies, combination therapy with MTX and CsA was investigated.[61,62] Mazzanti et al. reported the results of eight patients with PsA with both peripheral and axial involvement refractory to several DMARDs including MTX and CsA when given separately.[61] They were treated with a combination of MTX 10-15 mg/week and CsA 3-5 mg/kg/day. After 6 months the Ritchie index had decreased with 60%, pain with 55%, and ESR declined from 75 mm/h to 35 mm/h.

Espinoza et al. investigated combination therapy with MTX and CsA in 10 patients with PsA (mean duration of arthritis 10 years) who failed MTX

monotherapy with doses up to 50 mg/week.[62] They were switched to a combination of MTX (doses up to 20 mg/week) and CsA (2-3 mg/kg/day). After an average of 15 months follow up, clear improvements in skin involvement, Ritchie index, well-being, and a decrease in ESR from 55 mm/h to 18 mm/h were observed. However, radiographic deterioration occurred in 7 out of 10 patients, with development of new erosions.

The above-mentioned studies suggest that combination therapy with MTX and CsA may be of value in patients with severe treatment-refractory PsA, but sufficient evidence can only be elicited from properly conducted randomized controlled trials, which are currently lacking.

Other treatment options

The administration of colchicine, frequently used as an anti-inflammatory agent in gout, has also been investigated in patients with PsA in two randomized controlled cross-over trials in 25 and 15 patients, respectively.[63,64] Patients received either colchicine or placebo and switched to the other treatment after 2 months. Seideman et al. found significant improvements in Ritchie index, joint pain, joint size, and overall assessment in patients taking colchicine, but no changes in psoriatic skin lesions and laboratory values.[63] McKendry et al. failed to find significant changes in any of the outcome measures.[64] No large studies on colchicine treatment in PsA have been published to date.

The efficacy of etretinate, a vitamin A derivate that has shown to reduce psoriatic skin involvement, was also investigated for PsA.[65,66] In a double-blind randomized trial by Hopkins et al., patients received either etretinate (n=20) or ibuprofen (n=20).[65] Due to lack of efficacy, all but one patient from the ibuprofen group had discontinued at 24 weeks. No statistically significant between-group differences were found in the clinical parameters at any time point. ESR and CRP both decreased significantly in the etretinate group compared with the ibuprofen group. In addition, less skin involvement was observed in the etretinate group. In a pilot study, Klinkhoff et al. treated 40 patients with etretinate 50 mg/day.[66] Patients were followed up for 8-24 weeks. Significant improvements were observed for number of tender joints 50%, swollen joints 61%, morning stiffness 56%, pain 44%, patients' global well-being 43%, and ESR reduced from 39 mm/h to 23 mm/h.

The efficacy of gold salts in patients with PsA has been studied in two double-blind controlled multicenter trials.[67,68] No significant clinical changes were found after treatment with oral gold. However, treatment with intramuscular gold showed significant changes in pain score and Ritchie index.[68] In a comparative study on the efficacy of intramuscular gold in patients with PsA (n=14) and RA (n=42), Dorwart et al. showed a total of improvement (defined as >50% reduction in inflamed joints and disease activity) and remission in 71% of the patients with PsA, and in 60% of the patients with RA, suggesting that injections of gold may be effective in PsA.[69]

Several other drugs have been studied in the treatment of PsA. All of these studies were only performed once and comprised a small sample size. Vitamin D produced significant improvements of tender joint count, global well-being,

and ESR in a pilot study.[70] Azathioprine produced improvements in skin involvement, joint count, and morning stiffness compared with placebo.[71] D-penicillamine[72] and fumaric acid[73] did not show significantly better results compared with placebo. However, within-group analyses showed (significant) improvements in pain, stiffness, Ritchie index, patients' global well-being, and ESR. Anti-malarials have been assessed in a case-control study.[74] A trend towards more improvement in the number of inflamed joints in the active treatment group was seen, but failed to reach statistical significance. In addition, more exacerbations of psoriasis were observed in the anti-malarial group.[74]

Biologicals

To date, only a few studies have published results with anti-TNF-α in PsA.[75-77] In an open study, Antoni et al. treated 10 patients with PsA with infliximab 5 mg/kg on weeks 0, 2, and 6.[75] Eight patients continued MTX, one patient SSZ. At 10 weeks, all patients had improved according to the ACR50 criteria, and 7 of 10 fulfilled the ACR70 criteria. The number of swollen joints had decreased with 93% and tender joints with 90%. Physician's global assessment improved with 90% and functioning with 80%. CRP and ESR also dramatically improved. Yazici et al. treated 10 patients with active arthritis, and refractory to a mean of three DMARDs, with etanercept 25 mg twice weekly administered subcutaneously.[76] At 3 months, 5 of 10 patients had a complete remission of arthritis, and joint count had decreased with 82%. At 12 months, eight patients were still taking etanercept and maintained a good response; one patient discontinued due to side effects; and one patient discontinued due to increased disease activity. Mease et al. conducted a randomized controlled trial of etanercept versus placebo in 60 patients with PsA (disease duration 9 years).[77] Forty-seven percent of the patients continued MTX. After 12 weeks, response to treatment (based on non-validated criteria adapted from Clegg et al.[44] and including improvements in patients and physicians global assessment, joint tenderness, and joint swelling) was observed in 87% of the etanercept group, and in 23% of the placebo group; the ACR20 criteria were met by 73% of the etanercept group and 13% of the placebo group. Skin involvement only improved in the etanercept group.

Conclusive remarks

In contrast with the trends in RA, surprisingly little information on DMARD combination therapy is available with respect to patients with SpA. Reasons for this may be that the place of DMARD monotherapy in SpA has not yet been properly defined, that the prevalence of SpA is somewhat lower than that of RA, that up to now treatment response criteria were not defined for diseases of the SpA group, and that the long term course of diseases of the SpA group is considered generally better than that of RA.

DMARDs are frequently prescribed in addition to NSAIDs and also physical therapy plays an important role in the overall management of patients with SpA, but the combinations of these treatments were only sparsely described in the methodology sections of the papers studied.

SSZ is the most thoroughly studied DMARD for monotherapy in SpA. A number of randomized clinical trials on the efficacy of SSZ in patients with SpA in general, AS, and PsA is available.[12-15,24-26,40-44,78] The communis opinio is that SSZ is effective in suppressing inflammation, but only in peripheral joints rather than in axial joints.

Referring to the small numbers of studies available, it looks as if the popularity of MTX as a DMARD in PsA has been extrapolated to the entire SpA group. It is questionable whether this policy, which is not evidence based, is correct. It is recommended that the positive results of MTX elicited from small open studies be confirmed once in randomized controlled trials.

Only three studies were identified which reported the effects of treatment with more than one DMARD.[16,61,62] Two of these studies were small and not blinded (a draw back that has influenced the results of many trials in SpA).[61,62] Çalgüneri et al.[16] used a larger number of patients, but it could not be deduced whether the patients were randomly allocated to the treatment arms. However, all studies showed promising results, and larger randomized controlled studies are certainly justified.

Anti-TNF-α therapy might be a breakthrough in the treatment of SpA. Dramatic improvements in patients with SpA refractory to several established treatments have been reported in a number of open studies,[19,20,34,35,75-77] and one of the few randomized controlled trials yet available in PsA showed impressive efficacy of etanercept.[77] Anti-TNF-α might be effective in suppressing symptoms of inflammation of both axial and peripheral joints and have positive effects on skin manifestations. The results of well-performed randomized controlled trials are needed to define the role of anti-TNF-α in the treatment of SpA. In particular, combinations of DMARDs with anti-TNF-α therapy should be studied in patients with severe refractory SpA, in which anti-TNF-α therapy may induce, and conventional DMARDs (presumably MTX) may maintain clinical remission.

References

1. Amor B, Dougados M, Mijiyawa M. Criteria of the classification of spondylarthropathies [in French]. Rev Rhum Mal Osteoartic 1990;57:85-9.
2. Dougados M, van der Linden S, Juhlin R, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum 1991;34:1218-27.
3. van der Heijde D, Calin A, Dougados M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. J Rheumatol 1999;26:951-4.
4. Cheatum DE, Arvanitakis C, Gumpel M, Stead H, Geis GS. An endoscopic study of gastroduodenal lesions induced by nonsteroidal anti-inflammatory drugs. Clin Ther 1999;21:992-1003.
5. Singh G, Rosen Ramey D. NSAID induced gastrointestinal complications: the ARAMIS perspective--1997. Arthritis, Rheumatism, and Aging Medical Information System. J Rheumatol 1998;51 (suppl 51):8-16.
6. Jackson LM, Hawkey CJ. COX-2 selective nonsteroidal anti-inflammatory drugs. Do they really offer any advantages? Drugs 2000;59:1207-16.
7. Siegle I, Klein T, Backman JT, Saal JG, Nusing RM, Fritz P. Expression of cyclooxygenase 1 and cyclooxygenase 2 in human synovial tissue: differential elevation of cyclooxygenase 2 in inflammatory joint diseases. Arthritis Rheum 1998;41:122-9.
8. Maugars Y, Mathis C, Berthelot JM, Charlier C, Prost A. Assessment of the efficacy of sacroiliac corticosteroid injections in spondylarthropathies: a double-blind study. Br J Rheumatol 1996;35:767-70.
9. Braun J, Bollow M, Seyrekbasan F, et al. Computed tomography guided corticosteroid injection of the sacroiliac joint in patients with spondyloarthropathy with sacroiliitis: clinical outcome and followup by dynamic magnetic resonance imaging. J Rheumatol 1996;23:659-64.
10. Luukkainen R, Nissila M, Asikainen E, et al. Periarticular corticosteroid treatment of the sacroiliac joint in patients with seronegative spondylarthropathy. Clin Exp Rheumatol 1999;17:88-90.
11. Dougados M, van der Linden S, Leirisalo Repo M, et al. Sulfasalazine in the treatment of spondylarthropathy. A randomized, multicenter, double-blind, placebo-controlled study. Arthritis Rheum 1995;38:618-27.
12. Clegg DO, Reda DJ, Abdellatif M. Comparison of sulfasalazine and placebo for the treatment of axial and peripheral articular manifestations of the seronegative spondylarthropathies: a Department of Veterans Affairs cooperative study. Arthritis Rheum 1999;42:2325-9.
13. Taggart A, Gardiner P, McEvoy F, Hopkins R, Bird H. Which is the active moiety of sulfasalazine in ankylosing spondylitis? A randomized, controlled study. Arthritis Rheum 1996;39:1400-5.

14. Thomson GT, Thomson BR, Thomson KS, Ducharme JS. Clinical efficacy of mesalamine in the treatment of the spondyloarthropathies. J Rheumatol 2000;27:714-8.
15. Dekker Saeys BJ, Dijkmans BA, Tytgat GN. Treatment of spondyloarthropathy with 5-aminosalicylic acid (mesalazine): an open trial. J Rheumatol 2000;27:723-6.
16. Çalgüneri M, Çobankara V, Apras S, et al. Combination therapy in sero-negative spondyloarthropathies (SSA) [abstract]. Ann Rheum Dis 2000;59 (suppl 1):S56.
17. Beutler BA. The role of tumor necrosis factor in health and disease. J Rheumatol 1999;26 (suppl 57):16-21.
18. Braun J, Bollow M, Neure L, et al. Use of immunohistologic and in situ hybridization techniques in the examination of sacroiliac joint biopsy specimens from patients with ankylosing spondylitis. Arthritis Rheum 1995;38:499-505.
19. El Hassani S, Dougados M, Gombert B, Amor B, Breban M. Treatment of severe refractory spondylarthropathy (SPA) with thalidomide: results of an open study [abstract]. Arthritis Rheum 1999;42 (suppl):S373.
20. Van den Bosch F, Kruithof E, Baeten D, De Keyser F, Mielants H, Veys EM. Effects of a loading dose regimen of three infusions of chimeric monoclonal antibody to tumour necrosis factor alpha (infliximab) in spondyloarthropathy: an open pilot study. Ann Rheum Dis 2000;59:428-33.
21. Ward MM, Kuzis S. Treatments used by patients with ankylosing spondylitis. Comparison with the treatment preferences of rheumatologists. J Clin Rheumatol 1999;5:1-8.
22. Hidding A, van der Linden S, Boers M, et al. Is group physical therapy superior to individualized therapy in ankylosing spondylitis? A randomized controlled trial. Arthritis Care Res 1993;6:117-25.
23. van Tubergen A, Wolter N, Goei Thè H, et al. Spa therapy is effective in patients with ankylosing spondylitis [abstract]. Ann Rheum Dis 2000;59 (suppl):56-7.
24. Ferraz MB, Tugwell P, Goldsmith CH, Atra E. Meta-analysis of sulfasalazine in ankylosing spondylitis. J Rheumatol 1990;17:1482-6.
25. Kirwan J, Edwards A, Huitfeldt B, Thompson P, Currey H. The course of established ankylosing spondylitis and the effects of sulphasalazine over 3 years. Br J Rheumatol 1993;32:729-33.
26. Clegg DO, Reda DJ, Weisman MH, et al. Comparison of sulfasalazine and placebo in the treatment of ankylosing spondylitis. A Department of Veterans Affairs Cooperative Study. Arthritis Rheum 1996;39:2004-12.
27. Creemers MC, Franssen MJ, van de Putte LB, Gribnau FW, van Riel PL. Methotrexate in severe ankylosing spondylitis: an open study. J Rheumatol 1995;22:1104-7.
28. Biasi D, Carletto A, Caramaschi P, Pacor ML, Maleknia T, Bambara LM. Efficacy of methotrexate in the treatment of ankylosing spondylitis: a three-year open study. Clin Rheumatol 2000;19:114-7.
29. Ostendorf B, Specker C, Schneider M. Methotrexate lacks efficacy in the treatment of severe ankylosing spondylitis compared to rheumatoid and psoriatic arthritis. J Clin Rheumatol 1998;4:129-36.
30. Mintz G, Enriquez RD, Mercado U, Robles EJ, Jimenez FJ, Gutierrez G. Intravenous methylprednisolone pulse therapy in severe ankylosing spondylitis. Arthritis Rheum 1981;24:734-6.

31. Peters ND, Ejstrup L. Intravenous methylprednisolone pulse therapy in ankylosing spondylitis. Scand J Rheumatol 1992;21:134-8.
32. Maksymowych WP, Jhangri GS, Leclercq S, Skeith K, Yan A, Russell AS. An open study of pamidronate in the treatment of refractory ankylosing spondylitis. J Rheumatol 1998;25:714-7.
33. Steven MM, Morrison M, Sturrock RD. Penicillamine in ankylosing spondylitis: a double blind placebo controlled trial. J Rheumatol 1985;12:735-7.
34. Breban M, Gombert B, Amor B, Dougados M. Efficacy of thalidomide in the treatment of refractory ankylosing spondylitis. Arthritis Rheum 1999;42:580-1.
35. Brandt J, Haibel H, Cornely D, et al. Successful treatment of active ankylosing spondylitis with the anti-tumor necrosis factor alpha monoclonal antibody infliximab. Arthritis Rheum 2000;43:1346-52.
36. Gladman DD, Farewell VT, Nadeau C. Clinical indicators of progression in psoriatic arthritis: multivariate relative risk model. J Rheumatol 1995;22:675-9.
37. Gladman DD, Farewell VT. The role of HLA antigens as indicators of disease progression in psoriatic arthritis. Multivariate relative risk model. Arthritis Rheum 1995;38:845-50.
38. Jones G, Crotty M, Brooks P. Interventions for psoriatic arthritis. Cochrane Database Syst Rev 2000;Cd000212.
39. Chang DJ. A survey of drug effectiveness and treatment choices in psoriatic arthritis [abstract]. Arthritis Rheum 1999;42 (suppl):S372.
40. Farr M, Kitas GD, Waterhouse L, Jubb R, Felix Davies D, Bacon PA. Sulphasalazine in psoriatic arthritis: a double-blind placebo-controlled study. Br J Rheumatol 1990;29:46-9.
41. Fraser SM, Hopkins R, Hunter JA, Neumann V, Capell HA, Bird HA. Sulphasalazine in the management of psoriatic arthritis. Br J Rheumatol 1993;32:923-5.
42. Gupta AK, Grober JS, Hamilton TA, et al. Sulfasalazine therapy for psoriatic arthritis: a double blind, placebo controlled trial. J Rheumatol 1995;22:894-8.
43. Combe B, Goupille P, Kuntz JL, Tebib J, Liote F, Bregeon C. Sulphasalazine in psoriatic arthritis: a randomized, multicentre, placebo-controlled study. Br J Rheumatol 1996;35:664-8.
44. Clegg DO, Reda DJ, Mejias E, et al. Comparison of sulfasalazine and placebo in the treatment of psoriatic arthritis. A Department of Veterans Affairs Cooperative Study. Arthritis Rheum 1996;39:2013-20.
45. Black RL, O'Brien WM, Van Scott EJ, Auerbach R, Eisen AZ, Bunim JJ. Methotrexate therapy in psoriatic arthritis. JAMA 1964;189:743-47.
46. Willkens RF, Williams HJ, Ward JR, et al. Randomized, double-blind, placebo controlled trial of low-dose pulse methotrexate in psoriatic arthritis. Arthritis Rheum 1984;27:376-81.
47. Zachariae H, Zachariae E. Methotrexate treatment of psoriatic arthritis. Acta Derm Venereol (Stockh) 1987;67:270-3.
48. Espinoza LR, Zakraoui L, Espinoza CG, et al. Psoriatic arthritis: clinical response and side effects to methotrexate therapy. J Rheumatol 1992;19:872-7.
49. Pigatto PD, Gibelli E, Ranza R, Rossetti A. Methotrexate in psoriatic polyarthritis. Acta Derm Venereol (Stockh) 1994;186 (suppl):114-5.

50. Abu Shakra M, Gladman DD, Thorne JC, Long J, Gough J, Farewell VT. Longterm methotrexate therapy in psoriatic arthritis: clinical and radiological outcome. J Rheumatol 1995;22:241-5.
51. Gupta AK, Matteson EL, Ellis CN, et al. Cyclosporine in the treatment of psoriatic arthritis. Arch Dermatol 1989;125:507-10.
52. Steinsson K, Jonsdottir I, Valdimarsson H. Cyclosporin A in psoriatic arthritis: an open study. Ann Rheum Dis 1990;49:603-6.
53. Salvarani C, Macchioni P, Boiardi L, et al. Low dose cyclosporine A in psoriatic arthritis: relation between soluble interleukin 2 receptors and response to therapy. J Rheumatol 1992;19:74-9.
54. Wagner SA, Peter RU, Adam O, Ruzicka T. Therapeutic efficacy of oral low-dose cyclosporin A in severe psoriatic arthritis. Dermatology 1993;186:62-7.
55. Riccieri V, Sili Scavalli A, Spadaro A, Bracci M, Taccari E, Zoppini A. Short-term "cyclosporin A" therapy for psoriatic arthritis. Acta Derm Venereol (Stockh) 1994;186 (suppl):94-5.
56. Porzio F, Antonelli M, Antonelli S, et al. Cyclosporin A in the long-term treatment of psoriatic arthritis [letter]. Br J Rheumatol 1996;35:1331.
57. Mahrle G, Schulze HJ, Brautigam M, et al. Anti-inflammatory efficacy of low-dose cyclosporin A in psoriatic arthritis. A prospective multicentre study. Br J Dermatol 1996;135:752-7.
58. Salvarani C, Macchioni P, Olivieri I, et al. A comparison of cyclosporine, sulfasalazine, and symptomatic therapy in the treatment of psoriatic arthritis [abstract]. Ann Rheum Dis 2000;59 (suppl 1):S56.
59. Spadaro A, Riccieri V, Sili Scavalli A, Sensi F, Taccari E, Zoppini A. Comparison of cyclosporin A and methotrexate in the treatment of psoriatic arthritis: a one-year prospective study. Clin Exp Rheumatol 1995;13:589-93.
60. Spadaro A, Taccari E, Mohtadi B, Riccieri V, Sensi F, Zoppini A. Life-table analysis of cyclosporin A treatment in psoriatic arthritis: comparison with other disease-modifying antirheumatic drugs. Clin Exp Rheumatol 1997;15:609-14.
61. Mazzanti G, Coloni L, De Sabbata G, Paladini G. Methotrexate and cyclosporin combined therapy in severe psoriatic arthritis. A pilot study. Acta Derm Venereol (Stockh) 1994;186 (suppl):116-7.
62. Espinoza LR, Cuellar ML, Mendez E, Angulo J. Combination therapy of methotrexate (MTX) and cyclosporin-A (CYC-A) in refractory psoriatic arhtritis (PsA) [abstract]. Arthritis Rheum 1998;41 (suppl):S335.
63. Seideman P, Fjellner B, Johannesson A. Psoriatic arthritis treated with oral colchicine. J Rheumatol 1987;14:777-9.
64. McKendry RJ, Kraag G, Seigel S, al Awadhi A. Therapeutic value of colchicine in the treatment of patients with psoriatic arthritis. Ann Rheum Dis 1993;52:826-8.
65. Hopkins R, Bird HA, Jones H, et al. A double-blind controlled trial of etretinate (Tigason) and ibuprofen in psoriatic arthritis. Ann Rheum Dis 1985;44:189-93.
66. Klinkhoff AV, Gertner E, Chalmers A, et al. Pilot study of etretinate in psoriatic arthritis. J Rheumatol 1989;16:789-91.
67. Carette S, Calin A, McCafferty JP, Wallin BA. A double-blind placebo-controlled study of auranofin in patients with psoriatic arthritis. Arthritis Rheum 1989;32:158-65.

68. Palit J, Hill J, Capell HA, et al. A multicentre double-blind comparison of auranofin, intramuscular gold thiomalate and placebo in patients with psoriatic arthritis. Br J Rheumatol 1990;29:280-3.
69. Dorwart BB, Gall EP, Schumacher HR, Krauser RE. Chrysotherapy in psoriatic arthritis. Efficacy and toxicity compared to rheumatoid arthritis. Arthritis Rheum 1978;21:513-5.
70. Huckins D, Felson DT, Holick M. Treatment of psoriatic arthritis with oral 1,25-dihydroxyvitamin D3: a pilot study. Arthritis Rheum 1990;33:1723-7.
71. Levy J, Paulus HE, Barnett EV, Sokoloff M, Bangert R, Pearson CM. A double-blind controlled evaluation of azathioprine treatment in rheumatoid arthritis and psoriatic arthritis [abstract]. Arthritis Rheum 1972;15:116-7.
72. Price R, Gibson T. D-penicillamine and psoriatic arthropathy [letter]. Br J Rheumatol 1986;25:228.
73. Peeters AJ, Dijkmans BA, van der Schroeff JG. Fumaric acid therapy for psoriatic arthritis. A randomized, double-blind, placebo-controlled study [letter]. Br J Rheumatol 1992;31:502-4.
74. Gladman DD, Blake R, Brubacher B, Farewell VT. Chloroquine therapy in psoriatic arthritis. J Rheumatol 1992;19:1724-6.
75. Antoni C, Dechant C, Lorenz H-M, et al. Successful treatment of psoriatic arthritis with infliximab in a MRI controlled study [abstract]. Ann Rheum Dis 2000;59 (suppl 1):200.
76. Yazici Y, Erkan D, Lockshin MD. Etanercept in the treatment of severe, resistant psoriatic arthritis-one year follow-up [abstract]. Ann Rheum Dis 2000;59 (suppl 1):198.
77. Mease PJ, Goffe BS, Metz J, VanderStoep A, Finck B, Burge DJ. Etanercept in the treatment of psoriatic arthritis and psoriasis: a randomised trial. Lancet 2000;356:385-90.
78. Dougados M, Maetzel A, Mijiyawa M, Amor B. Evaluation of sulphasalazine in the treatment of spondyloarthropathies. Ann Rheum Dis 1992;51:955-8.

Chapter 3

A brief history of spa therapy

A van Tubergen, Sj van der Linden

Annals of the Rheumatic Diseases 2002; 61:273-5

Abstract

Bathing in thermal water has an impressive history and continuing popularity. In this paper a brief overview on the use of water in medicine over the centuries is given.

Nomenclature

The word "spa" may be derived from the Walloon word "espa" meaning fountain.[1,2] This, in turn, came from the name of the Belgian town Spa, where in the 14th century a curative, thermal spring was discovered. Spa may also originate from the Latin word "spagere" (to scatter, sprinkle, moisten) or may be an acronym of the Latin phrase "sanitas per aquas" (health through water).[2] In Britain, the word spa is still used, whereas in the rest of Europe the term "thermal waters" is preferred. Bathing in thermal water for therapeutic purposes has several descriptions (for example, taking the waters, balneotherapy, spa therapy, hydrotherapy), which will all be used throughout this paper, and are more or less interchangeable.

Ancient Greece and the Roman Empire

Taking the waters used to be a popular treatment modality for a wide range of diseases in classical times. The Greeks preferred baths in fresh water from natural resources, although bathing in the sea (thalassotherapy) was also applied. Initially, bathing was confined to the more wealthy people in private baths, but soon public baths were opened. The baths were considered sacred places and were dedicated to several deities.[2-4]

In Homeric times, bathing was primarily used for cleansing and hygienic purposes. By the time of Hippocrates (460-370 BC), bathing was considered more than a simple hygienic measure; it was healthy and beneficial for most diseases.[3] Hippocrates proposed the hypothesis that the cause of all diseases lay in an imbalance of the bodily fluids. To regain the balance a change of habits and environment was advised, which included bathing, perspiration, walking, and massages.[5] The baths were often combined with sports and education, the precursors of the gymnasium.

Influenced by the Greeks, the Romans built their own thermal baths at mineral and thermal springs. A military presence was often the key to development of such a spa resort.[3] Spas served not only for recuperation of wounded soldiers but also as rest and recreation centers for healthy soldiers. In contrast with the Greeks, who took the waters after intensive physical exercises, the Romans considered the baths more important than the gymnastics alone.[4] Besides cleansing, exercises, socializing, relaxation, and worship, medical treatment was also applied extensively. Spa treatment consisted of application of water to afflicted parts of the body, immersion of the whole body in the water (especially for rheumatic and urogenital diseases), and drinking excessive quantities of water.[3]

Asclepiades (c 124 BC), a Greek physician who practiced in Rome, introduced general hydrotherapy and drinking cures as treatments.[3] He

recommended bathing for both therapeutic and preventative purposes. Pliny the Elder (AD 23-79) assigned different properties and indications for cure to different types of waters.[3,6] Galen (AD 131-201) also advocated the use of water for the treatment of a variety of diseases.[6] He preferred cold water, a concept that is reconsidered periodically throughout the following ages.

In Rome three different types of baths developed: baths at home (balnea), private baths (balnea privata), and public baths (balnea publica) that were run by the state.[4] With the introduction of the aqueducts, the public baths later developed into huge and impressive edifices (thermae) with a capacity for thousands of people. During the heyday of the Roman bathing culture, the inhabitants of Rome used 1400 liters of water per person per day, mainly for bathing.[2,5] The Roman legions, far away from their home land, built their own baths at mineral and thermal springs in the newly conquered lands.[2,3] Examples are found all over Europe.

Throughout the years the Roman bathing culture gradually changed towards a place for relaxation and pleasure, rather than for medical treatment, although this was still provided. The Romans preferred to use the baths and very hot waters for renewing their appetites and thirst, and the baths became, rather, centers for various sexual practices.[2,6] Deterioration of morals became manifest, the hygienic and medical indications for bathing disappeared, and baths as a haunt for pleasure ruled.[4]

The dark and middle ages

With the fall of the Roman Empire in 476 and the rise of Christianity, the bathing culture fell into disrepute and bathing was officially prohibited. Faith in cure through worship and praying was regarded as more important than a medicinal bath.[5] Baths were redeveloped as churches, although some remained available for the aristocrats who were not affected by the church's decrees.[6] The aversion to bathing remained for many centuries. People abstained from bathing as long as possible, sometimes for years. From the 13th century onwards, baths gradually became into re-use, particularly in southern Europe under the influence of the Moors. Public baths were rebuilt and the entrance was usually free. The baths were often crowded and people bathed for hours, sometimes days in the same bath. Blood letting, enemas, and drinking cures (up to 10 liters a day) were prescribed, although relaxation and pleasure were most often the reasons for bathing.[5]

Renaissance

In the 16th century, the image of the public baths again deteriorated in many countries, which lead to closure of many public baths. They were considered to be a source of contagious diseases such as syphilis, plague, and leprosy, and

the baths became dangerous meeting places of political and religious dissidents.[5] In addition, due to a shortage of firewood, public baths became more expensive for a population that had already become impoverished by many wars. Nevertheless, the gentry continued to visit the baths, although they preferred to go to baths from natural sources with warm, mineral water instead of the public baths.

Taking the waters was now no longer a spontaneous activity, but it was increasingly prescribed under medical direction.[7] Several famous Italian doctors recovered lost texts on medical treatment from the ancient world, and the value of balneology as a therapeutic modality was reconsidered.[8] By this time, the first attempts to analyze the waters for their mineral components were made, although the results were often controversial.[8] It was equally important to recognize the quality of each mineral and its effect on the body, as to know which parts of the body might be influenced by taking the waters.[8] In 1553 an encyclopedic work, *De balneis omniae qua extant,* was published, containing an overview of ancient and modern literature on the use of medicinal water. In 1571, Bacci published *De thermis*, in which he taught the art of the baths from Galen and the Aristotelians. According to Bacci, taking the waters was not a matter of empiricism, but a sound discipline with its own rationale, institutes, and doctrine, which the learned physician alone was qualified to understand.[8] Minardo published in 1594 a compendium on the two baths of Caldiero in Verona. The first bath was used for drinking and bathing, the second was used by bathers with skin conditions, for bathing of animals, and for washing off therapeutic mud.[8] Seventy-eight conditions that might benefit from these baths were listed. The treatments consisted of drinking cures, bathing, purging, and application of mud. It was advised to follow this type of treatment for 15 days, and repeat it every year. According to Bacci, essential to the cure was a quiet orderly life in pleasant surroundings with good food and wine, and a maximum of comfort.[8] Therefore, he argued, the baths would do no good to the poor. Other, practical obstacles also restrained the poor from attending the baths: there was no time for leisure and the baths and mud were usually not free.[3,8]

The new bathing culture that had developed in Italy gradually spread over other parts of Europe, and was in particular popular with the elite. The development of spa treatment north of the Alps was mainly provided by the Paracelsians.[9] By the turn of the 17th century, many spas were rediscovered in France. Two types of spas existed: hot springs for drinking and bathing, and cold springs for drinking cures only.[9] Taking the waters in French spas was a serious activity and quite sober. Doctors created centers for treatment, not for leisure. Much attention was paid to purging, drinking cures, eating well-balanced diets, and bathing. In the afternoons some indoor leisure activities were provided. Late in the afternoon, people walked about on the promenade, and went to bed early in the evening.[9] This was in contrast with many other European countries, where in the evenings diverse leisure activities were offered such as theatre and dance.

The 19th and 20th century

Around 1800 interest in the bathing culture grew. Further attempts to analyze the mineral water were made, aiming at improving its use in medicine, and at preparing mixtures of water identical to those mineral waters famous for their curative properties.[10] Doctors were convinced that for each disease Mother Nature possessed an appropriate medicinal spring, which could be discovered through chemical analysis of the waters.[5] Priessnitz and Kneipp further developed the principles of balneotherapy (medicinal use of thermal water) and hydrotherapy (immersion of the body in thermal water for therapeutic purposes).[5,11] Individual treatments were prescribed, based on the composition and temperature of the water. Also, combinations of treatments were developed consisting of hot and cold baths, herbal baths, mud packs, active physical exercises, massages, and diets. Kneipp advocated a holistic approach to the treatment of a disease.[11] In contrast with the spa resorts, which aimed at the elite, Kneipp directed his attentions to the common man.[5]

The use of mineral waters and the development of hotels and guesthouses at the springs became prevalent throughout Europe and North America.[6,12] Every spa resort had its own theatre, casino, and promenades besides the bathing buildings. In Britain, Germany, Austria, and Belgium much importance was attached to ostentation. Grand hotels arose with casinos and dancing establishments surrounding the spa resorts. The spa resorts became a meeting center for the elite but also a place of creativity for painters, writers, and composers.[2] The baths were again crowded. Baden Baden (Germany) became the most glamorous resort in continental Europe. It was the place to see and be seen.[2]

However, in Britain use of the spa declined. The English spa resorts were run by amateurs, and the medical hydrology was poorly organized.[13,14] The resorts aimed more at pleasure, rather than medical treatment, and were exploited by estate developers with commercial interests.[14] Competition from seaside and foreign resorts, and an economic depression in the 1930s lead to a further decline. Eventually, spa therapy was excluded from the National Health Service, which meant that many spa resorts in Britain closed down.[13]

After the second world war and with the rise in welfare, spa treatment became available for the common man in many European countries, mainly owing to reimbursement by state medical systems.[2,5,15] Other activities and new treatments were introduced, and balneology, hydrotherapy, and physiotherapy underwent major developments.

In the past decades, a large change in the use of mineral water for the treatment of several diseases has taken place in continental Europe.[6] The medical significance of bathing is now acknowledged, especially by many rheumatologists and dermatologists, and this aspect is considered more important for a number of spa resorts than prestige and leisure. Bathing is usually combined with many other treatments such as physical exercises, hydrotherapy, and mud packs. The spa resorts are differentiated according to their location (for example, seaside, mountain area) and the chemical

composition of their mineral water (for example, sulphurous, bicarbonated, or sulphated).[16-18] Each spring has its own characteristics and related therapeutic properties.[17] However, a substantial number of spa resorts also direct more attention towards leisure. Steam baths, saunas, whirlpools, and solariums are standard equipment of many such spa resorts, with the main objective to relax and strengthen the body and mind, and prevent development of disease. In Britain, a revival of the spa culture may be expected, with the re-opening of the hot springs in Bath in 2002. This spa will offer facilities for medical treatment, but, in addition cater for a growing number of so-called health tourists, who combine their holidays with an investment in well-being.

Scientific evidence for efficacy of spa therapy

Despite the popularity of spa therapy, reported scientific evidence for its efficacy is sparse. A decade ago, Heywood reviewed well-documented records on spa treatment for lead poisoning in the 18th and 19th century in Bath.[19] Paralysis occurring as a result of chronic lead intoxication (colica pictonum) was a common problem in those days owing to the widespread use of lead in household ware, cosmetics, food colorants, wine, and salts for medicinal use. Already at the beginning of the 16th century, Bath was famous for curing paralysis even in those patients who were regarded as incurable. The treatment consisted of bathing, drinking cures, diet, and purges. Patients admitted to the Bath Hospital came from all over England, and often had already been treated for their paralysis elsewhere, without success. However, many of these presumed incurable patients were cured after their (months) stay in Bath.

An example can be found in the comparison of medical records of Bath and Exeter hospitals between 1762 and 1767.[19] During these 5 years, 285 patients with *colica pictonum* were admitted in Exeter and 281 patients in Bath. Seventy-three percent of the patients from Exeter were cured or improved, whereas the figure was 93% from Bath. Moreover, the group in Bath included some 80 patients referred from Exeter who had not been cured by treatment in Exeter.[19] From 1760 to 1879, 3,377 patients were admitted in Bath for paralysis due to lead intoxication. Forty-five percent were cured and 93% had at least improved.[19]

The high cure rates for paralysis by spa therapy in Bath may be attributed to several factors. Sitting in warm water produces diuresis, with increased excretion of sodium, potassium, calcium, and also lead.[19,20] Also the good food, exercises, removal from the source of lead, and the large quantities of water rich in calcium and iron contributed to the success of spa therapy in Bath.[19]

In the past decade several randomized controlled trials have studied the effects of spa therapy in rheumatoid arthritis[21-26] and osteoarthritis.[23,27-30] Patients were randomly allocated to receive spa therapy or sham/no therapy. The authors of a recent systematic review on the effects of spa therapy in rheumatic arthritis and osteoarthritis stated that a definite judgment about its efficacy is

impossible because of methodological flaws in these studies.[31] Overall, the results showed positive effects lasting for 3 to 9 months. Recently, a randomized controlled trial has shown that spa therapy is clearly effective in ankylosing spondylitis.[32] Two intervention groups followed a 3-week course of spa therapy at two different spa resorts, and were compared with a control group who stayed at home and continued standard treatment consisting of anti-inflammatory drugs and weekly group physical therapy. Significant improvements in function, pain, global well-being, and morning stiffness were found for both intervention groups until 9 months after spa therapy.[32]

Conclusion

Throughout the ages the interest in the use of water in medicine has fluctuated from century to century and from nation to nation. The (medical) world has viewed it with different opinions, from very enthusiastic to extremely critical, and from beneficial to harmful. Today, spa therapy is receiving renewed attention from many medical specialties and health tourists, and having a revival. However, the exact therapeutic potential of spa therapy still remains largely unknown. Better and more profound scientific evidence for its efficacy is therefore warranted, in particular for its effect on the musculoskeletal system.

