Efficacy and Safety of Willow Bark Extract in the Treatment of Osteoarthritis and Rheumatoid Arthritis: Results of 2 Randomized Double-Blind Controlled Trials

CLAUDIA BIEGERT, IRMELA WAGNER, RAINER LÜDTKE, INA KÖTTER, CLAUDIA LOHMÜLLER, ILHAN GÜNAYDIN, KATJA TAXIS, and LUTZ HEIDE

ABSTRACT. ***Objective.*** To investigate the efficacy and safety of a standardized willow bark extract in patients with osteoarthritis (OA) and rheumatoid arthritis (RA).

Methods. We studied 127 outpatients with hip or knee OA and a WOMAC pain score of at least 30 mm and 26 outpatients with active RA in 2 randomized, controlled, double-blind trials with followup for 6 weeks. OA trial: Patients were randomized to receive willow bark extract, corresponding to 240 mg of salicin/day, diclofenac 100 mg/day, or placebo (n = 43, 43, and 41, respectively). Main outcome measure was the pain subscore of the WOMAC OA Index. RA trial: Patients were randomized to receive willow bark extract, corresponding to 240 mg salicin/day (n = 13) or placebo (n = 13). Main outcome measure was the patient's assessment of pain rated on a 100 mm visual analog scale (VAS).

Results. OA trial: WOMAC pain scores decreased by 8 mm (17%) in the willow bark group and by 23 mm (47%) in the diclofenac group, compared with 5 mm (10%) in the placebo group. The difference between willow bark extract and placebo was not statistically significant (–2.8 mm; 95% CI –12.1 to 6.4 mm; p = 0.55, ANCOVA), but the difference between diclofenac and placebo was highly significant (–18.0 mm; 95% CI –27.2 to –8.8 mm; p = 0.0002, ANCOVA). RA trial: The mean reduction of pain on the VAS was –8 mm (15%) in the willow bark group compared with –2 mm (4%) in the placebo group. The difference was not statistically significant (estimated difference –0.8 mm; 95% CI –20.9 to 19.3 mm; p = 0.93, ANCOVA).

Conclusion. The OA study suggested that the willow bark extract showed no relevant efficacy in patients with OA. Similarly, the RA trial did not indicate efficacy of this extract in patients with RA. (J Rheumatol 2004;31:2121–30)

Key Indexing Terms:
OSTEOARTHRITIS
RHEUMATOID ARTHRITIS
WILLOW BARK
DICLOFENAC
DRUG THERAPY

The use of herbal medicines is popular in many industrialized countries. A 1997 survey estimated that 12.1% of adults in the United States had taken a herbal medicine in the previous 12 months[1]. In Germany, herbal drugs are used even more frequently, not only in self-medication but also in medical prescribing[2]. Despite the widespread use of herbal medicines there is still a lack of data on their efficacy and safety derived from well designed randomized controlled clinical trials[3-5].

For the treatment of chronic pain, German patients frequently use herbal remedies because they fear the risk of gastrointestinal (GI) side effects associated with nonsteroidal antiinflammatory drug (NSAID) treatment[6,7]. Preparations containing a standardized, highly dosed willow bark extract are marketed in Germany for the supportive treatment of rheumatic diseases[2], and they have been officially licensed by the federal health authorities for this indication. Three recent randomized controlled trials reported evidence for an analgesic efficacy of willow bark extract. A 4-week placebo controlled trial in patients with low back pain resulted in a remarkably high efficacy[8]. A later study by the same group compared willow bark with rofecoxib in

From the Pharmaceutical Institute, Eberhard Karls-Universität, Tübingen; the Department of Internal Medicine II (Haematology, Oncology, Immunology, and Rheumatology), University Hospital, Tübingen; and the Karl and Veronica Carstens Foundation, Essen, Germany.

Supported by Tübingen University and by Robugen GmbH, Esslingen, Germany. R. Lüdtke was supported by funding from the Karl and Veronica Carstens Foundation.

C. Biegert, PhD; I. Wagner, PhD; K. Taxis, PhD; L. Heide, PhD, Pharmaceutical Institute, Eberhard Karls-Universität Tübingen; I. Kötter, MD; C. Lohmüller, MD; I. Günaydin, MD, Department of Internal Medicine II, University Hospital, Tübingen; R. Lüdtke, Dipl. Stat., Karl and Veronica Carstens Foundation.

Dr. Biegert and Dr. Wagner contributed equally to this report.

Address reprint requests to Dr. L. Heide, Pharmaceutical Institute, University of Tübingen, Auf der Morgenstelle 8, 72076 Tübingen, Germany. E-mail: heide@uni-tuebingen.de

Submitted March 4, 2004; revision accepted May 21, 2004.

low back pain[9], and did not find a difference between efficacies of the 2 medications. Unfortunately, patients with much lower disease activity were enrolled in the rofecoxib controlled trial than in the preceding placebo controlled trial. A placebo controlled 2-week trial in patients with osteoarthritis (OA) carried out by our group[10] resulted in a moderate analgesic efficacy that just reached statistical significance ($p = 0.047$).

We conducted a 6-week, 3-arm, randomized controlled trial comparing the efficacy and safety of a standardized willow bark extract with diclofenac and placebo in patients with OA of the hip and the knee. The principal advantage of such a 3-armed study is that it simultaneously provides evidence of effectiveness (willow bark vs placebo) and of study sensitivity (diclofenac vs placebo)[11]. The design and the outcome measures of our study followed the current recommendations of the European Agency for the Evaluation of Medicinal Products (EMEA) and the Osteoarthritis Research Society and included all outcome measures recommended by the US Food and Drug Administration[12-14]. In addition, we carried out a small pilot trial (6 weeks, randomized and placebo controlled) in patients with rheumatoid arthritis (RA), in order to provide a first estimate of the efficacy of willow bark in the treatment of inflammatory rheumatic diseases.

MATERIALS AND METHODS

The trials were conducted in accord with the principles of good clinical practice and the revised Declaration of Helsinki. The study protocols were approved by the respective ethics committees, and all patients gave their written informed consent.

Study medication. An extract of the bark of *Salix daphnoides* cultivated in Germany (extraction solvent: ethanol 70% v/v, drug-extract ratio: 8–14:1) was obtained from Finzelberg, Andernach, Germany. The total salicin content was 15.0% after alkaline hydrolysis[15]. A detailed chemical characterization of the extract will be reported elsewhere. This extract was used in the form of coated tablets prepared by Robugen GmbH, Esslingen, Germany. Each tablet contained 393.24 mg extract (corresponding to 60 mg of salicin). Disintegration testing was carried out according to the European Pharmacopoeia.

