ARTHRITIS & RHEUMATISM
Vol. 52, No. 6, June 2005, pp 1756–1765
DOI 10.1002/art.21054

Nonsteroidal Antiinflammatory Drugs Reduce Radiographic Progression in Patients With Ankylosing Spondylitis

A Randomized Clinical Trial

Astrid Wanders,[1] Désirée van der Heijde,[1] Robert Landewé,[1] Jéhan-Michel Béhier,[2] Andrei Calin,[3] Ignazio Olivieri,[4] Henning Zeidler,[5] and Maxime Dougados[6]

Objective. **A 2-year randomized controlled trial was performed to test the hypothesis that long-term, continuous treatment with nonsteroidal antiinflammatory drugs (NSAIDs), in comparison with NSAID treatment on demand only, influences radiographic progression in patients with ankylosing spondylitis (AS).**

Methods. **Patients with AS (n = 215), who had previously participated in a 6-week, randomized, double-blind clinical trial that compared celecoxib, ketoprofen, and placebo, were randomly allocated to receive either continuous treatment with NSAIDs or on-demand treatment with NSAIDs for a period of 2 years. All patients began treatment with celecoxib, at a starting dosage of 100 mg twice daily; patients could increase this dosage to 200 mg twice daily or could switch to another NSAID while maintaining the same treatment strategy. Structural changes were assessed by radiographs of the lumbar and cervical spine and scored according to the modified Stoke Ankylosing Spondylitis Spine Score by one observer who was blinded to the treatment strategy and temporal order of the radiographs. Statistical analyses included a between-group comparison of 1) radiographic progression scores (by Mann-Whitney U test), 2) time-averaged values of variables reflecting signs and symptoms of AS (by linear regression analysis), and 3) the frequency of reported site-specific adverse events (by chi-square test or Fisher's exact test, as appropriate).**

Results. **Complete sets of radiographs were available for 76 of the 111 patients in the continuous-treatment group and for 74 of the 104 patients in the on-demand group. The mean ± SD scores for radiographic progression were 0.4 ± 1.7 in the continuous-treatment group and 1.5 ± 2.5 in the on-demand treatment group ($P = 0.002$). Parameters reflecting signs and symptoms were not statistically significantly different between groups. The between-group difference in radiographic progression did not disappear after adjusting for baseline values of radiographic damage or disease activity variables and for time-averaged values of disease activity variables, nor after imputation of missing data. Relevant adverse events tended to occur more frequently in the continuous-treatment group than in the on-demand group (for hypertension, 9% versus 3%; for abdominal pain, 11% versus 6%; for dyspepsia, 41% versus 38%), but the differences were not statistically significant.**

Conclusion. **A strategy of continuous use of NSAIDs reduces radiographic progression in symptomatic patients with AS, without increasing toxicity substantially.**

Supported by an unrestricted grant from Pharmacia.

[1]Astrid Wanders, MD, Désirée van der Heijde, MD, PhD, Robert Landewé, MD, PhD: University Hospital, Maastricht, The Netherlands; [2]Jéhan-Michel Béhier, MD: Pharmacia Medical Department, Paris, France; [3]Andrei Calin, MD: Royal National Hospital for Rheumatic Diseases, Bath, UK; [4]Ignazio Olivieri, MD: Ospedale San Carlo, Potenza, Italy; [5]Henning Zeidler, MD: Medizinische Hochschule, Hannover, Germany; [6]Maxime Dougados, MD: René Descartes University, Cochin Hospital, Paris, France.

Dr. Zeidler has received consulting fees and/or honoraria of less than $10,000 from Pfizer, Inc.

Address correspondence and reprint requests to Désirée van der Heijde, MD, University Hospital Maastricht, PO Box 5800, Maastricht 6202 AZ, The Netherlands. E-mail: dhe@sint.azm.nl.

Submitted for publication August 31, 2004; accepted in revised form March 1, 2005.

Numerous studies in patients with ankylosing spondylitis (AS) have demonstrated that nonsteroidal antiinflammatory drugs (NSAIDs) provide rapid relief of inflammatory back pain and stiffness and improve physical function (1–6). NSAIDs are among the most frequently prescribed drugs for AS, but toxic effects on the gastrointestinal tract limit their long-term use. Gastrointestinal adverse events are associated with the inhibition of cyclooxygenase 1 (COX-1), which is responsible for the production of cytoprotective prostaglandins in the gastric mucosa (7). The latest generation of NSAIDs selectively inhibit COX-2, which is upregulated under inflammatory conditions and is responsible for the production of proinflammatory prostaglandins, and leaves COX-1 relatively undisturbed. COX-2–selective NSAIDs are associated with a reduced risk of serious gastrointestinal complications (8,9) but at least have effectiveness in AS similar to that of conventional NSAIDs (2,3).

Because of the risk of serious adverse events, many physicians currently recommend that their patients with AS take NSAIDs only if necessary. It is, however, not established whether the use of NSAIDs can alter the long-term outcome of the disease. Most studies of the efficacy of NSAIDs have been short-term, with a duration up to 6 weeks. Measures of spinal mobility and levels of acute-phase reactants generally did not improve in these studies (3,6,10,11). A placebo-controlled 1-year study compared piroxicam with low-dose and high-dose meloxicam and with placebo (1). At 1 year, but not at 6 weeks, significantly more improvement in chest expansion and the C-reactive protein level was observed in patients who received NSAIDs compared with those who received placebo. These findings suggest that NSAIDs may, to some extent, control the disease process. Another study that points to the disease-controlling potential of NSAIDs is an uncontrolled, retrospective, observational study showing that phenylbutazone impaired ossification of the vertebral column in patients with AS (12).

The improved gastroprotective safety profile of COX-2–selective NSAIDs, as compared with unselective NSAIDs, justifies a formal test of the hypothesis that NSAIDs may alter the course of AS, thus disputing the current recommendation to take NSAIDs only as needed. In a 2-year randomized clinical trial, we compared the strategies of long-term continuous NSAID use and on-demand use of NSAIDs, with respect to their influence on radiographic progression.