References

1. Calin A. Royal National Hospital for Rheumatic Diseases--Bath. A 250th birthday party [editorial]. J Rheumatol 1988;15:733-4.
2. Croutier AL. Taking the waters: spirit, art, sensuality. New York: Abbeville Publishing Group 1992.
3. Jackson R. Waters and spas in the classical world. Med Hist Suppl 1990;10:1-13.
4. Schadewalt H. The history of Roman bathing culture [in Dutch]. Integraal 1989;4:25-35.
5. Looman J, Pillen H. The development of the bathing culture [in Dutch]. Integraal 1989;4:7-24.
6. Routh HB, Bhowmik KR, Parish LC, Witkowski JA. Balneology, mineral water, and spas in historical perspective. Clin Dermatol 1996;14:551-4.
7. Porter R. The medical history of waters and spas. Introduction. Med Hist Suppl 1990;10:vii-xii.
8. Palmer R. "In this our lightye and learned tyme": Italian baths in the era of the Renaissance. Med Hist Suppl 1990;10:14-22.
9. Brockliss LW. The development of the spa in seventeenth-century France. Med Hist Suppl 1990;10:23-47.
10. Coley NG. Physicians, chemists and the analysis of mineral waters: "the most difficult part of chemistry". Med Hist Suppl 1990;10:56-66.
11. van Wijk R. From the old towards the new Kneipp-cure [in Dutch]. Integraal 1989;4:36-38.
12. Benedetto AV, Millikan LE. Mineral water and spas in the United States. Clin Dermatol 1996;14:583-600.
13. Harley D. A sword in a madman's hand: professional opposition to popular consumption in the waters literature of southern England and the Midlands, 1570-1870. Med Hist Suppl 1990;10:48-55.
14. Cantor D. The contradictions of specialization: rheumatism and the decline of the spa in inter-war Britain. Med Hist Suppl 1990;10:127-44.
15. Tsankov NK, Kamarashev JA. Spa therapy in Bulgaria. Clin Dermatol 1996;14:675-8.
16. Titzmann T, Balda BR. Mineral water and spas in Germany. Clin Dermatol 1996;14:611-3.
17. Andreassi L, Flori L. Mineral water and spas in Italy. Clin Dermatol 1996;14:627-32.
18. Vassileva S. Mineral water and spas in Bulgaria. Clin Dermatol 1996;14:601-5.
19. Heywood A. A trial of the Bath waters: the treatment of lead poisoning. Med Hist Suppl 1990;10:82-101.
20. O'Hare JP, Heywood A, Summerhayes C, et al. Observations on the effect of immersion in Bath spa water. BMJ 1985;291:1747-51.
21. Sukenik S, Neumann L, Buskila D, et al. Dead Sea bath salts for the treatment of rheumatoid arthritis. Clin Exp Rheumatol 1990;8:353-7.

22. Sukenik S, Buskila D, Neumann L, et al. Sulphur bath and mud pack treatment for rheumatoid arthritis at the Dead Sea area. Ann Rheum Dis 1990;49:99-102.
23. Elkayam O, Wigler I, Tishler M, et al. Effect of spa therapy in Tiberias on patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 1991;18:1799-803.
24. Sukenik S, Neumann L, Flusser D, Kleiner Baumgarten A, Buskila D. Balneotherapy for rheumatoid arthritis at the Dead Sea. Isr J Med Sci 1995;31:210-4.
25. Hall J, Skevington SM, Maddison PJ, Chapman K. A randomized and controlled trial of hydrotherapy in rheumatoid arthritis. Arthritis Care Res 1996;9:206-15.
26. Franke A, Reiner L, Pratzel HG, Franke T, Resch KL. Long-term efficacy of radon spa therapy in rheumatoid arthritis--a randomized, sham-controlled study and follow-up. Rheumatology (Oxford) 2000;39:894-902.
27. Sylvester KL. Investigation of the effect of hydrotherapy in the treatment of osteoarthritic hips. Clin Rehabil 1989;4:223-8.
28. Green J, McKenna F, Redfern EJ, Chamberlain MA. Home exercises are as effective as outpatient hydrotherapy for osteoarthritis of the hip. Br J Rheumatol 1993;32:812-5.
29. Wigler I, Elkayam O, Paran D, Yaron M. Spa therapy for gonarthrosis: a prospective study. Rheumatol Int 1995;15:65-8.
30. Nguyen M, Revel M, Dougados M. Prolonged effects of 3 week therapy in a spa resort on lumbar spine, knee and hip osteoarthritis: follow-up after 6 months. A randomized controlled trial. Br J Rheumatol 1997;36:77-81.
31. Verhagen AP, de Vet HC, de Bie RA, et al. Balneotherapy for rheumatoid arthritis and osteoarthritis (Cochrane Review). In: The Cochrane Library, Issue 1. Oxford: Update Software 2001.
32. van Tubergen A, Landewé R, van der Heijde D, et al. Combined spa-exercise therapy is effective in patients with ankylosing spondylitis: a randomized controlled trial. Arthritis Rheum 2001;45:430-8.

Chapter 4

Combined spa-exercise therapy is effective in patients with ankylosing spondylitis: a randomized controlled trial

A van Tubergen, R Landewé, D van der Heijde, A Hidding, N Wolter, M Asscher, A Falkenbach, E Genth, H Goei Thè, Sj van der Linden

Arthritis and Rheumatism [Arthritis Care and Research] 2001; 45:430-8

Abstract

Objective. To determine the efficacy of combined spa-exercise therapy in addition to standard treatment with drugs and weekly group physical therapy in patients with ankylosing spondylitis (AS).

Methods. A total of 120 Dutch outpatients with AS were randomly allocated into three groups of 40 patients each. Group 1 (mean age 48±10 years; male:female 25:15) was treated in a spa resort in Bad Hofgastein, Austria; group 2 (mean age 49±9 years; male:female 28:12) in a spa resort in Arcen, The Netherlands. The control group (mean age 48±10 years; male:female 34:6) stayed at home, and continued their usual drug treatment and weekly group physical therapy during the intervention weeks. Standardized spa-exercise therapy of 3 weeks duration consisted of group physical exercises, walking, correction therapy (lying supine on a bed), hydrotherapy, sports, and visits to either the Gasteiner Heilstollen (Austria) or sauna (Netherlands). After spa-exercise therapy all patients followed weekly group physical therapy for another 37 weeks. Primary outcomes were functional ability, patient's global well-being, pain, and duration of morning stiffness, aggregated in a pooled index of change (PIC).

Results. Analysis of variance showed a statistically significant time-effect ($P<0.001$) and time-by-treatment-interaction ($P=0.004$), indicating that the three groups differed over time with respect to the course of the PIC. Four weeks after start of spa-exercise therapy, the mean difference in PIC between group 1 and controls was 0.49 (95% confidence interval [CI] 0.16-0.82, $P=0.004$), and between group 2 and controls 0.46 (95% CI 0.15-0.78, $P=0.005$). At 16 weeks, the difference between group 1 and controls was 0.63 (95% CI 0.23-1.02, $P=0.002$) and between group 2 and controls 0.34 (95% CI -0.05,0.73, $P=0.086$). At 28 and 40 weeks more improvement was found for group 1 compared with controls ($P=0.012$ and $P=0.062$ respectively), but not for group 2 compared with controls.

Conclusion. In patients with AS, a 3-week course of combined spa-exercise therapy, in addition to drug treatment and weekly group physical therapy alone, provides beneficial effects. These beneficial effects may last for at least 40 weeks.

Introduction

Since ancient times spa therapy -bathing in thermal water- has been applied to patients with several rheumatic conditions, including ankylosing spondylitis (AS). Nowadays, spa therapy is usually offered in combination with other treatments, such as active exercise therapy, massages, or mud packs. Despite the long history and popularity of spa therapy, only a few randomized controlled trials on use of such therapy in patients with rheumatic diseases have been conducted.[1-8] The authors of a recent systematic review on the effects of spa therapy in rheumatic diseases stated that a definite judgment about efficacy is impossible because of methodological flaws in these studies.[9] The efficacy of spa therapy in AS has been sparsely investigated. In English literature, one uncontrolled pilot study of spa therapy in AS reported positive but short-term effects.[10]

AS is a chronic inflammatory disease that predominantly affects the spine and may cause serious functional impairment. The prevalence of AS is approximately 0.1% of the Caucasoid population. Treatment of AS includes use of anti-inflammatory drugs to reduce pain and stiffness. In addition, patients are advised to exercise daily and to engage in weekly group physical therapy to maintain mobility of the spine and peripheral joints.[11]

The present randomized controlled trial was designed to assess the efficacy of spa therapy combined with exercise therapy in addition to standard treatment with anti-inflammatory drugs and weekly group physical therapy alone in patients with AS. The primary hypothesis was that a 3-week course of combined spa-exercise therapy along with drug treatment and weekly group physical therapy is superior to drug treatment and weekly group physical therapy alone with respect to inflammatory signs of AS, functional ability, and global well-being.

Patients and methods

Patients

A total of 1,646 members of Dutch AS patient societies received information about the study by mail, and 332 responded (Figure 4.1). Patients were eligible for the study if they [1] fulfilled the modified New York criteria for AS,[12] [2] reported pain and stiffness or functional limitations for at least 3 months before entry, and [3] were able to stay away from home and work for 3 pre-planned consecutive weeks. Exclusion criteria were inability or unwillingness to participate in weekly group physical therapy; pregnancy; claustrophobia; severe co-morbidity of heart, lung, liver, or kidneys; and a diagnosis of AS more than 20 years ago. Radiographs of the sacroiliac joints were checked for sacroiliitis according to the New York criteria.

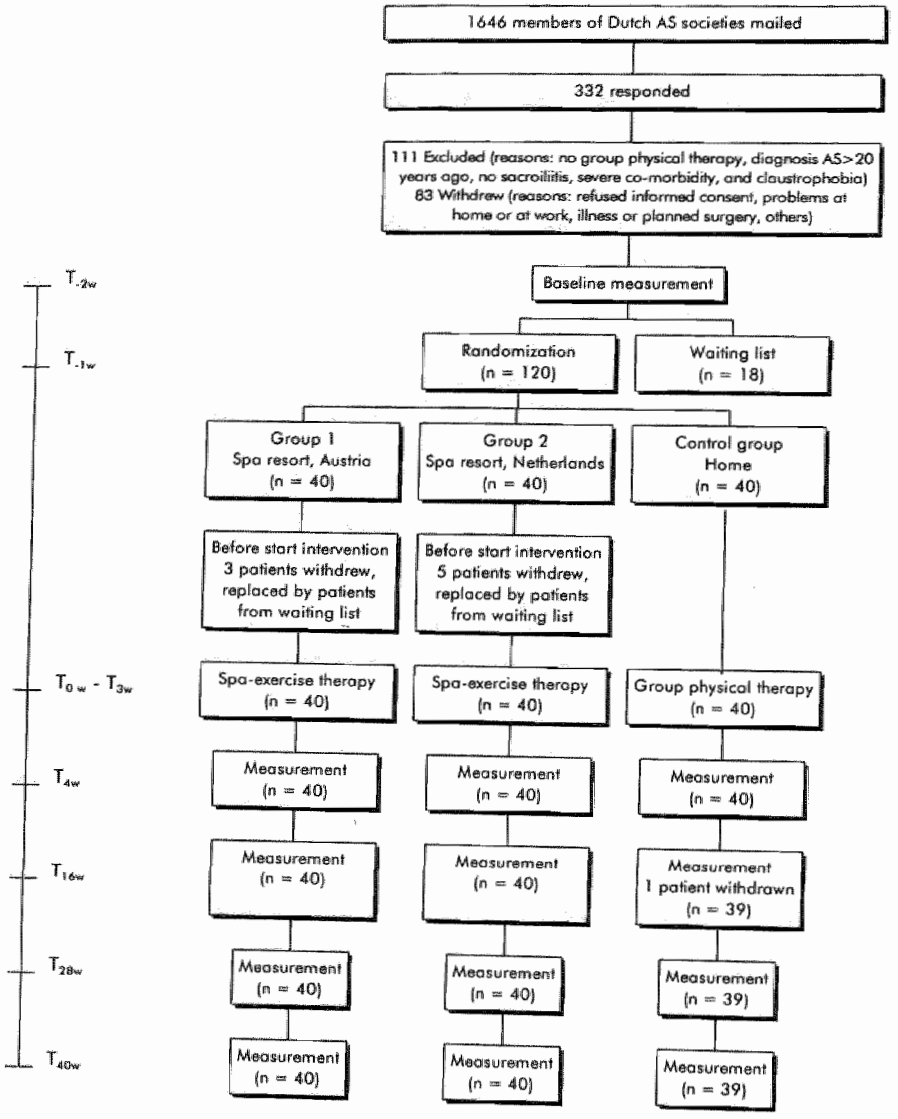

Figure 4.1 Flow chart of participants
AS = Ankylosing spondylitis

Of the 332 responding patients, 111 were not eligible, 83 declined to co-operate, and 138 patients signed informed consent and were placed consecutively on an inclusion list (Figure 4.1). One week before the intervention, the first 120 patients on this list were randomized with a computer-generated random-number list (prepared by a rheumatologist not further involved in the study) into three groups of 40 patients each: group 1 (spa therapy, Austria), group 2 (spa therapy, Netherlands), and a control group (home). The remaining 18 patients were placed on a waiting list. After randomization but before the start of the intervention, three patients from group 1 and five from group 2 withdrew for various reasons. They were randomly replaced by patients from the waiting list. The remaining 10 patients from the waiting list were excluded from the study.

Patients allocated to the control group were offered spa therapy at the end of the study in order to prevent withdrawal from the study and for ethical reasons. Permission to conduct the study was obtained from the medical ethical committee of the University Hospital of Maastricht.

Intervention

The intervention took place at two spa resorts: Bad Hofgastein, Austria (group 1), and Thermaalbad Arcen, The Netherlands (group 2), during 3 consecutive weeks in April 1999. Patients from group 1 traveled together to the spa resort by coach (12 hours), and stayed in the same hotel. Spa-exercise therapy was provided at the spa resort, and at the so-called Gasteiner Heilstollen. Patients from group 2 arrived at the spa resort either individually or by coach and stayed in the same hotel. Spa-exercise therapy was provided at the spa resort.

Patients in the intervention groups received spa-exercise therapy 5 days a week. Therapy programs were standardized for both spa resorts and were performed by trained physiotherapists not involved in the outcome assessment *and analyses of the study (Table 4.1)*. Every morning patients started with 1 hour of physical exercises, followed by 30 minutes of walking, and postural correction therapy by lying supine on bed (initially 14 minutes, but increasing daily by 2 minutes to a final period of 30 minutes a day). Every other afternoon, the patients in Bad Hofgastein visited the Gasteiner Heilstollen. The Gasteiner Heilstollen are former mine galleries with a climate characterized by temperatures from 38.0 to 41.5 °C, a humidity of 70% to 98%, and concentrations of radioactive radon in the air. Patients traveled by train to one of the treatment areas situated 2 kilometers inside the mountain, where they rested (naked) in the supine position on a bed for 1 hour. After the stay in the Heilstollen, the patients rested for 30 minutes at room temperature. Instead of visiting the Gasteiner Heilstollen, the patients in Arcen received a similar thermal treatment by visiting the sauna and the thermal baths. Either the Heilstollen or the sauna was visited a total of 10 times within a period of 3 weeks. In the Heilstollen, the radon progeny activity was measured in Working Level using the Instant Radon Progeny Meter (IRPM, Type TN-IR-21, Thomson & Nielsen, Canada).

Table 4.1. Treatments of the three study groups during the 3 intervention weeks

Group 1 Spa therapy Bad Hofgastein, Austria	Group 2 Spa therapy Arcen, The Netherlands	Control group Home
Morning - 1 hour group physical exercises - ½ hour walking - Postural correction therapy on bed (14 minutes to ½ hour)	Morning - 1 hour group physical exercises - ½ hour walking - Postural correction therapy on bed (14 minutes to ½ hour)	Once a week group physical therapy consisting of: - 1 hour group physical exercises - 1 hour sports - 1 hour hydrotherapy
Afternoon Alternately either: - 1h. visit to Gasteiner Heilstollen or: - ½ hour hydrotherapy - ½ hour bathing - 1 hour sports	Afternoon Alternately either: - 2 x 15 minutes to sauna - ½ hour bathing or: - ½ hour hydrotherapy - ½ hour bathing - 1 hour sports	

The intervention groups received spa therapy for 3 weeks, 5 days a week, plus an extra visit to the Heilstollen/sauna on a weekend day.

For the total 10-hour stay, patients were exposed to a cumulative dose of 0.536 Working Level Month (WLM). The other afternoons were spent with 30 minutes of intensive hydrotherapy and 30 minutes bathing in thermal water, followed by 1 hour of sports. Individual therapies were not allowed. During the weekends, patients were permitted to visit the thermal baths but were instructed not to exercise.

The control patients stayed at home and continued their usual activities and drug treatment and participated in weekly group physical therapy. Weekly group physical therapy consisted of 1 hour of physical exercises, 1 hour of sports, and 1 hour of hydrotherapy. After the intervention period, all patients from the three groups engaged in weekly group physical therapy sessions. During the intervention and the follow-up periods, all patients continued their usual drug treatment but were allowed to decrease or increase the amount of anti-inflammatory drugs based on present complaints during the study period. Another spa treatment was not permitted during the follow-up period.

Assessments of end points

In the absence of validated response criteria for AS, primary outcome measures were chosen on agreement by rheumatologists involved in the study during the design phase. Based on the preliminary core-set for physical therapy in AS, developed by the ASsessment in Ankylosing Spondylitis Working Group, measures that were clinically relevant and considered sensitive to change in the opinion of the investigators were included.[13] All outcome measures were elicited by self-assessment questionnaires.

Primary outcomes were functional ability (measured with Bath Ankylosing Spondylitis Functional Index [BASFI]),[14] patient's global well-being (measured on a 10 cm visual analog scale [VAS]), pain intensity (on a 10 cm VAS), and morning stiffness (in minutes). The BASFI contains 10 questions concerning activities of daily living and is scored on a 10 cm VAS, with anchors "easy" and "impossible" at either side. The mean of the items defines the final score. The scores on the BASFI, patient's global well-being, and pain intensity range from 0 (best) to 10 (worst).

Secondary outcomes were disease activity (measured with Bath Ankylosing Spondylitis Disease Activity Index [BASDAI]),[15] general health and function (measured with Health Assessment Questionnaire for Spondylarthropathies [HAQ-S]),[16] night pain (on a 10 cm VAS), quality of life (measured with Ankylosing Spondylitis Quality of Life questionnaire [ASQoL]),[17] and the intake of non-steroidal anti-inflammatory drugs (NSAIDs). The BASDAI consists of six questions answered on a VAS.[15] The questions are related to fatigue, back pain, pain and/or swelling of the peripheral joints, localized tenderness, and duration and severity of morning stiffness. The 10 cm horizontal VAS has the labels "none" (0) and "very severe" (10) at either end in the first five questions, and "0 hours" (0) and "2 or more hours" (10) for the duration of morning stiffness. The mean of the two questions on morning stiffness counts as one variable. The total score is calculated by taking the mean of the five items, ranging from 0 (best) to 10 (worst). The HAQ-S consists of eight subscales on health status and function

and is extended with five additional AS-specific items on function, divided into two subscales.[16] The questions are answered on Likert-formatted scales. The mean of all subscales defines the final score, ranging from 0 (best) to 3 (worst). The ASQoL is a new quality-of-life questionnaire specific to AS and was developed in co-operation with patients with AS.[17] The ASQoL contains 18 yes/no questions. Scores range from 0 to 18, with lower scores implying a better quality of life. The exact amount of NSAIDs taken in the previous week was registered and recalculated into standard dosages, equipotent to diclofenac.

All questionnaires, except for the newly developed ASQoL, are widely used instruments in the research of AS and have been shown to be valid, reliable, and sensitive to changes after active intervention.[14-23]

Spa-exercise therapy of 3 weeks duration took place from T_{0w} (start of spa therapy) to T_{3w} (3 weeks after start of spa therapy). The questionnaires were completed at baseline (2 weeks before spa therapy, T_{-2w}), and at 4 weeks (T_{4w}), 16 weeks (T_{16w}), 28 weeks (T_{28w}), and 40 weeks (T_{40w}) after the start of spa-exercise therapy.

Sample size

Data from the validation study of the BASFI were used to calculate the sample size.[14] In a group of patients who were following an inpatient course of intensive physiotherapy, the mean BASFI score changed with 1.1 ± 2.3 (20% improvement compared with baseline) on a 0 to 10 scale after 3 weeks, which we considered a minimum clinically important difference. Based on the assumption that combined spa-exercise therapy would be at least as effective as intensive physiotherapy, and that the mean BASFI would remain constant in the control group (mean change 0.0 ± 1.0), we calculated a sample size of 41 patients per group to be sufficient (two-sided $\alpha=0.05$; $\beta=0.20$) to detect a difference of at least 20% between one intervention group and the control group.

Analysis

The analyses were based on intention-to-treat. All returned questionnaires were checked for possible missing answers. If necessary, additional answers were obtained by telephone or by mail. In the few cases in which data were still missing, the instructions of the authors of the questionnaires were followed. If such information was lacking, the mean of the non-missing items of the patient or the group (in single-item questions) was filled in. Investigators not involved in the intervention of the groups performed the analyses.

The results of the primary outcomes were expressed in a pooled index of change (PIC).[24,25] For each component of the PIC, the change compared with baseline per time period was calculated for each patient. To obtain a standardized change score for each group, the mean change per time period of each group was divided by the pooled SD_{change} at T_{40w} of that instrument. The PIC was calculated as the unweighted mean of the four standardized scores per time period. Change in morning stiffness did not appear to be normally distributed,

and the change score was therefore logarithmically transformed before introducing it into the PIC.

In order to evaluate whether there was any overall effect of spa-exercise therapy compared with control therapy over time, analysis of variance for repeated measurements (general linear model) was performed, with the PIC as dependent variable and group allocation and time as factors. Post-hoc analyses for between-group differences were planned by doing Student's *t*-tests for unpaired observations. Non-normally distributed data, determined by probability plots (morning stiffness, pain at night, ASQoL, NSAIDs), were analyzed by Mann-Whitney U test. SPSS 10.0 software (Chicago, IL) was used for all analyses.

Results

Baseline characteristics of the groups are shown in Table 4.2. All characteristics except sex were well balanced among the groups. Relatively more men than women were allocated to the control group than to either intervention group. All patients completed 3 weeks of spa-exercise therapy; no adverse effects were reported. The mean attendance rate for each part of the intervention was 99% in both groups. One patient from the control group withdrew after 3 months, because he declined to co-operate. The last observation carried forward method was applied to this patient. After the follow-up period, 93% of the patients in the three groups still attended weekly group physical therapy.

Table 4.2. Baseline characteristics of the groups

	Group 1 Bad Hofgastein	Group 2 Arcen	Controls Home
Men/women	25/15	28/12	34/6
Age [years]*	48 (10)	49 (9)	48 (10)
Disease duration [years]*	11 (6)	12 (5)	10 (6)
Duration of complaints [years]*	19 (10)	19 (9)	15 (8)
NSAIDs [yes/no]†	34/6	33/7	36/4
Sulfasalazine [yes/no]	6/34	3/37	5/35
Uveitis [ever/never]	11/29	17/23	20/20
Inflammatory bowel disease [ever/never]	8/32	7/33	10/30
Psoriasis [ever/never]	4/36	5/35	4/36

*Mean (SD). †NSAIDs = Non-steroidal anti-inflammatory drugs. Disease duration is from time of diagnosis. Duration of complaints is from time complaints started.

Pooled index of change

Analysis of variance with adjustment for within correlation (repeated measurements) showed a statistically significant time-effect ($P<0.001$) and time-

by-treatment interaction (P=0.004), indicating that the three groups differed over time with respect to the course of the PIC. Post-hoc analyses demonstrated a significant time-by-treatment interaction for group 1 versus controls (P=0.005) and group 2 versus controls (P=0.022) but not for group 1 versus group 2 (P=0.132). Multivariate analysis of the PIC with sex as co-variate did not change the time-effect (P<0.001) and only slightly changed the time-by-treatment interaction (P=0.011).

Figure 4.2 shows the course of the PIC and individual variables over time. At T_{4w}, the mean (95% confidence interval [CI]) PIC was 0.56 (0.33-0.79) for group 1, 0.53 (0.32-0.74) for group 2, and 0.07 (-0.16,0.30) for the control group. The mean differences (95% CI) between group 1 and controls (0.49 [0.16-0.82]) and between group 2 and controls (0.46 [0.15-0.78]) were statistically significant (P=0.004 and P=0.005, respectively).

At T_{16w}, the PIC was 0.72 (0.46-0.98) for group 1, 0.43 (0.18-0.68) for group 2, and 0.09 (-0.20,0.38) for the control group. The difference between group 1 and controls (0.63 [0.23-1.02]) was statistically significant (P=0.002). The difference between group 2 and controls (0.34 [-0.05,0.73]) had lost statistical significance (P=0.086).

At T_{28w} and T_{40w}, the PIC gradually declined toward baseline values in the intervention groups. At T_{28w}, the difference between group 1 and controls was still statistically significant (0.48 [0.11-0.86], P=0.012), but at T_{40w}, the difference had just lost statistical significance (0.29 [-0.02,0.60], P=0.062).

Compared with group 2, more improvement was observed in group 1 at T_{16w}, T_{28w}, and T_{40w}, but the differences between group 1 and 2 were not statistically significant.

The course of the individual variables of the PIC is summarized in Table 4.3. A trend toward improvement in both intervention groups after spa-exercise therapy, with a gradual decrease toward baseline values at T_{40w}, was seen for all variables except morning stiffness. Maximum relative improvements (scores of group 1 or 2 minus those of controls) during follow-up (morning stiffness excluded) were 24% (group 1) and 19% (group 2) for BASFI, 24% (group 1) and 30% (group 2) for pain, and 33% (group 1) and 29% (group 2) for global well-being. At T_{40w}, group 1 still showed relative improvements of 12% in BASFI, 8% in pain, and 16% in global well-being.

Secondary outcomes

The results of the secondary outcome variables are summarized in Table 4.4. Most secondary outcome variables significantly improved in both intervention groups compared with controls after spa-exercise therapy, and a trend similar to that found in the primary outcomes was seen. On average, the maximum relative improvements achieved were 54% (group 1) and 35% (group 2) at T_{4w}, and 43% (group 1) and 16% (group 2) at T_{16w}. At T_{40w}, group 1 still showed a relative improvement of more than 10% in the BASDAI, HAQ-S, and pain at night, and up to 29% in the ASQoL.

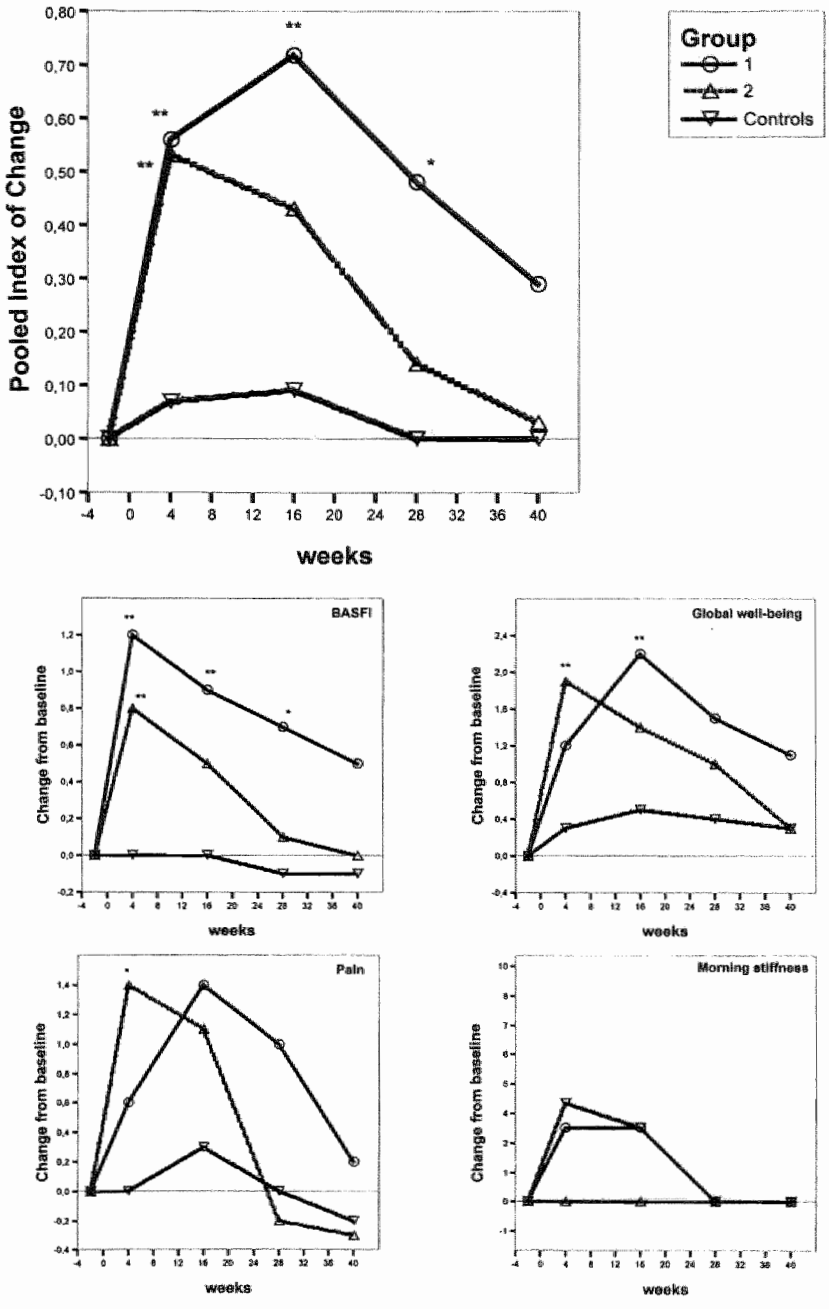

Figure 4.2 Pooled index of change (PIC) and individual variables of PIC
*$P<0.05$; **$P<0.01$.

Table 4.3 Primary outcomes

	Baseline value	Change from baseline			
Measure (range)	T_{-2w}	T_{4w}	T_{16w}	T_{28w}	T_{40w}
BASFI (0 – 10)					
Group 1	4.9 (1.8)	1.2 (1.4)**	0.9 (1.5)**	0.7 (1.5)*#	0.5 (1.3)
Group 2	4.3 (2.0)	0.8 (1.2)**	0.5 (1.1)	0.1 (1.2)	0.0 (1.1)
Control	4.2 (2.1)	0.0 (1.1)	0.0 (1.6)	-0.1 (1.7)	-0.1 (1.3)
Pain (0 – 10)					
Group 1	4.6 (2.5)	0.6 (2.4)	1.4 (2.7)	1.0 (2.9)	0.2 (2.5)
Group 2	4.6 (2.5)	1.4 (2.5)*	1.1 (2.7)	-0.2 (3.0)	-0.3 (2.8)
Control	4.8 (2.8)	0.0 (2.3)	0.3 (2.7)	0.0 (2.7)	-0.2 (2.1)
Global well-being (0 – 10)					
Group 1	5.1 (2.0)	1.2 (2.8)	2.2 (2.4)**	1.5 (2.6)	1.1 (2.6)
Group 2	5.4 (2.3)	1.9 (2.5)**	1.4 (3.0)	1.0 (2.8	0.3 (2.5)
Control	4.9 (2.5)	0.3 (2.9)	0.5 (3.0)	0.4 (2.7)	0.3 (2.8)
Morning stiffness (min)‡					
Group 1	30 (10;60)	3 (0;19)	3 (0;15)	0 (-1;15)	0 (-4;14)
Group 2	30 (15;60)	0 (-2;14)	0 (-5;14)	0 (-9;14)	0 (-10;15)
Controls	30 (10;60)	4 (0;10)	3 (-6;15)	0 (-15;10)	0 (-13;14)

Data are presented at baseline as mean (SD) and mean change (SD) from baseline. ‡Skewed data are presented as median (interquartile range) and median change (interquartile range). BASFI = Bath Ankylosing Spondylitis Functional Index. Positive changes imply improvement. *$P<0.05$ and **$P<0.01$ between intervention group and control; #$P<0.05$ between groups 1 and 2.

Discussion

The present randomized controlled trial indicates that a 3-week course of spa therapy combined with exercise therapy along with standard treatment with drugs and weekly group physical therapy has significant long-term benefits compared with standard treatment with drugs and weekly group physical therapy alone in patients with AS. The PIC in group 1 remained significantly different from that of controls until 28 weeks after start of the intervention and had just lost significance at 40 weeks ($P=0.062$). In group 2, significant improvement compared with controls had disappeared after 16 weeks, suggesting that the beneficial effects of spa-exercise therapy provided in Arcen, The Netherlands, were less persistent. In the primary outcomes, more improvement was observed in group 1 compared with group 2, but these differences were not statistically significant. Because this study did not have sufficient statistical power to detect such differences, a solid conclusion about differences in efficacy between the two intervention groups can not be drawn.

Table 4.4 Secondary outcomes

Measure (range)	Baseline value	Change from baseline			
	T_{-2w}	T_{4w}	T_{16w}	T_{28w}	T_{40w}
BASDAI (0 – 10)					
Group 1	4.7 (1.8)	1.0 (1.9)	1.9 (1.9)**	1.5 (1.6)*#	1.0 (1.6)
Group 2	5.1 (2.0)	1.2 (1.7)*	1.3 (1.9)	0.7 (1.8)	0.7 (1.5)
Control	4.5 (2.0)	0.3 (1.7)	0.6 (2.1)	0.8 (1.7)	0.4 (1.5)
HAQ-S (0 – 3)					
Group 1	0.92 (0.45)	0.18 (0.35)**	0.19 (0.31)**	0.15 (0.36)*	0.11 (0.38)
Group 2	0.87 (0.52)	0.20 (0.34)**	0.13 (0.36)*	0.06 (0.42)	0.06 (0.33)
Control	0.85 (0.50)	-0.05 (0.22)	-0.01 (0.24)	-0.06 (0.39)	-0.01 (0.30)
Pain at night (0 – 10)‡					
Group 1	2.9 (1.2;6.7)	1.2 (0.1;2.3)*	1.3 (0.1;2.8)*	0.4 (-1.0;3.1)	0.1 (-1.8;1.9)
Group 2	3.7 (1.4;7.0)	0.8 (0.0;2.2)*	0.7 (-0.3;2.2)	-0.1 (-1.9;1.2)	-0.2 (-1.5;0.7)
Control	3.9 (1.9;6.7)	-0.5 (-1.5;2.1)	0.1 (-1.8;1.8)	-0.3 (-1.8;2.4)	-0.3 (-2.3;0.7)
ASQoL (0 – 18)‡					
Group 1	7.0 (5.0;10.8)	2.0 (0.3;4.8)**	3.0 (0.0;4.0)**	2.5 (0.25;4.0)**	2.0 (-0.8;3.8)*
Group 2	9.0 (4.0;12.0)	1.3 (-1.0;3.8)	1.0 (0.0;3.8)*	1.0 (-0.8;3.0)	1.0 (-2.0;2.0)
Control	8.0 (3.0;11.8)	1.0 (-1.0;2.0)	0.0 (-1.1;1.8)	0.0 (-1.0;2.1)	0.0 (-1.0;1.8)
NSAIDs (mg)‡					
Group 1	100 (48;150)	0 (0;72)**	16 (0;63)*	21 (0;75)**	0 (0;36)
Group 2	75 (23;150)	0 (0;72)**	8 (0;75)	0 (-19;75)	0 (-11;75)
Control	150 (53;150)	0 (0;8)	0 (0;47)	0 (0;23)	0 (0;48)

Data are presented at baseline as mean (SD) and mean change (SD) from baseline. ‡Skewed data are presented as median (interquartile range) and median change (interquartile range). BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; HAQ-S = Health Assessment Questionnaire for Spondylarthropathies; ASQoL = Ankylosing Spondylitis Quality of Life questionnaire; NSAID = Non-steroidal anti-inflammatory drugs. Positive changes imply improvement. *$P<0.05$ and **$P<0.01$ between intervention group and control; #$P<0.05$ between groups 1 and 2.

