ARTHRITIS CARE & RESEARCH 45:430–438, 2001

ORIGINAL ARTICLE

Combined Spa–Exercise Therapy Is Effective in Patients With Ankylosing Spondylitis: A Randomized Controlled Trial

ASTRID VAN TUBERGEN,[1] ROBERT LANDEWÉ,[1,2] DÉSIRÉE VAN DER HEIJDE,[1,3] ALITA HIDDING,[4] NICO WOLTER,[1] MAX ASSCHER,[5] ALBRECHT FALKENBACH,[6] EKKEHARD GENTH,[7] HENK GOEI THÈ,[2] AND SJEF VAN DER LINDEN[1]

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 3 groups of 40 patients each. Group 1 (mean age 48 ± 10 years; male:female ratio 25:15) was treated in a spa resort in Bad Hofgastein, Austria; group 2 (mean age 49 ± 9 years; male:female ratio 28:12) in a spa resort in Arcen, The Netherlands. The control group (mean age 48 ± 10 years; male:female ratio 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 3 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 was 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 was 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.**

KEY WORDS. Randomized clinical trial; Ankylosing spondylitis; Spa therapy; Physical therapy; Hydrotherapy.

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 popu-

Major financial support was obtained from Land Salzburg, Gasteiner Tal Tourismusgesellschaft, Kurzentrum Thermentempel, and Gasteiner Heilstollen, Austria. Other sources were ZVN (Zorgvoorzieningen Nederland) N.V., IZA Insurance Company, Dick van Tol B.V. Insurance Company, and Yakult B.V., The Netherlands. In all parts of the study the investigators acted completely independent of the sponsors. Dr. A. Falkenbach is employed as a physician at the Gasteiner Heilstollen.

[1]Astrid van Tubergen, MD, Robert Landewé, MD, Désirée van der Heijde, MD, PhD, Nico Wolter, MD, and Sjef van der Linden, MD, PhD, Department of Medicine, Division of Rheumatology, University Hospital Maastricht, Maastricht, The Netherlands; [2]Robert Landewé, MD, and Henk Goei Thè, MD, PhD, Atrium Medical Center, Heerlen, The Netherlands; [3]Désirée van der Heijde, MD, PhD, University Center Diepenbeek, Diepenbeek, Belgium; [4]Alita Hidding, PT, MSc, PhD, Institute for Rehabilitation Research, Hoensbroek, The Netherlands; [5]Max Asscher, MD, ARBO Unie Oost-Brabant, Eindhoven, The Netherlands; [6]Albrecht Falkenbach, MD, Gasteiner Heilstollen Hospital and Gastein Research Institute, Bad Gastein, Austria; [7]Ekkehard Genth, MD, PhD, Rheuma Clinic and Research Institute, Aachen, Germany.

Address correspondence to Astrid van Tubergen, MD, Department of Internal Medicine, Division of Rheumatology, University Hospital Maastricht, PO Box 5800, 6202 AZ Maastricht, The Netherlands. E-mail: avantubergen@yahoo.com.

Submitted for publication October 10, 2000; accepted in revised form June 7, 2001.

larity 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 methodologic 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 antiinflammatory 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 antiinflammatory 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.

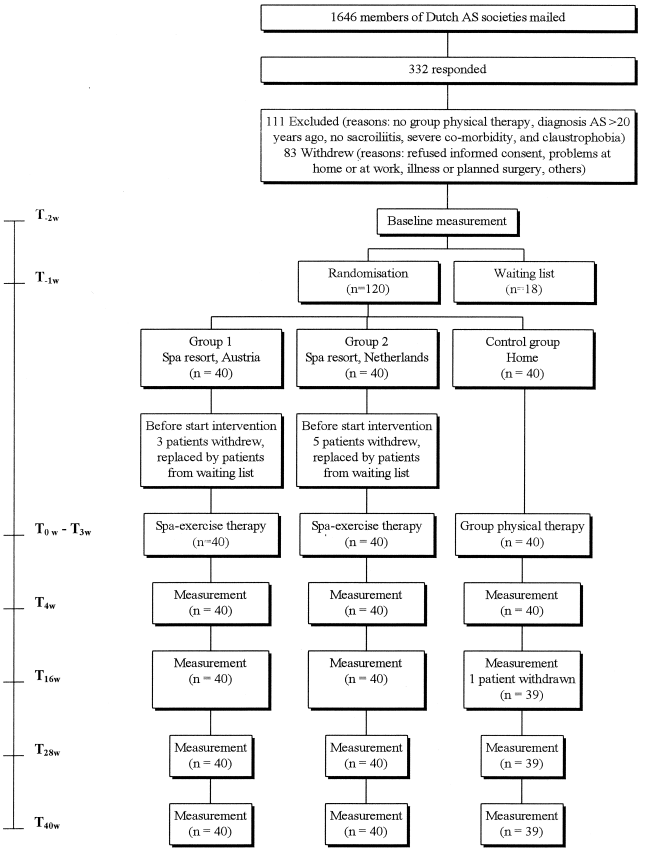

Figure 1. Flow chart of participants. AS = ankylosing spondylitis.