PATIENTS AND METHODS

Patients. The study was conducted from June 1998 to July 2001 by rheumatologists from 76 rheumatology centers (both hospitals and private practices) in France. The Ethics Committee of Cochin Hospital, Paris, approved the study. Two hundred fifteen outpatients fulfilling the modified New York criteria for AS (13), all of whom had previously participated in a 6-week randomized, double-blind clinical trial comparing celecoxib 100 mg twice daily with ketoprofen 100 mg twice daily and with placebo, were included after providing written informed consent (3). Criteria for inclusion in the 6-week trial were 1) daily NSAID intake during the month preceding the screening visit, 2) an NSAID washout period of 2–14 days before the baseline visit, and 3) a flare of the disease at baseline, defined by both an absolute score for pain of $\geq$40 mm on a 100-mm visual analog scale (VAS) and an increase in pain of at least 30% between the screening visit and the baseline visit.

Patients with peripheral arthritis, defined by the presence of active (with swelling) synovitis of a peripheral joint (excluding the shoulder) at the screening visit, and those with active inflammatory bowel disease were excluded, as were patients with concomitant severe medical illness. Patients who had received corticosteroids during the previous 6 weeks and/or any disease-modifying antirheumatic drug with a change in dosage during the previous 6 months were also excluded, as were patients with peptic ulcer confirmed by gastroduodenoscopy within the year preceding the screening visit.

Study design. The present study was a randomized, open-label, comparative trial of 2 strategies. At the final visit of the preceding 6-week trial, which is considered the baseline visit of the present study, patients were randomly allocated to receive either continuous treatment with NSAIDs or treatment on demand only. Randomization was performed using a computer-generated randomization list. Patients allocated to the continuous-treatment strategy received daily treatment with an NSAID, irrespective of symptoms, for a period of 2 years. Patients allocated to the on-demand treatment strategy were instructed to take their NSAID only when they had serious symptoms (e.g., pain, stiffness). Patients in both treatment groups started treatment with celecoxib at a dosage of 100 mg twice daily and were allowed to increase the dosage to 200 mg twice daily if necessary, at their discretion. If for any reason (inefficacy or adverse event) the initial treatment with celecoxib was discontinued, patients were allowed to continue the study with any other NSAID, but they were instructed to maintain the allocated treatment strategy (continuous treatment or treatment on demand only). Compliance was assessed by pill count.

Study visits. There were 10 planned visits (Figure 1), as follows: the baseline visit (month 0), a followup visit conducted after a 1-month interval, 7 followup visits conducted at 3-month intervals (months 4, 7, 10, 13, 16, 19, and 22), and a final visit at 24 months. At every visit, clinical signs and symptoms and adverse events were assessed. Laboratory tests were performed during the visits at months 1, 7, 13, 19, and 24. Radiography of the spine was performed at baseline and at 24 months.

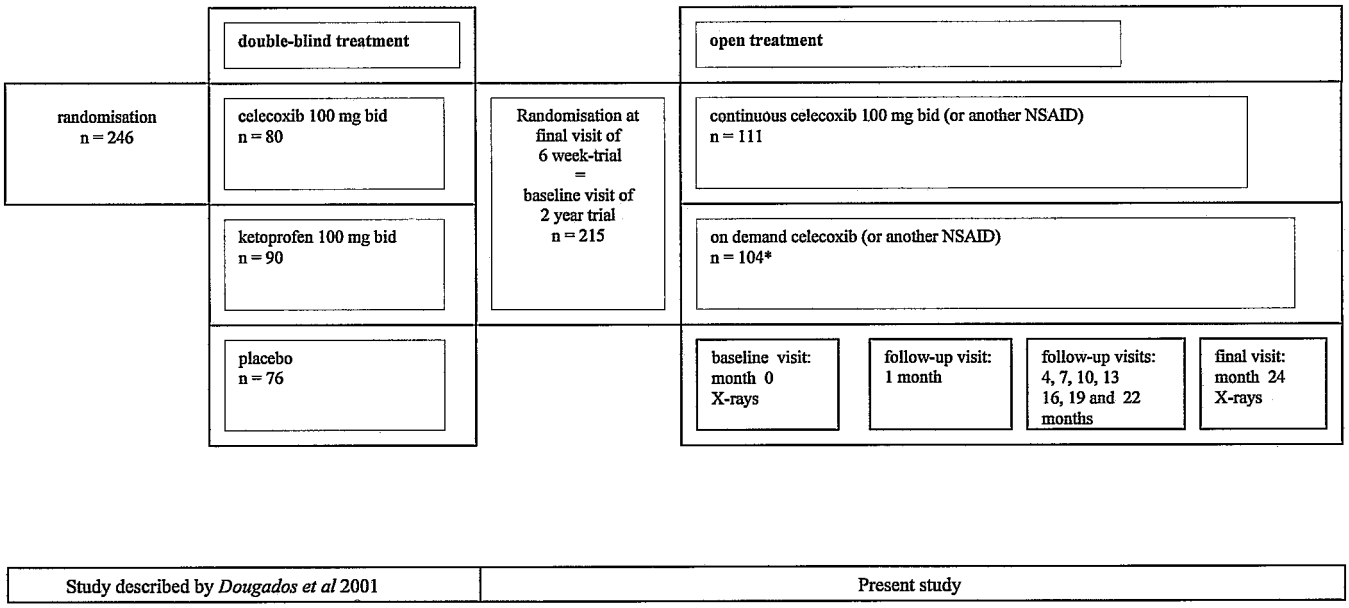

Figure 1. Flow chart of participation in the present study and the study described in ref. 3. bid = twice daily; NSAID = nonsteroidal antiinflammatory drug. * = One patient died in a car accident before treatment started; therefore, only 103 patients were analyzed.

Assessments. Structural damage was scored on radiographs of the lumbar and cervical spine, according to the modified Stoke Ankylosing Spondylitis Spine Score (SASSS) (14), by a single observer (AW) who was blinded to the treatment strategy and the temporal order of the radiographs. The difference between the modified SASSS at month 0 and month 24 was considered the progression score (range 0–72 modified SASSS units). Intraobserver and interobserver reliability of the modified SASSS was tested in a previous experiment. Intraclass correlation coefficients were 0.95 and 0.82, respectively (14).