The effects found in secondary outcomes mirrored those of the primary outcomes, with more improvement and more prolonged effects observed in group 1 compared with group 2.

How can the positive effects of spa-exercise therapy be explained? The present study was primarily designed to assess the overall efficacy of spa-exercise therapy in patients with AS rather than to investigate individual components of the intervention. It is likely that the beneficial effects of spa-exercise therapy are attributable to a combination of specific and non-specific effects.

Specific effects may be attributed to the intervention itself. Immersion of the body in warm (thermal) water produces many physiological effects, such as muscle relaxation and increased joint mobility.[26,27] Group physical therapy has been shown to improve fitness and mobility in patients with AS.[11] The radon exposure in the Gasteiner Heilstollen in group 1 may also contribute to positive effects. Radon is a noble gas arising from the decay of radium, which is normally present in the earth's crust. Radon has been reported to influence the

immune system and decrease the disease activity of several autoimmune diseases.[28] Clinical effects of radon include reduction of pain and other signs of inflammation.[29] The cumulative dose of radon to which the patients were exposed (0.536 WLM) was below the Austrian radiation protection regulations for employees at the Heilstollen, which allow radon exposure of maximum 4.1 WLM annually.

Many other non-specific factors may contribute to the observed differences, such as change of environment, the pleasant scenery, the non-competitive atmosphere with fellow-patients, and the absence of work duties.[9,30] True placebo effects caused by the belief in improvement by spa therapy and positive attention may certainly have contributed to the differences between the intervention groups and controls. It is, however, likely that non-specific effects rapidly extinguish once the specific treatment has finished. The increasing improvement seen at T_{16w} in group 1 and the persistent long-term effects point, in our opinion, to some specific effect of the intervention.

Except for one pilot study, no other studies on spa therapy in patients with AS have been published in the English literature. Tishler and colleagues described the results of an uncontrolled study in 14 patients with AS who stayed at the Tiberias spa in Israel for 2 weeks, with 3 months follow-up.[10] Significant improvements in morning stiffness and finger-floor distance and a decrease in use of NSAIDs were observed until the end of the study. Schober index, chest expansion, and laboratory tests did not significantly change. Randomized controlled studies on spa therapy in various other rheumatic diseases used patients' self-reports to determine outcome variables.[2-8] In these studies the effects of bathing in thermal water or the Dead Sea, with or without mud packs, were compared to no intervention or to sham therapy with tap water only. Significant improvements in pain, functioning, and quality of life until 3 to 6 months after the spa therapy were found in the intervention groups of most studies. However, these studies are difficult to compare with our study because of other patient populations and different co-interventions apart from bathing.

A number of methodological considerations should be addressed. First, the lack of blinding: true double-blind spa therapy trials are impossible to conduct. To prevent differences in non-specific effects, however, the control group should have stayed at the spa resort during the intervention period rather than at home. Because it was unclear beforehand whether spa-exercise therapy had any beneficial effect at all in patients with AS, we decided to leave the control patients at home and to test the efficacy of spa-exercise therapy in general rather than to find out at this stage which specific component may be responsible for an effect.

Second, the primary outcomes were based on patients' self-reports instead of observations by a blinded assessor. The primary outcomes of this study were recommended in the preliminary core-set for physical therapy trials in AS.[13] In addition, a study by Hidding and colleagues showed a high concordance between the self-reports of patients with AS and clinical observations.[31] Therefore, we think the present results reliably reflect the clinical situation.

Third, no adjustments for multiple testing were made. To limit the possibility of significant results arising by chance, we have defined before the start of the study a pooled index of change as the primary outcome. Advantages of expressing results in a PIC include reducing the number of tests required and increasing the power to detect change.[25] A disadvantage of using a PIC is the difficulty of interpretation, because there is no frame of reference.[24] We realize that some of the significant findings in the secondary outcomes may have occurred by chance. However, it is obvious that all significant differences have the same direction, and the maximum improvements of up to 54% are not only statistically significant but, in our opinion, also clinically relevant.

The ability to generalize the study findings is restricted by both the availability of spa resorts and the level of reimbursement. Spa therapy is usually considered expensive, and reimbursement by insurance companies is inconsistent. In The Netherlands partial reimbursement for spa therapy is provided by a few insurance companies for a limited number of diseases (usually rheumatoid arthritis and AS only), with a maximum reimbursement level. In the United States, where no reimbursement is supplied, spa therapy remains as yet unavailable to a large group of patients. Cost-effectiveness studies are needed to weigh the costs against the observed effects.

In conclusion, a 3-week course of combined spa-exercise therapy provides beneficial effects in addition to drug treatment and weekly group physical therapy alone in patients with AS. The beneficial effects may last for at least 40 weeks after the start of spa-exercise therapy.

References

1. Sukenik S, Neumann L, Buskila D, Kleiner Baumgarten A, Zimlichman S, Horowitz J. Dead Sea bath salts for the treatment of rheumatoid arthritis. Clin Exp Rheumatol 1990;8:353-7.
2. Sukenik S, Buskila D, Neumann L, et al. Sulphur bath and mud pack treatment for rheumatoid arthritis at the Dead Sea area. Ann Rheum Dis 1990;49:99-102.
3. Elkayam O, Wigler I, Tishler M, et al. Effect of spa therapy in Tiberias on patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 1991;18:1799-803.
4. Sukenik S, Neumann L, Flusser D, Kleiner Baumgarten A, Buskila D. Balneotherapy for rheumatoid arthritis at the Dead Sea. Isr J Med Sci 1995;31:210-4.
5. Wigler I, Elkayam O, Paran D, Yaron M. Spa therapy for gonarthrosis: a prospective study. Rheumatol Int 1995;15:65-8.
6. Nguyen M, Revel M, Dougados M. Prolonged effects of 3 week therapy in a spa resort on lumbar spine, knee and hip osteoarthritis: follow-up after 6 months. A randomized controlled trial. Br J Rheumatol 1997;36:77-81.
7. Franke A, Reiner L, Pratzel HG, Franke T, Resch KL. Long-term efficacy of radon spa therapy in rheumatoid arthritis--a randomized, sham-controlled study and follow-up. Rheumatology (Oxford) 2000;39:894-902.
8. Neumann L, Sukenik S, Bolotin A, et al. The effect of balneotherapy at the dead sea on the quality of life of patients with fibromyalgia syndrome. Clin Rheumatol 2001;20:15-9.
9. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Knipschild PG. Balneotherapy for rheumatoid arthritis and osteoarthritis (Cochrane Review). In: The Cochrane Library, Issue 1. Oxford: Update Software 2001.
10. Tishler M, Brostovski Y, Yaron M. Effect of spa therapy in Tiberias on patients with ankylosing spondylitis. Clin Rheumatol 1995;14:21-5.
11. Hidding A, van der Linden S, Boers M, et al. Is group physical therapy superior to individualized therapy in ankylosing spondylitis? A randomized controlled trial. Arthritis Care Res 1993;6:117-25.
12. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984;27:361-8.
13. van der Heijde D, Bellamy N, Calin A, Dougados M, Khan MA, van der Linden S. Preliminary core sets for endpoints in ankylosing spondylitis. Assessments in Ankylosing Spondylitis Working Group. J Rheumatol 1997;24:2225-9.
14. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.
15. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.

16. Daltroy LH, Larson MG, Roberts NW, Liang MH. A modification of the Health Assessment Questionnaire for the spondyloarthropathies. J Rheumatol 1990;17:946-50.
17. Helliwell P, Doward L, Whalley D, et al. Psychometric and scaling properties of a new quality of life instrument specific to ankylosing spondylitis (AS) [abstract]. Arthritis Rheum 1999;42 (suppl):S72.
18. Ward MM, Kuzis S. Validity and sensitivity to change of spondylitis-specific measures of functional disability. J Rheumatol 1999;26:121-7.
19. Ruof J, Stucki G. Comparison of the Dougados Functional Index and the Bath Ankylosing Spondylitis Functional Index. A literature review. J Rheumatol 1999;26:955-60.
20. Spoorenberg A, van der Heijde D, de Klerk E, et al. A comparative study of the usefulness of the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index in the assessment of ankylosing spondylitis. J Rheumatol 1999;26:961-5.
21. Calin A, Nakache JP, Gueguen A, Zeidler H, Mielants H, Dougados M. Outcome variables in ankylosing spondylitis: evaluation of their relevance and discriminant capacity. J Rheumatol 1999;26:975-9.
22. Ruof J, Sangha O, Stucki G. Comparative responsiveness of 3 functional indices in ankylosing spondylitis. J Rheumatol 1999;26:1959-63.
23. Calin A, Nakache JP, Gueguen A, Zeidler H, Mielants H, Dougados M. Defining disease activity in ankylosing spondylitis: is a combination of variables (Bath Ankylosing Spondylitis Disease Activity Index) an appropriate instrument? Rheumatology (Oxford) 1999;38:878-82.
24. Roberts RS. Pooled outcome measures in arthritis: the pros and cons. J Rheumatol 1993;20:566-7.
25. Goldsmith CH, Smythe HA, Helewa A. Interpretation and power of a pooled index. J Rheumatol 1993;20:575-8.
26. Machtey I. Hydrotherapy and balneotherapy: State of the art. In: Balint G, Gömör B, Hodinka L, editors. Rheumatology, State of the Art. Amsterdam: Excerpta Medica 1992. p. 390-2.
27. Becker B. The biologic aspects of hydrotherapy. J Back Musculoskel Rehabil 1994;4:255-64.
28. Soto J. Effects of radon on the immune system. In: Pratzel HG, Deetjen P, editors. Radon in der Kurortmedizin: zum Nutzen und vermeintlichen Risiko einer traditionelle medizinische Anwendung. Geretsried: I.S.M.H. Verlag Geretsried 1997. p. 103-113.
29. Jöckel H. Practical experiences in radon therapies. In: Pratzel HG, Deetjen P, editors. Radon in der Kurortmedizin: zum Nutzen und vermeintlichen Risiko einer traditionelle medizinische Anwendung. Geretsried: I.S.M.H. Verlag Geretsried 1997. p. 84-91.
30. Bell MJ. Spa therapy in arthritis: a trialist's view. J Rheumatol 1991;18:1778-9.
31. Hidding A, van Santen M, De Klerk E, et al. Comparison between self-report measures and clinical observations of functional disability in ankylosing spondylitis, rheumatoid arthritis and fibromyalgia. J Rheumatol 1994;21:818-23.

Chapter 5

Cost-effectiveness of combined spa-exercise therapy in ankylosing spondylitis: a randomized controlled trial

A van Tubergen, A Boonen, R Landewé, M Rutten-van Mölken,
D van der Heijde, A Hidding, Sj van der Linden

Arthritis and Rheumatism [Arthritis Care and Research] 2002 [in press]

Abstract

Objective. To evaluate the cost-effectiveness and cost-utility of a 3-week course of combined spa-exercise therapy in addition to standard treatment consisting of anti-inflammatory drugs and weekly group physical therapy in patients with ankylosing spondylitis (AS).

Methods. A total of 120 Dutch AS outpatients were randomly allocated into three groups of 40 patients each. Group 1 was treated at a spa resort in Bad Hofgastein, Austria; group 2 at a spa resort in Arcen, The Netherlands. A control group stayed at home and continued their usual activities and standard treatment during the intervention weeks. After the intervention all patients followed weekly group physical therapy. The total study period was 40 weeks. Effectiveness of the intervention was assessed by functional ability (Bath Ankylosing Spondylitis Functional Index [BASFI]). Utilities were measured with the EuroQoL (EQ-$5D_{utility}$). A time-integrated summary score (area-under-the-curve [AUC]) defined the clinical effects (BASFI-AUC) and utilities (EQ-$5D_{utility}$-AUC) over time. Both direct (health care and non-health care) and indirect costs were included. Resource utilization and absence from work were registered weekly by the patients in a diary. All costs were calculated from a societal perspective.

Results. A total of 111 patients completed the diary. The between-group difference (95% confidence interval) for the BASFI-AUC was 1.0 (0.4-1.6, P=0.001) for group 1 versus controls and 0.6 (0.1-1.1, P=0.020) for group 2 versus controls. The between-group difference for EQ-$5D_{utility}$-AUC was 0.17 (0.09-0.25, P<0.001) for group 1 versus controls and 0.08 (0.00-0.15, P=0.04) for group 2 versus controls. The mean total costs per patient (including costs for spa therapy) during the study period were €3,023 for group 1, €3,240 for group 2, and €1,754 for the control group. The incremental cost-effectiveness ratio per unit effect gained in functional ability (0-10 scale) was €1,269 (497-3,316) for group 1 and €2,477 (601-12,098) for group 2. The costs per QALY gained were €7,465 (3,294-14,686) for group 1 and €18,575 (3,678-114,257) for group 2.

Conclusion. Combined spa-exercise therapy in addition to standard treatment with drugs and weekly group physical therapy is more effective and shows favorable cost-effectiveness- and cost-utility ratios compared with standard treatment alone in patients with AS.

Introduction

Treatment of ankylosing spondylitis (AS) is aimed at reducing symptoms and preventing or at least minimizing spinal deformity and disability. Standard treatment in active AS consists of non-steroidal anti-inflammatory drugs (NSAIDs), which reduce pain and stiffness, and physical therapy which improves or maintains mobility, fitness, and strength, and prevents deformity.[1] Patients are advised to exercise daily and to follow weekly group physical therapy. In addition, patients may annually take courses of spa therapy in a spa resort. Spa therapy has a longstanding history, but apart from the opinions of participants, its value is largely unknown. It is usually considered expensive and reimbursement by insurance companies is inconsistent.

Recently in a randomized controlled trial in patients with AS, we have demonstrated considerable and persistent improvements in a variety of clinical outcomes after a 3-week course of spa therapy in combination with physical exercises in addition to standard treatment with anti-inflammatory drugs and weekly group physical therapy compared with standard treatment alone.[2] In this paper we present the results of incremental cost-effectiveness- and cost-utility analyses of spa-exercise therapy in patients with AS.

Patients and methods

Design

This economic evaluation, performed from a societal perspective, was designed alongside a randomized controlled trial on the efficacy of combined spa-exercise therapy in AS. The methodology of this trial has been published elsewhere.[2] In brief, 120 Dutch patients with active AS were recruited and randomly allocated into three groups of 40 patients each. Group 1 received spa therapy at a spa resort in Bad Hofgastein, Austria. Group 2 received spa therapy at a spa resort in Arcen, The Netherlands. The control group stayed at home and continued their usual activities, drug treatment, and followed weekly group physical therapy during the 3 intervention weeks. Spa-exercise therapy was standardized and consisted of group physical exercises, walking, postural correction therapy (lying supine on bed), hydrotherapy, sports, and thermal treatment by either visits to the so-called Heilstollen (Austria) or the sauna (The Netherlands). After the intervention period, all patients from the three groups continued weekly group physical therapy for another 37 weeks, consisting of 1 hour of physical exercises, 1 hour of sports and 1 hour of hydrotherapy. During the intervention and the follow-up period, all patients continued their usual drug treatment, but were allowed to change this.

Spa-exercise therapy took place from T_{0w} (start of spa therapy) to T_{3w} (3 weeks after start spa therapy). Assessments, consisting of questionnaires, were performed at baseline (2 weeks prior to spa therapy; T_{-2w}), and at 4 weeks (T_{4w}),

16 weeks (T_{16w}), 28 weeks (T_{28w}), and 40 weeks (T_{40w}) after the start of spa-exercise therapy.

Effectiveness measurement

The primary outcomes of the study were functional ability (assessed by the Bath Ankylosing Spondylitis Functional Index [BASFI]),[3] pain (10 cm visual analog scale [VAS]), global well-being (10 cm VAS), and duration of morning stiffness (in minutes). The BASFI contains 10 questions concerning activities of daily living, scored on a 10 cm VAS, with anchors "easy" and "impossible" at either side. The mean of the items defines the final score, ranging from 0 (best) to 10 (worst). The BASFI has been shown to be a valid, reliable, and responsive instrument for measuring function in AS.[3-7] Here, the results on the BASFI are used to express effectiveness of the intervention and to calculate the cost-effectiveness ratio. The scores on the pain and global well-being question range from 0 (best) to 10 (worst).

Utility measurement

Utilities refer to preferences individuals or society may have for any particular health state.[8] For the present study, the EuroQol (EQ-$5D_{utility}$) was chosen to assess utility from a societal perspective.[9] The EQ-$5D_{utility}$ measures quality of life in five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension can be divided into three levels of severity: "no problems", "some problems", and "severe problems". The levels for each dimension are combined into one health state. For each of the possible health states a utility is attributed from a societal point of view, based on a large population survey using the time-trade-off procedure.[10,11] The final utility score ranges from 0 (representing death) to 1 (full health). The results of a cost-utility analysis are expressed as quality-adjusted life years (QALYs). A QALY is a composite index that includes effects in terms of both quality of life (utility) and the duration of time in such a health state.[8]

Costs

In this incremental cost-effectiveness and cost-utility analysis the recommendations proposed by the Dutch Guidelines for Pharmaco-Economic Analyses were followed as closely as possible.[12] Direct (health care and non-health care) as well as indirect (non-health care) costs were included. Only costs with respect to AS were considered, including predefined AS-related disorders such as uveitis, inflammatory bowel disease, and psoriasis. The time horizon to measure costs was limited to the duration of the study (40 weeks).

Table 5.1. Categories, methods, and sources for valuation of unit costs

Categories costs	Method of valuation	Source of costs	Euro per unit
Direct health care			
Spa therapy in Bad Hofgastein, Austria	Market price	Spa resort Bad Hofgastein, Austria	1739
Spa therapy in Arcen, Netherlands	Market price	Spa resort Arcen, The Netherlands	1515
Weekly group physical therapy	Market price	Ankylosing spondylitis patient society	306
Visit to general practitioner	Cost per visit based on true cost calculations	DGEE	16.59
Outpatient visit to specialist in university hospital	Cost per visit based on true cost calculations	DGEE	40.84
Outpatient visit to specialist in non-university hospital	Cost per visit based on true cost calculations	DGEE	72.60
Alternative health care	Price per session or medication as registered by patient	Cost diary	
Physiotherapy	Cost per session based on true cost calculations	DGEE	18.15
Medication	Reimbursable price per prescription augmented with an extra contribution (€5.26) or price paid by patient in case of OTC medication	TAXE	
Additional diagnostic tests	Tariff	COTG	
Admission university hospital	Day-price based on true cost calculations	DGEE	331.72
Admission non-university hospital	Day-price based on true cost calculations	DGEE	235.97
Visit to emergency department	Integral (true) costs	Financial department hospital in question	199.66
Day care in rehabilitation center	Cost per hour based on true cost calculations	DGEE	79.87
Direct non-health care			
Formal help in household	Cost per hour as stated by patient or based on true cost calculations	Cost diary or DGEE	As stated or 8.53
Personal allowance	Integral (true) costs	Cost diary	
Indirect non-health care			
Absenteeism from work because of illness	Production value of working day based on gross earnings of patient	Cost diary	

OTC = Over-the-counter; DGEE = Dutch Guidelines for Economic Evaluations; TAXE = Official price list of the Royal Dutch Society of Pharmacists; COTG = Dutch official organisation for tariffs in health care.

Direct and indirect costs. All direct and indirect costs included are presented in Table 5.1. Acquisition of aids and appliances or adaptations to house or car were excluded, because they were considered not to be influenced by the intervention. For transport costs, only transport to and from the spa resort was considered (incorporated in the overall costs for spa-exercise therapy). Costs of informal (non-paid) help and patients' reported loss of time due to AS were not included in the base-case analysis. We assumed that the time spent on informal help was done in leisure time, i.e. not creating productivity loss of the helping person. Similarly, inactivity was also considered as taking place in leisure time, otherwise, this would have been registered as days of illness-related absence from work.

Resources. All resource utilization and number of days of illness-related absence from work were measured during the entire study period in a cost diary adapted from Goossens et al.[13] Data from the day the intervention groups left for spa-exercise therapy up to 40 weeks afterwards were used for analysis. Patients judged themselves whether costs were related to AS or not.

Valuation of costs. For the valuation of the health care costs, opportunity costs (representing market prices) are preferred from a theoretical point of view.[12] In this comparative study they are reflected best by integral (i.e. true) costs. In The Netherlands, integral costs are not available for all resources.[14] In these cases tariffs were used (prices negotiated by health insurance companies). In Table 5.1 the method of valuation and source of all cost categories are presented and, if applicable, with costs specified. The market price was used for the total costs of the spa-exercise therapy (see Table 5.2) and weekly group physical therapy (consisting of the honorarium of the physiotherapist(s), rent of sports- and swimming accommodation, materials, and administrative costs). For the costs due to loss of productivity attributable to AS or related disorders, the friction cost method was used.[15,16] In this method the period of production loss due to illness is limited to the time needed to replace a sick worker or to reorganize the production process (4 months). The production value of this friction period was based on the reported gross earnings of the patients, corrected for part time work and, if applicable, for the percentage of disability benefits.

Table 5.2. Market price for spa therapy at the two spa resorts

	Spa therapy, Bad Hofgastein, Austria (group 1)	Spa therapy, Arcen, The Netherlands (group 2)
Travel costs from home to spa resort	102	11
Hotel accommodation (all meals included) and attendant	842	1040
Travel costs from hotel to spa resort during treatment period	29	89
Therapy costs and medical treatment	766	375
Total costs	*1739*	*1515*

Costs per patient in Euro.

All costs were calculated in Euro currency. Costs of 1999 were used and if not available for 1999, they were adapted using the Dutch health-specific price index.[14]

Statistical analysis

The clinical trial was designed as an intention-to-treat analysis and the sample size was based on clinical outcomes.[2] Between-group differences in clinical outcomes and utilities were analyzed by Student's *t*-tests for unpaired observations. Non-normally distributed data were analyzed by Mann-Whitney U test. A time-integrated summary score, the area-under-the-curve (AUC), defined the effects and utilities over time. The AUC of the change from baseline was calculated with the trapezium rule standardized by the study duration.[17,18]

Resource utilization and costs are reported as arithmetic means (standard deviation) per patient per group.[19] The differences in volumes of resource utilization were analyzed by Mann-Whitney U test. The 95% confidence interval (95% CI) around the mean total costs per patient and the between-group differences in mean total costs were estimated with bootstrapping, repeating the analysis 10,000 times.[20]

The incremental cost-effectiveness ratio was calculated by dividing the extra costs for the intervention group by the extra effects derived from it. The incremental cost-utility ratio was calculated by dividing the extra costs for the intervention group by the extra utilities derived from it, and expressed as costs per QALY gained. The 95% CI of the ratios were estimated with bootstrapping. Because the time horizon of this study was less than 1 year, neither costs nor effects were discounted.

Sensitivity analysis

Four one-way sensitivity analyses were performed:

[1] Costs of informal (non-paid) help were added to the base-case analysis. The costs per hour were based on a shadow price of €7.94.

[2] Costs of patient-reported inactivity were added to the base-case analysis. In the cost diary only the number of days lost due to inactivity was registered. Therefore, two analyses were performed considering a day of inactivity being either 4 or 8 hours. The costs per hour were based on a shadow price of €7.94.

[3] The number of days of sick leave during spa-exercise therapy was varied. Only a few patients from the intervention groups reported sick leave during the 3 intervention weeks; most patients had voluntarily taken days off from work. The mean total costs and costs per QALY were recalculated first assuming that all patients would be on sick leave during the 3 intervention weeks and second assuming that all patients would need to voluntarily take days off from work.

[4] The outcome measures "global well-being" and "pain" were used as measure of effectiveness of the intervention in the incremental cost-effectiveness ratio. Morning stiffness was not investigated in a sensitivity analysis. First, because it has a different scale (0 to infinite) compared with the other outcome

measures (0 to 10), which could be misleading in cost-effectiveness ratios, and second because morning stiffness is not sensitive to change.

Results

Of the 120 participating patients, 9 did not complete the cost diary (2 from group 1, 4 from group 2, 3 from the control group), but they did not differ from the patient groups with respect to sex, age, work, and education. All analyses are based on the 111 (93%) patients who completed the cost diary. The baseline characteristics of the groups are presented in Table 5.3. All characteristics, except sex, were well balanced among the groups. Relatively less women than men were randomly allocated to the control group compared with both intervention groups.

Table 5.3. Baseline characteristics of the 111 patients who completed the diary

	Group 1 (n=38) Spa therapy Austria	Group 2 (n=36) Spa therapy The Netherlands	Controls (n=37) Group physical therapy, Home
Men/Women	24/14	24/12	31/6
Age [years]†	47 (10) (range 29 – 68)	48 (9) (range 25 – 66)	48 (10) (range 29 – 63)
Disease duration [years]†	11 (6)	12 (5)	10 (6)
Inflammatory bowel disease [yes/no]	7/31	5/31	9/28
Uveitis [yes/no]	10/28	15/21	18/19
Psoriasis [yes/no]	4/34	4/32	4/33
Education [years]†	12 (4)	12 (4)	13 (4)
Manual/non-manual profession	15/20	6/27	16/20
Paid work [yes/no]	21/17	17/19	23/14
Reasons no work:††			
Household	3	3	1
(Early) retirement	2	4	2
Work disability	13	12	12
Unemployment	0	1	0
Own choice	0	4	0

Absolute number of patients. †Mean (SD). ††Multiple answers were allowed.

Clinical outcomes and utilities

The clinical outcomes and utilities of the 111 patients are listed in Table 5.4. Improvements in all primary outcomes, except for morning stiffness, were found in both intervention groups after spa-exercise therapy. The between-group differences (95% CI) in the AUC for the BASFI of 1.0 (0.4-1.6) for group 1 versus controls and of 0.6 (0.1-1.1) for group 2 versus controls were both statistically significant (P=0.001 and P=0.020, respectively). After adjustment for the score of the control group, the mean improvement during the entire study period in BASFI was 20% compared with baseline for group 1 and 14% for group 2.

Table 5.4. Clinical outcomes of 111 patients

Measure (range)	Baseline value	Change from baseline				Mean AUC of change during study period
	T_{-2w}	T_{4w}	T_{16w}	T_{28w}	T_{40w}	
BASFI (0 – 10)						
Group 1	4.9 (1.8)	1.1 (1.5)**	1.0 (1.6)*	0.7 (1.5)*	0.4 (1.3)	1.0 (1.3)**
Group 2	4.2 (2.0)	0.8 (1.2)**	0.5 (1.1)	0.0 (1.2)	0.0 (1.1)	0.6 (1.0)*
Controls	4.2 (2.1)	0.0 (1.1)	0.0 (1.7)	-0.1 (1.7)	-0.1 (1.3)	0.0 (1.1)
Global well-being (0 – 10)						
Group 1	5.2 (2.0)	1.3 (2.8)	2.3 (2.4)*	1.4 (2.6)	1.1 (2.6)	1.5 (2.1)*
Group 2	5.3 (2.4)	1.8 (2.6)*	1.3 (3.0)	1.0 (2.9)	0.4 (2.6)	1.5 (2.5)
Controls	4.8 (2.4)	0.3 (3.0)	0.6 (3.1)	0.4 (2.8)	0.4 (2.8)	0.4 (2.6)
Pain (0 – 10)						
Group 1	4.6 (2.5)	0.7 (2.4)	1.4 (2.7)	0.8 (2.8)	0.1 (2.4)	1.0 (2.0)*
Group 2	4.6 (2.5)	1.4 (2.6)*	1.1 (2.7)	-0.3 (3.0)	-0.4 (2.8)	1.1 (2.3)
Controls	4.6 (2.7)	0.1 (2.3)	0.4 (2.8)	0.0 (2.8)	-0.2 (2.1)	0.1 (2.1)
Morning stiffness (min.) ‡						
Group 1	30 (10;60)	3 (0;21)	3 (0;17)	0 (-6;15)	0 (-9;11)	8.9 (21.0)
Group 2	30 (15;60)	0 (0;18)	0 (-5;15)	0 (-6;14)	0 (-15;12)	3.1 (24.5)
Controls	30 (10;56)	5 (0;10)	4 (-5;15)	0 (-13;10)	0 (-5;15)	-3.4 (30.4)
$EQ\text{-}5D_{utility}$ (0 – 1)						
Group 1	0.64 (0.22)	0.10 (0.24)**	0.12 (0.24)**	0.10 (0.21)**	0.03 (0.23)	0.11 (0.20)**
Group 2	0.65 (0.22)	0.02 (0.20)*	0.04 (0.21)*	-0.03 (0.23)	-0.01 (0.27)	0.02 (0.17)*
Controls	0.72 (0.10)	-0.06 (0.18)	-0.04 (0.19)	-0.08 (0.28)	-0.03 (0.19)	-0.06 (0.14)

Group 1 (n=38) spa therapy Austria, group 2 (n=36) spa therapy The Netherlands, control group (n=37) home. Data are presented as mean (SD) at baseline and as mean change (SD_{change}) from baseline. ‡Skewed data are presented as median (interquartile range) and median change (interquartile range). AUC = Area-under-the-curve; BASFI = Bath Ankylosing Spondylitis Functional Index; $EQ\text{-}5D_{utility}$ = EuroQoL 5 dimensions. Positive changes imply improvement. *$P<0.05$ and **$P<0.01$ between intervention group and control group.

The between-group difference in the AUC of the EQ-5D$_{utility}$ was 0.17 (0.09-0.25, $P<0.001$) for group 1 versus controls and 0.08 (0.00-0.15, $P=0.04$) for group 2 versus controls.

Health resource utilization and costs

The volumes of health care and non-health care utilization during the follow-up period are listed in Table 5.5. The number of prescribed and over-the-counter (OTC) medication was significantly lower in both intervention groups compared with controls ($P=0.006$ group 1 versus controls; $P=0.032$ group 2 versus controls). The number of hours and costs of formal help in household were higher in both intervention groups compared with the control group, but the differences were not statistically significant.

Table 5.5. Volumes of health care and non-health care utilisation during study period (40 weeks)

	Group 1 (n=38) Spa therapy Austria	Group 2 (n=36) Spa therapy The Netherlands	Control group (n=37) Home
Visits to general practitioner	1.6 (2.5)	1.3 (2.0)	1.7 (2.1)
Visits to specialists [outpatient]	2.0 (2.0)	1.9 (2.3)	2.5 (2.9)
Physiotherapy sessions	17.0 (20.9)	14.9 (21.9)	21.5 (24.3)
Prescribed and OTC medication	4.3 (6.0)**	4.5 (4.8)*	6.1 (4.5)
Prescribed and OTC alternative medication	0.5 (1.8)	0.2 (0.7)	0.1 (0.7)
Visits to alternative practitioner	0.4 (1.7)	0.3 (1.2)	0.8 (2.7)
Formal help in household [hours]	17.3 (45.0)	12.7 (31.7)	11.5 (34.8)

Mean number (SD) per patient. OTC = Over-the-counter. *$P<0.05$; **$P<0.01$ compared with the control group (Mann-Whitney U test).

Of the patients with a paid job, seven from group 1 (33%), four from group 2 (24%), and nine (39%) from the control group reported AS-related absence from work during the study period, including three patients from group 1 and four patients from group 2 who reported sick leave during the spa therapy period. The mean number of days of illness due to AS for the entire group was 2.5 (6.5), 6.4 (26.4), and 6.1 (15.8) for group 1, 2, and control group, respectively.

Direct and indirect costs per patient for each category are listed in Table 5.6. Mean total costs were €3,023 (2,728-3,359) for group 1, €3,240 (2,553-4,391) for group 2, and €1,754 (1,268-2,402) for the control group. The mean total incremental costs were €1,269 (565-1,867) for group 1 and €1,486 (501-2,707) for group 2.

Table 5.6. Results of costs during study period (40 weeks) of 111 patients

	Group 1 (n=38) Spa therapy Austria	Group 2 (n=36) Spa therapy The Netherlands	Control group (n=37) Home
Direct costs			
Health care costs			
Spa therapy	1739 (0)	1515 (0)	-
Group physical therapy	306 (0)	306 (0)	306 (0)
Health care professionals [GP and specialist]	112 (109)	107 (142)	163 (168)
Physiotherapy	308 (379)	271 (398)	391 (440)
Medications [prescription and OTC]	153 (239)	182 (234)	206 (177)
Additional diagnostic tests	38 (72)	24 (80)	23 (39)
Hospitalisation and day care rehabilitation	24 (174)	26 (157)	86 (525)
Alternative health care	24 (89)	3 (12)	35 (106)
Sum direct health care costs (excl. spa therapy)	964 (744)	919 (625)	1210 (785)
Sum direct health care costs (incl. spa therapy)	2703 (744)	2434 (625)	1210 (785)
Non-health care costs			
Formal help in household	106 (275)	100 (256)	72 (220)
Personal allowance	2 (10)	14 (83)	11 (67)
Indirect costs			
Absenteeism from work	211 (549)	693 (2864)	461 (1204)
Sum direct and indirect costs (excl. spa therapy)	1284 (1009)	1725 (2988)	1754 (1772)
Sum direct and indirect costs (incl. spa therapy)	3023 (1009)	3240 (2988)	1754 (1772)

Mean (SD) costs per patient in Euro. GP=General practitioner; OTC=Over-the-counter.

Cost-effectiveness and cost-utility

The incremental cost-effectiveness ratio per unit effect gained in function ability on a 0-10 scale (based on the AUC of the BASFI) was €1,269/1.0 = €1,269 (497-3,316) for group 1 and €1,486/0.6 = €2,477 (601-12,098) for group 2.

Assuming that the difference in AUC of the $EQ\text{-}5D_{utility}$ will be equal to 0 during the last 3 months of the year after a 3-week course of spa-exercise therapy, and that the costs will be equal in all three groups (since no extended effect of the intervention is expected, the cost will presumably equate the costs of the control group), the cost per QALY gained would be €1,269/0.17 = €7,465 (3,294-14,686) for group 1 and €1,486/0.08 = €18,575 (3,678-114,257) for group 2.

Sensitivity analysis

In Table 5.7 the mean total costs per patient per group and the mean costs per QALY gained are shown for each of the additional costs or alternatives calculated in the sensitivity analysis.

Informal help. The mean (SD) number of hours of informal help was 22.9 (67.7) in group 1, 25.4 (76.4) in group 2, and 19.9 (76.6) in the control group. Additional costs for informal help would be €182 (537) for group 1, €202 (607) for group 2, and €158 (608) for the control group. The costs per QALY gained would slightly increase (Table 5.7.)

Table 5.7. One-way sensitivity analyses

	Mean total costs			Incremental costs per QALY gained		Incremental costs per unit of effect gained on a 0-10 scale	
	Group 1	Group 2	Controls	Group 1	Group 2	Group 1	Group 2
Base-case analysis (with BASFI)	3023	3240	1754	7465	18575	1269	2477
Sensitivity analysis							
Informal help in household	3205	3442	1912	7606	19125	1293	2550
Inactivity 4 hours/day	3256	3386	2003	7371	17288	1253	2305
Inactivity 8 hours/day	3488	3533	2252	7271	16013	1236	2135
Absenteeism work during spa therapy	3523	3690	1754	10406	24200	1769	3227
Voluntarily days off during spa therapy	2965	3047	1754	7124	16163	1211	2155
CEA with global well-being as outcome measure	3023	3240	1754	7465	18575	1154	1351
CEA with pain as outcome measure	3023	3240	1754	7465	18575	1410	1486

Mean costs per patient in Euro. Group 1 = spa therapy Austria, group 2 = spa therapy The Netherlands, controls = home. BASFI = Bath Ankylosing Spondylitis Functional Index; CEA = Cost-effectiveness analysis; QALY = Quality-adjusted life years.