The diclofenac tablets contained 25 mg diclofenac in enteric coated form. They were produced by Robugen GmbH and tested in accord with the USP monograph for diclofenac sodium delayed release tablets[16].

Willow bark extract tablets, diclofenac tablets, and placebo tablets were of identical appearance, odor, and taste.

Osteoarthritis trial

Patient inclusion criteria. Subjects were required to be outpatients with OA of the hip or knee, verified according to the clinical, laboratory, and radiographic classification criteria of the American College of Rheumatology (ACR)[17,18], age over 18 years, with a Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC, VA 3.0) pain score[19,20] of at least 30 mm on Day 0.

Exclusion criteria. Criteria for exclusion were use of corticosteroids within the past 8 weeks (intraarticular in the target joint or systemic therapy); surgery of the target joint during the past 8 weeks; inflammatory joint diseases [erythrocyte sedimentation rate (ESR) > 40 mm/h]; known allergy to salicylates, willow bark extract preparations, or NSAID; abnormal renal or hepatic function [creatinine clearance < 40 ml/min[21]; AST > 35 U/l or ALT > 35 U/l or gamma-glutamyl transferase (GGT) > 50 U/l]; unexplained dyshemopoiesis; chronic obstructive airway diseases requiring prophylactic medication; GI ulcers (bleeding, discolored stool, or occult blood in the stool during the past 8 weeks); history of alcohol abuse; pregnancy and lactation; current therapy with an anticoagulant; malignant diseases; chronic heart failure (NYHA III and IV); participation in a clinical trial during the past 4 weeks; and WOMAC pain score on Day –7 < 23 mm without prior use of analgesics or NSAID.

No additional analgesics or NSAID or systemic or intraarticular corticosteroids were allowed during washout phase and the 6-week study phase. Aspirin was allowed up to 100 mg daily. Patients were allowed to continue any physical therapy, but such therapy had to remain unchanged during the study.

Study design. A double-blind, 3-arm, parallel randomized trial over 6 weeks was performed at 2 study centers in Germany from May 2001 to September 2002. Randomization was carried out in blocks of 6, using computer generated random numbers (Rancode™, IDV, Munich, Germany). Assessors and patients were blinded to the allocation.

Outpatients were recruited by orthopedists, internists, and general practitioners and by public advertisement. After a placebo washout phase of 4–10 days, depending on the half-life of the analgesic or NSAID (at least 7 days for long-acting NSAID, e.g., piroxicam), patients who were eligible for the study were randomly assigned to one of the following treatments: willow bark extract, 2 tablets twice daily, corresponding to a dose of 240 mg of salicin/day; or diclofenac, enteric coated tablets, 2 tablets twice daily, corresponding to 100 mg/day; or placebo, 2 tablets twice daily. Each medication was taken for 6 weeks in the mornings and evenings, half an hour before mealtimes.

Patients were assessed by a physician on Days –7, 0, 14, and 42. On Day 28, patients completed several questionnaires at home.

Outcome measures. The primary outcome measure was change in the pain subscore of the WOMAC OA Index from Day 0 to Day 42 [visual analog scale, VAS, ranging from 0 (best) to 100 mm (worst)][19,20]. We followed the translation of Stucki, *et al*[22] for the German phrasing of the questionnaire. Secondary outcome measures were as follows: (1) WOMAC stiffness subscore; (2) WOMAC function subscore; (3) WOMAC total index (the total index was calculated as pain 42%, stiffness 21%, and physical function 37%); (4) patient's and physician's assessment of overall efficacy measured by a 100 mm VAS; (5) quality of life assessment using the Short Form-36 (SF-36), which referred to patient's condition during the previous week, using the validated German translation[23]; (6) patient's assessment of tolerability measured by a 100 mm VAS; and (7) safety of the study drug.

On Day –7, medical history, physical examination, and laboratory tests (blood and urine samples) were conducted. The more painful hip or knee joint was identified as the target joint, i.e., this joint was used for future assessment in the study. Patients were given a diary to document the number of tablets taken each day, to record any adverse events, and to score their pain on a 100 mm VAS each evening.

Efficacy was assessed by the WOMAC on Day –7, 0, 14, and 42. On Day 28, an additional WOMAC questionnaire was filled in by patients independently at home. Quality of life was measured on Day 0 and Day 42. Patient's assessment of efficacy was evaluated on Day 14 and 42 as well as on Day 28 (take-home questionnaire); physician's assessment of efficacy was evaluated on Day 42. Safety evaluation was based on laboratory testing (blood and urine samples on Day 0 and 42) and reports of adverse events. Data on adverse events were collected by interviewing patients on each visit and by records in the patients' diaries. Onset, duration, end, and intensity (mild, moderate, or severe) of each adverse event, as well as the action taken and outcome, were recorded. The relationship between event and study medication was judged by the physician as none, unlikely, possible, probable, or certain. Adverse events were classified according to the World Health Organization terminology[24]. Patients assessed the tolerability of the study drug on the final visit on a 100 mm VAS.

Statistical analysis. Sample size calculation was based on a subset of the

results of a previous clinical trial[10] (WOMAC pain score, Δ = 9.4 mm, SD 15.4 mm). A sample size of 126 patients (42 per group) was calculated to give 80% power with a type I error rate of 5% (2-tailed).

The primary analysis of efficacy was based on the intention-to-treat population. The intention-to-treat population included all patients who were randomized on Day 0 and took at least one dose of study medication. The per-protocol population excluded patients with protocol violations, patients who withdrew from the study due to reasons other than pain before Day 14, and patients who were noncompliant. Patients were judged to be noncompliant if they had taken less than 80% of the study medication, according to pill counts at each followup visit.

Missing values were imputed by carrying forward the last observed value for patients who discontinued for reasons other than arthritic pain. Missing values for patients who withdrew due to pain (treatment failures) were replaced by the mean of the 5 worst values of patients in the respective treatment group.

The primary endpoint was analyzed by analysis of covariance (ANCOVA) with baseline values, center, and the use of physical therapy (yes/no; see Results) as covariates. Secondary endpoints were analyzed in the same way.

The principal comparison was between the willow bark and placebo groups; 95% confidence intervals (CI) were calculated for mean changes from Day 0 to Day 42. In terms of a priori ordered hypotheses, the willow bark and placebo groups were compared first, and the diclofenac and placebo groups subsequently.

