VU Research Portal

The Efficacy of Nonsteroidal Antiinflammatory Drugs (Nsaids) for Shoulder Complaints - A Systematic Review

van der Windt, D.A.W.M.; van der Heijden, G.J.M.G.; Scholten, R.J.P.M.; Koes, B.W.; Bouter, L.M.

published in

Journal of Clinical Epidemiology
1995

DOI (link to publisher)

[10.1016/0895-4356\(94\)00170-U](https://doi.org/10.1016/0895-4356(94)00170-U)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

van der Windt, D. A. W. M., van der Heijden, G. J. M. G., Scholten, R. J. P. M., Koes, B. W., & Bouter, L. M. (1995). The Efficacy of Nonsteroidal Antiinflammatory Drugs (Nsaids) for Shoulder Complaints - A Systematic Review. *Journal of Clinical Epidemiology*, 48(5), 691-704. [https://doi.org/10.1016/0895-4356\(94\)00170-U](https://doi.org/10.1016/0895-4356(94)00170-U)

General rights

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

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

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Review

THE EFFICACY OF NON-STEROIDAL ANTI-INFLAMMATORY DRUGS (NSAIDS) FOR SHOULDER COMPLAINTS. A SYSTEMATIC REVIEW

DANIËLLE A. W. M. VAN DER WINDT^{*1},
 GEERT J. M. G. VAN DER HEIJDEN,³ ROB J. P. M. SCHOLTEN,¹
 BART W. KOES¹ and LEX M. BOUTER¹²

¹Institute for Research in Extramural Medicine, Faculty of Medicine and ²Department of Epidemiology and Biostatistics, Vrije Universiteit, Amsterdam, The Netherlands and ³Department of Epidemiology, Faculty of Medicine and Health Sciences, University of Limburg, Maastricht, The Netherlands

(Received 27 October 1994)

Abstract—The medical literature was evaluated with respect to the efficacy of NSAIDs for shoulder complaints. Nineteen randomized clinical trials met the selection criteria and were included in this review. Each publication was independently scored by two blinded reviewers, according to a standardized set of 17 methodological criteria. The studies were ranked according to their total methods score (maximum 100 points). The methodological quality of the majority of the studies was rather disappointing; only 5 trials received a methods score exceeding 50 points. Furthermore, methodological criteria were often inadequately described, particularly those regarding the randomization procedure, co-interventions and control of compliance. The three trials with the highest methods scores demonstrated superior short-term efficacy of NSAIDs in comparison with placebo intervention. Fourteen trials comparing two types of NSAIDs showed no conclusive evidence in favour of a particular NSAID with respect to efficacy or tolerability. Future studies should compare the benefit-risk ratios of NSAIDs and analgesics for shoulder complaints in order to establish whether the use of NSAIDs is more favourable than analgesics, despite the higher risk of adverse reactions from NSAIDs.

Shoulder	Non-steroidal anti-inflammatory drugs	Blinded review	Efficacy
Adverse reactions	Randomized controlled trials	NSAIDs	

INTRODUCTION

A painful or stiff shoulder may be the result of a variety of diseases and conditions. Shoulder complaints occur frequently: epidemiological studies have reported point-prevalences ranging from 7% to 25% in western general populations [1-3]. The annual incidence is estimated at 10 per 1000, peaking at 25 per 1000 in the

age-category of 42-46 years [1]. Approximately 50% of these patients consult a physician for their shoulder complaints. The incidence of shoulder complaints in Dutch general practice is estimated at 13 per 1000 registered patients per year [4].

Shoulder complaints are usually a self-limiting condition of relatively short duration. In 48% of the patients, pain or stiffness will disappear or substantially improve within 1 month after consulting their general practitioner (GP) [unpublished observations]. Unfortunately, the

*All correspondence should be addressed to: Ms D.A.W.M. van der Windt, Institute for Research in Extramural Medicine, Vrije Universiteit, Van der Boechorststraat 7, 1080 BT Amsterdam, The Netherlands.

long-term outcome is not always this favourable. Persisting pain or restriction of movement, due to more chronic conditions, may last for several years [5–7].

A painful shoulder may be caused by articular or periarticular rheumatic conditions, or may be the result of neurological or vascular disorders, neoplasms, or referred pain from internal organs. Because of the complex functional anatomy of the shoulder girdle, even intrinsic complaints which emanate from the shoulder joint may constitute a diagnostic challenge. To date, there is no conclusive evidence on either the pathogenesis of shoulder disorders or the accuracy of diagnostic procedures such as X-ray, magnetic resonance imaging, computed tomography or ultrasonography [8, 9].