Inactivity. The mean number of days of inactivity was 7.3 (19.1) in group 1, 4.6 (12.5) in group 2, and 7.8 (21.3) in the control group. Applying 4 hours per day as the mean number of hours of inactivity, the additional costs would be €233 (606) for group 1, €146 (398) for group 2, and €249 (678) for the control group. In case a day of inactivity would represent 8 hours, the additional costs would be €465 (1,212) for group 1, €293 (795) for group 2, and €498 (1,356) for the control group. Due to higher reported inactivity in the control group, the costs per QALY gained would decrease for both intervention groups (Table 5.7).

Absenteeism from work. In case all patients who work would be on sick leave during spa therapy, the mean number of days of illness would be 7.5 (9.2) in group 1, and 10.9 (26.3) in group 2, both significantly more than the control group (P=0.020 group 1 versus controls, P=0.032 group 2 versus controls). The costs per QALY gained would increase to €10,406 for group 1 and €24,200 for group 2. In case all patients would need to voluntarily take days off to follow spa therapy, the mean number of days of illness would decrease to 1.9 (5.7) for group 1 and 4.7 (23.8) for group 2. The costs per QALY gained would decrease to €7,124 for group 1 and €16,163 for group 2.

Cost-effectiveness. When "global well-being" and "pain" are used as outcome measures in the cost-effectiveness study, the incremental costs per unit of effect gained are more or less similar for both intervention groups. Compared with the base-case analysis, group 2 showed considerable fewer incremental costs per unit of effect gained (Table 5.7).

Discussion

This economic evaluation shows that a 3-week course of spa-exercise therapy in addition to standard treatment has favorable cost-effectiveness and cost-utility ratios compared with standard treatment alone in patients with AS. Cost savings were found for both intervention groups compared with the control group with respect to visits to health care professionals, physiotherapy, and use of medication. In the sensitivity analyses the results were robust to variation in several cost categories, but were sensitive to variation in indirect costs. The incremental cost ratios would considerably increase for both intervention groups in case all patients with a paid job would be on sick leave during spa-exercise therapy instead of considering the treatment period as regular holidays.

Only one cost-analysis of spa therapy could be found in the English literature.[21] The study aimed at evaluating the cost-effectiveness of (undefined) spa therapy for a variety of rheumatic diseases, measured from a health insurance company's perspective. Effectiveness was assessed indirectly through determination of health care utilization before and after the intervention, based on reimbursement claims. No decrease in health care utilization was found in the groups that received spa therapy compared with controls. However, some concerns have been raised about the methodology of the study with respect to patient recruitment, lack of baseline description of the groups, the

randomization procedure, and inadequate statistical procedures.[22] Therefore, no reasonable comparison with the present study can be made.

Expressing the effects of an intervention in terms of "cost per QALY gained" has become increasingly popular in recent years. The primary argument was to facilitate the allocation of health care resources by decision makers, because the QALY-concept allows comparison across different interventions in different diseases.[23] Incremental cost-utility ratios can be listed in so-called league tables, in which the costs per QALY gained for each study are visible at a glance. There are, however, a number of disadvantages in using league tables.[24] League tables suggest homogeneity in, and comparability among, study designs. However, studies included in league tables might have been conducted in different years with consequently different technologies and relative prices; the approaches to measure utilities differ greatly among studies; some studies include only direct costs, whereas others also report indirect costs; and not all studies apply discounting, if necessary. In addition, the amount of costs per QALY gained is dependent on the alternative program. An intervention will show fewer costs per QALY gained if compared with "doing nothing" instead of with the "best alternative treatment".

The incremental costs per QALY gained for each of the intervention groups (€7,465 for group 1; €18,575 for group 2) are relatively high compared with the rather low prevalent costs for patients with AS (€1,754). This probably reflects the absence of effective (curative) treatment for AS. Drug treatment is mostly limited to NSAIDs. Also, the costs for weekly group physical therapy are relatively low.[25] These low prevalent costs contrast with the important impact the disease has with respect to pain, well-being, and work ability.[26,27] The latest developments in the treatment of AS comprise the administration of tumor necrosis factor alpha (TNF-α) inhibiting drugs, which have been reported to yield dramatic improvements.[28,29] Due to the huge costs of anti-TNF-α treatment (approximately €11,000/patient/year) it is not to be expected that cost-utility analyses with respect to these new drugs will reveal costs per QALY gained that are lower than demonstrated by us. These studies might probably show even larger discrepancies between the costs per QALY gained for the intervention groups, compared with the prevalent costs of the control group.

A few methodological issues should be considered. First, the data as reported by the patients in the diaries were not ascertained by consulting other data sources. It is conceivable that patients might have had different opinions on what is related to AS. Second, there were no data available on (non-)health care expenditures and days of illness-related absence from work during the pre-trial period. Therefore, no comparison with the previous year can be made and pre-existing differences in (non-)health care utilization between the groups can not be excluded. Third, the ability to generalize the study findings is restricted to Dutch patients with active AS who follow weekly group physical therapy and have a disease duration of <20 years. The results might differ when other spa resorts are visited or different programs are provided. Similarly, the costs are likely to be different for other countries with other health care systems.

In The Netherlands reimbursement for spa therapy is provided by a few insurance companies for a limited number of diseases (usually rheumatoid arthritis and AS only), with a maximum reimbursement level. It is estimated that 0.1% of the Dutch population suffers from AS (~16,000 patients).[30] Approximately 20% of them follow weekly group physical therapy. If reimbursement for spa therapy would be limited to these 3,200 patients, who have proved to be compliant with treatment by following weekly group physical therapy, then the incremental societal costs for providing spa therapy would annually be €4,408,000. In exchange of these costs, a large -although temporary- improvement in functional outcome and quality of life will be gained in a group of patients for whom no curative treatment is yet available.

In conclusion, a 3-week course of combined spa-exercise therapy in addition to standard treatment with anti-inflammatory drugs and weekly group physical therapy provides beneficial effects compared with standard treatment alone, and can be regarded as cost-effective in patients with AS.

References

1. van Tubergen A, Landewé R, van der Linden S. Spondylarthropathies: options for combination therapy. Springer Semin Immunopathol 2001;23:147-63.
2. van Tubergen A, Landewé R, van der Heijde D, et al. Combined spa-exercise therapy is effective in patients with ankylosing spondylitis: a randomized controlled trial. Arthritis Rheum 2001;45:430-8.
3. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.
4. Calin A, Nakache JP, Gueguen A, et al. Outcome variables in ankylosing spondylitis: evaluation of their relevance and discriminant capacity. J Rheumatol 1999;26:975-9.
5. Ruof J, Stucki G. Comparison of the Dougados Functional Index and the Bath Ankylosing Spondylitis Functional Index. A literature review. J Rheumatol 1999;26:955-60.
6. Ruof J, Sangha O, Stucki G. Comparative responsiveness of 3 functional indices in ankylosing spondylitis. J Rheumatol 1999;26:1959-63.
7. Spoorenberg A, van der Heijde D, de Klerk E, et al. A comparative study of the usefulness of the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index in the assessment of ankylosing spondylitis. J Rheumatol 1999;26:961-5.
8. Drummond M. Methods for the economic evaluation of health care programs. 2nd ed. Oxford: Oxford Medical Publications 1997.
9. The EuroQoL Group. Euroqol-a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.
10. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ 1996;5:141-54.
11. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095-108.
12. Riteco JA, de Heij LJM, van Luijn JCF, Wolff I. Guidelines for pharmaco-economic research [in Dutch]. Amstelveen: College voor zorgverzekeringen 1999.
13. Goossens ME, Rutten-van Molken MP, Vlaeyen JW, van der Linden SM. The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. J Clin Epidemiol 2000;53:688-95.
14. Oostenbrink JB, Koopmanschap MA, Rutten FFH. Guidelines for cost-analysis. Methods and guideline prices for economic evaluations in health care [in Dutch]. Amstelveen: College voor zorgverzekeringen 2000.
15. Koopmanschap MA, van Ineveld BM. Towards a new approach for estimating indirect costs of disease. Soc Sci Med 1992;34:1005-10.
16. Koopmanschap MA, Rutten FFH. A practical guide for calculating indirect costs of disease. PharmacoEconomics 1996;10:460-6.

17. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. BMJ 1990;300:230-5.
18. Pham B, Cranney A, Boers M, et al. Validity of area-under-the-curve analysis to summarize effect in rheumatoid arthritis clinical trials. J Rheumatol 1999;26:712-6.
19. Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? BMJ 2000;320:1197-200.
20. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman and Hall 1993.
21. Allard P, Deligne J, van Bockstael V, Duquesnoy B. Is spa therapy cost-effective in rheumatic disorders? Rev Rhum [Engl Ed] 1998;65:173-80.
22. Graber-Duvernay B. Letter to the editor about the article: Is spa therapy cost-effective in rheumatic disorders? Rev Rhum [Engl Ed] 1998;65:600-602.
23. Mason J, Drummond M, Torrance G. Some guidelines on the use of cost effectiveness league tables. BMJ 1993;306:570-2.
24. Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? Soc Sci Med 1993;37:33-40.
25. Bakker C, Hidding A, van der Linden S, van Doorslaer E. Cost effectiveness of group physical therapy compared to individualized therapy for ankylosing spondylitis. A randomized controlled trial. J Rheumatol 1994;21:264-8.
26. Zink A, Braun J, Listing J, Wollenhaupt J. Disability and handicap in rheumatoid arthritis and ankylosing spondylitis--results from the German rheumatological database. German Collaborative Arthritis Centers. J Rheumatol 2000;27:613-22.
27. Boonen A, Chorus A, Miedema H, et al. Employment, work disability, and work days lost in patients with ankylosing spondylitis: a cross sectional study of Dutch patients. Ann Rheum Dis 2001;60:353-8.
28. Brandt J, Haibel H, Cornely D, et al. Successful treatment of active ankylosing spondylitis with the anti-tumor necrosis factor alpha monoclonal antibody infliximab. Arthritis Rheum 2000;43:1346-52.
29. Van den Bosch F, Kruithof E, Baeten D, et al. Effects of a loading dose regimen of three infusions of chimeric monoclonal antibody to tumour necrosis factor alpha (infliximab) in spondyloarthropathy: an open pilot study. Ann Rheum Dis 2000;59:428-33.
30. van der Linden SM, Valkenburg HA, de Jongh BM, Cats A. The risk of developing ankylosing spondylitis in HLA-B27 positive individuals. A comparison of relatives of spondylitis patients with the general population. Arthritis Rheum 1984;27:241-9.

Part II

Outcome assessment in ankylosing spondylitis

Chapter 6

Comparison of statistically-derived ASAS improvement criteria for ankylosing spondylitis with clinically relevant improvement according to an expert panel

A van Tubergen, D van der Heijde, J Anderson, R Landewé, M Dougados, J Braun, N Bellamy, G Udrea, Sj van der Linden
For the ASAS Working Group

Annals of the Rheumatic Diseases 2002 [in press]

Abstract

Objective. To investigate whether the recently developed (statistically-derived) "ASsessment in Ankylosing Spondylitis Working Group"-improvement criteria (ASAS-IC) for ankylosing spondylitis (AS) reflect clinically relevant improvement according to the opinion of an expert panel.

Methods. The ASAS-IC consist of four domains: physical function, spinal pain, patient global assessment, and inflammation. Scores on these four domains of 55 patients with AS, who had participated in an NSAID efficacy trial, were presented to an international expert panel (consisting of patients with AS and members of the ASAS Working Group) in a 3-round Delphi-exercise. The number of (non-)responders according to the ASAS-IC was compared with the final consensus of the experts. The most important domains in the opinion of the experts were identified, and also selected with discriminant analysis. A number of provisional criteria sets that approximated the consensus of the experts were defined. Using other datasets, these clinically-derived criteria sets as well as the statistically-derived ASAS-IC were then tested for discriminative properties and for agreement with the end-of-trial efficacy by patient and physician.

Results. Forty experts completed the three Delphi-rounds. The experts considered twice as many patients as a responder compared with the ASAS-IC (n=42 versus n=21). Overall agreement between experts and ASAS-IC was 62%. Spinal pain was considered as the most important domain by the majority of the experts and was also selected as such by discriminant analysis. Provisional criteria sets with an agreement of ≥80% compared with the consensus of the experts showed high placebo response rates (27% to 42%), in contrast to the ASAS-IC with a pre-defined placebo response rate of 25%. All criteria sets and the ASAS-IC discriminated well between active and placebo treatment (chi-square value 36–45; $P<0.001$). Compared with the end-of-trial efficacy assessment, the provisional criteria sets showed an agreement of 71-82%, sensitivity of 67-83%, and specificity of 81-88%. The ASAS-IC showed an agreement of 70%, sensitivity of 62%, and specificity of 89%.

Conclusion. The ASAS-IC are strict in defining response, are highly specific, and consequently show lower sensitivity in comparison with the clinically-derived criteria sets. However, those patients who are considered as responders by applying the ASAS-IC are acknowledged as such by the expert panel as well as by patients' and physicians' judgments, and are therefore likely to be true responders.

Introduction

In the near future, many new therapies will be studied in patients with ankylosing spondylitis (AS). Besides demonstrating the mean efficacy of a therapy at a group level, it is also important to ascertain *how many* (and *which*) patients actually improve, since improvement of an entire group may be based on (small) improvements of many patients, as well as on a marked improvement of only a few patients with the majority showing no change.[1] Advantages of defining an individual response are that such changes are easier to understand; it may reduce the number of tests because usually several domains are combined; it facilitates comparison across different trials; it may help physicians decide whether a patient has responded adequately to therapy; and it may also be used to assess factors that predict response.[1-3] To establish an individual response, criteria for defining such "improvement" are essential.

In 1998, during the Outcome Measures in Rheumatoid Arthritis Clinical Trial Conference, the members of the ASsessment in Ankylosing Spondylitis (ASAS) Working Group have selected "core sets" of outcome measures to be used in different kinds of trials in AS, in order to create uniformity and allow comparison among these studies.[4] The ASAS Working Group is an international group of rheumatologists, epidemiologists, patients with AS, and representatives of the pharmaceutical industry from more than 20 countries, who share their expertise in the field of AS. More recently, a small group of members from the ASAS have developed a preliminary definition of short term improvement in AS, based on the core set of clinical trials with non-steroidal anti-inflammatory drugs (NSAIDs).[5] These ASAS improvement criteria (ASAS-IC) consist of four outcome domains: physical function (measured with the Bath AS Functional Index [BASFI]),[6] spinal pain (measured on a 100 mm visual analog scale [VAS]), patient global assessment (100 mm VAS), and inflammation (mean of the last two questions from the Bath AS Disease Activity Index (BASDAI),[7] concerning the intensity and duration of morning stiffness). Scores on each domain range from 0 (best) to 100 (worst). The ASAS-IC define improvement as ≥20% and ≥10 VAS points improvement on the 0-100 scale on at least three of the four domains, with on the fourth domain no worsening of ≥20% and ≥10 VAS points.

The composition and validation of the ASAS-IC were performed by using a statistical approach. It is, however, important to know to what degree these statistically-derived improvement criteria reflect the opinions of clinicians and patients with AS.[8] The aim of our study was to compare the ASAS-IC with clinically relevant improvement according to an expert panel in the field of AS by means of a 3-round consensus building Delphi-exercise.

Methods

Cases

For the present study the profiles of real patients with AS, who had participated in a randomized placebo controlled NSAID efficacy trial with a flare design, were used.[9] All patients fulfilled the modified New York criteria for AS.[10] In total 473 patients entered the study of which 363 patients completed the 6-week part of the trial. For each of the patients the absolute and relative change scores between last measurement and baseline on the four domains of the ASAS-IC (physical function, spinal pain, patient global assessment, and inflammation) were calculated, independently of treatment allocation or completion of the trial. Cases appropriate for the present study were selected manually by comparing the outcomes with the ASAS-IC. In total 55 cases were selected: 21 cases fulfilled the ASAS-IC, 23 cases approximated the ASAS-IC, but showed either too little improvement or too much worsening in the absolute or relative change score in one or two domains, and 11 cases were clearly non-responders showing either obvious worsening (>30%) in one or two domains while the others improved, or showed no/minor change in any domain.

Expert panel and Delphi exercise

The opinions of experts in the field of AS were examined by means of a 3-round consensus building Delphi-exercise.[11] The expert panel consisted of members of the ASAS Working Group extended with a group of patients with AS, who are actively involved in AS research. In total, 57 candidate participants with an e-mail address or fax number known to us, and who were not involved in the development of the ASAS-IC, were invited to participate in this study.

In the first Delphi-round, all 57 candidates were sent the profiles of the 55 cases, randomly presented in an Excel data spreadsheet. For each of the cases the values on each domain at baseline and at termination of the trial, as well as the absolute and relative change scores were provided. The experts were requested to indicate for each case whether they personally considered this patient as a responder (yes or no). After completion of the file, they e-mailed or faxed back their answers. In a second round, the same cases were sent to each participant, showing his/her initial judgment for each of the cases, including the anonymous, aggregated responses of the group. The experts were then again asked to supply their judgments for the cases, but with the possibility to modify the initial decisions, based on the group's responses. A final, third round was similar to that of the second round, at which the experts were provided their individual decisions and group's responses of the second round. In this final round all experts were also asked by questionnaire to describe how they came to a decision for judgments of the cases: whether they looked at both absolute and relative improvement or only either of them; how much improvement they considered at least necessary to certify a patient as a responder, and on how many domains; whether they considered some domains more important than

others, and which domains were the least important; and how much worsening they allowed at most, and on how many domains.

For each round, the experts were offered 2 weeks to judge the cases. If necessary, a reminder was sent. The exercise was performed before publication of the ASAS-IC.

Statistical analysis

The analyses were performed from both the viewpoint of the ASAS-IC and the final consensus of the experts. First, the number of patients improved and not improved according to the ASAS-IC was compared with the consensus of the experts in a 2 X 2 table. The percentage of agreement was calculated as the total number of corresponding responder and non-responder cases divided by the total number of cases.

Second, the experts' rationale on how they came to a decision for each of the cases was investigated, including the opinions on which domains had prevailed in the final decision. We also applied discriminant analysis to determine on a statistical basis which domains were considered most contributive for each participant with respect to both the absolute and the relative change scores. Dependent variable was the individual judgment of each participant for all 55 cases. Independent variables were either the absolute or the relative change scores on the four domains of the ASAS-IC. This exercise was done for all participants separately.

Third, because the consensus of the experts clearly differed from the ASAS-IC, a number of provisional criteria sets were defined in order to find criteria sets that approximated the consensus of the experts. These criteria sets and the ASAS-IC were analyzed for their properties to discriminate between active and placebo treatment in 2 X 2 tables with chi-square tests. For this purpose, the same data subset as for validation of the ASAS-IC was used.[5] In this validation set a random selection of three NSAID efficacy trials was taken (197 placebo, 408 actively treated).

Fourth, the ASAS-IC and the consensus of the experts were compared with the end-of-trial efficacy assessment by the physician and patient of the 55 selected cases, thereby considering the efficacy assessment as the gold standard. Also, the ASAS-IC and the provisional criteria sets that best represented the consensus of the experts were compared with the end-of-trial efficacy assessment of all patients (n=473) of the NSAID efficacy trial from which the 55 cases were initially selected.[9]

Results

In total 57 ASAS members and patients with AS were invited to participate in this 3-round Delphi exercise. Thirteen did not respond and two considered themselves not to have (enough) expertise to complete the exercises. Forty-two participants returned the questionnaire in the first round, 41 in the second, and

40 (of whom 4 patients with AS) in the last round. The reasons for the two non-responders during this exercise are unknown.

Agreement among the experts increased with each round for each of the cases. A median number of 4 (range 0 to 24) modifications were made in the second round by the experts, and a median number of 1 (range 0 to 7) in the third round. Agreement among the experts of ≥70% was finally reached in all 55 cases, ≥80% in 52/55, and ≥90% in 44/55 cases. In this paper, the consensus of ≥70% (including all cases) has been used for further analyses.

Agreement between ASAS improvement criteria and consensus of the expert panel

As stated, of the 55 cases by definition 21 (38%) fulfilled the ASAS-IC and were therefore considered as a responder (Table 6.1). Thirty-four (62%) cases did not fulfill the ASAS-IC and were therefore considered as a non-responder. The experts judged 42 (76%) cases as a responder and 13 (24%) as a non-responder. Sixteen of the 23 cases that approximated the ASAS-IC and 5 of the 11 clearly non-responder cases (which all showed large differences among the domains, with worsening up to 83% and 15 VAS points in one variable) were judged as a responder by the experts. The experts agreed on all 21 (100%) responder cases as defined by the ASAS-IC and on 13 (38%) ASAS-IC non-responder cases. Improvement according to the ASAS-IC was acknowledged by the experts in all cases, but the proportion of patients considered as a responder by the experts was twice as high. Overall agreement between the ASAS-IC and the consensus of the experts was 62%.

Table 6.1. Number of responders and non-responders according to the ASsessment in Ankylosing Spondylitis improvement criteria (ASAS-IC) for ankylosing spondylitis versus the consensus of the experts

	ASAS-IC		
Consensus experts	Responder	Non-responder	*Total*
Responder	21	21	*42*
Non-responder	0	13	*13*
Total	*21*	*34*	*55*

Judgment of the cases by the experts and weight of domains

In the last Delphi-round the experts were asked to describe how they came to a decision for each of the cases. The opinions to judge patients as a responder differed largely among the experts: relative improvement of 10%, 15%, 20%, 25%, or even more in ≥2 domains, with or without absolute improvement of 10 or 20 VAS points in ≥2 domains, and some experts requested spinal pain to be part of the improving domains. Ninety percent of the experts asserted that they looked at both absolute and relative improvement for their final judgment. Worsening was not included in most judgments. Spinal pain was considered as

the most important domain by 50% of the experts, patient global assessment by 26%, physical function by 11%, inflammation by 8%, and 5% considered all domains to be equal. Inflammation was considered as the least important domain by 43%, physical function by 30%, patient global assessment by 16%, spinal pain by 5%, and 5% considered all domains to be equal.

With discriminant analysis on both the absolute and the relative change scores the most important domains were selected for each of the 40 participants who had participated in the last round (Figure 6.1 and 6.2, respectively). Spinal pain was selected as step 1 of the discriminant analysis with respect to both absolute and relative change scores in 80% of the experts, whereby this becomes the most important domain for making a decision on improvement. Patient global assessment was selected as the most important second step for the absolute change scores (Figure 6.1), and inflammation for the relative change scores (Figure 6.2). For the absolute change scores only one discriminant was selected in six participants and for the relative change scores only one discriminant was selected in three participants.

Table 6.2 compares the opinions of the participants (n=38) regarding the prevailing domain with the results of the discriminant analysis. The results on the first step of the discriminant analysis for both the absolute and relative change scores were identical, except for one participant (who did not provide his personal judgment), and are therefore combined in Table 6.2. It can be seen that in 31 participants, spinal pain was selected by discriminant analysis as the domain with the highest influence on decision making, while only 17 participants had stated that spinal pain was the most important domain. The remaining 14 participants assumed that their judgments were mainly based on other domains (in particular patient global assessment).

Table 6.2. Prevailing domain in discriminating between improvement and no improvement according to each participant versus discriminant analysis for each participant (n=38)

	Prevailing domain according to participants					
Discriminant analysis	Pain	Inflammation	Global	Function	All domains equal	Total
Pain	17	1	9	2	2	*31*
Inflammation	1	2	0	1	0	*4*
Global	1	0	1	0	0	*2*
Function	0	0	0	1	0	*1*
Total	*19*	*3*	*10*	*4*	*2*	*38*

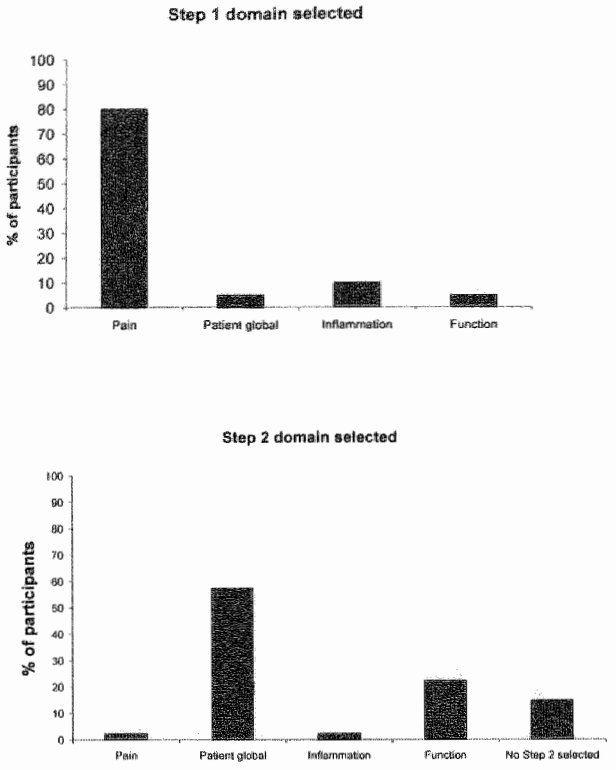

Figure 6.1 The percentage of participants per domain selected as the first and second step in the discriminant analysis on the *absolute* change scores is presented.

Provisional improvement criteria sets versus consensus of the expert panel

Because the consensus of the experts clearly differed from the ASAS-IC, a number of provisional criteria sets were defined in order to find criteria sets that approximated the consensus of the experts. Candidate criteria sets identical to those created for development of the ASAS-IC,[5] but also a number of provisional criteria sets adapted to the opinion of the experts, were tested for their agreement with the consensus of the experts. Because the majority of the experts did not consider worsening important in the judgments of the cases, a maximun in worsening was not included in most provisional criteria sets. These criteria sets always included relative improvement in at least two domains, with or without absolute improvement. Because spinal pain was selected by both the experts as well as by discriminant analysis to be the most important domain, a minimum of improvement on this domain was regarded as a prerequisite in many provisional criteria sets.

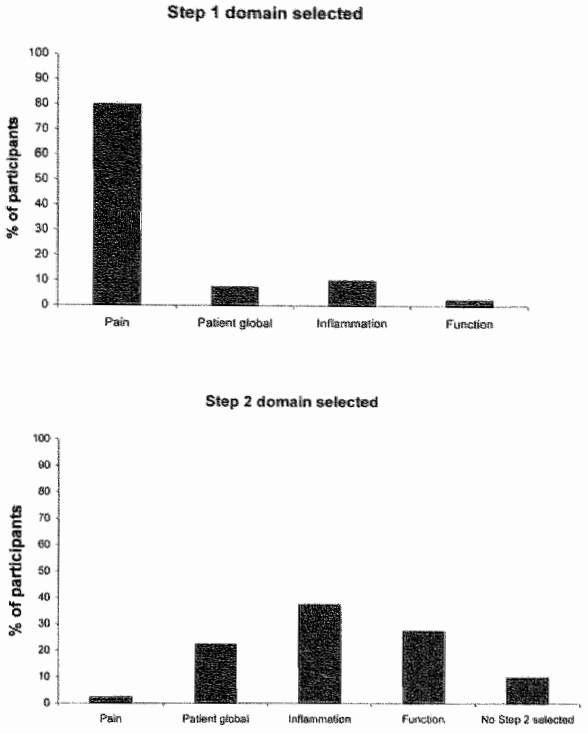

Figure 6.2 The percentage of participants per domain selected as the first and second step in the discriminant analysis on the *relative* change scores is presented.

In total 36 criteria sets were tested for their agreement with the consensus of the experts. In Table 6.3 only those criteria sets are shown with an agreement of ≥80% compared with the consensus of the experts, based on all 55 cases. The highest agreement found was 93%. The criteria sets were thereafter applied to the validation subset of the ASAS-IC in order to analyze their properties to discriminate between active and placebo treatment.[5] The percentage responders in the placebo treated group appeared to be large for many of the criteria sets (up to 42%). Ten of the criteria sets showed a placebo response rate of ≤30%. The ASAS-IC had by definition a placebo response rate of 25%.[5] All criteria sets and the ASAS-IC discriminated well between the placebo and active treated groups (chi-square 36-45; all $P<0.001$), and showed similar contrasts.

End-of-trial efficacy assessment

As a next stage, the ASAS-IC and the consensus of the experts were compared with the end-of-trial efficacy assessment by physician and patient for each of the 55 selected cases. Because in only 1 of the 55 cases, disagreement between physician and patient on efficacy assessment was noted, the results were

Table 6.3. Improvement criteria sets showing ≥80% agreement with the consensus of the experts and tested for their discriminative performances between placebo and active treated patients

Improvement definition	% Agreement compared with consensus of the experts (55 cases)	Validation subset			
		% Responders in placebo treated group (n=197)	% Responders in active treated group (n=408)	Difference	Chi-square value
Relative improvement					
30% in 2 of 4	82	33	60	27	40
20% in 3 of 4	84	27	55	28	41
20% in 2 of 4	91	42	69	27	42
10% in 3 of 4	91	36	62	26	38
Combinations in relative improvement					
40% in 2 of 4 or 20% in 3 of 4	89	30	58	28	44
30% in 2 of 4 or 20% in 3 of 4	93	34	62	28	42
Relative and absolute improvement					
30% and 10 VAS in 2 of 4	80	33	60	27	39
20% and 10 VAS in 2 of 4	93	40	67	27	42
Combinations in relative and absolute improvement					
40% in 2 of 4 or 20% in 3 of 4 and 10 VAS in 2 of 4	87	29	58	29	44
30% in 2 of 4 or 20% in 3 of 4 and 10 VAS in 2 of 4	91	33	61	28	43
Relative improvement including pain					
30% in 2 of 4, incl. pain	82	29	56	27	38
20% in 3 of 4, incl. pain	80	26	53	27	41
20% in 2 of 4, incl. pain	93	37	64	27	40
Combinations in relative improvement including pain					
40% in 2 of 4 or 20% in 3 of 4, all incl. pain	87	27	56	29	43
30% in 2 of 4 or 20% in 3 of 4, all incl. pain	91	30	59	29	43
Relative and absolute improvement including pain					
30% and 10 VAS in 2 of 4, all incl. pain	80	29	56	27	39
20% and 10 VAS in 2 of 4, all incl. pain	93	35	63	28	42
Combinations in relative and absolute improvement including pain					
40% in 2 of 4 or 20% in 3 of 4 and 10 VAS in 2 of 4, all incl. pain	85	27	55	28	43
30% in 2 of 4 or 20% in 3 of 4 and 10 VAS in 2 of 4, all incl. pain	89	30	59	29	45
ASAS Improvement criteria (ASAS-IC)					
20% and 10 VAS in 3 of 4, 4th no worsening of ≥20% and 10 VAS	62	25	51	26	36

VAS = Visual analog scale; ASAS-IC = Assessment in Ankylosing Spondylitis Improvement Criteria. All chi-square values showed *P* values of <0.001.

Table 6.4. Comparison of improvement criteria sets that best represented the consensus of the experts and ASAS-IC with the end-of-trial efficacy assessment by the physician and the patient

Improvement definition	% Agreement compared with efficacy assessment (n=473)*	Sensitivity	Specificity
Relative improvement			
30% in 2 of 4	77	73	84
20% in 3 of 4	77	67	88
20% in 2 of 4	82	83	81
10% in 3 of 4	80	79	84
Combinations in relative improvement			
40% in 2 of 4 or 20% in 3 of 4	76	72	85
30% in 2 of 4 or 20% in 3 of 4	77	75	83
Relative and absolute improvement			
30% and 10 VAS in 2 of 4	76	72	85
20% and 10 VAS in 2 of 4	81	80	83
Combinations in relative and absolute improvement			
40% in 2 of 4 or 20% in 3 of 4 and 10 VAS in 2 of 4	75	71	86
30% in 2 of 4 or 20% in 3 of 4 and 10 VAS in 2 of 4	77	74	84
Relative improvement including pain			
30% in 2 of 4, incl. pain	76	72	86
20% in 3 of 4, incl. pain	73	66	88
20% in 2 of 4, incl. pain	81	80	84
Combinations in relative improvement including pain			
40% in 2 of 4 or 20% in 3 of 4, all incl. pain	76	71	88
30% in 2 of 4 or 20% in 3 of 4, all incl. pain	77	73	85
Relative and absolute improvement including pain			
30% and 10 VAS in 2 of 4, all incl. pain	75	71	86
20% and 10 VAS in 2 of 4, all incl. pain	80	78	84
Combinations in relative and absolute improvement including pain			
40% in 2 of 4 or 20% in 3 of 4 and 10 VAS in 2 of 4, all incl. pain	75	70	88
30% in 2 of 4 or 20% in 3 of 4 and 10 VAS in 2 of 4, all incl. pain	76	72	85
ASAS Improvement criteria (ASAS-IC)			
20% and 10 VAS in 3 of 4, 4th no worsening of ≥20% and 10 VAS	70	62	89

VAS = Visual analog scale; ASAS-IC = Assessment in Ankylosing Spondylitis Improvement Criteria. *Ten patients were omitted from this analysis, because of a disagreement between the patient and the physician end-of-trial efficacy assessment. Another 27 patients were omitted from this analysis, because of missing data in domains of the ASAS-IC. Sensitivity (proportion of responders correctly classified) and specificity (proportion of non-responders correctly classified) were calculated considering the end-of-trial efficacy assessment as the gold standard.

combined and this case was omitted from the analyses. In total 44 patients were considered by themselves and by their physicians as responders, and 10 as non-responders. Agreement between the efficacy assessment and the consensus of the experts was 72%, sensitivity (proportion of responders correctly classified) was 80%, and specificity (proportion of non-responders correctly classified) was 40%. Agreement between the efficacy assessment and the ASAS-IC was 41%, sensitivity 36%, and specificity 60%.

Finally, the ASAS-IC and the provisional criteria sets that best represented the consensus of the experts were compared with the end-of-trial efficacy assessment of all patients (n=473) of the NSAID efficacy trial from which the 55 cases were initially selected (Table 6.4).[9] In 10 out of these 473 patients a disagreement between the physician and the patient on the efficacy was observed. These patients were omitted from the analyses. Another 27 patients were also omitted from the analyses due to missing data in one or more domains of the ASAS-IC. In total 307 patients were considered by themselves and by their physicians as responders (47 [43%] from the placebo group and 260 [80%] from the active group), and 129 patients as non-responders (63 [57%] from the placebo group and 66 [20%] from the active group). Agreement between the efficacy assessment and the criteria sets ranged from 71% to 82%, sensitivity from 67% to 83%, and specificity from 81% to 88%. Agreement between the efficacy assessment and the ASAS-IC was 70%, sensitivity 62%, and specificity 89%.

Discussion

In this study the recently, by statistical approach, developed ASAS-IC for AS were compared with the opinion of an expert panel in the field of AS by means of a 3-round Delphi-exercise, in order to assess the clinical relevance and applicability of the ASAS-IC. Advantages of using a Delphi-technique to obtain the opinions of experts over for instance a consensus meeting are that participants are offered the opportunity to modify their initial judgments, based on the opinion of the group in following rounds, and that all opinions are collected anonymously.[11] Most modifications by the experts were made in the second round. In the final round only one (median) modification was made compared with the previous round, suggesting that three rounds are sufficient for achieving consensus.

According to the ASAS-IC, 21 out of 55 patients were considered as a responder, whereas the experts judged 42 patients as a responder. Clearly, the experts personally used criteria for their judgments that were less strict than the ASAS-IC, but opinions among the experts varied largely. The majority of the experts considered improvement in only two domains sufficient to acknowledge a patient as a responder and worsening was not included in most judgments. Spinal pain was regarded as the most discriminative domain. This was also reflected in the discriminant analysis, in which pain was selected as the first step

in 80% of the experts. According to the experts' own opinion, pain was less often considered as the most important domain (in 50%). A number of participants defined other domains (particularly patient global assessment) as most discriminative, but apparently acted differently. This discrepancy in actual decision making and stated opinions on the importance of variables has been described previously by Kirwan et al.[12]

Because the consensus of the experts clearly differed from the ASAS-IC, a number of provisional criteria sets were defined in order to find criteria sets that best represented the consensus of the experts. Out of a total of 36 criteria sets, 19 sets showed an accuracy rate of ≥80% with the consensus of the experts, and were therefore tested for their discriminative properties by using the same validation subset as was used for development of the ASAS-IC. All provisional criteria sets, as well as the ASAS-IC, discriminated well between active and placebo treatment. The placebo response rate was, however, relatively high (up to 42%) for the clinically-derived criteria sets. For the development of the ASAS-IC, Anderson et al. made the condition that the placebo response rate should not exceed 25%.[5] No placebo response rates below 25% were found for the provisional criteria sets and only 10 criteria sets showed a placebo response rate of ≤30%. The predefined maximum placebo response rate explains why the ASAS-IC behaved relatively worse compared with some of the provisional criteria sets in terms of agreement and contrast between intervention and placebo. The ASAS-IC and the provisional criteria sets showed moderate to good agreement with the end-of-(NSAID)trial efficacy assessment by the patient and physician. Five criteria sets showed an agreement of ≥80%, with corresponding high sensitivity and specificity. The ASAS-IC showed lower agreement (70%), a relatively low sensitivity (62%), but the highest specificity (89%).