Demographic variables were compared using the chi-square test for categorical data and Kruskal-Wallis test for continuous data. Numbers and rates of adverse events (AE) were tabulated for each treatment group and analyzed using the Cochrane-Mantel-Haenszel test (numbers of AE) and Friedman test (rates of AE). In all statistical tests, the level of type I error (2-tailed) was set at 0.05. Reported p values for secondary outcome measures and AE were considered descriptive only. All analyses were performed with SAS version 8.02 software (SAS Institute Inc., Cary, NC, USA).

Rheumatoid arthritis trial

Patient inclusion criteria. Subjects were required to be outpatients with a diagnosis of RA defined by ACR criteria[25]; RA functional class I, II, or III[26]; age over 18 years; evidence of at least moderate disease activity at the randomization visit, i.e., at least 2 of the following signs and symptoms: ≥ 6 tender joints, ≥ 3 swollen joints, morning stiffness ≥ 45 min, ESR ≥ 28 mm/h.

Exclusion criteria. Criteria for exclusion were intraarticular, intramuscular, or intravenous injections of corticosteroids within one month before the study and during the study; chemical, radiologic, or surgical synovectomy in any large joint within 3 months before the study and during the study; known hypersensitivity to salicylates or willow bark extract; concomitant therapy with anticoagulants; concomitant severe cardiac, hepatic, renal, or hematologic disease; cancer; bronchial asthma; gastrointestinal ulcers; history of alcohol abuse; pregnancy and lactation; and participation in a clinical trial within the past 30 days.

Disease modifying antirheumatic drugs (except tumor necrosis factor inhibitors) were allowed as concomitant therapy, but they had to be taken since at least 6 months before, and their dosage had to be stable for 3 months before and during the study. Treatment with corticosteroids could not exceed prednisolone 7.5 mg/day or equivalent and had to be stable for one month before and during the study. NSAID and analgesics had to be discontinued before entering the trial. Patients were permitted to take low dose aspirin (up to 100 mg/day).

Study design. A double-blind, 2-arm, parallel randomized trial over 6 weeks was conducted at 4 centers in Germany from June 2001 to November 2002. Patients were randomly allocated to one of the following treatment groups: willow bark extract, 2 tablets twice daily, corresponding to 240 mg salicin per day; or placebo, 2 tablets twice daily. Each medication was taken for 6 weeks, as described for the OA trial.

A washout period of 4–10 days' duration was required, depending on the elimination half-life of the NSAID taken before the study. Patients were assessed by the investigator on Day –7, 0, 14, 28, and 42.

Outcome measures. Efficacy assessments included all components of the ACR core set of outcome measures[27]: the primary outcome measure was the change in patient's assessment of pain from Day 0 to Day 42 as recorded on a 100 mm VAS. Secondary outcome measures were the number of tender/painful and swollen joints (28 joint count)[28], physical function assessed with the Health Assessment Questionnaire (HAQ) Disability Index[29] (validated German translation[30]), patient's assessment of the severity of morning stiffness (100 mm VAS), patient's and physician's assessment of overall efficacy (100 mm VAS), quality of life assessed with the SF-36 index (validated German translation[23]), ESR, plasma concentrations of C-reactive protein (CRP), and the number of patients who met the ACR-20 criteria for improvement[31].

Assessment of tolerability and safety and statistical evaluation were carried out as described for the OA trial. Analysis was based on ANCOVA with baseline values and center as covariates.

RESULTS

Osteoarthritis trial

Patient characteristics. As shown in Figure 1, 298 patients were screened for inclusion in the trial; 68 patients did not meet entry criteria on Day –7; 6 dropouts occurred during the washout phase; and 97 patients did not fulfil entry criteria on Day 0. Therefore 127 patients were randomized and allocated to one of the 3 treatment groups. The treatment groups were comparable in all relevant demographic and clinical characteristics (Table 1). However, during the study only 12% of the patients in the willow bark group reported to receive physical therapy, in contrast to 40% in the diclofenac group and 27% in the placebo group. This difference was statistically significant (p = 0.01). Therefore, physical therapy was used as a covariate in the ANCOVA analysis of the outcome variables, as specified in the study protocol.

Overall, 106 patients (83%) completed the study. Lack of efficacy was the most common reason for withdrawal (5 willow bark group, 3 diclofenac, and 9 placebo); 2 patients in the diclofenac group withdrew due to adverse events and 2 patients in the placebo group withdrew since they decided to take a holiday abroad.

Important protocol violations were committed by 17 patients: 10 (4 willow bark, 2 diclofenac, 4 placebo) had taken additional NSAID or analgesics (because of reasons other than OA); 2 patients (both placebo) were considered noncompliant because they had taken less than 80% of the study medication. One patient (placebo) reported that his physical therapy was changed significantly during the study phase, and his final assessment (Day 42) was carried out outside the predefined visit window (± 3 days). Four patients were randomized although they did not fulfil the entry criteria (2 diclofenac patients and one placebo patient had taken NSAID during washout phase and one diclofenac patient had a WOMAC pain score less than the required minimum of 30 mm on Day 0). All these patients were therefore excluded from the per-protocol analysis.