Disease activity was measured using the 6-question, patient-reported Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (where 0 = no disease activity and 100 = highest level of disease activity) (15). Functional capacity was measured using the 10-question, patient-reported Bath Ankylosing Spondylitis Functional Index (BASFI) (where 0 = lowest level and 100 = highest level) (16). Pain was measured by 3 variables, as follows: 1) patient-reported global pain intensity on a 100-mm visual analog scale (VAS), 2) the 4-item spinal pain index, in which pain in the cervical spine, pain in the dorsal spine, pain in the lumbar spine, and pain in the sacroiliac joints were assessed on a 5-point Likert scale ranging from 0 (no pain) to 4 (unbearable pain); the spinal pain index was calculated as the sum score of the 4 items, ranging from 0 (no pain) to 16 (extreme pain), and 3) the percentage of days during the last 3 months on which the patient experienced pain, as measured on a 100-mm VAS (where 0 = no pain at all during the past 3 months and 100 = pain every day during the past 3 months). Inflammation was measured by 3 variables, as follows: 1) nocturnal pain during the last week as measured on a 100-mm VAS (where 0 = no nocturnal pain and 100 = extreme nocturnal pain), 2) duration and severity of morning stiffness during the last week (score represents the mean of the scores [on a VAS] for the fifth and sixth BASDAI questions), and 3) the C-reactive protein level and the erythrocyte sedimentation rate. Spinal mobility was measured by 4 variables, as follows: 1) the fingertip-to-floor distance, 2) the modified Schober test, 3) chest expansion, and 4) the occiput-to-wall distance (17). Fatigue was measured according to the score (on a VAS) for the first BASDAI question (where 0 = no fatigue and 100 = extreme fatigue). Global disease activity during the last week was measured by both the patient and the investigator on a 100-mm VAS (where 0 = no disease activity and 100 = severe disease activity).

Missing variable values. Missing values for variables assessing signs and symptoms were replaced by the last observation that was present, which was carried forward, provided that at least one value obtained while the patient was receiving treatment was available.

Safety variables. The investigator actively asked about adverse events at every visit. Laboratory assessments, including evaluation of the hemoglobin concentration, the platelet count, and the levels of serum creatinine, aspartate aminotransferase (AST), and alanine aminotransferase (ALT), were performed at 6-month intervals. Systolic and diastolic blood pressure were measured at every visit.

Statistical analysis. An exploratory analysis included time plots of all variables assessing signs and symptoms, stratified by treatment group, and probability plots of radiographic progression scores, stratified by treatment group. Probability plots present every patient's progression score against its cumulative frequency (expressed as a proportion and referred to as cumulative probability [18]).

The primary set of statistical analyses included a between-group comparison of 1) radiographic progression

Table 1. Baseline characteristics of all patients and the subgroup of patients with a complete set of radiographs, according to treatment group

Characteristic	All patients		Patients with a complete set of radiographs	
	Continuous use (n = 111)	On-demand use (n = 103)	Continuous use (n = 76)	On-demand use (n = 74)
Age, mean ± SD years	38.0 ± 10.7	40.1 ± 10.5	40.9 ± 9.8	37.9 ± 11.9
Male sex, %	67	72	66	70
Disease duration, mean ± SD years	11.9 ± 9.3	11.0 ± 9.4	13.0 ± 10.2	10.2 ± 9.3
HLA–B27 positive, %	86	87	88	88
DMARD use, %*	29	26	26	27
Sulfasalazine	25	22	24	27
Methotrexate	3	2	1	3
Other†	2	2	1	3
Analgesic use, %	10	9	11	9

* DMARD = disease-modifying antirheumatic drug.
† Gold or corticosteroids.

scores (by Mann-Whitney U test), 2) time-averaged values for variables reflecting signs and symptoms (by linear regression analysis, with treatment group and baseline values of these variables as independent variables), and 3) the frequency of reported site-specific adverse events (by chi-square test or Fisher's exact test, as appropriate). Other statistical analyses included 1) linear regression analysis of radiographic progression, with treatment group as the factor and baseline variables of disease activity and radiographic damage as covariates (to investigate whether baseline differences could account for between-group differences at the end of the trial), and 2) linear regression analysis of radiographic progression, with treatment group as the factor and time-averaged means of disease activity variables as covariates (to investigate whether a difference in radiographic progression could be explained by differences in disease activity during followup). Van der Waerden–

Table 2. Disease activity in all patients*

Variable	Baseline		Month 1		Time-averaged mean		
	Continuous (n = 111)	On-demand (n = 103)	Continuous (n = 111)	On-demand (n = 103)	Continuous (n = 111)	On-demand (n = 103)	P†
Disease activity							
BASDAI (0–100)	NA	NA	30 ± 19	32 ± 24	30 ± 18	32 ± 20	0.51
Function							
BASFI (0–100)	33 ± 25	38 ± 28	31 ± 24	32 ± 26	30 ± 21	31 ± 23	0.33
Pain							
Global pain (0–100)	50 ± 38	54 ± 37	37 ± 27	39 ± 27	37 ± 22	40 ± 23	0.44
Spinal pain index (0–16)	6.4 ± 3.6	6.9 ± 3.3	5.5 ± 3.6	6.1 ± 3.3	5.4 ± 3.2	5.7 ± 2.8	0.88
Percent painful days (0–100)	NA	NA	46 ± 33	52 ± 35	45 ± 27	49 ± 26	0.32
Inflammation							
Night pain (0–100)	38 ± 32	43 ± 34	25 ± 25	31 ± 28	27 ± 21	32 ± 24	0.91
Morning stiffness (0–100)	NA	NA	30 ± 22	34 ± 27	29 ± 20	32 ± 22	0.61
C-reactive protein, mg/liter	14.7 ± 17.9	12.7 ± 17.1	15.8 ± 22.6	12.0 ± 15.5	14.5 ± 17.8	12.3 ± 16.1	0.82
ESR, mm/hour	17.0 ± 13.8	17.0 ± 16.7	16.3 ± 13.7	16.9 ± 15.7	15.7 ± 11.4	15.8 ± 13.3	0.40
Spinal mobility							
Fingertip-to-floor distance, cm	21.2 ± 15.6	23.0 ± 14.7	19.4 ± 14.9	19.4 ± 14.1	19.0 ± 13.6	19.7 ± 13.2	0.48
Schober test, cm	3.2 ± 1.4	3.2 ± 1.4	3.2 ± 1.5	3.3 ± 1.4	3.2 ± 1.3	3.3 ± 1.3	0.97
Chest expansion, cm	4.7 ± 2.3	5.0 ± 2.3	4.8 ± 2.2	5.2 ± 2.1	5.0 ± 2.2	5.3 ± 2.1	0.65
Occiput-to-wall distance, cm	NA	NA	3.2 ± 4.3	3.0 ± 3.6	3.5 ± 4.5	2.8 ± 3.3	0.51
Fatigue (0–100)	NA	NA	38 ± 24	38 ± 28	38 ± 22	40 ± 24	0.53
Global assessment (0–100)							
Patient	43 ± 29	47 ± 31	37 ± 27	40 ± 26	37 ± 23	40 ± 22	0.94
Physician	42 ± 28	44 ± 29	32 ± 25	35 ± 25	32 ± 20	34 ± 19	0.60