A wide variety of treatment modalities has been introduced, including medication, physical therapy modalities, exercise therapy, corticosteroid injections and surgery. In primary health care, therapy is frequently initiated with the prescription of a non-steroidal anti-inflammatory drug (NSAID) [4]. A variety of NSAIDs is available to the GP, and new drugs are being introduced on a regular basis. NSAIDs are assumed to act by inhibiting prostaglandin synthesis, resulting in relief of pain and suppression of inflammatory processes in articular or peri-articular structures [10]. Adverse reactions include mild gastrointestinal symptoms which occur frequently, as well as more serious complications, such as gastrointestinal bleeding, renal insufficiency, hepatitis and bronchospasm. The efficacy of NSAIDs should be studied in terms of benefit–risk ratios, in order to determine whether the advantages of NSAIDs outweigh the risk of adverse reactions.

The efficacy of NSAIDs has been studied extensively, but most randomized clinical trials have included rheumatic conditions affecting various joints. The purpose of this paper was to evaluate the medical literature with respect to the efficacy of NSAIDs for patients with shoulder complaints only. There is still much confusion and corresponding lack of consensus regarding the classification of shoulder disorders. Periarthritis humeroscapularis or a painful stiff shoulder do not constitute clearly defined clinical entities. Diagnostic criteria may even differ between studies for disorders seemingly straightforward labelled as tendinitis or bursitis. We therefore prefer to present one overview of the use of NSAIDs for all indications. In this systematic review, conclusions with respect to the

efficacy of NSAIDs were made after blinded assessment of the methodological quality of the studies according to standardized methodological criteria.

METHODS

Search strategy and screening

A Medline literature search of the period (1966–1993) was conducted, using the keywords shoulder, analgesics, anti-inflammatory drugs, bursitis, tendinitis, periarthritis, capsulitis, comparative study and evaluation study. All Medline citations referring to review articles or randomized controlled trials were retrieved and examined. The references of all retrieved articles were screened for additional relevant publications.

Selection

Studies had to meet the following criteria to be included in this review:

1. The study population included patients with intrinsic shoulder complaints.
2. At least one of the interventions under study included NSAIDs.
3. Patients were randomly allocated to the interventions under study.
4. The publication was a full report. Letters and abstracts were excluded.
5. In publications concerning various rheumatic disorders, the results of patients with shoulder complaints were presented separately, or at least 90% of the study population had to be composed of patients with shoulder complaints.
6. The publication was written in English, German, French or Dutch.

Methodological quality assessment

The selected publications were blinded for author(s), source and results by removing sections of the text, tables and figures. Subsequently, two blinded reviewers (GJM/GH and RJP/MS) independently scored each publication according to a standardized set of 17 methodological criteria (A–Q), referring to the selection of the study population, description of interventions, outcome measurements, analysis and presentation of data (Table 1).

The reviewers rated each methodological criterion according to the following rules:

+ : Informative description of the criterion at issue; adequate design and conduct (prevention of bias).

–: Informative description, but inadequate design or conduct.

?: Lacking or insufficient information, assigning either ‘+’ or ‘–’ was not possible.

Disagreements between the reviewers were identified and subsequently discussed during a consensus meeting. If disagreements were not resolved during this meeting a third blinded reviewer (BWK) was consulted for a final judgement. Next, all methodological criteria rated ‘plus’ were scored using the weighting factors listed in Table 1. Each study was assigned a total methods score by summing up all weighted criteria. This evaluation finally resulted in a hierarchical order, ranking the studies according to methodological quality. The methodological criteria and matching weights are based on the criteria used by Koes *et al.* [11, 12], but were adjusted to fit the disorders and interventions included in this review.

Data abstraction

The success rate in each intervention group was determined by using information from text and tables of the original publications. Whenever possible, success was defined as recovery or substantial improvement from baseline, according to the patient. The exact definition of success

may somewhat differ between studies, and will be presented in tables summarizing the details and results of the studies. Success rates were calculated according to the intention-to-treat principle, dividing the number of documented successes at the end of the intervention period by the number of patients allocated to the intervention by randomization (irrespective of drop-outs and loss to follow-up during the study). There will probably be a number of successfully recovered patients among the drop-outs, but the reasons for drop-out are not always clearly described. Consequently, for some studies our calculations will be an underestimation of the real success rate and produce somewhat lower success rates than the results reported in the original publications. The differences between the success rates in groups treated with NSAIDs and reference groups were calculated, along with the corresponding 95% confidence intervals (95% CI).