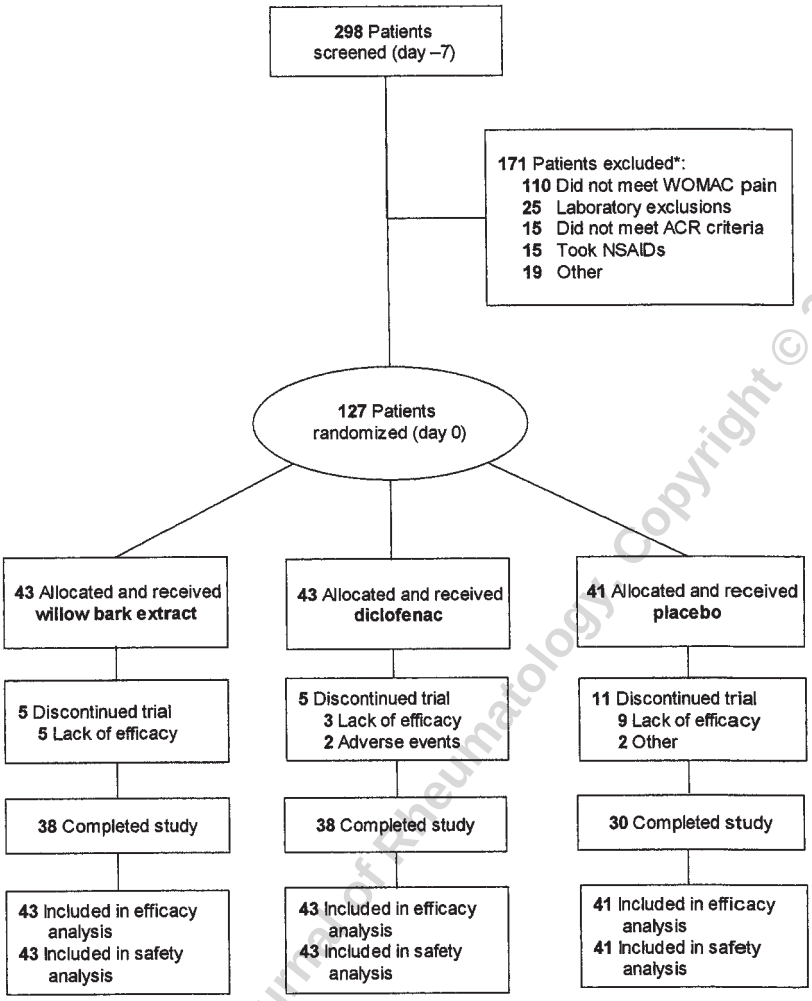

Figure 1. OA trial: Disposition of the study patients.

Table 1. Osteoarthritis trial: baseline demographic and clinical characteristics of patients. Data are expressed as number for nominal data and as mean (SD) for continuous data.

Characteristics	Willow Bark, n = 43	Diclofenac, n = 43	Placebo, n = 41
Age, yrs	62.9 (7.2)	61.2 (6.6)	62.4 (8.9)
Sex, M/F	22/21	15/28	16/25
Body weight, kg	81.7 (15.4)	79.9 (14.1)	79.6 (12.9)
Height, cm	171.5 (8.4)	167.9 (7.3)	168.1 (7.1)
OA of the hip/knee	8/35	15/28	14/27
OA unilateral/bilateral	9/34	11/32	13/28
Duration of OA, yrs	8.8 (6.4)	8.5 (6.2)	10.9 (8.6)
Prior use of analgesics/ NSAID, yes/no	24/19	28/15	26/15
Physical therapy during the study, yes/no*	5/38	17/26	11/30

* Difference statistically significant (p = 0.01, chi-square test).

Efficacy results. Table 2 shows the mean values for the outcome measures. Baseline values for each efficacy variable were similar for all treatment groups and deteriorated only slightly during the washout phase. At the end of the treatment phase, all groups had improved in nearly all indicators.

Primary outcome measure — WOMAC pain score. Over 6 weeks, WOMAC pain scores decreased by 8 mm in the willow bark group, by 5 mm in the placebo group, and by 23 mm in the diclofenac group. The ANCOVA using baseline values, center, and physical therapy as covariates showed no statistically significant difference in the reduction of WOMAC pain scores between willow bark and placebo group (intention-to-treat population: –2.8 mm; 95% CI –12.1 to 6.4 mm; p = 0.55). This was confirmed by the results of the per-protocol population (difference –1.0 mm; 95% CI –11.2 to 9.2 mm; p = 0.85, ANCOVA).

Table 2. Efficacy of willow bark and diclofenac in the treatment of OA. Data are mean values for the intention-to-treat population.

Variable	Group*	Baseline Assessment (± SD)			Final Assessment (± SD)	Difference Baseline vs Final Assessment[1] (± SD)	Difference Willow Bark vs Placebo[2] (ANCOVA)	Difference Diclofenac vs Placebo[3] (ANCOVA)
		Day (–7)	Day 0	Day 14	Day 28	Day 42		
WOMAC (100 mm VAS; best value: 0 mm)								
Pain	W	44	48 (12)	42	39	41 (22)	–8 (21)	–2.8	–18.0
	P	48	50 (17)	44	45	45 (27)	–5 (23)	(p = 0.55)	(p = 0.0002)
	D	47	49 (14)	26	26	26 (21)	–23 (20)		
Stiffness	W	47	50 (21)	48	45	46 (24)	–4 (24)	1.1	–18.4
	P	52	51 (17)	43	46	46 (25)	–5 (26)	(p = 0.82)	(p = 0.0004)
	D	54	53 (21)	31	30	28 (23)	–24 (25)		
Physical Function	W	44	48 (15)	43	43	40 (22)	–8 (18)	–2.5	–16.4
	P	48	50 (14)	45	46	45 (23)	–5 (20)	(p = 0.54)	(p = 0.0001)
	D	50	49 (17)	28	28	28 (22)	–21 (19)		
Total Score	W	45	49 (13)	43	41	42 (21)	–7 (19)	–1.7	–17.6
	P	49	50 (13)	44	45	45 (24)	–5 (21)	(p = 0.69)	(p < 0.0001)
	D	50	50 (15)	28	28	27 (21)	–23 (19)		
SF-36 (0 to 100 score; best value: 100)								
Physical Component	W		31.4 (9.1)			33.1 (10.3)	1.2 (9.1)	2.36	7.94
Summary	P		31.9 (8.2)			30.7 (10.3)	–1.2 (9.6)	(p = 0.24)	(p = 0.0001)
	D		30.9 (7.9)			37.7 (9.2)	6.8 (8.1)		
Mental Component	W		56.7 (9.9)			53.7 (11.6)	–2.3 (7.6)	–1.81	–0.64
Summary	P		53.7 (14.0)			53.2 (12.0)	–0.5 (10.9)	(p = 0.41)	(p = 0.77)
	D		54.5 (11.1)			53.3 (12.5)	–1.1 (10.6)		
Overall assessment of efficacy (100 mm VAS; best value: 0 mm)								
Patient	W		50[4]	50	48	46 (24)	–4	1.0	–18.4
	P		50[4]	47	47	45 (23)	–5	(p = 0.84)	(p = 0.0002)
	D		50[4]	27	27	26 (19)	–24		
Physician	W		50[4]			45 (20)	–5	4.1	–7.6
	P		50[4]			41 (15)	–9	(p = 0.30)	(p = 0.05)
	D		50[4]			33 (16)	–17		

[1] Within-group baseline vs final assessment, difference of means. [2] Between-group baseline vs final assessment, difference willow bark–placebo, estimated by ANCOVA. [3] Between-group baseline vs final assessment, difference diclofenac–placebo, estimated by ANCOVA. [4] Baseline value corresponding to an unchanged state. Negative within-group differences are consistent with improvement, except for the SF-36 scores. Negative between-group differences are consistent with a tendency in favor of willow bark or diclofenac, respectively, except for the SF-36 scores. * W: willow bark extract group, P: placebo, D: diclofenac.