* Values are the mean ± SD. BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; NA = not available; BASFI = Bath Ankylosing Spondylitis Functional Index; ESR = erythrocyte sedimentation rate.
† Between-group differences were analyzed by linear regression, with treatment group and baseline values as independent variables (if baseline values were not available, month 1 values were used).

Table 3. Disease activity in patients with a complete set of radiographs available*

Variable	Baseline		Month 1		Time-averaged mean		
	Continuous (n = 76)	On-demand (n = 74)	Continuous (n = 76)	On-demand (n = 74)	Continuous (n = 76)	On-demand (n = 74)	P†
Disease activity							
BASDAI (0–100)	NA	NA	29 ± 18	31 ± 25	26 ± 17	32 ± 20	0.17
Function							
BASFI (0–100)	30 ± 23	36 ± 28	28 ± 22	31 ± 26	27 ± 19	30 ± 23	0.72
Pain							
Global pain (0–100)	46 ± 37	52 ± 37	35 ± 25	37 ± 26	33 ± 19	39 ± 22	0.13
Spinal pain index (0–16)	5.7 ± 3.3	6.6 ± 3.0	5.0 ± 3.3	5.9 ± 3.3	4.6 ± 2.7	5.8 ± 2.8	0.07
Percent painful days (0–100)	NA	NA	45 ± 33	54 ± 35	39 ± 23	48 ± 25	0.12
Inflammation							
Night pain (0–100)	34 ± 31	41 ± 33	21 ± 22	32 ± 29	22 ± 18	32 ± 24	0.01
Morning stiffness (0–100)	NA	NA	29 ± 21	33 ± 27	26 ± 19	31 ± 21	0.25
C-reactive protein, mg/liter	13.1 ± 15.3	12.2 ± 17.5	13.7 ± 22.8	10.8 ± 13.4	12.8 ± 14.8	16.5 ± 14.4	0.82
ESR, mm/hour	16.5 ± 13.1	17.5 ± 18.1	16.3 ± 13.6	17.1 ± 16.7	15.0 ± 10.4	12.2 ± 15.8	0.40
Spinal mobility							
Fingertip-to-floor distance, cm	18.0 ± 14.4	21.3 ± 13.2	17.7 ± 14.7	17.9 ± 13.3	16.1 ± 13.0	18.3 ± 12.3	0.76
Schober test, cm	3.3 ± 1.4	3.2 ± 1.4	3.3 ± 1.4	3.2 ± 1.4	3.4 ± 1.3	3.2 ± 1.3	0.31
Chest expansion, cm	4.9 ± 2.3	5.2 ± 2.3	5.0 ± 2.2	5.3 ± 2.1	5.3 ± 2.2	5.5 ± 2.0	0.68
Occiput-to-wall distance, cm	NA	NA	2.8 ± 3.7	2.7 ± 3.7	2.9 ± 4.2	2.5 ± 3.3	0.45
Fatigue (0–100)	NA	NA	36 ± 25	36 ± 28	33 ± 19	38 ± 23	0.02
Global assessment (0–100)							
Patient	39 ± 27	44 ± 31	34 ± 24	39 ± 27	31 ± 20	39 ± 22	0.07
Physician	38 ± 26	41 ± 29	30 ± 24	34 ± 25	27 ± 17	34 ± 19	0.05
Modified SASSS score	7.9 ± 14.7	9.3 ± 15.2	—	—	—	—	—

* Values are the mean ± SD. BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; NA = not available; BASFI = Bath Ankylosing Spondylitis Functional Index; ESR = erythrocyte sedimentation rate; SASSS = Stoke Ankylosing Spondylitis Spine Score.
† Between-group differences were analyzed by linear regression, with treatment group and baseline values as independent variables (if baseline values were not available, month 1 values were used).

normalized data were used in the linear regression analyses if variables had a non-normal distribution pattern. Residual plots were checked for homogeneity and linearity.

A sensitivity analysis included reanalysis of the primary results after imputation of missing data by 2 means: 1) imputation by the mean value of the total population (which overestimates true progression), and 2) imputation by 0 (which underestimates true progression).

RESULTS

Patients. The study group comprised 215 patients, 111 of whom were randomized to the continuous-treatment group, and 104 of whom were randomized to the on-demand treatment group. One patient in the on-demand group was excluded from the analysis because he died in a car accident before starting the trial.

In the continuous-treatment group, 96 patients completed the study (68 completed the study while taking celecoxib, and 28 patients completed the study while taking a different NSAID). The reasons for withdrawal of the 15 patients in this group were inefficacy (n = 8), adverse events (n = 2), moving to another city or country (n = 2), and unknown (n = 3).