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 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 preplanned 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.

Of the 332 responding patients, 111 were not eligible, 83 declined to cooperate, and 138 patients signed informed consent and were placed consecutively on an inclusion list (Figure 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 3 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, 3 patients from group 1 and 5 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 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 2 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 resorts.

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 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 a 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 char-

Table 1. Treatments of the 3 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) Afternoon Alternately either: —1 hour visit to Gasteiner Heilstollen or: —½ hour hydrotherapy —½ hour bathing —1 hour sports	Morning —1 hour group physical exercises —½ hour walking —Postural correction therapy on bed (14 min. to ½ hour) Afternoon Alternately either: —2 × 15 minutes to sauna —½ hour bathing or: —½ hour hydrotherapy —½ hour bathing —1 hour sports	Once a week group physical therapy consisting of: —1 hour group physical exercises —1 hour sports —1 hour hydrotherapy

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.

acterized 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). 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 3 groups engaged in weekly group physical therapy sessions. During the intervention and the followup periods, all patients continued their usual drug treatment but were allowed to decrease or increase the amount of antiinflammatory drugs based on present complaints during the study period. Another spa treatment was not permitted during the followup 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 the 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 nonsteroidal antiinflammatory drugs (NSAIDs). The BASDAI consists of 6 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 scale has the labels "none" (0) and "very severe" (10) at either end in the first 5 questions, and "0 hours" (0) and "2 or more hours" (10) for the duration of morning stiffness. The mean of the 2 questions on morning stiffness counts as one variable. The total score is calculated by taking the mean of the 5 items, ranging from 0 (best) to 10 (worst). The HAQ-S consists of 8 subscales on health status and function and is extended with 5 additional AS-specific items on function, divided into 2 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

Table 2. Baseline characteristics of the groups

	Group 1 Bad Hofgastein	Group 2 Arcen	Controls Home
Male/female	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 = nonsteroidal antiinflammatory drugs. Disease duration is from time of diagnosis. Duration of complaints is from time complaints started.

developed in cooperation 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 the 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 BASFI changed 1.1 ± 2.3 (mean ± SD) (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 (2-sided α = 0.05; β = 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 nonmissing 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 4 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 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 cooperate. The last observation carried forward method was applied to this patient. After the followup period, 93% of the patients in the 3 groups still attended weekly group physical therapy.