However, there was a large and statistically significant difference in the reduction of WOMAC pain scores between diclofenac and placebo groups (intention-to-treat population: –18.0 mm; 95% CI –27.2 to –8.8 mm; p = 0.0002. Per-protocol population: –16.2 mm; 95% CI –26.6 to –5.8 mm; p = 0.003). As shown in Figure 2, most of the improvement was already achieved after 2 weeks of treatment.

Secondary outcome measures. The WOMAC stiffness scores and the WOMAC physical function scores showed strong improvement in the diclofenac group that was significantly superior to placebo (Table 2). In contrast, the improvements under willow bark treatment were similar to those under placebo.

In the SF-36 quality of life index, the diclofenac group showed a statistically significant improvement in comparison to placebo in the subscales of body pain, vitality, physical functioning, and physical role (all p values < 0.02, data not shown). The subscales for general health, social functioning, emotional role, and mental health showed smaller and nonsignificant improvements. In total, this resulted in a significant superiority of diclofenac over placebo in the physical component summary (Table 2). As may be expected, diclofenac did not result in a significant improvement of the mental component summary of the SF-36.

The willow bark group experienced a statistically significant improvement in the physical functioning subscale in comparison to placebo (difference 9.32; 95% CI 0.29 to 18.35; p = 0.04, ANCOVA). Three other subscales of the SF-36 showed nonsignificant improvement, and 4 subscales nonsignificant deterioration under willow bark treatment in comparison to placebo. Correspondingly, neither the physical nor the mental component summary showed an efficacy for the willow bark preparation.

In the assessments of overall efficacy (Table 2), both patients and physicians reported strong improvements with diclofenac treatment, but only minimal improvements with willow bark and with placebo. Patient's assessment showed a highly significant superiority of diclofenac over placebo,

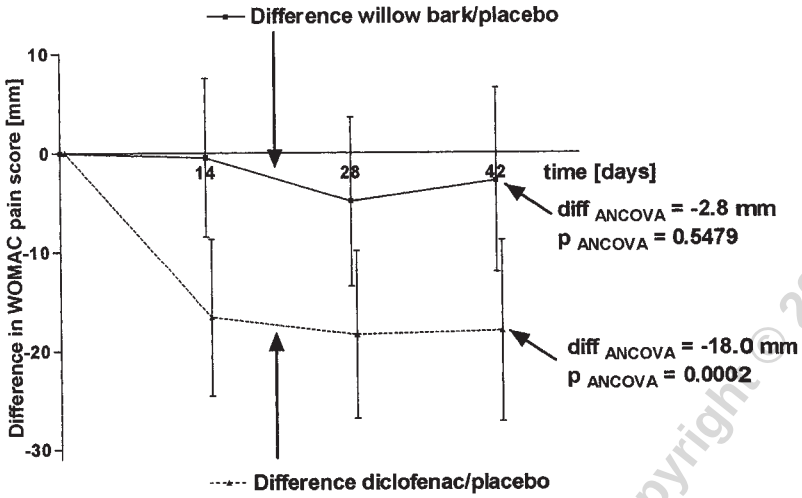

Figure 2. Influence of willow bark extract and diclofenac on the WOMAC pain score. Difference between treatment group and placebo group by ANCOVA with 95% CI. Intention-to-treat analysis (n = 127).

while in the physician's assessment this superiority just missed statistical significance. This may indicate the well known fact that patient's assessment is a more valid and sensitive outcome indicator than physician's assessment.

At the end of the study the patients were asked which type of medication they believed they had received in the preceding 6 weeks (willow bark, diclofenac, or placebo). In the willow bark group, 8 patients guessed their medication correctly, versus 5 patients in the diclofenac group and 14 in the placebo group. Thus, blinding was apparently successful during the study.

A power analysis of the results showed that a true difference for WOMAC pain reduction between willow bark and placebo of 15 mm or more can be excluded with a probability of more than 99%. For a difference of 10 mm or more, the probability is 94%.

Rheumatoid arthritis trial

The trial described above failed to show an efficacy of willow bark extract in OA, the most common form of degenerative joint disease. It may be argued that willow bark extract could be more efficacious in inflammatory rather than in degenerative diseases, as possibly indicated by a study with a nonstandardized willow bark preparation[32].

To test this hypothesis, we decided to carry out a first pilot trial in patients with RA, the most common inflammatory rheumatic disease. In order to maximize the power of this small study, a 2-arm rather than a 3-arm design was used (randomized, placebo controlled, 6-week, double-blind trial). The outcome variables followed the recommendation of the ACR[27], and included measurements of swollen joints, ESR, and CRP as indicators for inflammatory processes.

Patient characteristics. As shown in Figure 3, 89 patients were screened and 26 met eligibility criteria and were randomized. Patients were excluded mainly because they did not meet the ACR criteria for diagnosis of RA or the disease activity criteria; 6 patients withdrew before completion of the study due to lack of efficacy (3 willow bark extract, 3 placebo), and one patient in the willow bark group withdrew due to an adverse event (disc prolapse).

All 26 patients were included in the intention-to-treat efficacy and safety analysis; 18 patients (11 placebo, 7 willow bark extract) were included in the per-protocol analysis. Patients were excluded from the per-protocol population mainly because they had study visits outside the predefined visit window of ± 3 days of the visits on Days 14, 28, and 42.

Treatment groups were comparable with respect to patient demographics and medical history (Table 3). However, patients in the willow bark group showed a more active disease in all baseline arthritis assessments (Table 4). Mean baseline pain values in the willow bark group were 10 mm higher than in the placebo group (55 mm vs 45 mm; p = 0.22, Kruskal-Wallis test).