In the on-demand treatment group, 86 patients completed the study (67 patients completed the study while taking celecoxib, and 19 patients completed the study while taking a different NSAID). The reasons for withdrawal of the 17 patients in this group were inefficacy (n = 8), adverse events (n = 3), moving to another city or country (n = 2), and unknown (n = 4).

Complete sets of radiographs were available for 76 patients in the continuous-treatment group and for 74 patients in the on-demand group. The baseline characteristics of all randomized patients as well as patients for whom a complete set of radiographs was available are shown in Table 1. With respect to all randomized patients, between-group differences at baseline were small and negligible. Among patients for whom a complete set of radiographs was available, age and disease duration were lower in patients in the on-demand group than in those in the continuous-treatment group, but these differences were not statistically significant. Other baseline variables were similar.

Table 2 shows the variables reflecting disease activity at baseline and during the trial for all randomized patients, and Table 3 shows the same variables in patients for whom a complete set of radiographs was

available. Overall, disease activity tended to be lower in the continuous-treatment group as compared with the on-demand group; this difference was more pronounced among patients for whom a complete set of radiographs were available, but the between-group differences were not statistically significant.

Mean daily dose of celecoxib. Based on pill counts, the mean ± SD daily dose of celecoxib was 243 ± 59 mg in the continuous-treatment group and 201 ± 93 mg in the on-demand group. The mean difference of 42 mg (95% confidence interval 21–63) was statistically significant ($P = 0.0001$).

Radiographic progression. At baseline, the mean ± SD radiographic damage score in the 2 treatment groups was similar (7.9 ± 14.7 modified SASSS units in the continuous-treatment group and 9.3 ± 15.2 units in the on-demand treatment group). The probability plot for radiographic progression, based on use of the modified SASSS, in the 2 treatment groups over 24 months (Figure 2) showed radiographic progression (>0 units) in a greater proportion of patients in the on-demand treatment group (45%) compared with the continuous-treatment group (22%). Using a cutoff value of ≥3 units, again twice as many patients in the on-demand group compared with the continuous-treatment group showed this level of progression (23% versus 11%). The maximum progression scores were 4 units and 9 units in the continuous-treatment group and the on-demand group, respectively. The curve for the on-demand group lies left of the curve for the continuous-treatment group along the entire range, reflecting a higher level of radiographic progression. The mean ± SD scores for radiographic progression after 2 years were 0.4 ± 1.7 modified SASSS units in the continuous-treatment group and 1.5 ± 2.5 modified SASSS units in the on-demand group. The between-group difference was statistically significant ($P = 0.002$).

A sensitivity analysis showed that imputation of missing data by different means did not influence the direction of the between-group difference ($P = 0.002$ for the between-group difference after imputation with the entire group mean, and $P = 0.077$ after imputation with the value 0).

Signs and symptoms. The course of 4 variables of disease activity, which can be considered representative of the other variables including those reflecting spinal mobility, is shown in Figure 3. It is obvious that disease activity was stable over time in both groups, after some decrease in global pain between baseline and the first assessment. The general impression is that disease activity, as measured by any of these variables, was somewhat higher in the on-demand group than in the continuous-treatment group. Time-averaged values of all variables reflecting signs and symptoms are shown in Tables 2 and 3. The results confirm the impression of somewhat higher disease activity in the on-demand group as compared with the continuous-treatment group, but the differences were not statistically significant for any variable. When the analysis was limited to patients with a complete set of radiographs (Table 3), the time-averaged values for pain at night ($P = 0.01$), fatigue ($P = 0.02$), and physician's global assessment ($P = 0.05$) were significantly worse in the on-demand treatment group. Adjustments for multiple testing were not performed.

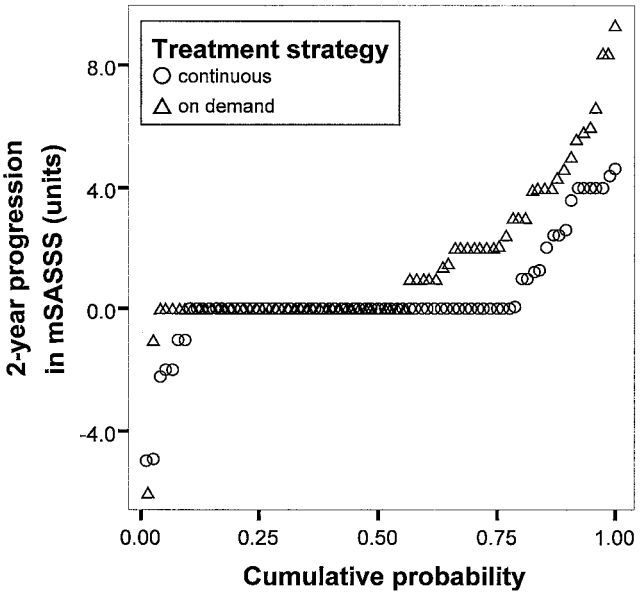

Figure 2. Probability plot of modified Stoke Ankylosing Spondylitis Spine Score (mSASSS) progression over 24 months.

Confounding. In the group of patients for whom a complete set of radiographs was available, we investigated whether the observed baseline differences in signs and symptoms could explain the between-group difference in radiographic progression. We performed linear regression analysis of radiographic progression, with treatment as the factor and baseline values of variables reflecting signs and symptoms (all entered separately) as covariates. The factor treatment remained statistically significant in all analyses, and the regression coefficient for treatment did not change substantially (always <10%), which indicates that the between-group differ-