The ASAS-IC appeared to be strict in defining a patient as a responder, but those patients who are responders according to the ASAS-IC were also acknowledged as such by the expert panel. In addition, compared with the efficacy assessment by patient and physician of the NSAID trial, the specificity of the ASAS-IC was high, implying that the number of patients incorrectly classified as responder by the ASAS-IC is very low (Table 6.4). Thus, those patients who are classified as a responder by the ASAS-IC are most likely to be true responders, because they are responders in the opinion of the experts, the patients themselves, and their treating physicians, independently of each other. This makes the ASAS-IC particularly valid for clinical trials, in which true responders are important to detect. Using analogues reasoning, applying the ASAS-IC in clinical practice is much more complicated; the ASAS-IC are not sensitive enough to pick up changes that are considered important in the opinion of the experts. With regard to this aspect, the ASAS-IC can be compared with classification criteria (characterized by high specificity) and the consensus of the experts as diagnostic criteria (characterized by high sensitivity). In our opinion, the ASAS-IC should, however, not be substituted by any of the provisional criteria sets that best represented the consensus of the experts, because these criteria sets showed too high placebo response rates, which is an undesired effect in the assessment of potentially active treatment. We would,

therefore, recommend to apply the ASAS-IC particularly in clinical trials, in which the false-positive rates should be kept low, and apply the ASAS-IC with reserves in daily practice, in which sensitivity is of greater importance.

The ASAS-IC have been validated in NSAID efficacy trials with "flare" as a selection criterion. Studies with flare-based designs may be associated with large placebo response rates.[13] It is therefore arguable whether the ASAS-IC would behave similarly in NSAID efficacy trials without this criterion. In addition, it is yet unknown how the ASAS-IC behave in other clinical trials assessing for instance the efficacy of disease modifying drugs, anti-TNF-α therapy, or physical therapy. Also, the opinion of the experts on the degree of improvement and prevailing domains may vary for these different trial settings. It should be kept in mind that the opinion of experts on what they consider improvement may largely be determined by their clinical perception; if more effective treatment will become available, then the opinion of the experts may also change on what they call improvement. In other words: experts' opinion and thus criteria sets derived from this are context-specific (i.e. based on current state of the art). In contrary, the ASAS-IC are statistically-derived and consequently more "timeless" (i.e. less influenced by current opinions). It may be expected with the introduction of more effective treatment that the experts will change more into the direction of the ASAS-IC. Future studies will be needed to validate the current ASAS-IC or to develop new improvement criteria for other kind of trials.

In conclusion, the ASAS-IC are strict in defining patients as a responder, however, those patients classified as responder are acknowledged as such by the expert panel as well as by patients themselves and their treating physicians, independently of each other. Therefore, the opinion of the patients and the physicians is well reflected by the ASAS-IC.

References

1. van Riel PL, van de Putte LB. DC-ART: what proportion of response constitutes a positive response? J Rheumatol 1994;21 (suppl 4):54-5.
2. Giannini EH, Ruperto N, Ravelli A, et al. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40:1202-9.
3. Dougados M, Leclaire P, van der Heijde D, et al. Response criteria for clinical trials on osteoarthritis of the knee and hip: a report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials response criteria initiative. Osteoarthritis Cartilage 2000;8:395-403.
4. van der Heijde D, Calin A, Dougados M, et al. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. J Rheumatol 1999;26:951-4.
5. Anderson JJ, Baron G, van der Heijde D, Felson DT, Dougados M. Ankylosing spondylitis assessment group preliminary definition of short-term improvement in ankylosing spondylitis. Arthritis Rheum 2001;44:1876-86.
6. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.
7. Garrett S, Jenkinson T, Kennedy LG, et al. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.
8. Ward MM. Response criteria and criteria for clinically important improvement: separate and equal? Arthritis Rheum 2001;44:1728-9.
9. Dougados M, Gueguen A, Nakache JP, et al. Ankylosing spondylitis: what is the optimum duration of a clinical study? A one year versus a 6 weeks non-steroidal anti-inflammatory drug trial. Rheumatology (Oxford) 1999;38:235-44.
10. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984;27:361-8.
11. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995;311:376-80.
12. Kirwan JR, Chaput de Saintonge DM, Joyce CR, Currey HL. Clinical judgment in rheumatoid arthritis. II. Judging 'current disease activity' in clinical practice. Ann Rheum Dis 1983;42:648-51.
13. Scott Lennox JA, McLaughlin Miley C, Lennox RD, et al. Stratification of flare intensity identifies placebo responders in a treatment efficacy trial of patients with osteoarthritis. Arthritis Rheum 2001;44:1599-607.

Chapter 7

The use of a numerical rating scale as an answer modality in ankylosing spondylitis specific questionnaires

A van Tubergen, I Debats, L Ryser, J Londoño, R Burgos-Vargas, MH Cardiel, R Landewé, G Stucki, D van der Heijde

Arthritis and Rheumatism [Arthritis Care and Research] 2002; 47:242-8

Abstract

Objectives. To determine the agreement of scores on the original visual analog scale (VAS) or Likert-scale of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI) and Dougados Functional Index (DFI) with scores on a numerical rating scale (NRS); To assess the reproducibility and responsiveness of the instruments with the original answer scale and the NRS.

Methods. Five hundred thirty-six patients with ankylosing spondylitis from The Netherlands, Mexico, and Switzerland completed a questionnaire in which all questions from the BASDAI, BASFI, and DFI were presented twice in random order with an 11-point NRS and either a 10 cm VAS (BASDAI and BASFI) or a 5-point Likert-scale (DFI). Agreement of scores using Bland-Altman plots and intraclass correlation coefficients [ICCs]), reproducibility using ICCs, and responsiveness were assessed.

Results. Large variability between the scores on the original scales and the NRS was found in individual questions of all three questionnaires, although total scores showed ICCs of at least 0.88. Reproducibility of all answer modalities showed low ICCs in individual questions, but moderate to good ICCs in total scores (Dutch group 0.62-0.89; Mexican group 0.53-0.72). Moderate to large effects (0.48-1.04) were found in responsiveness scores in the three questionnaires. No major differences in reproducibility and responsiveness between the answer modalities were found.

Conclusion. Although large variability between the scores on the original answer scales and the NRS was observed, the BASDAI, BASFI, and DFI can be administered with an NRS which does not show important differences compared with the original answer scales.

Introduction

The Bath Ankylosing Spondylitis Disease Activity Index (BASDAI),[1] the Bath Ankylosing Spondylitis Functional Index (BASFI),[2] and the Dougados Functional Index (DFI)[3] are well established, widely used instruments to evaluate disease activity and functioning in patients with ankylosing spondylitis (AS). The BASDAI and BASFI are completed on a visual analog scale (VAS), and the DFI on a Likert-scale.

The VAS is a commonly used scale consisting of a 10 cm horizontal line anchored with two extremes at either end. It has proven to be a valid and reliable measure for subjective feelings such as pain and function.[4-8] Disadvantages of the VAS are that many patients experience difficulties in completing the VAS and that the VAS can only be administered in a written form, which is a limitation for illiterate or visually impaired patients.[9-15] Furthermore, there is a risk for measurement error: random errors that can occur during measuring the distance to the mark on the line, and systemic errors during the reproduction of the VAS questionnaires, because photocopying may alter the length of the line.

The Likert-scale (or verbal rating scale) consists of several categories, most commonly five or seven with adjectives representing degrees of, for instance, functional ability. Subjects mark the adjective that best describes their impairment. Advantages of the Likert-scale are that it is easy to understand, simple to complete, and it can be administered in either a written or a verbal form.[11] Disadvantages are the potential discrepancy between the patient's feelings and the descriptions on the scale, the different interpretations that can be attributed to the adjectives of the scale, and the unequal intervals between the categories.[4]

Another type of scale is the numerical rating scale (NRS). The NRS is usually an 11-, 21-, or (rarely) 101-point scale, with numbers in boxes that are anchored with two extremes at either end. Subjects mark their answer by putting a cross through the appropriate number. The NRS is simple to complete and score, and can be administered in both written and verbal form.[11]

Although no major differences in practical use of these answer modalities have been found, the NRS seems to be slightly preferred, since it is easy to complete and appropriate for all groups of patients.[10,11,16,17] Furthermore, the presumed high sensitivity of the VAS, because of its infinite number of possibilities for answers, has been disproven by Jensen et al., who showed that little information was lost when a 101-point NRS was transformed to an 11- or 21-point NRS.[18] The NRS and the Likert-scale, both ordinal scales, have inherent problems that do not differ from those of a VAS because in practice, clusters are formed on the scale, which limit the actual number of responses.[12] Consequently, the VAS does not behave as a true continuous scale.

Instruments for research in rheumatology should be valid in all their aspects. To standardize the nomenclature of validity, the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) Filter has been proposed.[19] The three domains of the OMERACT Filter are truth (validity), discrimination

(reproducibility and responsiveness), and feasibility. One of the criteria for feasibility is the appropriateness of the answer scales used in questionnaires. Because some patients may experience difficulties with VAS or Likert-scales, and because the NRS is slightly preferred in the literature, we decided to assess the discrimination and feasibility properties of the BASDAI, BASFI, and DFI with an NRS. Our first objective was to study the agreement of scores on the original scales of the BASDAI, BASFI, and DFI with scores on an NRS. Second, the reproducibility and responsiveness of the BASDAI, BASFI, and DFI on the original answer scale and on the NRS were assessed. Attention was paid to both scores on single items as well as the total scores of these questionnaires. The first was done to supply more insight into the properties of single-item questionnaires, sometimes used to assess certain aspects of a disease. To enhance the generalizability of the study results, we decided to investigate all objectives across different languages and cultures, and in clinical trial patients and outpatients with varying disease duration and disease activity.

Patients and Methods

Patients

A total of 536 patients with AS from The Netherlands (n=182 [34%]), Mexico (n=166 [31%]), and Switzerland (n=188 [35%]) entered the study. Table 7.1 shows an overview of the groups of patients that participated in the various parts of the study, as well as baseline characteristics of each group.

From The Netherlands, 120 outpatients participating in a randomized controlled trial to assess the efficacy of spa therapy in patients with AS completed the questionnaires. The patients were randomly allocated to receive either spa therapy in Austria (group 1; n=40) or The Netherlands (group 2; n=40), or to a control group (group 3; n=40) that stayed at home and continued weekly group physical therapy. All patients completed the questionnaires twice at home: at baseline (2 weeks prior to the intervention) and 1 week after 3 consecutive weeks of spa therapy or after 3 weeks of weekly group physical therapy. Furthermore, a convenience sample of 62 Dutch patients from a secondary and a tertiary outpatient clinic completed the questionnaire in the hospital after their regular outpatient visit (group 4).

In Mexico, 166 consecutive outpatients from two referral centers completed the questionnaire in the hospital (group 5). At one of the centers, teenagers were also enrolled in the study. Thirty-eight patients from group 5, who visited the outpatient clinic again within the study period, completed the same questionnaire a second time with a mean interval of 4 weeks.

Table 7.1. Patients used for several aspects of the study and characteristics of the groups

	Group 1 Netherlands (spa therapy Austria) (n=40)	Group 2 Netherlands (spa therapy Netherlands) (n=40)	Group 3 Netherlands (spa therapy control) (n=40)	Group 4 Netherlands (outpatients) (n=62)	Group 5 Mexico (outpatients) (n=166)	Group 6 Switzerland (members AS society) (n=188)
Comparison VAS or Likert with NRS	X	X	X	X	X	X
Responsiveness	X	X	X	-	-	-
Reproducibility	-	-	X	-	38 patients	-
% Men	63	70	85	77	65	70
Age [years]†	48 (10)	49 (9)	48 (10)	47 (11)	31 (12)	50 (12)
Disease duration [years]†	11 (6)	12 (5)	10 (6)	12 (7)	5 (7)	18 (10)
Duration of complaints [years]†	19 (10)	19 (9)	15 (8)	20 (9)	11 (10)	24 (11)

†Mean (SD). VAS = Visual analog scale; NRS = Numerical rating scale; AS = Ankylosing spondylitis; X = Participated in this aspect of the study.

In Switzerland, a random sample of the members from the Swiss AS association was drawn. In total, 418 patients received the questionnaire at home, which was completed by 188 patients (group 6).

Instruments

The BASDAI consists of six questions on fatigue, pain of the spine, pain and/or swelling of the peripheral joints, localized tenderness, and severity and duration of morning stiffness.[1] The questions are answered on a 10 cm VAS, anchored with the labels "none" and "very severe" at either end of the first five questions, and with "0 hours" and "2 hours" in the question on duration of morning stiffness. The mean of the two questions on morning stiffness counts as one variable. The final score is defined by calculating the mean of the five items. Scores range from 0 (best) to 10 (worst).

The BASFI contains 10 questions concerning activities of daily living and is scored on a 10 cm VAS, with anchors "easy" and "impossible" at either side.[2] The mean of the items defines the final score, with scores ranging from 0 (best) to 10 (worst).

The DFI consists of 20 Likert-formatted scales, and includes activities of daily living.[3] Originally, the DFI was completed on a 3-point Likert-scale and later modified to a 5-point Likert-scale. In the present study, we have applied the 5-point Likert-scale. For each item, possible scores are 0, 0.5, 1, 1.5, and 2. Total scores range from 0 (best) to 40 (worst), but to facilitate the comparability with scores of the other questionnaires, we have converted the scores on the DFI to a 0-10 scale, by dividing the scores by 4.

We used instruments that were translated and validated in all three languages by other researchers according to proposed guidelines.[20] Some of these validation studies have been published.[21,22] Most of these instruments have been used extensively in clinical trials and epidemiological studies in AS. All questionnaires are self-reported.

Procedure

All questions from the BASDAI, BASFI, and DFI with both the original answer modality and an 11-point NRS with figures in boxes were mixed up and presented in random order in the questionnaire. It was checked that the same question with the two different answer modalities did not appear on the same page. Instructions to complete the questionnaire were given on the first page. On the last page, the preference of the participants was assessed with the question: "Which of the three answer scales did you like best?", with the possibility to mark one of the three answer scales. Questionnaires were administered in Dutch (The Netherlands), Spanish (Mexico), and German (Switzerland).

Missing values were only dealt with in calculating the total scores of the questionnaires, according to Creusen et al.[23] At most 1 of 6 questions from the BASDAI, 2 of 10 from the BASFI, and 8 of 20 questions from the DFI were substituted by the patient's mean.

Statistical analysis

Feasibility. Bland-Altman plots were made to visualize the difference between the results on the original answer scale and on the NRS against the mean.[24] Intraclass correlation coefficients (ICCs) were calculated to assess the concordance of the total scores on the different scales of each questionnaire. ICCs higher than 0.75 were considered relevant.[25]

Reproducibility. Data from the control group of the spa therapy trial (group 3) and the 38 patients from Mexico who completed the questionnaire twice were used for assessing reproducibility. ICCs were calculated for both individual questions and total scores of each instrument. Because response criteria for outcome in AS trials were not yet available, the results on the reproducibility tests were used to determine cut-off levels for improvement or worsening on an individual patient level, according to the 95% limits of agreement method described by Bland and Altman.[24] This method helps to distinguish between true changes and variability due to measurement errors. The smallest detectable difference (SDD) is calculated as $1.96 * SD_{change}$ of the mean total score of each instrument in the control group. Patients from both the intervention groups (n=80) and the control group (n=40) with a change score exceeding the positive value of the SDD were considered truly improved, those with a change score exceeding the negative value of the SDD were considered worsened.

Responsiveness. Responsiveness or sensitivity to change was assessed with the data from both intervention groups of the spa therapy trial (n=80) compared with the control group (n=40). Because no consensus exists on which method is preferred, we determined the responsiveness with the effect size (ES),[26] the standardized response mean (SRM),[27] and the method described by Guyatt et al.[28] The results were interpreted according to the Cohen's effect size index, in which 0.2 refers to a small change, 0.5 to moderate, and 0.8 or more to large change.[29] The ES is calculated as the mean change after treatment compared with baseline, divided by standard deviation (SD) of the baseline scores.[26] The SRM is similar to the ES, but with the SD of the change score as denominator.[27] In the Guyatt method, the mean change score in the treatment group is divided by the SD of the change score of the control group.[28] Consequently, a responsiveness score for the control group can not be calculated.

Results

Feasibility

Of the 536 patients, 14 were unable to complete questions on a VAS, 2 patients had difficulties with the NRS, and 1 with the Likert-scale. Forty-nine percent of the patients preferred to answer on a Likert-scale, 38% on an NRS, 9% on a VAS, and 4% did not have a particular preference.

The results for the total scores on the questionnaires are presented in Table 7.2. For the BASDAI and BASFI, no major differences in scores between the VAS

Table 7.2. Results of total scores on different answer modalities for BASDAI, BASFI, and DFI, split per group

	BASDAI		BASFI		DFI	
	VAS	NRS	VAS	NRS	Likert	NRS
Group 1	4.7 (1.8)	4.6 (1.8)	4.9 (1.8)	5.0 (2.1)	2.9 (1.4)	3.7 (1.7)
Group 2	5.1 (2.0)	5.0 (2.0)	4.3 (2.0)	4.4 (2.0)	2.7 (1.3)	3.6 (1.8)
Group 3	4.5 (2.0)	4.2 (2.1)	4.2 (2.1)	4.2 (2.2)	2.8 (1.4)	3.4 (1.9)
Group 4	3.3 (1.8)	3.5 (1.7)	3.5 (1.9)	3.6 (2.1)	2.4 (1.4)	3.0 (1.7)
Group 5	4.2 (2.3)	4.1 (2.4)	3.6 (2.4)	3.6 (2.5)	2.4 (1.9)	2.7 (2.1)
Group 6	4.3 (2.2)	4.4 (2.3)	2.8 (2.0)	2.7 (2.1)	1.8 (1.4)	2.1 (1.6)

Mean (SD). BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; BASFI = Bath Ankylosing Spondylitis Functional Index; DFI = Dougados Functional Index; VAS = Visual analog scale; NRS = Numerical rating scale.

and NRS were found. However, consistently lower scores on the DFI were observed with the Likert-scales compared with scores on the NRS.

Bland-Altman plots of the difference plotted against the mean of every question answered by NRS and either a VAS or Likert-scale showed a wide distribution in all questions (representative examples are shown in Figure 7.1). The DFI scored on a Likert-scale gave lower scores compared with the NRS. This was not present for VAS scores compared with NRS scores. The greatest variability was found in the middle part of the scoring range. Total scores of each of the instruments showed better concordance between the original answer scales and the NRS; the ICCs calculated were 0.95 for the BASDAI, 0.97 for the BASFI, and 0.88 for the DFI.

Reproducibility

To assess reproducibility of individual questions, question 5 of the BASDAI on the level of morning stiffness appeared twice with an NRS in the questionnaire with two pages in between. Although the ICC was 0.85, an SDD of 2.8 could be calculated. Reproducibility of the three questionnaires is shown in Table 7.3. Because consistently lower ICCs in the Mexican group compared with the Dutch group were found, we decided to present the results separately. Low to moderate ICCs were found for individual questions of all three questionnaires. Total scores showed higher ICCs, but remained low for all questionnaires in the Mexican group and for the BASDAI in the Dutch group. The differences between the answer modalities, however, were minor. Only in the Mexican group was reproducibility of the individual questions of the BASDAI clearly lower on the VAS compared with NRS. In group 3, lower ICCs were found in individual questions on the Likert-scale compared with NRS on the DFI. There was no difference in reproducibility with respect to total scores.

To assess reproducibility dichotomously, the SDDs of each instrument on each scale were calculated (Table 7.4). No major differences between the original scales and the NRS were found for each instrument.

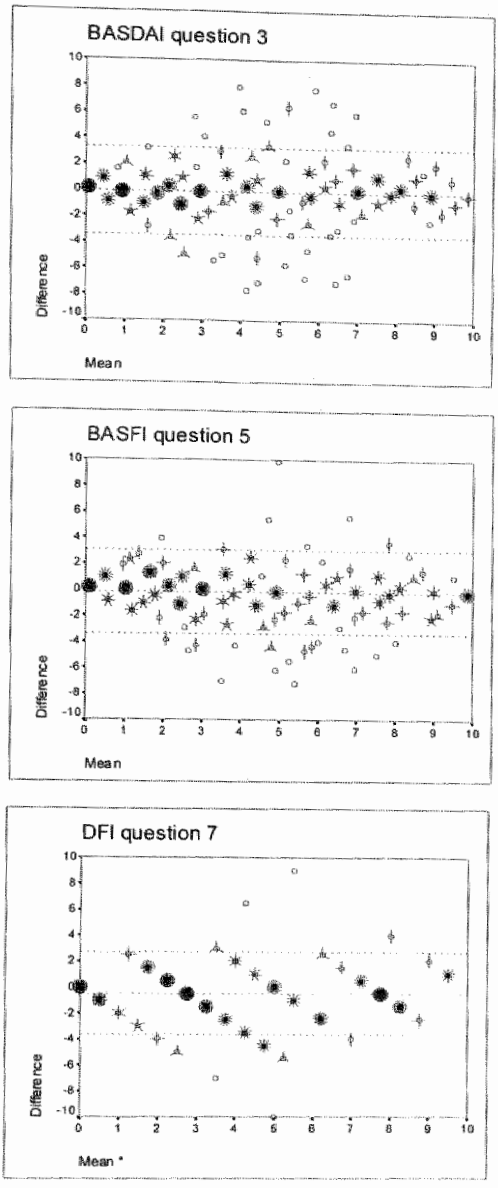

Figure 7.1. Bland-Altman plots

The scores on the original answer scales subtracted by the scores on the numerical rating scale plotted against the mean of these scores (95% level of agreement). Each center of a flower and each petal represents one case. BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; BASFI = Bath Ankylosing Spondylitis Functional Index; DFI = Dougados Functional Index. *Scores from the DFI were recoded to a 0-10 scale.

Table 7.3. Reproducibility assessed with intraclass correlation coefficients (ICCs) for individual questions and the total score on the BASDAI, BASFI, and DFI for the different answer modalities in Dutch and Mexican patients

	BASDAI		BASFI		DFI	
	VAS	NRS	VAS	NRS	Likert	NRS
Group 3 (controls spa therapy)						
Individual questions†	0.61 (0.43–0.72)	0.59 (0.41–0.69)	0.76 (0.54–0.88)	0.75 (0.57–0.84)	0.62 (0.41–0.84)	0.78 (0.57–0.90)
Total scores	0.64	0.62	0.88	0.89	0.85	0.88
38 patients from Mexico						
Individual questions†	0.47 (0.29–0.57)	0.58 (0.36–0.75)	0.62 (0.34–0.82)	0.62 (0.38–0.87)	0.62 (0.34–0.83)	0.58 (0.20–0.85)
Total scores	0.53	0.56	0.70	0.72	0.72	0.70

†Mean of ICCs (range). For abbreviations see Table 7.2.

Table 7.4. Smallest detectable difference (SDD) of the BASDAI, BASFI, and DFI on different answer modalities, and number of patients improved/worsened based on the corresponding SDD used as cut-off level in Dutch patients from the spa therapy trial

	BASDAI		BASFI		DFI	
	VAS	NRS	VAS	NRS	Likert	NRS
SDD	3.35	3.57	2.08	1.98	1.65	1.86
Intervention group improved†	9	8	15	17	10	10
Intervention group worsened†	0	1	0	1	0	0
Control group improved†	2	2	1	1	0	1
Control group worsened†	1	1	2	2	2	2

†Absolute number of patients provided: intervention group n=80; control group n=40. For abbreviations see Table 7.2.

Responsiveness

Table 7.5 shows the scores of all patients taking part in the spa therapy trial before and after the intervention. Statistically significant improvement ($P<0.001$) for the intervention group compared with baseline was found in all questionnaires using Wilcoxon signed rank test.

All questionnaires showed moderate to large responsiveness in the intervention group, and the control group remained reasonably stable (Table 7.6). Higher responsiveness scores were found with the method described by Guyatt. Responsiveness with the ES method showed the lowest scores. For the BASDAI, the responsiveness score was slightly higher on the VAS compared with NRS on all responsiveness methods. This was not observed in the BASFI. Except for the control group, no major differences in responsiveness on the Likert-scale and NRS were found in the DFI.

Table 7.5. Results on BASDAI, BASFI, and DFI from the intervention groups and control group of the spa therapy trial before and after the intervention on different answer modalities

Answer modality	Intervention group n=80	Control group n=40
BASDAI VAS		
Before	4.9 (1.9)	4.5 (2.0)
After	3.8 (2.2)*	4.1 (2.0)
BASDAI NRS		
Before	4.8 (1.9)	4.2 (2.1)
After	3.9 (2.3)*	4.0 (2.1)
BASFI VAS		
Before	4.6 (1.9)	4.2 (2.1)
After	3.6 (2.1)*	4.2 (2.2)
BASFI NRS		
Before	4.7 (2.0)	4.2 (2.2)
After	3.6 (2.2)*	4.2 (2.2)
DFI Likert		
Before	2.8 (1.3)	2.8 (1.4)
After	2.1 (1.4)*	2.7 (1.6)
DFI NRS		
Before	3.7 (1.8)	3.4 (1.9)
After	2.9 (1.9)*	3.5 (1.9)

Mean (SD). For abbreviations see Table 7.2. *$P<0.001$ compared with baseline.

Table 7.6. Responsiveness of BASDAI, BASFI, and DFI on different answer modalities calculated with three different responsiveness methods

	BASDAI		BASFI		DFI	
	VAS	NRS	VAS	NRS	Likert	NRS
SRM intervention group	0.60	0.51	0.76	0.74	0.95	0.85
ES intervention group	0.57	0.51	0.52	0.52	0.51	0.48
Guyatt method intervention group	0.64	0.54	0.93	1.04	0.81	0.89
SRM controls	0.22	0.08	-0.01	-0.02	0.07	-0.13
ES controls	0.19	0.07	0.00	-0.01	0.04	-0.07

SRM = Standardized response mean; ES = Effect size; See Table 7.2 for additional abbreviations. Positive changes imply improvement

Discussion

In this study, three commonly used AS-specific questionnaires, administered with both their original answer modality and an NRS, were judged with respect to feasibility (appropriateness of the answer modalities) and discrimination (reproducibility and responsiveness) criteria of the OMERACT Filter.

More patients were found to have difficulties in completing a VAS than either the Likert-scale or NRS. Eighty-seven percent of the patients preferred to answer on either an NRS or a Likert-scale, and only 9% on a VAS. A greater preference for the Likert-scale was also described by Kremer et al. in 57% of the patients studied.[10]

Bland-Altman plots showed a major variety with respect to answering the same question on different scales (Figure 7.1). The variation was not solely due to the different answer modalities: the question that was answered twice on the NRS also showed a substantial variability. The huge differences in scores give the impression that some patients did not fully understand or properly read the anchors of the scales. From the Bland-Altman plots it can be deduced that the variability is random. However, the variability of the total scores of the questionnaires answered on different answer scales was less impressive, as can be expected by aggregating different answers into one score, but was still substantial. The ICCs of the total scores were relatively high, implying a high degree of concordance between scores on different answer modalities.

The reproducibility of individual questions and total scores of the BASDAI, BASFI, and DFI was much lower than expected (Table 7.3). Large differences in all questionnaires were found between the scores obtained from the outpatients in Mexico and the control group of the spa therapy trial in The Netherlands. It is arguable whether age, cultural aspects, or the level of education were the basis for these differences. In the Dutch group, the mean ICCs of both individual and total scores of the BASDAI were lower than the minimum of 0.75; in the Mexican group none of the ICCs from both individual and total scores of the three questionnaires reached the minimum of 0.75. Only the total scores of the BASFI and DFI in the Dutch group showed acceptable ICCs. However, no major differences between the ICCs with respect to different answer modalities were found for all questionnaires.

The lack of reproducibility found in individual questions can have major implications for single-item questionnaires. For instance, the two questions on "pain" and "patient's global", both single-item questions and selected as specific outcome instruments in research in patients with AS, may lack a sufficient degree of responsiveness on an individual patient level, due to low reproducibility.[30] This lack of reproducibility merits further investigation.

Garrett et al. reported a test-retest reliability with a Pearson's correlation of 0.93 for the BASDAI.[1] Pearson's correlations were also published by Calin et al. for the BASFI ($r=0.89$) and DFI ($r=0.96$).[2] Whereas the Pearson's correlation is a measure of association, the ICC gives information about the concordance of the results, which is the degree to which the same results are found in the same, stable subjects at repeated measurements. Consequently the Pearson's correlation coefficients give higher results, which are difficult to interpret. The ICC is preferred in calculating the reproducibility.[28] Dougados et al. reported an ICC of 0.86 for the DFI, which is similar to our result in the Dutch group.[3]

Responsiveness was assessed with three different methods. Moderate to large effects were found in the intervention group with all three responsiveness methods, independent of the answer modality used. The method described by

Guyatt showed higher responsiveness than either the SRM or ES method in all questionnaires, with exception of the DFI answered on the Likert-scale. Ruof et al. also found that the Guyatt method showed higher responsiveness scores in their comparative study of the BASFI and DFI, in which they also stated that the BASFI was more responsive than the DFI.[31] The results of our study do not confirm the latter; the superiority of one of these questionnaires with respect to responsiveness appeared to be dependent on the responsiveness method applied.

Bolton and Wilkinson showed in their study that the responsiveness of measures was higher when using the NRS compared with VAS and Likert, although NRS and VAS were closely related.[17] In the present study, minor differences between the scales were found, with only the BASDAI showing consistently higher scores on the VAS compared with the NRS. In general, the different scales seem to have reasonably similar properties with respect to responsiveness.

In conclusion, although a major variability in individual questions on the original answer scales of the BASDAI, BASFI, and DFI compared with the NRS was found, total scores showed a high level of agreement. These results were found for all three countries, and for both clinical trial- and outpatients. The variability between the scales may entirely be explained by the low reproducibility of individual questions found in each of these questionnaires. All three questionnaires showed good responsiveness on both the original scales and on the NRS. The original answer modalities of the BASDAI, BASFI, and DFI can all be replaced by an NRS that maintains the properties of the original scales.

References

1. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.
2. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.
3. Dougados M, Gueguen A, Nakache JP, Nguyen M, Mery C, Amor B. Evaluation of a functional index and an articular index in ankylosing spondylitis. J Rheumatol 1988;15:302-7.
4. Ohnhaus EE, Adler R. Methodological problems in the measurement of pain: a comparison between the verbal rating scale and the visual analogue scale. Pain 1975;1:379-84.
5. Scott J, Huskisson EC. Graphic representation of pain. Pain 1976;2:175-84.
6. Huskisson EC, Jones J, Scott PJ. Application of visual-analogue scales to the measurement of functional capacity. Rheumatol Rehabil 1976;15:185-7.
7. Scott PJ, Huskisson EC. Measurement of functional capacity with visual analogue scales. Rheumatol Rehabil 1977;16:257-9.
8. Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. Pain 1983;17:45-56.
9. Dixon JS, Bird HA. Reproducibility along a 10 cm vertical visual analogue scale. Ann Rheum Dis 1981;40:87-9.
10. Kremer E, Atkinson JH, Ignelzi RJ. Measurement of pain: patient preference does not confound pain measurement. Pain 1981;10:241-8.
11. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986;27:117-26.
12. Bird HA, Dixon JS. The measurement of pain. Baillieres Clin Rheumatol 1987;1:71-89.
13. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. J Chronic Dis 1987;40:1129-33.
14. Ferraz MB, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022-4.
15. Eyres S, van der Heijde D, Dougados M, Tennant A. The Visual Analogue Scale: Deception on a 'VAS!' scale? [Submitted]
16. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. Ann Rheum Dis 1978;37:378-81.
17. Bolton JE, Wilkinson RC. Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. J Manipulative Physiol Ther 1998;21:1-7.

18. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? Pain 1994;58:387-92.
19. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology [editorial]. J Rheumatol 1998;25:198-9.
20. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417-32.
21. Creemers MC, Van 't Hof MA, Franssen MJ, Van de Putte LB, Gribnau FW, Van Riel PL. A Dutch version of the functional index for ankylosing spondylitis: development and validation in a long-term study. Br J Rheumatol 1994;33:842-6.
22. Ruof J, Sangha O, Stucki G. Evaluation of a German version of the Bath Ankylosing Spondylitis Functional Index (BASFI) and Dougados Functional Index (D-FI). Z Rheumatol 1999;58:218-25.
23. Creusen E, van der Heijde D, Spoorenberg A, et al. How to deal with missing answers in self-assessment questionnaires in ankylosing spondylitis: BASDAI-BASFI-ASFI? [Submitted]
24. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
25. Streiner D, Norman G. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press 1989.
26. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178-89.
27. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632-42.
28. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171-8.
29. Cohen J. Statistical power analysis for behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum Associates 1988.
30. van der Heijde D, Calin A, Dougados M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. J Rheumatol 1999;26:951-4.
31. Ruof J, Sangha O, Stucki G. Comparative responsiveness of 3 functional indices in ankylosing spondylitis. J Rheumatol 1999;26:1959-63.

Chapter 8

Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis

A van Tubergen, J Coenen, R Landewé, A Spoorenberg, A Chorus, A Boonen, Sj van der Linden, D van der Heijde

Arthritis and Rheumatism [Arthritis Care and Research] 2002; 47:8-16

Abstract

Objectives. To investigate whether the single-item fatigue question of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and the Multidimensional Fatigue Inventory MFI) are appropriate instruments to measure fatigue in ankylosing spondylitis (AS); To identify factors that influence fatigue in AS; To assess how fatigue in all its aspects is associated with quality of life in AS.

Methods. A total number of 812 patients with AS were included. Patients completed questionnaires on disease activity (BASDAI), functional ability (Bath Ankylosing Spondylitis Functional Index [BASFI]), global well-being (Bath Ankylosing Spondylitis Global Score [BASG]), overall perceived health (EuroQoL visual analog scale), and quality of life (Ankylosing Spondylitis Quality of Life Questionnaire and Short Form 36 [SF-36]). Patients were dichotomized into an F+ group (fatigue = major symptom) if the BASDAI fatigue score was >5.0 and an F- group (fatigue = minor symptom) if the fatigue score was <5.0. Reproducibility was assessed with intraclass correlation coefficients, and responsiveness was calculated according to three different methods. Logistic regression analysis was used to determine which factors were associated with fatigue. Multiple regression analysis was used to investigate whether fatigue contributes in explaining quality of life.

Results. Fifty-three percent of the patients were categorized to the F+ group. They scored significantly worse compared with the F- group with respect to each dimension of the MFI and to all other questionnaires studied (all $P<0.001$). The BASDAI fatigue question, as well as each separate dimension of the MFI, showed moderate to good reproducibility (0.57-0.75) and responsiveness (0.23-0.96). Fatigue was, in the opinion of the patients, highly associated with pain (70% of patients) and stiffness (54% of patients). Logistic regression analysis showed that scores on BASDAI, BASFI, BASG, and mental health status (SF-36) were independently associated with fatigue ($R^2=0.52$). Multiple regression analysis showed that scores on the BASDAI fatigue question were significantly associated with quality of life. With the five MFI dimensions as explanatory variables, different aspects of fatigue were associated with different domains of quality of life.

Conclusion. Fatigue as a major symptom of AS can effectively be measured with both a single-item question on the intensity of fatigue and with the MFI. The MFI, however, provides more insight in specific dimensions of fatigue. Fatigue appears to be associated with the level of disease activity, functional ability, global well-being, and mental health status. In addition, fatigue negatively influences different aspects of quality of life.