Efficacy results. Efficacy results are summarized in Table 4. The mean reduction of pain VAS values was 8 mm (15%) in the willow bark group compared with 2 mm (4%) in the placebo group. However, statistical analysis with adjustment for baseline values and study center (ANCOVA), as predefined in the study protocol, resulted in an estimated difference of only –0.8 mm (95% CI –20.9 to 19.3 mm; p = 0.93). Similar results were obtained in the per-protocol analysis (estimated difference of 0.5 mm; 95% CI –25.0 to 26.1 mm; p = 0.97). Similarly, no secondary outcome measure revealed statistically significant differences between willow bark extract and placebo groups. While some outcome measures (HAQ, morning stiffness, SF-36, ESR, and CRP) showed a slight tendency in favor of willow bark extract, others showed a tendency in favor of placebo (ten-

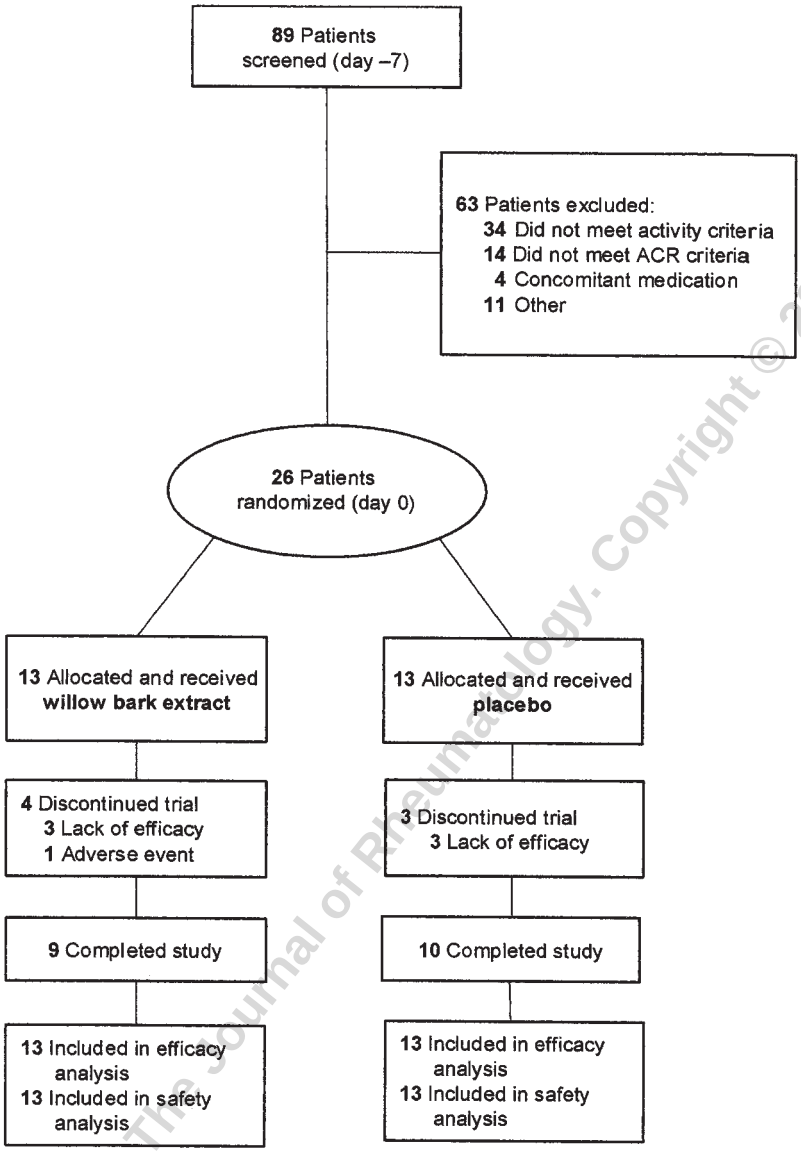

Figure 3. RA trial: Disposition of the study patients.

der and swollen joints, patient and physician overall assessment of efficacy). Therefore, the observed between-group differences may be attributed to chance. Only 2 patients in the willow bark extract group and one in the placebo group were classified as ACR-20 responders[31]. At the final visit, patients were asked which type of medication (placebo or willow bark extract) they believed they had received during the preceding weeks. Only 4 out of 26 patients believed they had taken willow bark extract (one correctly, 3 mistakenly). Apparently, most patients did not feel a clear treatment benefit.

A power estimate of the study showed that a true difference in pain reduction between willow bark extract and placebo of 15 mm (suggested as the minimum clinically relevant difference[33]) or more can be excluded with a probability of 93%. A difference of 10 mm or more can be excluded with a probability of 83%.

Safety and tolerability. There were 173 adverse events reported in the OA trial (Table 5), most in the diclofenac group. Seven GI adverse events were reported in the willow bark group, 19 in the placebo group, and 35 in the diclofenac group. The difference between willow bark and diclofenac was statistically significant ($p = 0.001$, Friedman test).

One patient in the willow bark group developed an itching exanthema after exposure to the sun on Day 30, which improved until study termination.

Two patients in the diclofenac group withdrew from the study because of adverse events, one because of indigestion and another due to typical symptoms of diclofenac intolerability (heartburn, exertional dyspnea, blood in urine and stool). The latter patient also experienced 2 serious adverse events 12 days after withdrawal from the study when he was

Table 3. Rheumatoid arthritis trial: Baseline demographic and clinical characteristics of patients. Data are presented as number for nominal data and as mean (SD) for continuous data.

Characteristics	Willow Bark, n = 13	Placebo, n = 13
Age, yrs	56.5 (8.9)	60.1 (11.0)
Sex, M/F	3/10	1/12
Body weight, kg	70.9 (11.7)	68.5 (11.3)
Height, cm	166.5 (6.9)	165.1 (8.3)
Duration of RA, yrs	9.4 (5.3)	13.8 (11.5)
Functional class		
I	1	2
II	9	8
III	3	3
Rheumatoid factor positive/negative	10/3	9/4
Radiologic evidence of RA, yes/no	10/3	12/0*
DMARD treatment, yes/no	8/5	10/3
Corticosteroid treatment, yes/no	5/8	6/7
Physical therapy during the study, yes/no	7/6	6/7
Prior use of analgesics/NSAID, yes/no	9/4	8/5

* Value missing for one patient. DMARD: disease modifying antirheumatic drug.

hospitalized for suspected gastritis and deep vein thrombosis of the legs.

Changes in hematology and clinical chemistry showed statistically significant reductions of red blood cell count, hematocrit, and hemoglobin in the diclofenac group ($p \leq 0.01$, Kruskal-Wallis test). The liver enzymes ALT, AST, and GGT increased significantly under diclofenac therapy ($p \leq 0.001$). In the placebo group, there was a statistically significant increase in serum glucose ($p = 0.02$). In the willow bark group, no statistically significant changes in laboratory values were observed.

Patient's assessment of tolerability was rated on a 100 mm VAS, 0 mm being the best value. Mean values recorded in the OA trial were 13 mm for willow bark, 15 mm for placebo, and 16 mm for diclofenac.

In the RA trial, there were 7 adverse events reported in each group (data not shown), none of which was classified as "serious." Causality was assessed as "possible" for 2 adverse events in the placebo group and one adverse event in the willow bark extract group. The latter was mild itching on the arms, which resolved completely without discontinuation of the study drug and without additional treatment. No patient experienced clinically significant changes in laboratory indicators. Patients rated the tolerability of the study drug with 14 mm (willow bark extract) and 9 mm (placebo) on the 100 mm VAS described above.