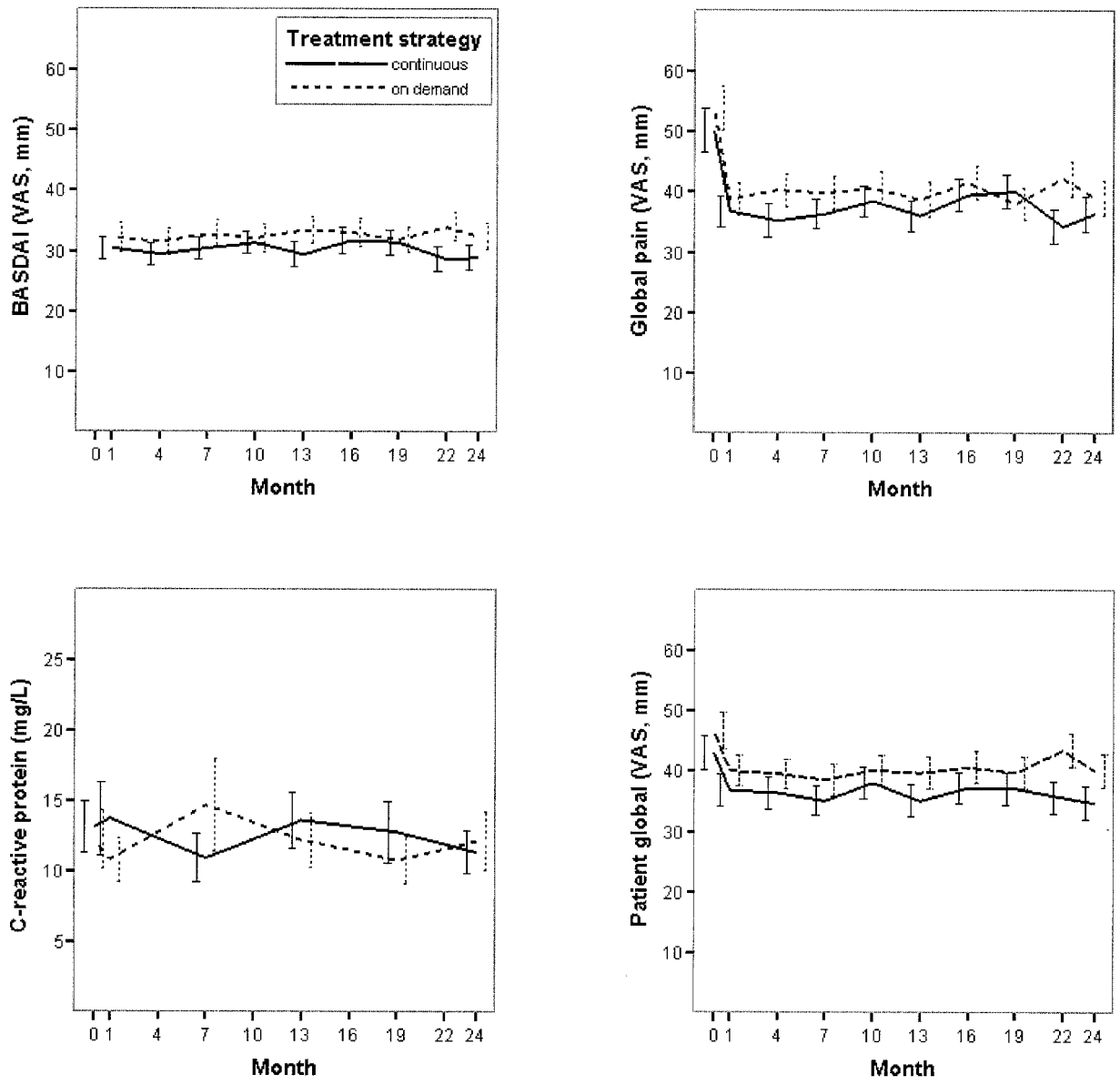

Figure 3. The course of 4 variables of disease activity in patients with a complete set of radiographs, according to treatment strategy. BASDAI = Bath Ankylosing Spondylitis Disease Activity Index; VAS = visual analog scale. Bars show the mean ± SD.

ence in radiographic progression cannot be explained by baseline differences between the 2 groups.

We further investigated whether differences in signs and symptoms *during* treatment could explain the between-group difference in radiographic progression, by entering the time-averaged values as covariates in the linear regression analysis described above. Again, the regression coefficient for the factor treatment was not substantially influenced by any of the time-averaged variables, suggesting that differences in signs and symptoms during followup could not explain the observed difference in radiographic progression.

Safety. Serious adverse events were reported 22 times by 22 separate patients (19.8%) in the continuous-treatment group, and 25 times in 16 separate patients (15.5%) in the on-demand group. Only one of these serious adverse events (a case of severe abdominal pain requiring hospital admission in the on-demand group) was considered by the treating physician to be related to the study medication.

Table 4. Adverse events in the 2 treatment groups*

Adverse event	Continuous use (n = 111)	On-demand use (n = 103)
Cardiovascular		
Hypertension	10	3
Angina pectoris	0	2
Coronary artery disorder	0	1
Myocardial infarction	0	2
Edema	4	3
Neurologic		
Headache	13	13
Vertigo	7	5
Gastrointestinal		
Abdominal pain	12	6
Diarrhea	21	13
Duodenal ulcer	1	0
Dyspepsia	46	39
Gastroenteritis	17	13
Esophageal symptoms	5	4
Laboratory abnormalities		
Increased liver enzymes	3	0
Increased serum creatinine level	1	0
Neoplastic		
Undifferentiated carcinoma	0	2
Gastrointestinal malignant neoplasm	0	1
Pseudomononucleosis	1	0
Platelet, bleeding, and clotting disorders	1	3
Psychiatric		
Anxiety	7	6
Symptoms of depression	15	4
Sleeping disorder	8	5
Respiratory		
Upper respiratory tract infection	45	44
Lower respiratory tract infection	14	17
Skin		
Rash	5	8
Pruritus	9	6

* Values are the number of patients.

The most important and frequently occurring adverse events are shown in Table 4. Gastrointestinal adverse events occurred most often, followed by respiratory tract infections. Numerically, some relevant adverse events occurred more frequently in the continuous-treatment group compared with the on-demand group, including hypertension (10 patients and 3 patients, respectively; $P = 0.12$), abdominal pain (12 patients and 6 patients, respectively; $P = 0.28$), diarrhea (21 patients and 13 patients, respectively; $P = 0.28$), and dyspepsia (46 patients and 39 patients, respectively; $P = 0.65$), but the differences were not statistically significant. Symptoms of depression occurred more frequently in the continuous-treatment group (in 15 patients versus 4 patients in the on-demand group), and this difference was statistically significant ($P = 0.03$).