In order to estimate the ability to detect a clinically relevant difference, the power of each study was determined [13]. For these calculations a difference in success rate of 25% between intervention groups was considered to be clinically relevant.

The conclusions of the studies regarding the

Table 1. Criteria list for a methodological assessment of randomized clinical trials for shoulder complaints

Methodological aspect	Weighting
<i>Study population</i>	35
A. Selection and restriction: homogenous study population	4
B. Adequate randomization procedure	3
C. Prognostic comparability at baseline	7
D. Drop-outs	3
E. Loss to follow-up <20% or loss to follow-up <10%	1
F. Smallest study group ≥25 subjects or smallest study group ≥50 subjects or smallest study group ≥75 subjects	3
	6
	9
	15
<i>Interventions</i>	25
G. Informative description of each intervention	10
H. Placebo-controlled study	5
I. Pragmatic study	5
J. Co-interventions avoided	5
<i>Measurement of outcome</i>	25
K. Blinding of patient and physician	7
L. Relevant outcome measures	8
M. Blinded assessment of outcome measures	5
N. Adequate follow-up period	5
<i>Analysis and presentation of data</i>	15
O. Analysis according to intention-to-treat principle	5
P. Adequate presentation of data	5
Q. Adjustment for confounding variables and/or differences of prognostic indicators at baseline.	5

A more detailed description of the methodological criteria is given in the Appendix.

Table 2. Methodological quality of randomized clinical trials on the efficacy of NSAIDs for shoulder complaints

First author [ref.]	Methodological criteria*																	Methods score	Inadequate information†	Indication	Conclusion‡	
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q					
	4	3	7	3	3	15	10	5	5	5	7	8	5	5	5	5	5	100				
Petri [14]	4	3	2	—	3	6	8	5	—	—	5	8	5	3	—	5	—	57	24	Painful shoulder	Positive at 2 weeks Negative at 4 weeks	
Adebajo [15]	4	3	5	3	3	—	8	5	—	—	5	4	3	3	5	5	—	56	20	Tendinitis (<3 months)	Positive	
Mena [16]	4	—	3	2	3	6	10	5	—	—	3	7	3	3	—	5	—	54	23	Tendinitis/bursitis (<4 days)	Positive	
Zuinen [17]	4	—	6	—	1	15	8	—	5	—	3	7	2	3	—	—	—	54	23	Tendinitis/bursitis (<3 days)	Negative	
Berry [18]	1	—	5	3	3	—	6	5	5	—	2	5	4	3	5	5	—	52	24	Rotator cuff lesions	Negative	
Menkes [19]	1	—	3	1	—	9	8	—	—	—	5	6	3	3	5	5	—	49	28	Periarthritis	Negative	
Friis [20]	—	—	3	2	3	9	10	—	—	3	5	4	2	3	—	5	—	49	19	Tendinitis	Negative	
Duke [21]	1	—	3	2	—	6	8	—	—	—	5	8	4	3	—	5	—	45	18	Various conditions	Negative	
Thumb [22]	—	—	5	2	1	—	8	—	5	—	3	7	4	3	—	5	—	43	21	Periarthritis	Negative	
Ginsberg [23]	—	—	3	1	1	—	8	—	—	5	5	7	4	3	—	5	—	42	23	Tendinitis/bursitis	Negative	
Ginsberg [23]	—	—	3	1	1	—	8	—	—	5	5	7	4	3	—	5	—	42	23	Tendinitis/bursitis	Negative	
Famaey [24]	—	—	3	2	3	6	8	—	5	—	3	7	1	3	—	—	—	41	23	Various conditions	Negative	
White [25]	4	—	6	1	—	—	8	—	—	—	5	7	4	—	—	5	—	40	17	Tendinitis (<12 weeks)	Negative	
Huskisson [26]	1	—	3	2	3	—	8	—	—	5	5	7	3	3	—	—	—	40	12	Painful stiff shoulder	Negative	
Wielandt [27]	—	—	3	1	—	—	8	5	—	3	7	3	3	—	5	—	—	38	22	Tendinitis	Negative	
Smith [28]	1	—	3	2	1	—	10	—	—	—	5	5	1	3	—	5	—	36	24	Chronic painful shoulder	Negative	
Yamamoto [29]	1	—	—	—	—	9	8	—	—	—	5	2	—	3	—	5	—	33	43	Periarthritis	Negative	
Rhind [30]	—	—	1	2	—	—	6	—	—	—	5	7	3	3	—	5	—	32	23	Periarthritis	Negative	
Valtonen [31]	—	—	—	1	3	—	8	—	—	—	5	4	2	3	—	5	—	31	36	Various conditions	Positive at 1 week Negative at 2 weeks	

*Points are assigned to criteria rated '+', '—' (no points for '? or '—').