DISCUSSION

We found no evidence for a relevant analgesic or antiinflammatory efficacy of the investigated willow bark extract in patients with OA and RA.

The sensitivity of the OA study was clearly demonstrated by the superiority of diclofenac over placebo observed in all outcome measures. The study confirmed the well known efficacy as well as the known side effects of diclofenac. Despite the small sample size (diclofenac, n = 43; placebo, n = 41), superiority of diclofenac over placebo could be shown with very high significance (total WOMAC score: $p < 0.0001$). This shows the excellent sensitivity of the WOMAC index for the evaluation of OA treatments.

Our results are in conflict with the positive outcomes of earlier trials with willow bark extract using identical doses. Two of these trials were carried out in patients with low back pain[8,9]. The strong effect reported especially from the first of these studies[8] is in striking contrast with our findings, although a direct comparison is difficult due to the different diseases investigated. A confirmation of the efficacy

Table 5. Adverse events reported under treatment with willow bark, diclofenac, and placebo in the OA trial. Adverse events are summarized according to the WHO terminology[24]. The same patient may be listed under different adverse events.

Organ System	Willow Bark, n = 43	Placebo, n = 41	Diclofenac, n = 43	Total, n = 127
GI system	7	19	35	61
Central and peripheral nervous system	7	15	16	38
Body as a whole—general disorders	7	4	11	22
Respiratory system	5	4	3	12
Musculoskeletal system	4	3	3	10
Skin and appendages	3	2	3	8
Psychiatric	1	0	4	5
Cardiovascular disorders, general	1	1	3	5
Urinary system	1	2	2	5
Vision disorders	1	0	2	3
Platelet, bleeding, and clotting disorders	1	0	1	2
Vascular (extracardiac) disorders	0	1	1	2
Total no. of adverse events	38	51	84	173
Total no. of patients experiencing adverse events	19	20	30	69

Table 4. Efficacy of willow bark extract in the treatment of RA. Data are mean values for the intention-to-treat population.

Variable	Group*	Baseline Assessment (± SD) Day (–7)	Day 0	Day 14	Day 28	Final Assessment (± SD) Day 42	Difference Baseline vs Final Assessment[1] (± SD)	Difference Willow Bark vs Placebo[2] (ANCOVA)
Primary endpoint								
Patient's assessment of pain	W	51	55 (20)	49	52	47 (24)	–8 (24)	–0.8
(100 mm VAS; best value: 0 mm)	P	44	45 (25)	47	48	43 (30)	–2 (27)	(p = 0.93)
Secondary endpoints								
Tender joint count (total	W	9.9	11.0 (6.8)	8.6	9.2	10.0 (7.0)	–1.0 (6.7)	2.30
28 joints)	P	6.8	7.9 (5.9)	7.2	6.5	5.8 (4.7)	–2.1 (2.8)	(p = 0.25)
Swollen joint count (total	W	8.3	8.2 (7.6)	7.7	7.4	7.5 (7.9)	–0.7 (7.4)	0.82
28 joints)	P	5.3	6.2 (5.4)	5.8	5.2	5.0 (4.4)	–1.2 (3.2)	(p = 0.69)
HAQ Disability Index	W		1.2 (0.6)	1.2	1.2	1.2 (0.6)	0.0 (0.3)	–0.05
(0–3 score; best value: 0)	P		0.9 (0.6)	0.9	0.8	0.9 (0.6)	0.1 (0.2)	(p = 0.60)
Morning stiffness (100	W	42	50 (29)	42	47	39 (2.6)	–11 (26)	–4.4
mm VAS; best value: 0 mm)	P	45	45 (32)	45	39	42 (31)	–3 (36)	(p = 0.69)
Patient's assessment of	W		50[4]	46	46	53 (17)	3 (17)	1.3
efficacy (100 mm VAS; best value: 0 mm)	P		50[4]	56	54	52 (25)	2 (25)	(p = 0.89)
Physician's assessment	W		50[4]			58 (23)	8 (23)	9.5
of efficacy (100 mm	P		50[4]			48 (16)	–2 (16)	(p = 0.26)
VAS; best value: 0 mm)								
ESR, mm/h	W	25.6	27.8 (19.4)			26.5 (19.9)	–1.3 (5.0)	–5.95
	P	21.5	21.0 (10.2)			25.8 (14.3)	4.0 (12.3)	(p = 0.14)
CRP, mg/dl	W	2.4	1.9 (1.4)			1.7 (1.6)	–0.2 (1.4)	–0.74
	P	1.8	2.0 (1.7)			2.5 (1.9)	0.5 (1.7)	(p = 0.24)
ACR-20 responders, n	W					2		
	P					1		p = 0.55[3]
SF-36 (0 to 100 score; best value: 100)								
Physical Component	W		30.8 (7.4)			33.1 (6.8)	2.6 (5.4)	3.0
Summary	P		41.5 (9.2)			39.7 (10.2)	–1.3 (5.8)	(p = 0.38)
Mental Component	W		52.2 (11.1)			51.5 (9.2)	0.6 (9.0)	5.4
Summary	P		54.4 (8.5)			50.0 (9.0)	–3.3 (4.7)	(p = 0.16)

[1] Within-group baseline vs. final assessment, difference of means. [2] Between-group baseline vs. final assessment, difference willow bark–placebo, estimated by ANCOVA. [3] Cochrane-Mantel-Haenszel test. [4] Baseline value corresponding to an unchanged state. Negative within-group differences are consistent with improvement, except for the SF-36 scores. Negative between-group differences are consistent with a tendency in favor of willow bark, except for the SF-36 scores. * W: willow bark extract, P: placebo.

of willow bark extract in low back pain by a 3-arm trial, in comparison to placebo and a standard analgesic drug, may be desirable in this situation. Our own previous trial, carried out in OA patients in a hospital setting for 2 weeks, had shown moderate superiority of willow bark extract over placebo in the WOMAC pain score (–6.5 mm; p = 0.047), but no significant improvement in the WOMAC stiffness and physical function scores[10]. Our present study again showed a slight tendency in favor of willow bark in the WOMAC pain score, but the observed difference compared to placebo was neither statistically significant nor clinically relevant. The absolute reduction of the WOMAC pain score from baseline to Day 14 in the previous trial[10] was similar to that observed in the present study. However, the lower placebo response in the previous trial was part of the reason this moderate effect reached statistical significance in comparison to placebo.