Introduction

Fatigue has been acknowledged as an important symptom in ankylosing spondylitis (AS).[1-4] Although fatigue is a widely used term, no uniform accepted definition is available. Belza described fatigue as "the enduring, subjective sensation of generalized tiredness or exhaustion".[5] Fatigue appears to be multifactorial and multidimensional: physiological, psychological, social, and personal factors influence the presence and experience of fatigue.[5-9]

During the Outcome Measures in Rheumatoid Arthritis Clinical Trials Conference in 1998 -in follow-up of the Assessments in Ankylosing Spondylitis workshop in 1995- fatigue was considered to be an important domain in the core set for disease controlling anti-rheumatic therapy in AS, but a specific instrument to measure fatigue was not selected, because little information was available on assessing fatigue in AS.[10] The Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) is a widely used instrument to measure disease activity in patients with AS, and includes one single question concerning the degree of fatigue.[2] Many questionnaires can be used for measuring fatigue multidimensionally, but none of them have been validated in AS.[6,11-14] The Multidimensional Fatigue Inventory (MFI) is such a multidimensional questionnaire.[12] The psychometric properties of the MFI have been validated and tested in several patient populations as well as in healthy subjects, but not yet in patients with AS.[12,15-17]

The first aim of this study was to investigate whether the single-item BASDAI fatigue question and the MFI are appropriate instruments to measure fatigue in AS. Important aspects of outcome measures such as reproducibility and responsiveness were investigated. After this fist stage, both instruments were used to investigate the impact of fatigue in AS.

The second aim of this study was to assess which factors influence fatigue in AS. This was investigated by obtaining the opinion of patients with AS. We hypothesized that fatigue could be influenced, among other factors, by disease activity, sleep disturbance, weather, medication, and work. In addition, we wanted to identify factors, measured by AS-specific instruments, that are most strongly associated with fatigue.

A final aim was to assess how fatigue in all its aspects influences quality of life in patients with AS. To date, little data have been reported on the relationship between fatigue and quality of life in patients with AS.[4]

Patients and Methods

Patients

A total of 812 patients with AS were derived from three sources: [1] Five hundred sixty-eight patients from 15 Dutch outpatient departments of rheumatology, who completed a postal questionnaire. These centers participated in the Standardized Diagnosis Register (SDR), a nation-wide diagnosis register of

rheumatic diseases.[18] Patients from the OASIS cohort study (see source 3 below) who also participated in this study were excluded, to avoid double counting. [2] One hundred twenty Dutch outpatients participating in a randomized controlled trial to assess the efficacy of a 3-week course of spa therapy in patients with AS.[19] Patients were randomly allocated to receive either spa therapy in Austria (arm 1, n=40) or in The Netherlands (arm 2, n=40), or to a control arm (n=40) that stayed at home and continued standard treatment (anti-inflammatory drugs and weekly group physical therapy). [3] One hundred twenty-four out of 137 Dutch patients participating in an international longitudinal observational study on outcome in AS (OASIS). Table 8.1 shows the distribution of the patients for the present study.

Table 8.1. Patient populations used for various aspects of the study

		Spa therapy trial			
	SDR (n=568)	Arm 1 Spa therapy Austria (n=40)	Arm 2 Spa therapy Netherlands (n=40)	Arm 3 Controls (n=40)	OASIS (n=124)
Comparison F+/F- scores questionnaires	X	X	X	X	X
Reproducibility fatigue questionnaires	-	-	-	X	-
Responsiveness fatigue questionnaires	-	X	-	(X)†	-
Factors influencing fatigue according to the patients	-	-	-	-	X
Factors implicitly influencing fatigue	X	-	-	-	X
Correlation fatigue with scores questionnaires	X	X	X	X	X
Influence fatigue on quality of life	X	-	-	-	X

SDR = Standardized diagnosis register of rheumatic diseases; OASIS = Outcome in ankylosing spondylitis international study; X = Participating in this aspect of the study. †Only as control arm in the method according to Guyatt (1987).

Questionnaires

The self-reports contained socio-demographic and disease-related questionnaires, specifically the MFI, BASDAI, Bath AS Functional Index (BASFI),[20] Bath AS Global Score (BASG),[21] and Short Form 36 (SF-36).[22]

The MFI consists of 20 items covering five dimensions: general fatigue, physical fatigue, reduced motivation, reduced activity, and mental fatigue. Each dimension has four statements, that are directed either positively or negatively. The answers are given on a 5-point scale (1 to 5), with the labels "yes, that is true", and "no, that is not true" at either end. Scores range from 4 to 20 on each dimension, with higher scores indicating more fatigue. A summary score of

the 20 items is discouraged. If a single score on fatigue is required, the dimension "general fatigue" should be used.[12]

The BASDAI, BASFI, and BASG are well established and widely used instruments in clinical trials and epidemiological studies to evaluate disease activity, functioning, and global well-being in patients with AS. They have been shown to be valid, reliable, and discriminatory.[2,20,23-29] The BASDAI consists of six questions answered on a visual analog scale (VAS).[2] The questions are related to fatigue, back pain, pain and/or swelling of the peripheral joints, localized tenderness, and duration and severity of morning stiffness. The 10 cm horizontal VAS has the labels none (=0) and very severe (=10) at either end for the first five questions, and 0 hours (=0) and 2 or more hours (=10) for the duration of morning stiffness. The mean of the two questions on morning stiffness counts as one variable. The mean of the five items is the total score. The BASFI contains 10 questions on functional ability, completed on a 10 cm horizontal VAS, with the labels easy (=0) and impossible (=10) at either end of the scale.[20] The mean of the 10 items is the final score. The BASG consists of two questions on the effect of AS on well-being over the past week and over the past 6 months respectively, and is answered on a 10 cm horizontal VAS, with the labels none (=0) and very severe (=10) at either end.[21] The mean of the two items is the final score. The ranges of the final scores for BASDAI, BASFI, and BASG are 0 (best) to 10 (worst).

The SF-36 is a widely applied generic instrument for measuring health status, and consists of eight domains: physical functioning, social functioning, role limitations (physical problem), role limitations (emotional problem), mental health, vitality, bodily pain, and general health perceptions.[22] Scores range from 0 (worst) to 100 (best).

Patients participating in the spa therapy trial and the OASIS cohort study also completed a disease-specific quality of life questionnaire (ASQoL),[30] and the VAS question from the EuroQoL (EQ-$5D_{VAS}$) on a person's overall perceived health.[31] The ASQoL is a newly developed disease-specific quality of life instrument containing 18 yes/no questions.[30] Scores range from 0 to 18, with lower scores implying a better quality of life. EQ-$5D_{VAS}$ scores range from 0 to 100, with higher scores implying a better health.

The question "Which factors increase your fatigue due to AS?" was administered only to the 124 OASIS patients and could be answered by the terms pain, stiffness, problems with a good sleep posture, the weather, exercise, poor sleep, work, medication, none of these factors, or other factors (with space to write down any factors). Multiple answers were allowed.

Co-morbidity was assessed in the SDR and OASIS study by asking patients to select co-morbid diseases from a list provided in the questionnaire or to add diseases not included in the list.

Statistical Analysis

Patients pooled from all three sources were divided into two groups. Patients with scores >5.0 on the BASDAI fatigue question were labeled as experiencing fatigue as a major symptom (F+). Patients with scores <5.0 were labeled as

experiencing fatigue as a minor symptom (F-). Data from the patients scoring exactly 5.0 on the first BASDAI question were omitted. The data were re-analyzed with other cut-off points (F- group <4.0, F+ group >6.0), but the results were similar (data not shown).

Student's *t*-tests were applied to compare continuously distributed characteristics between the F+ and F- groups. Dichotomous variables were analyzed by chi-square test.

To define reproducibility of the BASDAI fatigue question and the dimensions of the MFI, intraclass correlation coefficients (ICCs) were calculated for the control arm of the spa therapy trial, with a 6-week interval between the measurements without change of therapy. ICCs >0.75 were considered acceptable.[32]

Responsiveness of the BASDAI fatigue question and the dimensions of the MFI were determined by three different methods: the effect size (ES),[33] the standardized response mean (SRM),[34] and the method described by Guyatt.[35] The results were interpreted according to Cohen's effect size index in which 0.2 refers to a small change, 0.5 to a moderate change, and 0.8 or more to a large change.[36] The ES is calculated as the mean change after treatment compared with baseline, divided by the standard deviation (SD) of the baseline scores.[33] The SRM is calculated as the mean change after treatment compared with baseline, divided by the SD of the change score.[34] The method described by Guyatt is calculated as the mean change score in the treatment group divided by the SD of the change score in the control group.[35]

The distribution of factors believed to influence fatigue in F+ and F- patients was analyzed by chi-square test. Pearson's correlations were used to determine the relationship between both the BASDAI fatigue question and the dimensions of the MFI with other questionnaires. Logistic regression analysis was used to assess which factors were associated with fatigue in patients with AS. Independent variables comprised age, sex, disease duration, co-morbidity, mental health status (measured with the mental health domain from the SF-36), and the disease-specific measures BASDAI (without the fatigue question), BASFI, and BASG.

Multiple regression analysis was used to investigate whether fatigue contributes in explaining quality of life, independent of demographic and disease-related factors. Dependent variables were the dimensions of the SF-36 separately, and the ASQoL. Independent variables were either the BASDAI fatigue question (VAS score) or the five dimensions of the MFI. Adjustments were made for age, sex, co-morbidity, disease duration, mental health status, and BASFI. Because the BASDAI (without fatigue), BASFI, and BASG showed a high level of collinearity (tolerance <0.4), only the BASFI score was used as co-variate, and BASDAI and BASG were excluded.

Results

Table 8.2 shows the patient characteristics of the total study population and the three studies separately. Eighteen patients (14 from the SDR study and 4 from the spa therapy trial) were excluded, because they answered exactly 5.0 on the BASDAI fatigue question. An additional 18 patients (17 from the SDR and 1 from the OASIS study) did not complete the BASDAI fatigue question and were therefore also excluded. Overall, 53% (n=415) of the patients were included in the F+ group, and 47% (n=361) in the F- group. Both groups were similar with respect to age, disease duration, and duration of complaints. Significantly more women and significantly more co-morbidity were found in the F+ group compared with the F- group in the total population ($P=0.043$ and $P<0.001$, respectively).

Table 8.3 shows how patients in the F+ and F- groups performed on the MFI and other questionnaires. Patients from the F+ group had significantly worse scores on all questionnaires as compared with the F- group (all $P<0.001$).

Responsiveness and reproducibility

Six weeks after baseline measurement, and after the intervention had taken place, all patients from the spa therapy trial completed a second questionnaire. Reproducibility was calculated using the results of the control arm (n=40). The ICCs showed moderate to good concordance: the ICC for the BASDAI fatigue question was 0.60, and the ICC for the MFI dimension "general fatigue" was 0.67, for "physical fatigue" 0.57, for "reduced activity" 0.66, for "reduced motivation" 0.75, and for "mental fatigue" 0.75.

Table 8.4 shows the scores on the BASDAI fatigue question and on the MFI dimensions of the three spa therapy trial arms (n=40 for each arm). Improvements in the BASDAI fatigue question and in all dimensions of the MFI were observed after spa therapy in both intervention arms compared with controls. However, because arm 1 showed a greater and more prolonged improvement after the intervention compared with arm 2 (consistent with the results on the primary outcomes of the spa therapy trial),[19] we decided to calculate responsiveness scores only for arm 1 with the results at 3 months after spa therapy (time point of maximum effect) compared with baseline (Table 8.5). The BASDAI fatigue question showed high responsiveness with all three methods. For the MFI, the highest responsiveness was found for the dimensions "general fatigue" and "physical fatigue" irrespective of the method. The other dimensions showed moderate responsiveness scores.

Table 8.2. Patients characteristics of the total study population and the different study groups separately, divided into F+ and F- groups

	All patients (n=812)		SDR (n=568)		Spa therapy trial (n=120)		OASIS (n=124)	
	F+ (n=415)†	F- (n=361)†	F+ (n=290)	F- (n=247)	F+ (n=72)	F- (n=44)	F+ (n=53)	F- (n=70)
% Men	67*	74*	67*	75*	71	75	66	69
Age [years]	45 (9)	45 (11)	44 (9)	43 (10)	47 (10)	50 (10)	48 (11)	49 (12)
Disease duration [years]	12 (8)	13 (8)	12 (8)	13 (8)	11 6)	11 (6)	14 (10)	15 (8)
Duration of complaints [years]	22 (12)	23 (11)	n.a.	n.a.	18 (9)	18 (9)	27 (13)	26 (11)
% of patients with co-morbidity	48**	33**	50**	33**	n.a.	n.a.	40	33

Mean (SD). F+ = fatigue score on the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) >5.0, F- = fatigue score on BASDAI <5.0 (Data from 18 patients scoring exactly 5.0 as well as 18 patients with missing data on the BASDAI fatigue question are omitted from the F+ and F- groups); SDR = Standardized diagnosis register of rheumatic diseases; OASIS = Outcome in ankylosing spondylitis international study; n.a.= data not available. †Patient characteristics are calculated on available data. *$P<0.05$, **$P<0.001$ between F+ and F- groups.

Table 8.3. Scores of the F+ and the F- groups on the dimensions of the Multidimensional Fatigue Inventory and several questionnaires

Questionnaire (range)	All patients	
	F+ (n=415)	F- (n=361)
MFI (4 – 20)		
General fatigue	15.5 (3.3)	10.1 (3.5)
Physical fatigue	14.2 (3.3)	9.9 (3.7)
Reduced activity	11.5 (4.0)	8.6 (3.5)
Reduced motivation	10.4 (3.6)	8.0 (3.1)
Mental fatigue	10.1 (4.3)	7.8 (3.6)
BASDAI (0 – 10)	5.6 (1.8)	2.4 (1.5)
BASDAI without fatigue question (0 – 10)	5.2 (2.1)	2.5 (1.7)
BASDAI fatigue question (0 – 10)	7.4 (1.4)	2.4 (1.5)
BASFI (0 – 10)	5.0 (2.2)	2.6 (2.0)
BASG (0 – 10)	5.9 (2.1)	2.9 (2.1)
ASQoL (0 – 18)†	9.6 (3.8)	4.3 (3.4)
EQ-$5D_{VAS}$ (0 – 100)†	53.7 (18.3)	70.0 (14.8)
SF-36 (0 – 100)		
Physical functioning	55.8 (23.4)	75.6 (18.7)
Social functioning	64.6 (24.6)	84.3 (19.4)
Role limitation (physical)	35.6 (38.1)	73.4 (36.8)
Role limitation (emotional)	68.5 (41.3)	85.8 (29.5)
Mental health	65.5 (18.1)	77.3 (15.5)
Vitality	42.6 (16.5)	64.5 (16.9)
Bodily pain	48.3 (19.2)	69.8 (17.9)
General health	42.8 (20.2)	59.9 (18.7)

Mean (SD). MFI = Multidimensional Fatigue Inventory; BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; BASFI = Bath Ankylosing Spondylitis Functional Index; BASG = Bath Ankylosing Spondylitis Global; ASQoL = Ankylosing Spondylitis Quality of Life questionnaire; EQ-$5D_{VAS}$ = EuroQol visual analogue scale; SF-36 = Short Form 36. F+ = fatigue score on the BASDAI >5.0, F- = fatigue score on the BASDAI <5.0. †Data available on spa therapy trial and OASIS patients only (F+ n=125; F- n=114). All differences between scores in the F+ and F- groups were statistically significant at the 0.001 level.

Factors influencing and associated with fatigue

More than 50% of all OASIS patients (n=124) associated fatigue with pain (70%) and stiffness (54%) (Table 8.6). The factors perceived to be associated with fatigue, and presented in a hierarchical order, were grossly similar in the F+ and F- groups, but pain, poor sleep, and work were significantly more often considered important in the F+ group.

Table 8.4. Results from the intervention arms and control arm of the spa therapy trial on the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) fatigue question and the dimensions of the Multidimensional Fatigue Inventory before and after the intervention

	BASDAI fatigue question	Dimensions Multidimensional Fatigue Inventory				
		General fatigue	Physical fatigue	Reduced activity	Reduced motivation	Mental fatigue
Arm 1 (n=40)						
Baseline	5.6 (2.7)	13.6 (4.0)	12.7 (4.3)	10.8 (3.9)	9.5 (3.3)	9.5 (5.2)
1 week after end spa therapy	5.7 (3.4)	12.2 (5.3)	10.6 (5.1)**	10.1 (4.1)	8.6 (3.2)	8.0 (4.0)*
3 months after end spa therapy	3.2 (2.3)*	10.2 (4.7)**	9.2 (4.0)**	9.7 (3.9)	7.7 (3.0)**	7.5 (3.9)*
Arm 2 (n=40)						
Baseline	5.9 (2.5)	14.8 (4.7)	13.0 (5.0)	10.9 (4.4)	8.6 (3.2)	10.2 (4.7)
1 week after end spa therapy	5.9 (3.1)	13.5 (5.2)	10.6 (5.1)**	10.4 (4.9)	8.8 (3.8)	10.2 (4.9)
3 months after end spa therapy	4.8 (2.7)	12.9 (4.7)	11.4 (5.0)	9.7 (4.1)*	8.8 (3.5)	10.2 (5.0)
Control arm (n=40)						
Baseline	5.9 (2.7)	13.6 (4.4)	12.0 (3.9)	10.4 (3.6)	8.6 (3.8)	9.2 (4.4)
1 week	5.3 (2.6)	13.5 (3.7)	12.5 (4.1)	10.6 (3.6)	8.6 (3.5)	9.2 (4.1)
3 months	4.9 (2.4)	13.3 (4.6)	12.1 (4.1)	10.8 (4.1)	9.0 (3.7)	9.2 (4.6)

Mean (SD). Arm 1 = spa therapy in Austria; Arm 2 = spa therapy in The Netherlands; Control arm = continued standard treatment at home. See Patients and Methods section for possible score ranges. *$P<0.05$ and **$P<0.01$ compared with controls.

Table 8.5. Responsiveness of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) fatigue question and the dimensions of the Multidimensional Fatigue Inventory calculated with three different responsiveness methods

	BASDAI fatigue question	Dimensions Multidimensional Fatigue Inventory				
		General fatigue	Physical fatigue	Reduced activity	Reduced motivation	Mental fatigue
Effect size	0.89	0.82	0.81	0.28	0.54	0.38
Standardized response mean	0.89	0.70	0.82	0.23	0.51	0.49
Guyatt method	0.92	0.86	0.96	0.30	0.50	0.57

For calculation of the different responsiveness methods see Patients and Methods section of the text. Responsiveness was measured with the data of the patients from arm 1 of spa therapy trial (n=40). Positive changes imply improvement.

Table 8.6. Factors influencing fatigue in the opinion of the patients from the OASIS study

	All patients (n=124)	F+ (n=53)	F- (n=70)	P value (χ^2-test)
Pain	87 (70)	44 (83)	42 (60)	0.011
Stiffness	67 (54)	31 (58)	35 (50)	0.456
Weather	55 (44)	28 (53)	26 (37)	0.128
Poor sleep	49 (40)	30 (57)	18 (26)	<0.001
Sleep posture	45 (36)	24 (45)	20 (29)	0.088
Work	27 (22)	19 (36)	8 (11)	0.003
Exercises	10 (8)	5 (9)	4 (6)	0.667
Medication	9 (7)	5 (9)	4 (6)	0.667
None of these factors	2 (2)	0 (0)	2 (3)	0.603
Other factors	0 (0)	0 (0)	0 (0)	-

Absolute number (%) of patients. OASIS = Outcome in ankylosing spondylitis international study; F+ = fatigue score on the BASDAI >5.0, F- = fatigue score on the BASDAI <5.0. Multiple answers were allowed.

The scores of five disease-specific questionnaires as well as the EQ-$5D_{VAS}$, and all dimensions of the SF-36 were significantly correlated with the BASDAI fatigue question and all dimensions of the MFI separately (Table 8.7). No major differences in correlations between these questionnaires and either the BASDAI fatigue question or the "general fatigue" and "physical fatigue" dimensions of the MFI were found. The dimensions "reduced activity", "reduced motivation", and "mental fatigue" were correlated to a lesser degree. When using the "vitality" domain of the SF-36 as an external validation to measure fatigue, similar correlations were found between the above mentioned questionnaires and the "vitality" domain, confirming the robustness of the results (data not shown).

To investigate which domains of the MFI contributed independently in explaining "fatigue" as elicited by the BASDAI fatigue question, logistic regression analysis was performed with F+/F- (n=812) as dependent variable, and the five MFI dimensions as well as age, sex, and disease duration as explanatory variables. General fatigue ($P<0.001$), physical fatigue ($P=0.02$), reduced activity ($P=0.042$), and reduced motivation ($P=0.008$) were independently associated with F+/F-, whereas mental fatigue ($P=0.93$) was not. This confirmed the hypothesis that fatigue is a multidimensional attribute.

In a separate logistic regression analysis (n=776), the scores of the disease-specific questionnaires BASDAI (without fatigue), BASFI, and BASG were assessed for their independent association with F+/F-, after controlling for age, sex, co-morbidity, mental health status, and disease duration (Table 8.8). The BASDAI-, BASFI-, and BASG score, mental health status, and age were selected as independently associated with fatigue. The age effect was only marginally statistically significant, suggesting that fatigue mainly relates to disease activity, functional ability, global well-being, and mental health status. In a multiple regression analysis model with the "vitality" domain of the SF-36 as the dependent factor, similar results were found (data not shown).

Table 8.7. Correlations between several questionnaires and both the BASDAI fatigue question and dimensions of the MFI

	BASDAI fatigue question	Dimensions Multidimensional Fatigue Inventory				
		General fatigue	Physical fatigue	Reduced activity	Reduced motivation	Mental fatigue
BASDAI	.79	.60	.61	.44	.37	.33
BASDAI without fatigue question	.67	.51	.56	.40	.34	.31
BASFI	.57	.51	.60	.45	.37	.28
BASG	.68	.56	.63	.44	.37	.32
ASQoL†	.67	.71	.73	.59	.47	.46
EQ-5D$_{VAS}$†	-.53	-.52	-.60	-.41	-.36	-.25
SF-36						
Physical functioning	-.51	-.47	-.59	-.44	-.35	-.28
Social functioning	-.49	-.49	-.55	-.52	-.42	-.39
Role limitation (physical)	-52	-.50	-.51	-.42	-.29	-.30
Role limitation (emotional)	-.29	-.31	-.32	-.37	-.27	-.35
Mental health	-.40	-.44	-.42	-.47	-.49	-.49
Vitality	-.64	-.73	-.65	-.59	-.53	-.42
Bodily pain	-.59	-.53	-.60	-.45	-.36	-.28
General health	-.48	-.53	-.64	-.45	-.39	-.30

†Correlations were measured in the entire study population (n=812), except for the ASQoL and EQ-5D$_{VAS}$, for which data of spa therapy trial and OASIS patients only were available (n=244). For abbreviations see Table 8.3. All correlations were significant at the 0.01 level.

Fatigue and quality of life

Multiple regression analysis was used to assess whether fatigue contributed in explaining quality of life, as assessed by dimensions from the SF-36 and the ASQoL, after controlling for age, sex, co-morbidity, disease duration, mental health status, and BASFI (n=776). Dependent variables of the SF-36 were physical functioning, social functioning, vitality, bodily pain, and general health, respectively. Role limitation (physical) and role limitation (emotional) were excluded, because neither were normally distributed, nor after log transformation. Mental health was excluded as a dependent variable, because this was used in the analyses as an independent variable to control for mood disorders.

Fatigue as elicited by the BASDAI fatigue question was significantly associated with all separate dimensions of the SF-36 studied and with the ASQoL (Table 8.9). Fatigue as elicited by the MFI was related to quality of life in a different pattern. Different aspects of fatigue (measured with the five MFI dimensions) explained different domains of quality of life measured with the SF-36 (Table 8.9).

Table 8.8. Logistic regression model for explaining fatigue

Variable	RR	95% CI	*P* value
BASDAI (without fatigue)	1.47	1.28 to 1.69	<0.001
BASG	1.34	1.19 to 1.50	<0.001
BASFI	1.16	1.03 to 1.31	0.018
Mental health	0.98	0.97 to 1.00	0.012
Age	0.98	0.95 to 1.00	0.046
Sex	1.12	0.71 to 1.76	0.627
Disease duration	0.99	0.96 to 1.01	0.294
Co-morbidity	0.84	0.53 to 1.31	0.435

Data of the SDR and OASIS study were used (n=776). Dependent variable: F+/F- group. All independent variables were entered simultaneously into the model. R^2=0.52. RR = Relative risk; CI = Confidence interval; See Table 8.3 for additional abbreviations.

Table 8.9. Multiple regression analyses to study whether fatigue, measured with either the BASDAI fatigue question or the dimensions of the MFI, contributes in explaining quality of life as assessed by the SF-36 and the ASQoL.

	BASDAI	Dimensions Multidimensional Fatigue Inventory				
Dependent variable†	fatigue question	General fatigue	Physical fatigue	Reduced activity	Reduced motivation	Mental fatigue
SF-36						
Physical functioning	-0.061*	0.139	-1.119***	-0.413*	0.589**	0.123
Social functioning	-0.066*	-0.003	-0.522	-0.936***	0.265	0.121
Vitality	-0.258***	-1.875***	-0.346	-0.488**	-0.516**	0.138
Bodily pain	-0.179***	-0.458*	-1.082***	-0.062	0.010	0.244
General health	-0.118***	-0.082	-2.045***	-0.118	-0.129	0.260
ASQoL	0.055***	0.261***	0.226**	0.038	-0.040	-0.039

Beta coefficients are shown for each of the models. †Dependent variables were five of the eight dimensions of the SF-36 separately, and the ASQoL. Independent variables were either the BASDAI fatigue question or the five MFI dimensions. Adjustments were made for age, sex, co-morbidity, disease duration, mental health status, and BASFI score. For the separate dimensions of the SF-36, the data of the SDR and OASIS study (n=776) were used; for the ASQoL, only data of the OASIS study were available (n=124). See Table 8.3 for abbreviations. *$P<0.05$, **$P<0.01$, ***$P<0.001$.

Discussion

Fatigue has been described as a major symptom in AS, with up to 65% of the patients reporting it.[1-4] In the present study 53% of all patients experienced fatigue, defined as a BASDAI fatigue score of >5.0.

We assessed the properties of both the single-item BASDAI fatigue question and a multidimensional fatigue questionnaire, the MFI. The advantage of a single question is that it is easy to complete. Important disadvantages are that

detailed information is lost with respect to the type of fatigue, and that it does not take into account differences in fatigue experience.[12] The advantage of a multidimensional questionnaire is that it deals with different aspects of fatigue, i.e. it provides a profile of fatigue.

Within the field of rheumatology the MFI has been used to investigate fatigue in patients with primary Sjögren syndrome compared with patients with rheumatoid arthritis (RA) and healthy controls.[37] No differences in fatigue scores between patients with primary Sjögren syndrome and RA were found, and both groups reported significantly more fatigue than healthy controls. Depression appeared to be an important confounder. After controlling for depression, significant differences with respect to the dimensions "reduced motivation" and "mental fatigue" disappeared between the patient groups and healthy controls. In the present study, a specific questionnaire to measure depression was not administered to the patients. To address this issue, we controlled for mood disorders in the regression analyses by using the mental health domain from the SF-36. However, analyses with and without the mental health domain as an independent factor were not substantially different (data not shown).

The reproducibility of the BASDAI fatigue question and each dimension of the MFI was moderate to good according to accepted standards, and more or less similar for both questionnaires. However, the ICCs never exceeded 0.75. This might be due to the relatively long interval between the measurements (6 weeks). Although reproducibility was measured in a group of patients presumed to be stable, it may have been influenced by fluctuating behavior of fatigue in AS. The responsiveness of the BASDAI fatigue question and of each dimension of the MFI was moderate to good, and comparable for both instruments (Table 8.5). From the present study it can be concluded that the BASDAI fatigue question and the MFI have equal properties with respect to reproducibility and responsiveness.

The patients reported a considerable number of factors that, in their opinion, were associated with fatigue. More than 50% of the patients reported that pain and stiffness negatively influenced their fatigue. Fatigue due to pain may be explained by several mechanisms.[38] Dealing with pain may require both mental and physical energy, more energy may be necessary to perform daily tasks in a less painful way, and pain may cause sleep disturbances with consequent daytime fatigue. Except for pain, poor sleep, and work, however, none of the other factors studied discriminated between F+ and F-. In a study by Jones et al. the weather was reported most often by patients with AS to negatively influence fatigue.[3] Exercises were reported to have an increasing effect on fatigue by 37% of the strongly fatigued patients, and by 10% of the less fatigued patients. In our study, exercises were not reported to be of substantial negative influence on fatigue by both groups. Contrary, the level of fatigue had decreased after a 3-week course of spa therapy with an intensive physical exercise program (Table 8.4). Patients who were strongly fatigued before the intervention did not show a worsening of the symptoms after the spa therapy. In contrast, they showed more improvement in fatigue than the patients from the F- group (data not shown).

Fatigue implicitly appeared to be related to disease activity, functional ability, global well-being, mental health status, and age, explaining 52% of the variance. In agreement with these findings, Jones et al. found that pain, stiffness, and functional ability were significantly associated with the level of fatigue, explaining 42% of the variance.[3] In Jones' study, age was not significantly associated with fatigue. In the present study, the age effect was only marginally statistically significant, and not considered clinically relevant. The pain and stiffness components of Jones' study were incorporated in our study in the BASDAI, and therefore not selected separately. Global well-being and mental health status were not entered into the regression models by Jones et al., but were found to be important factors in explaining fatigue in our study.

The final aim of the study was to investigate whether fatigue contributes in explaining quality of life. Scores on the BASDAI fatigue question were significantly associated with scores on several dimensions of the SF-36 and with the ASQoL, suggesting that quality of life is influenced by the degree of fatigue. Multiple regression analysis further suggested that different aspects of fatigue explained different domains of quality of life. Such an interrelationship is missed if a single-item questionnaire is used. This differential assessment of various aspects of the SF-36 further validates the MFI.

In conclusion, fatigue as a major symptom of AS can appropriately be measured with both a single-item question on the intensity of fatigue and with the MFI. The MFI, however, provides more insight in specific dimensions of fatigue. Fatigue appears to be associated with the level of disease activity, functional ability, global well-being, and mental health status. In addition, fatigue negatively influences different aspects of quality of life.

References

1. Calin A, Edmunds L, Kennedy LG. Fatigue in ankylosing spondylitis--why is it ignored? J Rheumatol 1993;20:991-5.
2. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.
3. Jones SD, Koh WH, Steiner A, Garrett SL, Calin A. Fatigue in ankylosing spondylitis: its prevalence and relationship to disease activity, sleep, and other factors. J Rheumatol 1996;23:487-90.
4. Ward MM. Health-related quality of life in ankylosing spondylitis: a survey of 175 patients. Arthritis Care Res 1999;12:247-55.
5. Belza B. The impact of fatigue on exercise performance. Arthritis Care Res 1994;7:176-80.
6. Piper BF, Lindsey AM, Dodd MJ, Ferketich S, Paul SM, Weller S. The development of an instrument to measure the subjective dimension of fatigue. In: Funk SG, Tournquist EM, Champagne MT, Copp LA, Weise RA, editors. Key aspects of comfort: Management of pain, fatigue, and nausea. New York: Springer 1989. p. 199-208.
7. Belza Tack B. Self-reported fatigue in rheumatoid arthritis. Arthritis Care Res 1990;3:154-7.
8. Aaronson LS, Teel CS, Cassmeyer V, et al. Defining and measuring fatigue. Image J Nurs Sch 1999;31:45-50.
9. Tiesinga LJ, Dassen TW, Halfens RJ. Fatigue: a summary of the definitions, dimensions, and indicators. Nurs Diagn 1996;7:51-62.
10. van der Heijde D, Calin A, Dougados M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. J Rheumatol 1999;26:951-4.
11. Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. Clin Infect Dis 1994;18 (suppl 1):S79-83.
12. Smets EM, Garssen B, Bonke B, De Haes JC. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. J Psychosom Res 1995;39:315-25.
13. Belza BL. Comparison of self-reported fatigue in rheumatoid arthritis and controls. J Rheumatol 1995;22:639-43.
14. Piper BF, Dibble SL, Dodd MJ, Weiss MC, Slaughter RE, Paul SM. The revised Piper Fatigue Scale: psychometric evaluation in women with breast cancer. Oncol Nurs Forum 1998;25:677-84.
15. Smets EM, Garssen B, Cull A, de Haes JC. Application of the multidimensional atigue inventory (MFI-20) in cancer patients receiving radiotherapy. Br J Cancer 1996;73:241-5.

16. Schneider RA. Reliability and validity of the Multidimensional Fatigue Inventory (MFI-20) and the Rhoten Fatigue Scale among rural cancer outpatients. Cancer Nurs 1998;21:370-3.
17. Meek PM, Nail LM, Barsevick A, et al. Psychometric testing of fatigue instruments for use with cancer patients. Nurs Res 2000;49:181-90.
18. Miedema HS, van der Linden SM, Rasker JJ, Valkenburg HA. National database of patients visiting rheumatologists in The Netherlands: the Standard Diagnosis Register of Rheumatic Diseases. A report and preliminary analysis. Br J Rheumatol 1998;37:555-61.
19. van Tubergen A, Landewé R, van der Heijde D, et al. Combined spa-exercise therapy is effective in patients with ankylosing spondylitis: a randomized controlled trial. Arthritis Rheum 2001;45:430-8.
20. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.
21. Jones SD, Steiner A, Garrett SL, Calin A. The Bath Ankylosing Spondylitis Patient Global Score (BAS-G). Br J Rheumatol 1996;35:66-71.
22. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
23. Ward MM, Kuzis S. Validity and sensitivity to change of spondylitis-specific measures of functional disability. J Rheumatol 1999;26:121-7.
24. Ruof J, Stucki G. Comparison of the Dougados Functional Index and the Bath Ankylosing Spondylitis Functional Index. A literature review. J Rheumatol 1999;26:955-60.
25. Spoorenberg A, van der Heijde D, de Klerk E, et al. A comparative study of the usefulness of the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index in the assessment of ankylosing spondylitis. J Rheumatol 1999;26:961-5.
26. Calin A, Nakache JP, Gueguen A, Zeidler H, Mielants H, Dougados M. Outcome variables in ankylosing spondylitis: evaluation of their relevance and discriminant capacity. J Rheumatol 1999;26:975-9.
27. Ruof J, Sangha O, Stucki G. Comparative responsiveness of 3 functional indices in ankylosing spondylitis. J Rheumatol 1999;26:1959-63.
28. Calin A, Nakache JP, Gueguen A, Zeidler H, Mielants H, Dougados M. Defining disease activity in ankylosing spondylitis: is a combination of variables (Bath Ankylosing Spondylitis Disease Activity Index) an appropriate instrument? Rheumatology Oxford 1999;38:878-82.
29. van Tubergen A, Debats I, Ryser L, et al. The use of a numerical rating scale as an answer modality in ankylosing spondylitis specific questionnaires. Arthritis Rheum [In press].
30. Helliwell P, Doward L, Whalley D, et al. Psychometric and scaling properties of a new quality of life instrument specific to ankylosing spondylitis (AS) [abstract]. Arthritis Rheum 1999;42:S72.
31. The EuroQoL Group. Euroqol-a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.

32. Streiner D, Norman G. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press 1989.
33. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178-89.
34. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632-42.
35. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis 1987;40:171-8.
36. Cohen J. Statistical power analysis for behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum Associates 1988.
37. Barendregt PJ, Visser MR, Smets EM, et al. Fatigue in primary Sjogren's syndrome. Ann Rheum Dis 1998;57:291-5.
38. Belza BL, Henke CJ, Yelin EH, Epstein WV, Gilliss CL. Correlates of fatigue in older adults with rheumatoid arthritis. Nurs Res 1993;42:93-9.

Chapter 9

Assessment of disability with the WHODAS II in patients with ankylosing spondylitis

A van Tubergen, R Landewé, L Heuft-Dorenbosch, A Spoorenberg,
D van der Heijde, H van der Tempel, Sj van der Linden

Annals of the Rheumatic Diseases 2002 [in press]

Abstract

Objectives. To investigate in ankylosing spondylitis (AS) whether the newly developed World Health Organization Disability Assessment Schedule II (WHODAS II) is a useful instrument for measuring disability, to assess its responsiveness in relation to other, traditional disease-specific instruments, and to identify factors that are associated with both short-term and long-term scores on the WHODAS II.