The mean blood pressure was similar in both groups and remained at a constant level. The hemoglobin concentration and the levels of AST, ALT, and serum creatinine all remained at a constant level and were similar in both groups.

DISCUSSION

The main conclusion of this study, which compared the strategies of continuous use of NSAIDs and use of NSAIDs on demand only, is that a strategy of continuous treatment reduces radiographic progression despite a similar effect of both strategies on signs and symptoms (pain, inflammation, spinal mobility), whereas a strategy of continuous use of NSAIDs is not associated with significantly more toxicity. This observation provides a strong indication that NSAIDs may have disease-controlling properties.

The ASsessment in Ankylosing Spondylitis Working Group selected radiographic evaluation as an obligatory outcome assessment to prove disease-controlling properties of a medication (19). Until recently, however, and in contrast with the situation in rheumatoid arthritis (RA), only very few studies had included radiology as an outcome parameter. A literature search of the potential of NSAIDs to reduce radiographic progression revealed only one article, by Boersma (12), who performed an uncontrolled cohort study, the results of which suggested that continuous use of phenylbutazone may reduce ossification of the vertebral column in patients with AS. The best explanation for the scarcity of data is that NSAIDs have been considered symptom modifiers rather than drugs that may control the course of disease. This picture changed after the demonstration that tumor necrosis factor (TNF)–blocking drugs, which inhibit radiographic progression in RA, are also highly effective in alleviating the symptoms of AS (20,21), and it is expected that as in RA, use of TNF-blocking drugs in patients with AS can inhibit radiographic progression.

Another reason for the absence of radiographic data in AS clinical trials is the lack of a reliable scoring method. Only recently, in a method-comparing cohort study, we demonstrated the superiority (with respect to sensitivity to change) of the modified SASSS for detecting 2-year changes (14), and we decided to use this scoring method in the present study. Indeed, it appeared to be possible to detect progression after 2 years in a significant proportion of patients in both groups by using the modified SASSS, with radiographs scored by a single reader who was blinded to the temporal order of the radiographs. We believe that concealment of the reading order is as important as scoring blinded for treatment

allocation, because it prevents expectation bias, and the occurrence of negative scores gives a good impression about the level of measurement error (under the assumption that true-negative scores are impossible) (18). As such, concealment of both the time order and treatment allocation adds to the validity of the results and to their credibility.

The finding that NSAIDs reduce radiographic progression requires a proper biologic explanation. COX-2 is relevant in bone formation: both COX-2–knockout mice and mice treated with COX-2–inhibiting drugs showed reduced callus formation after a fracture, which is attributable to suppression of osteoblasts (22). In an immunohistochemical analysis comparing synovial tissue samples obtained from patients with AS, patients with osteoarthritis, patients with RA, and patients with psoriatic arthritis, COX-2 expression appeared to be highest in the samples from patients with AS (23). If an up-regulated level of COX-2 in AS is indeed responsible for increased osteoblastic bone formation (syndesmophytes), inhibition of COX-2 by NSAIDs may be a rational approach to preventing the occurrence of syndesmophytes. Both nonselective and COX-2–selective NSAIDs inhibit COX-2, and the radiographic effects of COX-2 inhibition in AS can therefore be expected with selective and nonselective NSAIDs. The clinical finding that (conventional) NSAIDs may reduce the risk of heterotopic bone formation after hip arthroplasty by 50–65% (24,25) is consistent with the observations in animal models, as well as with the results of the present study.

It could be argued that if we had strived for tighter control of disease activity in the continuous-treatment group, the between-group differences might have been even greater. However, we noticed a remarkable lack of association between radiographic progression and variables reflecting disease activity (pain and inflammation), which is in contrast with the situation in RA (26,27). The data from our study suggest—but do not prove—that inflammation and progression of structural damage are 2 separate processes in AS. Results of a study by Lussier and de Medicis in a rat experimental model (28) provide support for such a dissociation: of 3 different NSAIDs, the drug exhibiting almost no antiinflammatory activity (phenylbutazone) appeared to be the best inhibitor of ossification.

We found that the incidence of several relevant adverse events in the continuous-treatment group was higher than that in the on-demand treatment group, although the difference was not statistically significant. It is important to mention, however, that this study was not designed to provide a reliable picture of adverse events associated with long-term NSAID use: patients were not blinded to the treatment strategy, and knowledge of the treatment strategy may well have influenced the rate of reporting adverse events. Given the rather low incidence of relevant adverse events in the context of the design of the study, and given the absence of drug-related serious adverse events, we conclude that both strategies have an acceptable long-term toxicity profile.

This study may evoke a number of concerns. First, a trial comparing strategies in which the drugs and dosages are not completely fixed is susceptible to bias, because patients and physicians try to minimize the level of pain within the limits of the protocol (confounding by indication). However, this type of confounding will generally reduce between-group differences and cannot be responsible for the contrast in radiographic progression.

Second, only the dosage of celecoxib was recorded; the dosage of other NSAIDs (after a switch) was not recorded. Therefore, if patients switched NSAIDs, we were unable to determine whether patients were still compliant with the allocated treatment strategy. However, a bias caused by violation of the allocated treatment strategy would have obscured rather than revealed a true between-group contrast in radiographic progression. The between-group difference in the mean dose of celecoxib was rather small (although highly statistically significant). It should be noted, however, that a mean dose does not appropriately reflect the pattern of drug use: a strategy involving moderate doses of an NSAID taken daily and a strategy involving high doses of an NSAID taken every other day both may arrive at the same mean dose but may have different pharmacodynamic consequences. As such, the mean dose of celecoxib is not the best parameter with which to explain the differences in radiographic progression. Additional information on the timing of dosing would have shed more light on this aspect. However, this information cannot be deduced from pill counts, and the only valid instrument for measuring this is an electronic monitoring device that registers the dates and times at which the pill bottle is opened. Such a technique was not applied in the present study.