In our previous study, we used an extract of a *Salix purpurea x daphnoides* hybrid. In the present study, we used the extract from *Salix daphnoides*, which is currently contained in the most widely used commercial willow bark preparations in Germany. Although both extracts were standardized for the same total salicin content, we found clear differences in regard to other constituents (unpublished data).

Willow bark contains salicin derivatives that are metabolized *in vivo* to salicylic acid. In a recent pharmacokinetic trial[34], we observed that serum salicylate concentrations in human volunteers after oral administration of current therapeutic doses of willow bark extract were too low to explain a relevant analgesic effect. Willow bark extract has recently been reported to inhibit the COX-2-mediated release of prostaglandin E_2 *in vitro* as well as the release of tumor necrosis factor-α (TNF-α) and interleukin 1ß (IL-1ß)[35]. However, we have shown that these effects are only

observed *in vitro* and not after oral ingestion of the same extract by human volunteers, indicating that the active substances do not reach effective serum concentrations[36].

The pharmacokinetic, pharmacological, and clinical data we obtained do not support the hypothesis of a relevant efficacy of currently used doses of willow bark extract in the treatment of rheumatic diseases.

ACKNOWLEDGMENT

We thank Doctors Bernhard Heilig, Doris Lassak-Siedl, Markus Müller, Hartmut Rapp, Michaela Geiger, and Constanze Richter for their cooperation. We thank Robugen GmbH, Esslingen, Germany, for preparation of the study medication and for their constructive cooperation.

REFERENCES

1. Eisenberg DM, Davis RB, Ettner SL, et al. Trends in alternative medicine use in the United States, 1990-1997: results of a follow-up national survey. JAMA 1998;280:1569-75.
2. Tyler VE. Foreword. In: Blumenthal M, editor. The complete German Commission E monographs: therapeutic guide to herbal medicines. Austin, Texas: American Botanical Council; 1998.
3. Angell M, Kassirer JP. Alternative medicine — the risks of untested and unregulated remedies [editorial]. N Engl J Med 1998;339:839-41.
4. De Smet PA. Herbal remedies. N Engl J Med 2002;347:2046-56.
5. Straus SE. Herbal medicines — what's in the bottle? N Engl J Med 2002;347:1997-8.
6. Wolfe MM, Lichtenstein DR, Singh G. Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. N Engl J Med 1999;340:1888-99.
7. FitzGerald GA, Patrono C. The coxibs, selective inhibitors of cyclooxygenase-2. N Engl J Med 2001;345:433-42.
8. Chrubasik S, Eisenberg E, Balan E, Weinberger T, Luzzati R, Conradt C. Treatment of low back pain exacerbations with willow bark extract: a randomized double-blind study. Am J Med 2000;109:9-14.
9. Chrubasik S, Künzel O, Model A, Conradt C, Black A. Treatment of low back pain with a herbal or synthetic anti-rheumatic: a randomized controlled study. Willow bark extract for low back pain. Rheumatology Oxford 2001;40:1388-93.
10. Schmid B, Lüdtke R, Selbmann HK, et al. Effectiveness and tolerance of standardized willow bark extract in osteoarthritis patients. Randomized, placebo controlled double-blind study. Z Rheumatol 2000;59:314-20.
11. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Ann Intern Med 2000;133:455-63.
12. European Agency for the Evaluation of Medicinal Products–Committee for Proprietary Medicinal Products. Points to consider on clinical investigation of medicinal products used in the treatment of osteoarthritis. CPMP/EWP/784/97. London: EMEA; 1998. Internet. [cited July 14, 2004]. Available from: http://www.emea.eu.int/pdfs/human/ewp/078497en.pdf
13. Altman R, Brandt K, Hochberg M, et al. Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Results from a workshop. Osteoarthritis Cartilage 1996;4:217-43.
14. Food and Drug Administration. Guidance for industry: Clinical development programs for drugs, devices, and biological products intended for the treatment of osteoarthritis (OA). Washington, DC: Federal Register 64 (135), July 15, 1999; docket 98D-0077.
15. Meier B, Sticher O, Bettschart A. Weidenrinden-Qualität. Deutsche Apotheker Zeitung 1985;125:341-7.
16. US Pharmacopeia 25. Diclofenac sodium delayed release tablets. In: The United States Pharmacopeia. Philadelphia: National Publishing; 2002:554-5.
17. Altman R, Alarcon G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505-14.
18. Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986;29:1039-49.
19. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
20. Bellamy N. WOMAC Osteoarthritis Index — A user guide. London, ON: University of Western Ontario; 1995.
21. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron 1976;16:31-41.
22. Stucki G, Meier D, Stucki S, et al. Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Osteoarthritis Index. Z Rheumatol 1996;55:40-9.
23. Bullinger M. German translation and psychometric testing of the SF-36 Health Survey: preliminary results from the IQOLA Project. International Quality of Life Assessment. Soc Sci Med 1995;41:1359-66.
24. WHO. Adverse reaction terminology. Geneva: World Health Organization, The Uppsala Monitoring Centre; 2001.
25. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
26. Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. Arthritis Rheum 1992;35:498-502.
27. Felson DT, Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum 1993;36:729-40.
28. Fuchs HA, Brooks RH, Callahan LF, Pincus T. A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis. Arthritis Rheum 1989;32:531-7.
29. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
30. Lautenschlager J, Mau W, Kohlmann T, et al. Comparative evaluation of a German version of the Health Assessment Questionnaire and the Hannover Functional Capacity Questionnaire. Z Rheumatol 1997;56:144-55.
31. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
32. Mills SY, Jacoby RK, Chacksfield M, Willoughby M. Effect of a proprietary herbal medicine on the relief of chronic arthritic pain: a double-blind study. Br J Rheumatol 1996;35:874-8.
33. Bellamy N. Musculoskeletal clinical metrology. Dordrecht: Kluwer Academic Publishers; 1993.
34. Schmid B, Kötter I, Heide L. Pharmacokinetics of salicin after oral administration of a standardised willow bark extract. Eur J Clin Pharmacol 2001;57:387-91.
35. Fiebich BL, Chrubasik S. Effects of an ethanolic salix extract on the release of selected inflammatory mediators in vitro. Phytomedicine 2004;11:135-8.
36. Wagner I, Greim C, Laufer S, Heide L, Gleiter CH. Influence of willow bark extract on cyclooxygenase activity and on tumor necrosis factor alpha or interleukin 1 beta release in vitro and ex vivo. Clin Pharmacol Ther 2003;73:272-4.