Methods. Altogether 214 patients with AS from a randomized controlled trial assessing the efficacy of spa therapy (n=117) and from a 5-year longitudinal observational study (n=97) participated. The patients completed several self-administered questionnaires, including the WHODAS II. After a 3-week course of spa therapy, 31 patients completed again all questionnaires in order to assess responsiveness. To determine to what degree the WHODAS II reflects a number of AS-oriented measures on disease activity, physical functioning, and quality of life, correlation coefficients between the WHODAS II and these other questionnaires were calculated. Responsiveness was calculated by effect size (ES) and standardized response mean (SRM). Linear regression analysis was performed to explore which factors might be associated with short-term changes on the WHODAS II and to investigate (in the observational study) which factors might predict disability (WHODAS II), 5 years later.

Results. Mean score on the WHODAS II was 23.9±15.5 (range 0.0 to 76.1). Scores on the WHODAS II were significantly correlated with all disease-specific questionnaires measured (all $P<0.001$). The WHODAS II showed a comparable short-term responsiveness score (SRM 0.41; ES 0.39). In the regression analysis, these short-term changes on the WHODAS II were significantly associated with changes in physical functioning (beta coefficient 4.25 [95% confidence interval 1.24-7.26], $P=0.007$). In the observational study, disease activity (beta coefficient 0.35 [0.17-0.53], $P<0.000$) as well as physical functioning (beta coefficient 0.23 [0.09-0.38], $P=0.002$) appeared to significantly predict disability (WHODAS II) after 5 years.

Conclusion. The WHODAS II is a useful instrument for measuring disability in AS in that it truthfully reflects disease-specific instruments and that it shows similar responsiveness scores. In AS, a short-term change on the WHODAS II is associated with a change in physical function. At the group level, disease activity and physical functioning may predict disability after 5 years.

Introduction

Ankylosing spondylitis (AS) is a chronic rheumatic disease that may cause serious functional impairment. In the past decade, several questionnaires have been developed to measure this functional impairment in AS.[1-4] "Disability" in AS is often equated with functional impairment and is measured with the same questionnaires. Disability, however, encompasses more than loss of physical function alone. It is a multidimensional construct in which problems at the bodily, person, and social level are united.[5] Recently, the World Health Organization (WHO) has revised the International Classification of Impairments, Disabilities, and Handicaps (ICIDH) towards a more bio-psycho-social model: the International Classification of Functioning, Disability, and Health (ICF), in order to comprehend human functioning at the bodily, personal, and social levels.[5-7] To assess functioning and disability, the WHO has developed a new instrument that is conceptually compatible with the ICF: the World Health Organization Disability Assessment Schedule II (WHODAS II). The WHODAS II is a multidimensional questionnaire, which can be used for measuring the level of disability across a variety of conditions and interventions. It includes six domains: understanding and communicating, getting around, self care, getting along with others, household and work activities, and participation in society. The WHODAS II is translated and validated in many languages. A number of versions of the WHODAS II is available through the WHO: a fully structured self-administered version, an interviewer-administered version and a proxy version, and of each a short and a long item version exists. A semi-structured questionnaire is under development.

Because the WHODAS II, as a generic instrument, should be applicable to all kinds of diseases, we wanted to assess to what degree the WHODAS II would also be a useful instrument for measuring disability in a single disease, in our case AS. In order to be such a useful instrument, it should reflect usual disease-specific instruments. Therefore, the first aim of this study was to assess to what degree the WHODAS II correlates with AS-oriented instruments, and to define its responsiveness after an efficacious intervention in relation to these other instruments. As an external criterion we used the widely used, generic instrument the Short Form 36 (SF-36), which is also frequently applied in patients with AS.[8,9] As a second validation step, we wanted to assess which -if any- disease-specific or socio-demographic factors were associated with short-term changes in WHODAS II scores after an intervention. We hypothesized that a short-term change on the WHODAS II would be independent of socio-demographic factors and factors that are subject to short-term fluctuations, but is dependent on structural changes. The last aim of our study was to investigate at the group level whether there are AS-oriented instruments that may predict outcome on disability, as measured by the WHODAS II, 5 years later.

Patients and Methods

Patients

Altogether 214 patients with AS from two different studies participated: Study group A consisted of 117 out of 120 Dutch outpatients who had participated in a randomized controlled trial to assess the efficacy of a 3-week course of spa-exercise therapy in AS. They completed the WHODAS II, a number of AS-oriented questionnaires, and the SF-36 at the end of the study period (January 2000). From this trial it was concluded that spa-exercise therapy is an efficacious treatment for patients with AS.[10] Patients from the "control group" of this trial were also offered a 3-week course of spa therapy after completion of the study period. Thirty-one patients from this group completed all questionnaires a second time 1 week after the end of the spa therapy to assess responsiveness of these questionnaires. Study group B consisted of 97 out of 137 Dutch patients participating in an ongoing international longitudinal observational study on outcome in AS (OASIS). They completed the WHODAS II, a number of AS-oriented questionnaires, and the SF-36, and underwent physical examination, laboratory tests, and radiographs of the axial skeleton in the summer of 2001. Patients from this study have been followed up since the summer of 1996. During this 5-year follow-up period, 40 patients had withdrawn from the observational study for various reasons.

Methods

All patients completed self-reports containing socio-demographic questions and a number of (AS-oriented) questionnaires (see below). We used the self-administered 36-item version of the WHODAS II in Dutch. The WHODAS II consists of 36 Likert-formatted questions, divided into six domains. The final score is calculated with an SPSS syntax (available through the WHO). We used two different syntax versions: the 36-version if the patient has a job, and the 32-version if patient does not have a job. In the latter version, the four questions concerning work ability are omitted. The final score ranges from 0 (best) to 100 (worst).

Physical function was measured with the Bath Ankylosing Spondylitis Functional Index (BASFI),[3] the Dougados Functional Index (DFI),[2,11] and the Health Assessment Questionnaire for Spondylarthropathies (HAQ-S).[1] The BASFI contains 10 questions concerning activities of daily living, and is scored on a 10 cm visual analog scale (VAS), with anchors "easy" and "impossible" at either side. The mean of the items defines the final score, with scores ranging from 0 (best) to 10 (worst). The DFI consists of 20 Likert-formatted scales, and includes activities of daily living. Originally, the DFI is completed on a 3-point Likert-scale, and later modified to a 5-point Likert-scale. In the present study we have applied the 5-point Likert-scale. For each item possible scores are 0, 0.5, 1, 1.5, and 2. Total scores range from 0 (best) to 40 (worst). The HAQ-S consists of eight subscales on health status and physical function, and is extended with five

additional AS-specific items on physical function, divided into two subscales. The questions are answered on Likert-scales. The mean of all subscales defines the final score, ranging from 0 (best) to 3 (worst). The HAQ-S was administered to the patients from the spa therapy trial only.

Disease activity was measured with the Bath AS Disease Activity Index (BASDAI),[12] which consists of six questions on fatigue, pain of the spine, pain and/or swelling of the peripheral joints, localized tenderness, and severity and duration of morning stiffness. The questions are answered on a 10 cm VAS, anchored with the labels "none" and "very severe" at either end of the first five questions, and with "0 hours" and "2 hours" in the question on duration of morning stiffness. The mean of the two questions on morning stiffness counts as one variable. The final score is defined by calculating the mean of the five items. Scores range from 0 (best) to 10 (worst).

Patient global well-being and pain of the spine were measured with a 10 cm VAS. The scores range from 0 (best) to 10 (worst). Quality of life was measured with the ASQoL[13] and the Short Form-36 (SF-36).[8] The ASQoL is a newly developed AS-oriented quality of life instrument containing 18 yes/no questions. Scores range from 0 to 18 with lower scores implying a better quality of life. The SF-36 is a widely applied generic instrument for measuring health status and consists of eight domains: physical functioning, social functioning, role limitations (physical problem), role limitations (emotional problem), mental health, vitality, bodily pain, and general health perceptions. Scores range from 0 (worst) to 100 (best).

The patients from the OASIS study also underwent physical examination, and laboratory and radiographic assessments at regular intervals. Measures for determining spinal mobility consisted of the Schober index, finger-to-floor distance, chest expansion, and occiput-to-wall distance. The Schober test was measured as the increase in the distance after maximal forward flexion on a 10 cm segment marked on the patient's back. Finger-to-floor distance was measured as the distance between the patients' fingertips and the floor after forward flexion with knees straight and feet together. Chest expansion was measured as the difference in chest circumference between full inspiration and full expiration at the level of the lower sternum. Occiput-to-wall distance was measured as the distance between the patients' occiput and the wall when the patients' heels and back were against the wall and chin was held level. Laboratory variables assessed included erythrocyte sedimentation rate (ESR) and C-Reactive Protein (CRP). Radiographs of axial skeleton (anteroposterior (AP) pelvis, AP and lateral lumbar spine, and lateral cervical spine) were taken. The radiographs of the sacroiliac joints, lumbar spine, cervical spine, and hips were graded 0 to 4 (for normal, suspicious, mild, moderate, and severe, respectively). The scores were combined according to the Bath AS Radiology Index-Total (BASRI-t) for the spine and hips.[14] The score of the BASRI-t ranges from 0 (best) to 16 (worst). At physical examination, disease activity by both patient and physician (on a 10 cm VAS) were also measured.

Statistical Analysis

Data from the two different study groups were pooled and scores are presented using descriptive statistics. In order to assess whether the WHODAS II would reflect AS-oriented questionnaires, Pearson's correlations were used. The same exercise was done for the WHODAS II versus the SF-36, and as an external criterion, the SF-36 versus all AS-oriented questionnaires. Scatter plots were made to visualize scores on the WHODAS II compared with scores on AS-oriented questionnaires.

Responsiveness was assessed with data from the 31 patients who had received a 3-week course of spa therapy and completed a second questionnaire thereafter. Responsiveness was calculated by two different methods: the effect size (ES)[15] and the standardized response mean (SRM).[16] The ES is calculated as the mean change after treatment compared with baseline, divided by the standard deviation (SD) of the baseline scores. The SRM is similar to the ES, but with the SD of the change score as denominator. The results were interpreted according to the Cohen's effect size index, in which 0.2 refers to a small change, 0.5 to moderate, and 0.8 or more to large change.[17]

Pearson's correlations between the WHODAS II and AS-oriented questionnaires were also calculated on change scores (difference in scores before and after the 3-week course of spa therapy). Forward stepwise selection, linear regression analysis was performed to investigate which factors are associated with this short-term change on the WHODAS II. Dependent variable was the change score on the WHODAS II, independent variables were change scores on the BASFI, BASDAI, spinal pain, patient global well-being, and ASQoL. Adjustments were made for age, gender, and disease duration. Because the BASFI, DFI, and HAQ-S showed a high level of collinearity (tolerance less than 0.4), only the BASFI score was used as co-variate, and DFI and HAQ-S were excluded.

Forward stepwise, linear regression analysis was also used to investigate which factors may predict disability, as measured by the WHODAS II, 5 years later. The 1996 baseline data from the OASIS study were used for this purpose. Dependent variable was the WHODAS II score. Independent variables were not only AS-oriented questionnaires (BASDAI, BASFI, patient global assessment, and spinal pain), but also measures of physical examination (occiput-to-wall distance, finger-to-floor distance, Schober's test, chest expansion), laboratory values (ESR and CRP), scores on radiographs (BASRI-t), and disease activity by both patient and physician. Adjustments were made for age, gender, and disease duration. Because the BASFI and DFI again showed a high level of collinearity (tolerance less than 0.4), only the BASFI score was used as co-variate, and DFI was excluded. Occiput-to-wall distance, ESR, CRP, and disease activity by the physician were non-normally distributed and therefore log-transformed before introducing into the regression model.

Results

Table 9.1 shows some characteristics of the study groups. Patients from the spa therapy trial showed lower disease duration (time since diagnosis) and duration of complaints compared with the OASIS group. Sixty-three percent of the patients from the total study population had a paid or unpaid job. The scores of the WHODAS II in this group of patients were therefore calculated with the 36-item version SPSS syntax, and the scores in the group of patients without a job were calculated with the 32-item version SPSS syntax.

Table 9.1. Patients characteristics of the study groups

	Total population (n=214)	Spa therapy trial (n=117)	OASIS study (n=97)
% Men	69%	73%	65%
Age†	50 (11; 23–82)	49 (10; 26–69)	51 (12; 23–82)
Time since diagnosis†	14 (8; 2–46)	12 (6; 2–21)	17 (10; 5–46)
Duration of complaints†	23 (11; 3–56)	19 (9; 3–41)	29 (12; 8–55)
% Patients with a (un)paid job	63%	64%	61%

†Mean (standard deviation; range) in years. OASIS = Outcome in ankylosing spondylitis international study.

In Table 9.2, the final scores of the total study population on the WHODAS II as well as results on the AS-oriented questionnaires and the SF-36 are presented. The WHODAS II showed a mean score of 23.9±15.5 (median 24.5; range 0.0 to 76.1). The other questionnaires showed comparable scores on the lower one-third range of the scale, respectively the upper one-third range of the scale for the SF-36, because this is scored in the opposite direction.

The scores on the WHODAS II were significantly correlated with AS-oriented questionnaires on physical functioning, disease activity, and quality of life (all $P<0.001$, Table 9.3). The WHODAS II and all AS-oriented questionnaires also correlated well with all domains of the generic health status questionnaire SF-36 (all $P<0.001$). Two examples of scatter plots, in which the WHODAS II is plotted against the BASFI and the ASQoL respectively, are shown in Figure 9.1.

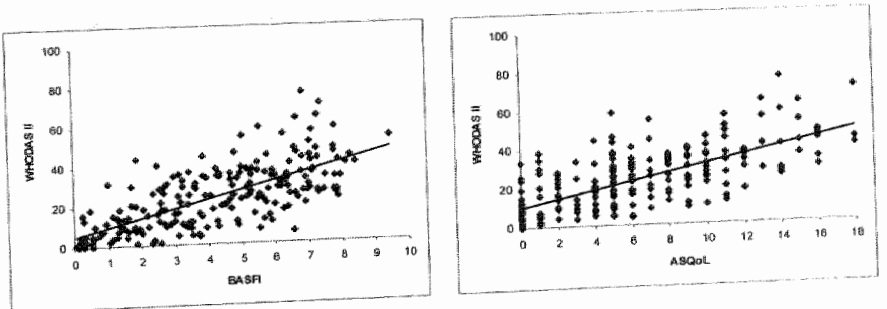

Figure 9.1. Scatter plots of the WHODAS II plotted against the BASFI and the ASQoL For abbreviations see Table 9.2.

Table 9.2. Scores on the WHODAS II, AS-oriented questionnaires, and the SF-36 in the total study population (n=214)

Questionnaire (range)	Score
WHODAS II (0 – 100)	23.9 (15.5)
BASFI (0 – 10)	4.3 (2.3)
DFI (0 – 40)	10.9 (7.0)
HAQ-S (0 – 3)*	0.82 (0.55)
BASDAI (0 – 10)	3.8 (2.2)
Patient global (0 – 10)	4.2 (2.5)
Spinal pain (0 – 10)	4.2 (2.6)
ASQoL (0 – 18)	6.4 (4.8)
SF-36 (0 – 100)	
Physical functioning	64.6 (22.8)
Social functioning	73.1 (22.8)
Role limitation (physical)	50.1 (42.0)
Role limitation (emotional)	75.4 (37.6)
Mental health	72.9 (18.8)
Vitality	55.8 (21.3)
Bodily pain	58.1 (20.3)
General health	49.4 (19.9)

Mean (standard deviation). WHODAS II = World Health Organization Disease Ability Score II; AS = Ankylosing spondylitis; BASFI = Bath Ankylosing Spondylitis Functional Index; DFI = Dougados Functional Index; HAQ-S = Health Assessment Questionnaire for Spondylarthropathies; BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; ASQoL = Ankylosing Spondylitis Quality of Life questionnaire; SF-36 = Short Form 36. *Data from spa therapy trial patients only.

In Table 9.4, the scores on the questionnaires before and after spa therapy from the 31 patients that followed a 3-week course of spa therapy are shown, including the responsiveness scores. In all variables an improvement was seen, with moderate to large responsiveness scores. The WHODAS II showed a comparable, moderate responsiveness score (SRM 0.41 and ES 0.39).

The correlation coefficients between the change score on the WHODAS II and the change scores on the AS-oriented questionnaires are presented in Table 9.5. It can be seen that, on the short term, a change in WHODAS II is significantly correlated with a change in physical functioning (BASFI and DFI), but not with a change in disease activity or quality of life. With forward stepwise linear regression analysis the change score on the BASFI was selected as the only variable to be associated with short-term change on the WHODAS II (adjusted R^2=0.21, beta coefficient for BASFI = 4.25 [95% confidence interval 1.24 to 7.26] P=0.007, beta coefficient for intercept = 1.59 [95% confidence interval -3.19 to 6.37] P=0.501).

Forward stepwise linear regression analysis was also performed to investigate whether clinical and demographic variables could predict disability, as measured by the WHODAS II, after a 5-year period. The variables BASDAI and BASFI at baseline were selected as independently associated with WHODAS II scores after 5 years (adjusted R^2=0.48, Table 9.6).

Table 9.3. Correlations between WHODAS II and several ankylosing spondylitis specific questionnaires

		Short Form 36							
	WHODAS II	Physical function	Social function	Role physical	Role emotional	Mental health	Vitality	Bodily pain	General health
WHODAS II	-	-.70	-.62	-.60	-.46	-.52	-.57	-.64	-.60
BASFI	.67	-.79	-.46	-.52	-.23	-.29	-.42	-.62	-.58
DFI	.70	-.82	-.54	-.54	-.29	-.39	-.49	-.64	-.62
HAQ-S*	.61	-.81	-.49	-.53	-.43	-.44	-.54	-.62	-.60
BASDAI	.57	-.56	-.55	-.58	-.32	-.44	-.62	-.76	-.60
Pt Global	.58	-.52	-.56	-.54	-.29	-.45	-.58	-.73	-.63
Spinal pain	.54	-.51	-.50	-.51	-.27	-.38	-.50	-.72	-.55
ASQoL	.67	-.66	-.71	-.72	-.49	-.58	-.71	-.76	-.70

All correlation coefficients were calculated with Pearson's correlation tests and are significant at the 0.01 level. *Data from spa therapy trial patients only. For abbreviations see Table 9.2.

Table 9.4. Scores on the WHODAS II, several AS-oriented questionnaires, and the domains of the SF-36 before and after a 3-week course of spa therapy as well as the responsiveness scores are presented (n=31)

Questionnaire	Score before intervention†	Score after intervention†	Standardized response mean	Effect size
WHODAS II (0 – 100)	24.9 (12.1)	20.2 (12.9)	0.41	0.39
BASFI (0 – 10)	4.4 (2.3)	3.5 (2.3)	0.64	0.37
DFI (0 – 40)	10.4 (5.9)	8.5 (5.2)	0.54	0.33
HAQ-S (0 – 3)	0.85 (0.54)	0.72 (0.48)	0.41	0.26
BASDAI (0 – 10)	3.9 (1.8)	3.1 (2.1)	0.35	0.37
Patient global (0 – 10)	4.2 (2.6)	3.3 (2.5)	0.28	0.32
Spinal pain (0 – 10)	4.8 (2.2)	3.3 (2.6)	0.61	0.68
ASQoL (0 – 18)	7.3 (5.0)	6.2 (4.6)	0.24	0.22
SF-36 (0 – 100)				
Physical functioning	66.1 (17.8)	70.8 (15.8)	0.35	0.26
Social functioning	69.8 (21.6)	76.6 (19.0)	0.27	0.32
Role limitation (physical)	48.4 (44.7)	62.9 (41.3)	0.30	0.32
Role limitation (emotional)	75.3 (41.2)	80.6 (36.3)	0.13	0.13
Mental health	73.9 (16.8)	80.5 (16.0)	0.44	0.39
Vitality	54.7 (20.9)	66.8 (20.5)	0.50	0.58
Bodily pain	57.9 (19.6)	69.2 (14.5)	0.53	0.58
General health	43.5 (17.1)	55.3 (18.7)	0.83	0.69

†Mean (SD). For abbreviations see Table 9.2. Positive changes imply improvement.

Table 9.5. Correlations between change scores (Δ) in WHODAS II and several AS-oriented questionnaires calculated in the 31 patients who had followed a 3-week course of spa therapy

Questionnaire	Correlation coefficient	*P* value
Δ BASFI	.50	0.005
Δ DFI	.43	0.016
Δ BASDAI	.34	0.065
Δ Pt Global	.23	0.230
Δ Spinal pain	.20	0.294
Δ ASQoL	.16	0.389
Δ HAQ-S	.16	0.399

For abbreviations see Table 9.2.

Discussion

In this study a new generic instrument for measuring disability and functioning, the WHODAS II that is conceptually compatible with the recently revised WHO's ICF, has been investigated in patients with AS. The questionnaire was administered to patients that had participated in an intervention trial and to patients participating in an ongoing longitudinal observational study. Both

Table 9.6. Multiple stepwise regression analysis to study whether demographic and clinical outcome measures contribute in explaining disability, as assessed by the WHODAS II, 5 years later

Variable	Beta coefficient	95% confidence interval	*P* value
Intercept	1.61	-4.68 to 7.89	0.611
BASDAI	0.35	0.17 to 0.53	0.000
BASFI	0.23	0.09 to 0.38	0.002

Analyses were performed in the OASIS study group only (n=97). Variables not selected were patient global well-being, spinal pain, occiput-to-wall distance, finger-to-floor distance, Schober's test, chest expansion, erythrocyte sedimentation rate, C-reactive protein, Bath Ankylosing Spondylitis Radiology Index-Total, and disease activity by both patient and physician. Adjustments were made for age, sex, and disease duration. Adjusted R^2=0.48. For abbreviations see Table 9.2.

groups together showed a wide spectrum of disease duration and duration of complaints. The advantage of this approach is to enhance the ability to generalize the results. A large range in scores on the WHODAS II was found (0 to 76). A relevant ceiling effect of the WHODAS II score is not to be expected because -despite the problems in activities of daily living- patients with AS are seldom completely disabled, not even after a long disease duration.[18]

Being a generic instrument, the WHODAS II would particularly be useful for studies designed to compare across different interventions and different disease entities (e.g. in economic evaluations), similar to other generic instruments such as the SF-36 and the EuroQoL (EQ-5D).[8,19] However, the WHODAS II distinguishes itself from the other measures of health status in that it has been based on an international classification system; it is cross-culturally applicable; and it treats all disorders at parity when determining the level of functioning.[20] In addition, the WHODAS II can also be applied as an instrument to measure overall disability at a disorder-specific level, as we have investigated in AS. The bodily, personal, and social levels of disability are well represented in the WHODAS II in our study: scores on the WHODAS II were significantly correlated with all AS-oriented questionnaires on disease activity, physical functioning, and quality of life. These findings were similar to the external criterion, the SF-36, thereby further validating the WHODAS II.

We also investigated which variables are associated with a short-term change in the WHODAS II and which variables may predict disability over a 5-year period. A short-term change in the WHODAS II was to some degree associated with a short-term change in physical function (BASFI; R^2=0.21), and not with a change in disease activity. This finding further strengthens the validation of the WHODAS II in that it is not influenced by measures that may show -daily- fluctuating pattern (such as disease activity), but that it is influenced by factors that induce structural changes. Variables that best predict disability (WHODAS II scores) after 5 years were physical function (BASFI) and disease activity (BASDAI; R^2=0.48). Apparently, WHODAS II scores are reasonably associated with early predictors of outcome in AS, which further improves its quality as an outcome measure. However, it can be seen from the relatively low adjusted R^2 that other factors, which have not been assessed in our disease-oriented study, may also influence disability (WHODAS II) on the long term.

A few methodological issues should be considered. We did not administer the WHODAS II twice to a group of patients that remained stable. Therefore, we could not assess the reproducibility of the WHODAS II in AS. Future studies will be needed to establish this. However, in extensive testing, the WHODAS II has been shown to have a test-retest reliability ranging from 0.91-0.95 (intra-class correlations) across respondents from different geographical regions and from physical disorder, general population, drug, alcohol, and mental disorder samples.[21]

Second, as a consequence, we have no information on to what degree the responsiveness scores of the WHODAS II may be determined by a potential degree of lack of reproducibility. However, the level of responsiveness was similar to the AS-oriented questionnaires that have been shown to be reproducible and responsive in previous studies.[1-3,12,13,22-25] Therefore, we think the present results reliably reflect clinical findings. In patient populations with schizophrenia, depression, and alcohol dependence, the effect size of the WHODAS II was 1.46, 1.05, and 1.35, respectively, showing that it is able to detect even larger changes than the observed moderate, although comparable, effect size in our study.[26]

In conclusion, the WHODAS II is a useful instrument for measuring disability in AS in that it truthfully incorporates results of AS-oriented instruments on disease activity, physical functioning, and quality of life. It shows responsiveness scores that are comparable with AS-oriented instruments. In AS, a short-term change on the WHODAS II is partly associated with a change in physical function. At the group level, disease activity and physical functioning seem to be strong predictors of disability, as measured by the WHODAS II, after 5 years.

References

1. Daltroy LH, Larson MG, Roberts NW, Liang MH. A modification of the Health Assessment Questionnaire for the spondyloarthropathies. J Rheumatol 1990;17:946-50.
2. Dougados M, Gueguen A, Nakache JP, et al. Evaluation of a functional index and an articular index in ankylosing spondylitis. J Rheumatol 1988;15:302-7.
3. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21:2281-5.
4. Abbott CA, Helliwell PS, Chamberlain MA. Functional assessment in ankylosing spondylitis: evaluation of a new self-administered questionnaire and correlation with anthropometric variables. Br J Rheumatol 1994;33:1060-6.
5. Chatterji S, Ustun B, Bickenbach JE. What is disability after all? Disabil Rehabil 1999;21:396-8.
6. Gray DB, Hendershot GE. The ICIDH-2: developments for a new era of outcomes research. Arch Phys Med Rehabil 2000;81:S10-4.
7. World Health Organization. International Classification of Functioning, Disability and Health (ICF). Geneva: WHO 2001.
8. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
9. Ward MM. Health-related quality of life in ankylosing spondylitis: a survey of 175 patients. Arthritis Care Res 1999;12:247-55.
10. van Tubergen A, Landewé R, van der Heijde D, et al. Combined spa-exercise therapy is effective in patients with ankylosing spondylitis: a randomized controlled trial. Arthritis Rheum 2001;45:430-8.
11. Creemers MC, Van 't Hof MA, Franssen MJ, et al. A Dutch version of the functional index for ankylosing spondylitis: development and validation in a long-term study. Br J Rheumatol 1994;33:842-6.
12. Garrett S, Jenkinson T, Kennedy LG, et al. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.
13. Doward LC, Spoorenberg A, Cook SA, Whalley D, Helliwell PS, Kay LJ, McKenna SP, Tennant A, van der Heijde D, Chamberlain MA. The development of the ASQoL: A quality of life instrument specific to ankylosing spondylitis. Ann Rheum Dis 2002 [In press].
14. Calin A, Mackay K, Santos H, Brophy S. A new dimension to outcome: application of the Bath Ankylosing Spondylitis Radiology Index. J Rheumatol 1999;26:988-92.
15. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178-89.
16. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632-42.
17. Cohen J. Statistical power analysis for behavioral sciences. 2nd ed. Hillsdale: Lawrence Erlbaum Associates 1988.
18. Carette S, Graham D, Little H, Rubenstein J, Rosen P. The natural disease course of ankylosing spondylitis. Arthritis Rheum 1983;26:186-90.

19. The EuroQoL Group. Euroqol-a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.
20. Ustun TB, Chatterji S, Rehm J, et al. World Health Organization Disability Assessment Schedule II (WHODAS II): development, psychometric testing, and applications [Submitted].
21. Epping-Jordan JE, Prieto L, Doyle P, Chatterji S, Ustun B. Reliability of the World Health Organization Disability Assessment Schedule II (WHODAS II): subgroup analysis [Submitted].
22. Ward MM, Kuzis S. Validity and sensitivity to change of spondylitis-specific measures of functional disability. J Rheumatol 1999;26:121-7.
23. Ruof J, Sangha O, Stucki G. Comparative responsiveness of 3 functional indices in ankylosing spondylitis. J Rheumatol 1999;26:1959-63.
24. Spoorenberg A, van der Heijde D, de Klerk E, et al. A comparative study of the usefulness of the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index in the assessment of ankylosing spondylitis. J Rheumatol 1999;26:961-5.
25. Calin A, Nakache JP, Gueguen A, et al. Outcome variables in ankylosing spondylitis: evaluation of their relevance and discriminant capacity. J Rheumatol 1999;26:975-9.
26. Chisholm D, Chatterji S, Ustun TB, Rehm J, von Korff M. Sensitivity to change of the World Health Organization Disability Schedule II (WHODAS II) [Submitted].

Chapter 10

Summary and general discussion

Ankylosing spondylitis and the spondylarthropathies

Ankylosing spondylitis (AS) is a chronic inflammatory disease, in which predominantly the sacroiliac joints and the spine are affected.[1] AS is a key disorder of the spondylarthropathies (SpA), a family of rheumatic diseases classified together because of common features, including the absence of rheumatoid factor, the presence of mono- or oligoarthritis predominantly of the lower limbs, and inflammatory back pain.[1] The SpA are strongly associated with the class-I molecule HLA B27.

In the first part of this thesis (chapters 2-5) some therapeutic aspects of the SpA, and AS in particular, have been presented. In the second part of this thesis (chapters 6-9) a number of aspects of outcome assessment in AS have been discussed in detail.

Part I: Therapeutic aspects of ankylosing spondylitis and other spondylarthropathies

Combination therapy for the spondylarthropathies

In chapter 2, an overview on drug treatment for SpA has been described. A distinction has been made in evidence for treatment of SpA in general, and for AS and psoriatic arthritis (PsA) in particular. Although SpA are classified together because of common features, subgroups may respond differently to similar treatments. Much, if not most, clinical research has been performed in relatively homogeneous patient-groups such as in patients with AS or PsA, as compared with the more heterogeneous spectrum of patients with SpA.

Although the main purpose of the overview was to find evidence for efficacy of combinations of treatments for SpA, only a limited number of studies explicitly reported these combinations. Unlike diseases such as rheumatoid arthritis, in which combination therapy refers to combinations of DMARDs, it is not clear in SpA what is meant by combination therapy, because many combinations can be made. In SpA, potential DMARDs are frequently prescribed in addition to NSAIDs and also physical therapy plays an important role in the overall management in patients with SpA, but the combinations of these treatments have only sparsely been described. Therefore, in this chapter a more general overview on treatment for SpA has been presented.

Overall, despite their common use, only a limited number of randomized controlled trials on several treatment modalities have been performed in patients with SpA, and AS and PsA in particular. A trend favoring a number of DMARDs over placebo treatment was found. However, due to insufficient statistical power,

sound conclusions cannot be drawn for many of these DMARDs. Some preliminary studies suggested that combination therapy may certainly be a valuable option in patients with SpA and it seems likely that well-powered clinical trials may confirm the efficacy of combining treatment strategies. However, it is to be recommended that before exploring the potential of combination therapy in SpA, first of all the place of several DMARDs as monotherapy should be established by sufficiently large, well-performed randomized controlled trials.

History of spa therapy

In chapter 3, a brief general history of spa medicine has been described. Spa therapy has a longstanding history as a treatment modality for many rheumatic disorders. Throughout the ages, the interest towards the use of water in medicine has been fluctuating from century to century and from nation to nation. The (medical) world has viewed it with different opinions from very enthusiastic to extremely critical, and from beneficial to harmful. Today, spa therapy receives renewed attention from many medical specialties, but also from health tourists, and the health care industry or health insurances. Spa therapy is now no longer confined to bathing in water, but is increasingly offered in combination with other treatments, such as active exercise therapy, massages, or mud packs.

Efficacy of spa therapy in ankylosing spondylitis

Despite the long history and popularity of spa therapy, only a few randomized controlled trials on the efficacy of spa therapy in several rheumatic diseases have been conducted in the past decades. A definite judgment about efficacy of spa therapy was, however, impossible to establish, because of methodological flaws in these studies.[2] Many patients with AS annually take courses of spa therapy in which, besides bathing, also exercises play a central role. The efficacy of spa therapy in AS, however, has never been properly investigated. In chapter 4, a randomized controlled trial on the efficacy of spa therapy in combination with exercises in AS has been described. The aim of this study was to determine whether a 3-week course of spa-exercise therapy, in addition to standard treatment with drugs and weekly group physical therapy, has beneficial effects on function, pain, stiffness, and global well-being. A total of 120 patients with AS were randomly allocated into three groups of 40 patients each: group 1 (spa therapy in Bad Hofgastein, Austria), group 2 (spa therapy in Arcen, The Netherlands), and a control group (continued their usual drug treatment and weekly group physical therapy at home during the intervention weeks). After the intervention, all patients continued drug treatment (if necessary) and weekly group physical therapy for another 37 weeks. The results of the primary

outcomes were summarized in a pooled index. The pooled index in group 1 showed significantly more improvement until 28 weeks after start of the intervention and had just lost significance at 40 weeks in comparison with the control group ($P=0.062$). In group 2, the significant improvements on the pooled index compared with controls had disappeared after 16 weeks, suggesting that the beneficial effects of spa-exercise therapy provided in Arcen, The Netherlands, were less persistent. Except for morning stiffness, each of the primary outcome measures separately showed statistically significant improvements in both intervention groups after spa therapy. Maximum relative improvements (mean scores of group 1 or 2 minus those of controls) during follow-up were 24% (group 1) and 19% (group 2) for function, 24% (group 1) and 30% (group 2) for pain, and 33% (group 1) and 29% (group 2) for global well-being. Remarkably, at 40 weeks group 1 still showed relative improvements of 12% in function, 8% in pain, and 16% in global well-being.

This trial showed that a 3-week course of spa-exercise therapy in addition to standard treatment has significant long-term benefits in patients with AS. The study was primarily designed to assess the overall efficacy of spa-exercise therapy in patients with AS rather than to investigate individual components of the intervention. It is likely that the beneficial effects of spa-exercise therapy are attributable to many factors, both related and unrelated to the intervention. Specific effects may be attributed to the intervention itself such as muscle relaxation, decrease in pain, and increase in mobility induced by immersion of the body in thermal water;[3,4] increase in mobility and fitness by exercising;[5] and, possibly, a decrease of inflammatory activity by exposure to radon.[6,7] Non-specific factors may be attributed to a change of environment, the pleasant scenery, the non-competitive atmosphere with fellow-patients, and the absence of work duties.[2,8,9] Placebo effects caused by the belief in improvement by spa therapy and positive attention may also have contributed to the differences between the intervention groups and controls. One would, however, imagine that non-specific effects would rapidly extinguish once the specific treatment has finished. The increasing improvement seen at 16 weeks in group 1 and the persistent long-term effects seem to favor some specific effects of the intervention itself. However, to control for possible placebo effects or any other non-specific effects, ideally, the control group should have stayed at the spa resort during the intervention period. Similarly, to investigate the value of individual components of spa therapy, different control groups should have been used in which these components have been omitted. These issue should be addressed in future studies.

Costs of spa therapy in ankylosing spondylitis

Alongside the clinical trial an economic evaluation was performed, in which the incremental cost-effectiveness and cost-utility ratios of the 3-week course of spa-exercise therapy besides standard treatment has been investigated (chapter 5).

Direct (health care and non-health care) as well as indirect (non-health care) costs were included. All resource utilization and number of days of illness-related absence from work were measured during the entire study period in a cost diary completed by the patients. The results on the Bath AS Functional Index (BASFI) were used to express effectiveness of the intervention and to calculate the cost-effectiveness ratio. The EuroQol ($EQ\text{-}5D_{utility}$) was applied to assess utility from a societal perspective. The results from the cost-utility analysis were expressed as quality-adjusted life years (QALYs).

Cost savings were found for both intervention groups compared with the control group with respect to visits to health care professionals, physiotherapy, and use of medication during the study period. Group 1, but not group 2, showed less absence from work due to AS compared with controls. The mean total costs per patient during the study period were €3,023 for group 1 (including €1,739 for spa therapy), €3,240 for group 2 (including €1,515 for spa therapy), and €1,754 for the control group. The mean total *incremental* (= additional) costs were per patient €1,269 for group 1, and €1,486 for group 2. The incremental cost-effectiveness ratio per unit effect gained in functional ability on a 0-10 scale was €1,269 for group 1 and €2,477 for group 2. The costs per QALY gained were €7,465 for group 1 and €18,575 for group 2. In addition to these analyses, sensitivity analyses were also performed, in which outcome measures were varied and in which several other cost categories were added. No substantial deviations in the cost ratios were found in these analyses.