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

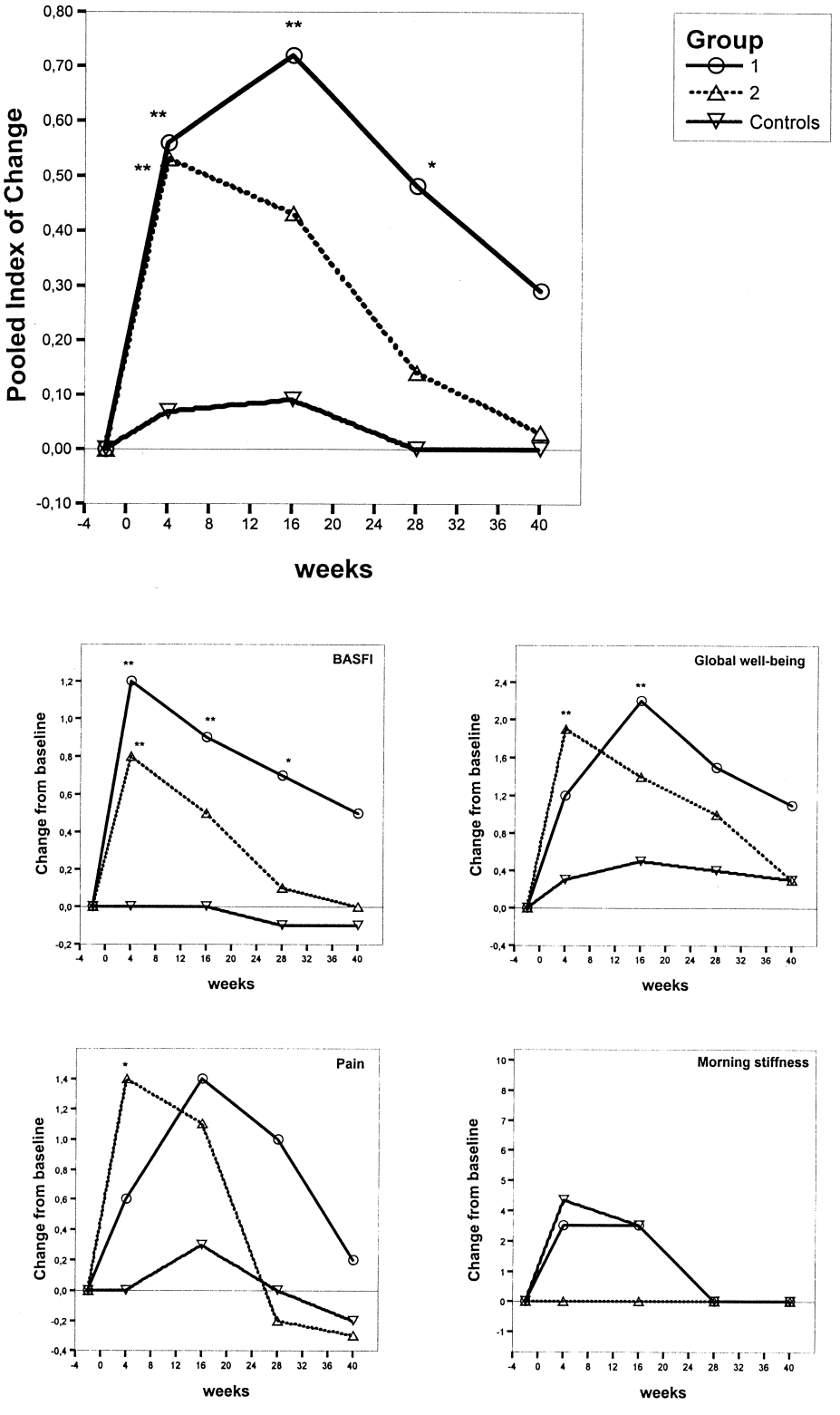

Figure 2. Pooled index of change (PIC) and individual variables of PIC. *$P < 0.05$; **$P < 0.01$

that the 3 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 covariate did not change the time–effect ($P < 0.001$) and only slightly changed the time-by-treatment interaction ($P = 0.011$).

Figure 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

Table 3. Primary outcomes

Measure	Baseline value T_{-2w}	Change from baseline 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 (−5; 15)	0 (−4; 14)
Group 2	30 (15; 60)	0 (−2; 14)	0 (−5; 14)	0 (−9; 14)	0 (−10; 15)
Control	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. BASFI = Bath Ankylosing Spondylitis Functional Index.
* $P < 0.01$.
† $P < 0.05$ between intervention group and control.
‡ $P < 0.05$ between groups 1 and 2. Postive changes imply improvement.
§ Skewed data are presented as median (interquartile range) and median change (interquartile range).

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 groups 1 and 2 were not statistically significant.

The course of the individual variables of the PIC is summarized in Table 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 followup (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. 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.

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 the 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 2 intervention groups cannot be drawn. 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.

Table 4. Secondary outcomes

Measure	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. BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; HAQ-S = Health Assessment Questionnaire for Spondylarthropathies; ASQOL = Ankylosing Spondylitis Quality of Life; NSAID = Nonsteroidal antiinflammatory drugs.
* $P < 0.01$, † $P < 0.05$ between intervention group and control; ‡ $P < 0.05$ between groups 1 and 2. Postive changes imply improvement.
§ Skewed data are presented as median (interquartile range) and median change (interquartile range).

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 nonspecific effects.

Specific effects may be attributed to the intervention itself. Immersion of the body in warm (thermal) water produces many physiologic 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 nonspecific factors may contribute to the observed differences, such as change of environment, the pleasant scenery, the noncompetitive 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 nonspecific 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 followup (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. Schöber 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 with no intervention or 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 methodologic considerations should be addressed. First, the lack of blinding: True double-blind spa therapy trials are impossible to conduct. To prevent differences in nonspecific 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 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.

ACKNOWLEDGEMENTS

We thank the participating patients, D. Martens and B. van Albada from the Dutch League of Associations of Patients with Rheumatic Disorders, all coordinators of the local ankylosing spondylitis patient societies, J. Daanen from Thermaalbad Arcen, M. Schafflinger from Kurzentrum Thermentempel Bad Hofgastein, F. Hartlieb from the Gasteiner Heilstollen, Jan Thomassen from Institute of Rehabilitation Research, and all others who have contributed to a successful completion of this study.

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, 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.
3. Elkayam O, Wigler I, Tishler M, Rosenblum I, Caspi D, Segal R, 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, Abu-Shakra M, Amir M, Flusser D, 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, Gielen X, de Witte L, Kester A, 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, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21: 2281–5.
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, Tennant A, McKenna S, Reynolds S, et al. Psychometric and scaling properties of a new quality of life instrument specific to ankylosing spondylitis (AS). Arthritis Rheum 1999;42: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, Dougados M, de

Vlam K, Mielants H, et al. A comparative study of the usefulness of the Bath Ankylosing Spondylitis Functional Index and the Dougados Functional Index in the assessment of ankylosing spondylitis. J Rheumatol 1999;26:961–5.
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. 103–13.
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, Gielen X, Boers M, Geenen R, 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.