Third, one may think that the observation of reduced progression is coincidental (Type I error) or is biased by baseline differences and/or missing observations in 30% of the patients. Obviously, a Type I error can never be entirely excluded, and our observation awaits confirmation from other studies. Nonetheless,

several arguments point to a true effect. For example, the sensitivity analyses by various means of missing value imputation all arrived at the same result. Moreover, the treatment contrast did not disappear after adjustment for any of the potential confounders at baseline. Finally, the literature provides arguments for the biologic plausibility of our observation, which may add to its validity.

Thus, we conclude that a strategy of continuous use of NSAIDs decreases radiographic progression in patients with AS without substantially increasing toxicity. While awaiting confirmation of these results, we carefully recommend that if patients need treatment with NSAIDs to reduce the signs and symptoms of AS, they should take NSAIDs continuously instead of as needed based on symptoms. Currently, data underscoring such a recommendation in asymptomatic patients are lacking.

REFERENCES

1. Dougados M, Gueguen A, Nakache JP, Velicitat P, Veys EM, Zeidler H, 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.
2. Van der Heijde D, Baraf H, Ramos-Remus C, Calin A, Weaver A, Schiff M, et al. Evaluation of the efficacy of etoricoxib in ankylosing spondylitis: results of a 52-week randomized controlled study. Arthritis Rheum. In press.
3. Dougados M, Behier JM, Jolchine I, Calin A, van der Heijde D, Olivieri I, et al. Efficacy of celecoxib, a cyclooxygenase 2–specific inhibitor, in the treatment of ankylosing spondylitis: a six-week controlled study with comparison against placebo and against a conventional nonsteroidal antiinflammatory drug. Arthritis Rheum 2001;44:180–5.
4. Dougados M. Treatment of spondyloarthropathies: recent advances and prospects in 2001. Joint Bone Spine 2001;68:557–63.
5. Calin A, Grahame R. Double-blind cross-over trial of flurbiprofen and phenylbutazone in ankylosing spondylitis. Br Med J 1974;4:496–9.
6. Dougados M, Nguyen M, Caporal R, Legeais J, Bouxin-Sauzet A, Pellegri-Guegnault B, et al. Ximoprofen in ankylosing spondylitis: a double blind placebo controlled dose ranging study. Scand J Rheumatol 1994;23:243–8.
7. Dougados M, Dijkmans B, Khan M, Maksymowych W, van der Linden S, Brandt J. Conventional treatments for ankylosing spondylitis. Ann Rheum Dis 2002;61 Suppl 3:iii40–50.
8. Langman MJ, Jensen DM, Watson DJ, Harper SE, Zhao PL, Quan H, et al. Adverse upper gastrointestinal effects of rofecoxib compared with NSAIDs. JAMA 1999;282:1929–33.
9. Simon LS, Weaver AL, Graham DY, Kivitz AJ, Lipsky PE, Hubbard RC, et al. Anti-inflammatory and upper gastrointestinal effects of celecoxib in rheumatoid arthritis: a randomized controlled trial. JAMA 1999;282:1921–8.
10. 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.
11. Dougados M, Caporal R, Doury P, Thiesce A, Pattin S, Laffez B, et al. A double blind crossover placebo controlled trial of ximoprofen in AS. J Rheumatol 1989;16:1167–9.
12. Boersma JW. Retardation of ossification of the lumbar vertebral column in ankylosing spondylitis by means of phenylbutazone. Scand J Rheumatol 1976;5:60–4.
13. 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.
14. Wanders AJ, Landewe RB, Spoorenberg A, Dougados M, van der Linden S, Mielants H, et al. What is the most appropriate radiologic scoring method for ankylosing spondylitis? A comparison of the available methods based on the Outcome Measures in Rheumatology Clinical Trials filter. Arthritis Rheum 2004;50:2622–32.
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. 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.
17. Bellamy N. Clinimetric concepts in outcome assessment: the OMERACT filter. J Rheumatol 1999;26:948–50.
18. Landewe R, van der Heijde D. Radiographic progression depicted by probability plots: presenting data with optimal use of the individual values. Arthritis Rheum 2004;50:699–706.
19. Van der Heijde D, van der Linden S, Bellamy N, Calin A, Dougados M, Khan MA. Which domains should be included in a core set for endpoints in ankylosing spondylitis? Introduction to the ankylosing spondylitis module of OMERACT IV. J Rheumatol 1999;26:945–7.
20. Brandt J, Haibel H, Cornely D, Golder W, Gonzalez J, Reddig J, et al. Successful treatment of active ankylosing spondylitis with the anti–tumor necrosis factor α monoclonal antibody infliximab. Arthritis Rheum 2000;43:1346–52.
21. Gorman JD, Sack KE, Davis JC Jr. Treatment of ankylosing spondylitis by inhibition of tumor necrosis factor α. N Engl J Med 2002;346:1349–56.
22. Zhang X, Schwarz EM, Young DA, Puzas JE, Rosier RN, O'Keefe RJ. Cyclooxygenase-2 regulates mesenchymal cell differentiation into the osteoblast lineage and is critically involved in bone repair. J Clin Invest 2002;109:1405–15.
23. 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.
24. Ijiri K, Matsunaga S, Fukuda T, Shimizu T. Indomethacin inhibition of ossification induced by direct current stimulation. J Orthop Res 1995;13:123–31.
25. Neal B, Rodgers A, Dunn L, Fransen M. Non-steroidal anti-inflammatory drugs for preventing heterotopic bone formation after hip arthroplasty. Cochrane Database Syst Rev 2000;3:CD001160.
26. Wolfe F. The prognosis of rheumatoid arthritis: assessment of disease activity and disease severity in the clinic. Am J Med 1997;103:12S–8S.
27. Lindqvist E, Jonsson K, Saxne T, Eberhardt K. Course of radiographic damage over 10 years in a cohort with early rheumatoid arthritis. Ann Rheum Dis 2003;62:611–6.
28. Lussier A, de Medicis R. Correlation between ossification and inflammation using a rat experimental model. J Rheumatol Suppl 1983;11:114–7.