This economic evaluation showed that a 3-week course of spa-exercise therapy in addition to standard treatment has favorable cost-effectiveness and cost-utility ratios compared with standard treatment alone in patients with AS. In The Netherlands, reimbursement for spa therapy is provided by a few insurance companies for a limited number of diseases (usually rheumatoid arthritis and AS only), with a maximum reimbursement level often not sufficient to cover all costs. At least 16,000 patients with AS live in The Netherlands and approximately 20% of them follow weekly group physical therapy. If reimbursement for spa therapy would be limited to these 3,200 patients, who clearly showed to be compliant to treatment by regularly following weekly group physical therapy, then the incremental societal costs for providing spa therapy would annually be €4,408,000. In exchange of these costs, a large -although temporary- improvement in functional outcome and quality of life will be gained in a group of patients for whom no curative treatment is yet available. Future studies will be necessary to investigate whether patients who annually follow a 3-week course of spa-exercise therapy, starting in an early phase of the disease, will experience even more benefit with respect to functional outcome, health care utilization, and ability to work on the long-term.

Part II: Outcome assessment in ankylosing spondylitis

Improvement criteria for ankylosing spondylitis

In 1998, the members of an international working group who share their expertise in the field of AS, the ASsessment in Ankylosing Spondylitis (ASAS) Working Group, have selected "core sets" of outcome measures to be used in different kinds of trials in AS, in order to create uniformity and allow comparability among these studies.[10] As a next step, by statistical approach, a preliminary definition of short term improvement in AS has been developed, based on the core set of clinical trials with NSAIDs.[11] In chapter 6, these statistically-derived improvement criteria for AS (ASAS-IC) have been evaluated for their clinical relevance as assessed by the opinion of an expert panel in the field of AS. The ASAS-IC consist of four domains: physical function, spinal pain, patient global well-being, and inflammation. Scores on these four domains of 55 patients with AS, who had participated in an NSAID efficacy trial, were presented to the experts in a 3-round Delphi-exercise.

Compared with the ASAS-IC, the experts considered twice as many patients as a responder (42 versus 21). Overall agreement between experts and ASAS-IC was 62%. Pain was considered as the most important domain by the majority of the experts and was also selected as such by discriminant analysis. Newly defined, provisional criteria sets that represented the consensus of the experts showed higher placebo response rates than the ASAS-IC, however, all criteria sets discriminated well between active and placebo treatment. In comparison with the ASAS-IC, these criteria sets showed higher agreements with the end-of-trial efficacy assessment completed by the patient and physician, and also higher sensitivity rates, but somewhat lower specificity rates.

From this study it can be concluded that the ASAS-IC are strict in defining a patient as a responder. However, those patients who were classified as a responder by the ASAS-IC are most likely to be true responders; all experts as well as the patients themselves and the treating physicians, independently of each other, acknowledge these patients as a responder. Moreover, because of the high specificity, the number of patients incorrectly classified to the responder group is low. This makes the ASAS-IC particularly valid for clinical trials, in which it is important to detect true responders and to minimize the false-positive rate.

Although the pain domain appeared to be the most important domain on which decisions with respect to response were made, it is not to be recommended to use only one single domain as an outcome measure.[11] Single outcome measures often show relatively poor reliability due to large measurement errors and, subsequently, patients will need to show major changes before they can be regarded as a true responder. Besides, AS is differently expressed and experienced among patients. Applying only one single outcome measure would under represent the entire group of patients by

excluding those patients for whom this single outcome measure is not the major symptom of their disease. Using a combination of domains also has the advantage that if changes occur in several domains simultaneously, it would be more likely that these change are real.

The ASAS-IC have been developed in NSAID efficacy trials. It is yet unknown how the ASAS-IC behave in other clinical trials assessing the efficacy of other treatment modalities. Also, the opinion of the experts on the degree of improvement and prevailing domains may vary for different trial settings. Future studies will be needed to validate the current ASAS-IC or to develop new improvement criteria for other kind of trials.

Use of the numerical rating scale in ankylosing spondylitis specific questionnaires

Several forms of answer scales can be used in questionnaires. Because some patients experience difficulties with visual analog scales (VAS) or Likert-scales, the numerical rating scale (NRS) is slightly preferred in the literature.[12-20] In chapter 7, three widely used AS-specific questionnaires on disease activity (Bath AS Disease Activity Index [BASDAI]) and function (BASFI and Dougados Functional Index [DFI]) were investigated with both their original answer scale (VAS or Likert) and an NRS format, and assessed for their discriminative properties (reproducibility and responsiveness).

A large variation between the scores on the original answer scales and the NRS was found in *individual* questions of all questionnaires, although *total* scores showed a high level of agreement. Reproducibility of all answer scales showed low intraclass correlation coefficients (ICCs) in individual questions, but moderate to good ICCs in total scores of the questionnaires (0.53-0.89). The total scores of the questionnaires showed moderate to large responsiveness (0.48-1.04).

The variability between the NRS and the original answer scales may be explained by the low reproducibility of individual questions in each of these questionnaires. This finding may have large implications for single-item questionnaires which may, due to low reproducibility, consequently lack responsiveness. The variability of the total scores of the questionnaires answered on the different answer scales was less impressive as can be expected by aggregating a number of answers into one score, but was still substantial. No major differences in reproducibility and responsiveness between the answer scales were found. Therefore, it may be concluded that the answer scales of the three AS-specific questionnaires can all be replaced by an NRS, which is easier to complete and score, but has similar discriminative properties as the original answer scales.

Fatigue in ankylosing spondylitis

Fatigue has been acknowledged as an important symptom in AS.[21-24] The ASAS Working Group did not select a specific instrument to measure fatigue, because little information was available on assessing fatigue in AS.[10] In chapter 8, the value of a single-item questionnaire and a multidimensional questionnaire (MFI) for measuring fatigue in AS has been studied. In addition, factors associated with fatigue in AS have been identified and the impact of fatigue on quality of life was investigated.

The study patients were dichotomized into an F+ group (fatigue a major symptom) and an F- group (fatigue a minor symptom). The F+ group scored significantly worse compared with the F- group on all disease-specific and generic questionnaires studied. The single-item question as well as each of the five dimensions of the MFI separately showed equally moderate to good reproducibility and responsiveness.

The patients most often reported that pain and stiffness negatively influenced (i.e. increased) fatigue. With logistic regression analysis, fatigue appeared to be mainly related to disease activity, functional ability, global well-being, and mental health status, and to a lesser extent to age, explaining 52% of the variance. Multiple linear regression analysis showed that quality of life is negatively influenced by the degree of fatigue and that different aspects of fatigue have an impact on different domains of quality of life.

From this study it may be concluded that fatigue in AS can appropriately be measured by both a single-item question on the intensity of fatigue and the MFI. The MFI, however, provides more insight into specific dimensions of fatigue and may be preferred in, for instance, studies specifically designed to assess the impact of a particular intervention on fatigue.

Disability in ankylosing spondylitis

The World Health Organization (WHO) has recently developed an instrument to measure disability, the WHO Disability Assessment Schedule II (WHODAS II). In chapter 9, the value of measuring disability with the WHODAS II in AS has been studied. In addition, factors associated with short-term change have been identified as well as factors that predict scores on this questionnaire 5 years later.

Scores on the WHODAS II were statistically significantly correlated with AS-oriented questionnaires on disease activity, physical functioning, and quality of life. The WHODAS II showed a moderate responsiveness score, in agreement with responsiveness scores on several AS-oriented and generic questionnaires. Regression analysis showed that a short-term change on the WHODAS II is significantly associated with a change in physical functioning (BASFI; $R^2=0.21$). Variables that best predict WHODAS II scores after 5 years were physical functioning (BASFI) and disease activity (BASDAI; $R^2=0.48$).

This study showed that the WHODAS II is an useful instrument to measure disability in AS and is sensitive to detect changes. In AS, short-term change on the WHODAS II is associated with a change in physical functioning. At the group level, physical functioning and disease activity may predict disability after 5 years.

The WHODAS II is a generic instrument and would therefore particularly be useful for studies designed to compare across different interventions and different disease entities (e.g. in economic evaluations), similar to other generic instruments such as the SF-36 and the EuroQoL. However, the WHODAS II distinguishes itself from the other measures of health status in that it has been based on an international classification system; it is cross-culturally applicable; and it treats all disorders at parity when determining the level of functioning.[25] From this study it can be concluded that the WHODAS II may also be applied as an instrument to measure overall disability at a disorder-specific level, such as in AS.

Improvement by movement

Physical exercises and movement are important in the overall management of AS to maintain or improve mobility, fitness, and strength, and to prevent deformity. Patients should consider exercising as part of their daily routine. In spa therapy, exercises also play a central role. In the first part of this thesis it has been shown that a 3-week course of combined spa-exercise therapy has beneficial effects up to 40 weeks after the start of therapy and can be considered as cost-effective.

In order to reliably measure improvement after any intervention, appropriate, well-validated instruments are a prerequisite. The ASAS Working Group has selected several instruments in the core sets of outcome measures for different kinds of trials in AS, and has defined criteria to measure improvement at an individual level in patients with AS. In the second part of this thesis, the improvement criteria and several questionnaires have been investigated for their clinical relevance, usefulness, and/or discriminative properties. Future studies will be needed to further validate these and other instruments.

In conclusion, improvement requires movement. To achieve more improvement in the future, we should keep on moving...

References

1. van der Linden S, van der Heijde D. Ankylosing spondylitis. Clinical features. Rheum Dis Clin North Am 1998;24:663-76.
2. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Knipschild PG. Balneotherapy for rheumatoid arthritis and osteoarthritis (Cochrane Review). In: The Cochrane Library, Issue 1. Oxford: Update Software 2001.
3. Becker B. The biologic aspects of hydrotherapy. J Back Musculoskel Rehabil 1994;4:255-64.
4. Machtey I. Hydrotherapy and balneotherapy: State of the art. In: Balint G, Gömör B, Hodinka L, editors. Rheumatology, State of the Art. Amsterdam: Excerpta Medica; 1992. p. 390-2.
5. Hidding A, van der Linden S, Boers M, et al. Is group physical therapy superior to individualized therapy in ankylosing spondylitis? A randomized controlled trial. Arthritis Care Res 1993;6:117-25.
6. Jöckel H. Practical experiences in radon therapies. In: Pratzel HG, Deetjen P, editors. Radon in der Kurortmedizin: zum Nutzen und vermeintlichen Risiko einer traditionelle medizinische Anwendung. Geretsried: I.S.M.H. Verlag Geretsried 1997. p. 84-91.
7. Soto J. Effects of radon on the immune system. In: Pratzel HG, Deetjen P, editors. Radon in der Kurortmedizin: zum Nutzen und vermeintlichen Risiko einer traditionelle medizinische Anwendung. Geretsried: I.S.M.H. Verlag Geretsried 1997. p. 103-13.
8. Bell MJ. Spa therapy in arthritis: a trialist's view. J Rheumatol 1991;18:1778-9.
9. Fam AG. Spa treatment in arthritis: a rheumatologist's view. J Rheumatol 1991;18:1775-7.
10. van der Heijde D, Calin A, Dougados M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. J Rheumatol 1999;26:951-4.
11. Anderson JJ, Baron G, van der Heijde D, Felson DT, Dougados M. Ankylosing spondylitis assessment group preliminary definition of short-term improvement in ankylosing spondylitis. Arthritis Rheum 2001;44:1876-86.
12. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. Ann Rheum Dis 1978;37:378-81.
13. Kremer E, Atkinson JH, Ignelzi RJ. Measurement of pain: patient preference does not confound pain measurement. Pain 1981;10:241-8.
14. Dixon JS, Bird HA. Reproducibility along a 10 cm vertical visual analogue scale. Ann Rheum Dis 1981;40:87-9.
15. Jensen MP, Karoly P, Braver S. The measurement of clinical pain intensity: a comparison of six methods. Pain 1986;27:117-26.
16. Bird HA, Dixon JS. The measurement of pain. Baillieres Clin Rheumatol 1987;1:71-89.

17. Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. J Chronic Dis 1987;40:1129-33.
18. Ferraz MB, Quaresma MR, Aquino LR, Atra E, Tugwell P, Goldsmith CH. Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. J Rheumatol 1990;17:1022-4.
19. Bolton JE, Wilkinson RC. Responsiveness of pain scales: a comparison of three pain intensity measures in chiropractic patients. J Manipulative Physiol Ther 1998;21:1-7.
20. Eyres S, van der Heijde D, Dougados M, Tennant A. The Visual Analogue Scale: Deception on a 'VASt' scale? [Submitted].
21. Calin A, Edmunds L, Kennedy LG. Fatigue in ankylosing spondylitis--why is it ignored? J Rheumatol 1993;20:991-5.
22. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21:2286-91.
23. Jones SD, Koh WH, Steiner A, Garrett SL, Calin A. Fatigue in ankylosing spondylitis: its prevalence and relationship to disease activity, sleep, and other factors. J Rheumatol 1996;23:487-90.
24. Ward MM. Health-related quality of life in ankylosing spondylitis: a survey of 175 patients. Arthritis Care Res 1999;12:247-55.
25. Ustun TB, Chatterji S, Rehm J, et al. World Health Organization Disability Assessment Schedule II (WHODAS II): development, psychometric testing, and applications [Submitted].

Chapter 11

Samenvatting

Ziekte van Bechterew

De ziekte van Bechterew (ook wel spondylitis ankylopoetica genaamd) is een chronisch reumatische aandoening die gepaard gaat met ontstekingen in voornamelijk de wervelkolom en bekkengewrichten. De ziekte van Bechterew maakt deel uit van een groep aandoeningen, die op grond van gemeenschappelijke kenmerken gezamenlijk aangeduid worden als de spondylarthropathieën (SpA). Er bestaat een sterke relatie met het voorkomen van de erfelijk bepaalde factor HLA B27. Tot deze groep van SpA behoren behalve de ziekte van Bechterew ook andere ziektebeelden zoals het syndroom van Reiter en reactieve artritis, alsmede artritis psoriatica en de gewrichtsaandoeningen die voorkomen bij chronische darmontstekingen zoals de ziekte van Crohn en colitis ulcerosa.

De oorzaak van de ziekte van Bechterew is onbekend. Bij circa 90% van de Bechterewpatiënten is HLA B27 aanwezig in vergelijking met slechts 8% van de gewone bevolking in Nederland. De ziekte wordt 2 tot 3 keer vaker bij mannen dan bij vrouwen gezien en begint vaak op vrij jonge leeftijd (tussen het 20e en 30e levensjaar). In Nederland zijn er minimaal 16000 mensen met de ziekte van Bechterew.

In het eerste gedeelte van dit proefschrift worden enkele therapeutische mogelijkheden voor de SpA en de ziekte van Bechterew in het bijzonder besproken. In het tweede gedeelte worden enkele aspecten van uitkomstmetingen ("outcome measures") onderzocht.

Deel I. Therapeutische mogelijkheden voor de ziekte van Bechterew en andere spondylarthropathieën

Combinatietherapie voor de spondylarthropathieën

In hoofdstuk 2 wordt een overzicht gegeven van bewijs voor de werkzaamheid van diverse medicamenteuze behandelingen bij de SpA, waarbij een onderscheid is gemaakt tussen studies bij de SpA als groep en studies bij de ziekte van Bechterew en artritis psoriatica afzonderlijk. Hoewel het belangrijkste doel van dit onderzoek was een overzicht te geven van bewijs voor werkzaamheid van het combineren van diverse behandelingen bij SpA, bleek dat slechts enkele studies dit expliciet hadden vermeld. Daarom wordt in dit hoofdstuk meer ingegaan op bewijs voor de werkzaamheid van de afzonderlijke behandelmogelijkheden. Over het algemeen was de hoeveelheid bewijs aan de magere kant. Het wordt dan ook aanbevolen om, alvorens de mogelijkheden van het combineren van behandelingen voor de SpA te onderzoeken, eerst de

plaats van de diverse medicijnen afzonderlijk vast te stellen door middel van goed opgezette klinische studies.

Geschiedenis van het kuren

In hoofdstuk 3 wordt in het kort de geschiedenis van het kuren beschreven. Al sinds de oudheid worden kuurbehandelingen toegepast bij reumatische aandoeningen. Door de eeuwen heen is het gebruik ervan -en geloof erin- zeer wisselend geweest. De (medische) wereld is er erg enthousiast over geweest, maar ook extreem kritisch en heeft het zowel als heilzaam, maar ook als schadelijk beschouwd. Vandaag de dag krijgt de kuurbehandeling weer nieuwe belangstelling van vele medische specialismen -maar ook van gezondheids-toeristen- en wordt het baden vaak gecombineerd met andere behandelingen zoals intensieve oefentherapie, massages of modderpakkingen.

Werkzaamheid van kuurbehandelingen bij de ziekte van Bechterew

Ondanks de lange geschiedenis en de populariteit van het kuren, zijn slechts enkele gerandomiseerde gecontroleerde studies naar de werkzaamheid van kuurbehandelingen bij diverse reumatische aandoeningen uitgevoerd, maar niet bij de ziekte van Bechterew. In hoofdstuk 4 wordt de werkzaamheid van een 3 weken durende kuurbehandeling in combinatie met oefentherapie onderzocht in een gerandomiseerde gecontroleerde studie bij de ziekte van Bechterew. Doel van deze studie was om na te gaan of kuurbehandelingen een verbetering in functioneren, pijn, stijfheid en algemeen welbevinden kunnen geven bovenop de standaardbehandeling met ontstekingsremmers en wekelijkse groepstherapie. In totaal 120 patiënten werden via het lot verdeeld over drie groepen van 40 personen elk. Groep 1 ging 3 weken kuren in Bad Hofgastein (Oostenrijk), groep 2 ging 3 weken kuren in Arcen (Nederland) en een controlegroep bleef thuis tijdens de kuurperiode en ging door met hun normale activiteiten en standaardbehandeling. Na de kuur continueerden alle patiënten de wekelijks groepsoefentherapie en -indien nodig- hun ontstekingsremmers. De belangrijkste uitkomstmaten (functioneren, pijn, stijfheid en algemeen welbevinden) werden samengevat en uitgedrukt in één getal. De controlegroep veranderde gemiddeld genomen niet of nauwelijks gedurende de hele studieperiode. Groep 1 liet tot 28 weken na begin van de kuur een statistisch significante verbetering zien t.o.v. de controlegroep. Op 40 weken na de kuur was de nog steeds aanwezige verbetering in groep 1 niet meer significant t.o.v. de controlegroep. In groep 2 waren de significante verbeteringen t.o.v. de controlegroep op 16 weken na de kuur verdwenen, wat suggereert dat de gunstige effecten van kuurbehandelingen in Arcen minder lang aanhouden in

vergelijking tot Bad Hofgastein. De verschillen tussen groep 1 en 2 waren echter niet statistisch significant en kunnen dus op toeval berusten.

Met deze studie is aangetoond dat een 3 weken durende kuur-oefenbehandeling naast standaardbehandeling met ontstekingsremmers en wekelijkse groepsoefentherapie tot 40 weken na het begin van de kuur gunstige effecten kan hebben in patiënten met de ziekte van Bechterew.

Kosten van kuurbehandelingen bij de ziekte van Bechterew

Behalve naar de werkzaamheid is ook gekeken naar de kosten van kuurbehandelingen, met name in relatie tot de gevonden effecten (hoofdstuk 5). Diverse kostenposten op het gebied van de gezondheidszorg, maar ook kosten voor bijvoorbeeld hulp in huis en afwezigheid van het werk i.v.m. de ziekte van Bechterew werden onderzocht. De patiënten hielden deze kosten wekelijks bij in een boekje. Met deze gegevens zijn de kosteneffectiviteitsratio (kosten afgezet tegen de gevonden effecten) en de kostenutiliteitsratio (kosten afgezet tegen de winst in utiliteiten, ofwel "kwaliteit van leven") van kuurbehandelingen t.o.v. de standaardbehandeling voor Bechterew berekend. De gemiddelde kosten tijdens de studieperiode van 40 weken waren, per patiënt, €3023 voor groep 1 (inclusief €1739 voor de kuurbehandeling), €3240 voor groep 2 (inclusief €1515 voor de kuurbehandeling) en €1754 voor de controlegroep. De kosteneffectiviteitratio was als volgt berekend: om 1 eenheid op de 0 tot 10 schaal van functioneren te winnen waren de extra kosten van een kuurbehandeling t.o.v. de controlegroep voor groep 1 €1269 en voor groep 2 €2477. De kostenutiliteitsratio (uitgedrukt in kosten per gewonnen QALY) was voor groep 1 €7465 en voor groep 2 €18585 per gewonnen QALY.

Met deze economische evaluatie is aangetoond dat een 3 weken durende kuur-oefenbehandeling bovenop de standaardbehandeling voor Bechterew zowel gunstige kosteneffectiviteits- als kostenutiliteitsratio's heeft.

Deel II: Uitkomstmetingen voor de ziekte van Bechterew

Criteria voor verbetering voor de ziekte van Bechterew

Onlangs zijn er criteria samengesteld om "verbetering" bij de ziekte van Bechterew vast te stellen. Deze criteria zijn ontwikkeld met behulp van statistische technieken. Het was de vraag of deze criteria ook de mening van de artsen en de patiënten goed weergeeft. In hoofdstuk 6 worden deze criteria getoetst aan de consensus van een groep experts op het gebied van Bechterew. Voor dit onderzoek is gebruik gemaakt van de profielen van patiënten die aan een studie naar de werkzaamheid van ontstekingsremmers hebben meegedaan.

Uit het onderzoek blijkt dat de ontwikkelde criteria voor verbetering streng (= specifiek) zijn als het gaat om een patiënt als verbeterd te rekenen in vergelijking tot de mening van de experts. Echter, de patiënten die als verbeterd beoordeeld werden door deze criteria zijn waarschijnlijk ook echt verbeterd: zij werden allen als verbeterd beschouwd door de experts, maar ook door de studiepatiënten zelf alsmede hun behandelaars. Deze criteria voor verbetering zijn juist door hun strengheid (= specificiteit) met name geschikt voor klinische studies, waarbij het belangrijk is om zo min mogelijk mensen onterecht tot de groep van verbeteraars te rekenen.

Gebruik van een "numerical rating scale" bij vragenlijsten voor Bechterew

Diverse antwoordschalen kunnen worden gebruikt bij vragenlijsten: een "visual analog scale" (VAS), een Likert schaal en een "numerical rating scale" (NRS). Omdat sommige patiënten moeite hebben met de VAS en de Likert schaal, gaat in de literatuur soms de voorkeur uit naar een NRS. In hoofdstuk 7 worden drie veelvuldig gebruikte vragenlijsten bij Bechterewonderzoek, om de mate van ziekteactiviteit en functioneren te meten en die gescoord worden op een VAS of Likert schaal, ook onderzocht op een NRS. In deze studie werd een grote variatie in scores op de oorspronkelijke antwoordschalen in vergelijking tot de NRS gevonden bij de individuele vragen van de drie vragenlijsten, hoewel de totaalscores minder variatie lieten zien. Matig tot goede betrouwbaarheid en gevoeligheid voor verandering werden gevonden voor de totaalscores van de vragenlijsten, waarbij er nauwelijks verschillen waren tussen de oorspronkelijke antwoordschalen en de NRS. Hieruit kan worden geconcludeerd dat de antwoordschalen van de drie vragenlijsten door de NRS vervangen kunnen worden, die makkelijker te beantwoorden en te scoren is en dezelfde onderscheidende eigenschappen bezit als de oorspronkelijke antwoordschalen.

Moeheid bij de ziekte van Bechterew

Moeheid is een belangrijk symptoom bij de ziekte van Bechterew. In hoofdstuk 8 is de waarde van een enkele vraag over moeheid alsmede een multidimensionele vragenlijst over moeheid (MFI) bij de ziekte van Bechterew onderzocht. Ook is gekeken naar factoren die gerelateerd zijn aan moeheid en naar de invloed van moeheid op kwaliteit van leven bij Bechterew. Zowel de enkele vraag over moeheid als de MFI lieten matig tot goede betrouwbaarheid en gevoeligheid voor verandering zien. Moeheid was met name gerelateerd aan de mate van ziekteactiviteit, functioneren, algemeen welbevinden, mentale gezondheidsstatus en in mindere mate leeftijd. Moeheid bleek kwaliteit van leven negatief te beïnvloeden en verschillende aspecten van moeheid hadden een invloed op verschillende deelgebieden van kwaliteit van leven. Uit deze

studie kan geconcludeerd worden dat moeheid zowel met een enkele vraag alsook met de multidimensionele MFI gemeten kan worden. Echter, wanneer meer inzicht in de verschillende aspecten van moeheid gewenst is, geniet de MFI de voorkeur.

"Disability" bij de ziekte van Bechterew

De Wereld Gezondheidsorganisatie (WHO) heeft recentelijk een generieke vragenlijst ontwikkeld om "disability" (te vergelijken met invaliditeit of handicap) te meten: de WHODAS II. In hoofdstuk 9 wordt de waarde van deze vragenlijst bij de ziekte van Bechterew onderzocht. Daarnaast worden factoren die gerelateerd zijn aan een korte termijn verandering op de WHODAS II vastgesteld, evenals factoren die een score op de WHODAS II na 5 jaar kunnen voorspellen. De WHODAS II bleek statistisch significant te correleren met vragenlijsten over ziekteactiviteit, functioneren en kwaliteit van leven. De WHODAS II liet een matige gevoeligheid voor verandering zien, overeenkomstig met andere afgenomen vragenlijsten. Een korte termijn verandering op de WHODAS II bleek gerelateerd te zijn aan verandering in functioneren. Scores op de WHODAS II na 5 jaar worden het best voorspeld door functioneren en ziekteactiviteit 5 jaar daarvoor. Uit deze studie kan geconcludeerd worden dat de generieke WHODAS II ook als ziektespecifieke vragenlijst goed bij de ziekte van Bechterew kan worden toegepast om disability te meten.

Verbetering door beweging ("improvement by movement")

Lichaamsoefeningen en bewegen zijn speerpunten in de behandeling van Bechterew om beweeglijkheid, fitheid en kracht te behouden of te verbeteren en verkromming van de wervelkolom te voorkomen. Patiënten moeten het doen van oefeningen als een dagelijkse routine beschouwen. Ook in kuurbehandelingen staan oefeningen centraal. In het eerste gedeelte van dit proefschrift is aangetoond dat een 3 weken durende gecombineerde kuur-oefenbehandeling gunstige effecten heeft tot 40 weken na begin van de kuur en ook als kosteneffectief kan worden beschouwd.

Om betrouwbaar een verbetering na een interventie te kunnen meten zijn geschikte, goed gevalideerde instrumenten een voorwaarde. In het tweede gedeelte van dit proefschrift zijn criteria om zo'n verbetering aan te tonen onderzocht op hun klinische relevantie en zijn diverse vragenlijsten getoetst op hun toepasbaarheid en onderscheidend vermogen bij Bechterewpatiënten. In de toekomst zullen nog veel meer van dergelijke onderzoeken noodzakelijk zijn.

Concluderend, zonder beweging is geen verbetering mogelijk. Om meer verbetering in de toekomst te bereiken is het nodig dat we (in brede zin) blijven bewegen...

Chapter 12

Dankwoord

Dankwoord

Veel dank ben ik verschuldigd aan al diegenen die op enigerlei wijze hebben bijgedragen aan dit proefschrift. In de wetenschap dat ik onvolledig ben, wil ik een aantal mensen graag met name noemen.

Allereerst mijn promotoren, Sjef van der Linden en Désirée van der Heijde. Beste Sjef en Désirée, precies één avond kreeg ik van jullie om te beslissen of, en zo ja, welk onderzoek ik wilde gaan doen. Ik koos het kuuronderzoek, een project van slechts 1,5 jaar. Het werden er uiteindelijk drie. Ik wil jullie dan ook hartelijk danken dat jullie me de mogelijkheid hebben gegeven om gedurende en na afronding van het kuuronderzoek nog verder onderzoek te doen, dat uiteindelijk tot dit proefschrift heeft geleid. Ik heb er geen dag spijt van gehad. Jullie begeleiding was uitermate stimulerend en motiverend. Sjef, een speciale dank aan je beide zonen, die ook nauw betrokken zijn geweest bij het kuurproject. Met veel plezier denk ik terug aan de beginperiode waarin we avond na avond samen met wat studenten en Sietze heel Nederland zijn doorgereisd om patiënten te screenen en te meten. Als er maar (Bastogne)koeken waren, dan vonden jullie het allang best. Veel dank ook aan Sjoerd die een prachtige website heeft gemaakt over het kuuronderzoek (www.kuurtherapie.info).

Mijn co-promotor Robert Landewé, Beste Robert, jij bent ongetwijfeld mede verantwoordelijk voor de snelle totstandkoming van dit proefschrift. Met veel geduld heb je me de beginselen van onderzoek doen uitgelegd en goede suggesties voor verbeteringen van mijn (in jouw ogen) soms 'Libelle'-achtige manuscripten gegeven. Deze werden in een rap tempo geretourneerd en hoewel je commentaar dus niet altijd even genuanceerd was, heb ik er vaak hard om moeten lachen, maar vooral ook veel van geleerd. Een cursus Latijn kreeg ik er ook nog eens gratis bij.

Alle (oud) medeonderzoekers Anneke, Erik, Simone, Liesbeth, Annelies, Karin, Guy en Astrid en de onderzoeksassistenten Anita, Jolanda en Mieke. Ik ben jullie dankbaar voor de gezellige lunches en goede samenwerking. In het bijzonder wil ik mijn kamergenoten, Liesbeth en Anita, bedanken voor al jullie wijsheden en corrigerende opmerkingen. Het provisorium lag ver weg van de rest, maar we hebben genoten van de zon, het flirten met de parkeerwachters en de dagelijks e-mail grapjes... Ik mis jullie gezelschap nog dagelijks. Annelies wil ik bedanken voor de betrokkenheid bij het kuuronderzoek, die veel verder ging dan alleen begeleiding van het kostengedeelte: je tips over de Gasteiner Sonate van Schubert en de kuurreizen van Renoir heb ik zeer gewaardeerd.

Een speciale dank wil ik richten aan Nico Wolter. Als Bechterewpatiënt en als ervaringdeskundige heb jij de basis gelegd voor het hele kuuronderzoek. Nico, jouw gedrevenheid en betrokkenheid bij dit onderzoek waren enorm. Zonder jou zou dit onderzoek nooit plaats hebben gevonden. Heel erg veel dank voor je

inzet en al die zaken die je 'even geregeld' hebt. Ook wil ik graag de andere direct betrokkenen bij het kuuronderzoek bedanken voor al hun ideeën en bijdragen voor de totstandkoming en uitvoering van dit onderzoek: Alita Hidding, Max Asscher, Henk Goei Thè en namens de Reumapatiëntenbond Dick Martens. Jos Daanen wil ik bedanken voor de mogelijkheden die hij heeft geboden om patiënten van het kuuronderzoek in Thermaalbad Arcen te laten kuren en voor zijn verdere betrokkenheid bij dit onderzoek. From Austria, I would like to thank Albrecht Falkenbach for his active involvement in the spa therapy project. Working together with you has been enjoyable and stimulating. My German has not seen any progress, despite all your attempts, but your Dutch is absolutely brilliant. Thank you for convincing me that visiting the Gasteiner Heilstollen is not something claustrophobic, but that it can actually be relaxing. Also many thanks to Franz Hartlieb from the Gasteiner Heilstollen and Michael Schafflinger from the Thermentempel in Gastein. Your personal efforts, your conviction in the efficacy of spa therapy, and your financial support were essential to the successful course of the spa project.

A special thanks to all other contributors of the several papers: Maureen Rutten, Astrid Chorus, John Londoño, Ekkehard Genth, Maxime Dougados, Nicholas Bellamy, Gaby Udrea, Hille van der Tempel, Jennifer Anderson, Jürgen Braun, Gerold Stucki, Ruben Burgos-Vargas, Mario Cardiel, Liliane Ryser, Iris Debats, and Jolanda Coenen. Collaborating with you has been joyful and stimulating.

Mijn paranimfen, Mijke Vankan en Nicole Zwaga, Lieve Mijke en Nicole, precies 10 jaar geleden leerden we elkaar kennen in Maastricht. De contacten zijn in al die jaren wisselend van intensiteit geweest met als hoogtepunten de jaarlijks terugkerende vaardigheidstoets (waarbij we uiteindelijk toch liever naar de film gingen) en wanneer weer een van ons in het buitenland verbleef en er ellenlange brieven (later e-mails) geschreven werden met onzinnige krantenartikels en geplette koekjes. Dank voor al die gezellige jaren en dat jullie mijn paranimfen willen zijn. Ik wens jullie allebei veel geluk toe met jullie carrière en (gezins)leven in het Haagse!

Mijn ouders en broer, Lieve Elly, Joop en Frank, hoewel ik verzekerd had jullie niet te noemen om alle clichés te vermijden, wil ik jullie toch graag bedanken voor zoals jullie zijn en altijd alle ontwikkelingen met veel belangstelling en trots hebben gevolgd!

Lieve Rogier, het boekje is helaas af! In die jaren heb je kunnen genieten van mijn vroege thuiskomst, mijn kookkunsten en de weekenden en avonden dat ik er (bijna) altijd was. We weten dat dit nu radicaal veranderd is; veel reizen, lange werkdagen en weekenden delen met de nieuwe werkgevers... De komende jaren zullen even afzien worden, maar zoals je reeds zei... you can only attain improvement by movement. Dank voor deze allereerste en tevens laatste inbreng in mijn proefschrift: de alles samenvattende titel!

Curriculum Vitae

Curriculum Vitae

Astrid van Tubergen werd geboren op 1 december 1973 te Amsterdam. In 1991 behaalde zij het VWO diploma aan het Hooghe Landt College te Amersfoort. In 1992 startte zij met de studie Geneeskunde aan de Universiteit Maastricht, alwaar zij in 1996 het doctoraalexamen (cum laude) en in 1998 het artsexamen (cum laude) behaalde. Na het artsexamen is zij 3 maanden werkzaam geweest als arts-assistent bij de afdeling plastische chirurgie in het academisch ziekenhuis Maastricht. Op 1 februari 1999 startte ze als artsonderzoeker bij de afdeling interne geneeskunde (hoofd prof. dr. H. Hillen), werkgroep reumatologie in het academisch ziekenhuis Maastricht onder leiding van prof. dr. Sj. van der Linden. De onderzoeken die in deze periode werden verricht, zijn grotendeels verwerkt in dit proefschrift. Voor de onderzoeken beschreven in hoofdstukken 4 en 5 ontving de auteur de Boxberger Preis Bad Kissingen 2001 en voor het onderzoek beschreven in hoofdstuk 6 de Novartis Rheumatology Grant 2001 van de Nederlandse Vereniging voor Reumatologie. Op 1 januari 2002 is zij gestart met de vooropleiding interne geneeskunde in het Máxima Medisch Centrum te Eindhoven (opleider dr. H. Haak).

Cover illustration, page 15, and 85: Strigils and sponges (1879), water-colour by Sir Lawrence Alma-Tadema, and reproduction by Paul Adolphe Rajon.
© The British Museum, London.

Sir Lawrence Alma-Tadema was born in 1836 in a little village in Friesland, The Netherlands. He was trained at the art academy of Antwerp and subsequently worked as a classicist painter. To improve his art of painting, Alma-Tadema moved around many European countries. In 1870, he settled in Great Britain, where his worked was highly appreciated. He became a member of the prestigious Royal Academy and was internationally rewarded. He was renowned for his technique, archaeologically detailed paintings, and his unusual themes, mainly because he illustrated the daily home life of the Ancient rather than important historical matters. His work was often copied for distribution among the less wealthy. Alma-Tadema, however, showed an obsessive concern about the quality of these reproductions, illustrated by the numerous detailed suggestions for improvement in the marginal lines of a proof etching by Rajon (page 85). He was also very critical to his own work, which he had destroyed or repainted more than once. Alma-Tadema died in 1912 in a spa resort in Germany, leaving behind over 400 pieces of art.