†Amount of points assigned to criteria rated '—'.

‡“Positive” if treatment with NSAIDs (treatment i) in Table 4 or 5 is significantly ($p < 0.05$) more effective than a reference treatment with respect to success rate or at least two other relevant outcome measures. “Negative” if treatment with NSAIDs is less effective than a reference treatment or if there are no significant differences.

Table 3. Methods scores and ranking* of the studies according to their methodological quality under three different assumptions

First author [ref.]	Weighted scores† (ranking I)	Maximal attainable scores‡ (ranking II)	Unweighted scores§ (ranking III)
Petri [14]	57 (1)	81 (1)	24 (1)
Adebajo [15]	56 (2)	76 (5)	23 (2)
Mena [16]	54 (3)	77 (2)	23 (2)
Zuinen [17]	54 (3)	77 (2)	22 (4)
Berry [18]	52 (5)	76 (5)	22 (4)
Menkes [19]	49 (6)	77 (2)	20 (6)
Friis [20]	49 (6)	68 (8)	19 (10)
Duke [21]	45 (8)	63 (13)	20 (6)
Thumb [22]	43 (9)	64 (11)	20 (6)
Ginsberg [23]	42 (10)	65 (10)	19 (10)
Famaey [24]	41 (11)	64 (11)	17 (13)
White [25]	40 (12)	57 (16)	20 (6)
Huskisson [26]	40 (12)	52 (18)	19 (10)
Wielandts [27]	38 (14)	60 (14)	17 (13)
Smith [28]	36 (15)	60 (14)	16 (15)
Yamamoto [29]	33 (16)	76 (5)	10 (18)
Rhind [30]	32 (17)	55 (17)	13 (16)
Valtonen [31]	31 (18)	67 (9)	13 (16)

*Studies with equal scores are given the same ranking in the hierarchy.

†Weighted methods scores as presented in (Table 2).

‡Methods scores under the assumption that all criteria assigned a question-mark would have received points if relevant information had been available. Points reflecting "insufficient information" (Table 2) have been added to the weighted methods score.

§Methods scores under the assumption that all methodological criteria are equally important: methods scores are calculated without the use of weighting factors.

efficacy of NSAIDs were marked as being 'positive' or 'negative'. A study was marked 'positive' if NSAIDs proved to be significantly ($p < 0.05$) more effective than a reference treatment with respect to at least two relevant outcome measures (e.g. severity of pain or functional limitation) or with respect to our calculations of differences in success rates (95% CI excludes a difference of '0'). Conclusions were marked 'negative' if treatment with NSAIDs was less effective than a reference treatment, or if there were no statistically significant differences.

RESULTS

The literature search resulted in the identification of 63 publications on the efficacy of NSAIDs for shoulder complaints. Forty four publications were excluded, being either non-randomized controlled trials, abstracts, or studies concerning several joints or disorders without a separate presentation of the outcome for intrinsic shoulder complaints. Eighteen publications met the selection criteria and were included in this review [14–31], one of which reported on two different studies which have been reviewed separately [23].

The two blinded reviewers initially agreed on 840 of the 931 items scored (90%). Nearly all disagreements were due to reading errors or a difference in interpretation of the methodologi-

cal criteria. Since all disagreements were resolved during the consensus meeting, the third reviewer did not have to be consulted.

Methodological quality

In Table 2, scores assigned to each methodological criterion, together with the final methods score, are presented for each study. Only 5 trials [14–18] received a methods score exceeding 50 points, indicating a rather poor methodological quality of most of the evaluated publications. Design or conduct was often inadequate (rated '—') with respect to loss to follow-up (E), sample size (F) and analysis according to the intention-to-treat principle (O).

A considerable amount of relevant information on methodological criteria was lacking or inadequate, representing a total of 12 [26] to 43 points [29] (Table 2). Description of the following criteria was often considered to be inadequate: selection criteria and restriction to a homogeneous study population (A), randomization procedure (B), co-interventions (J), prognostic comparability at baseline (C), control of compliance (G4) and adjustment for confounding variables (Q).

Table 3 presents the ranking of the studies according to their methods score, as well as the ranking under the assumption that the study was in fact adequately performed with respect to

insufficiently described methodological criteria. The maximum attainable scores would range from 52 points [26] to 81 points [14], provided that more informative description of design and conduct did not reveal any bias. This assumption did not result in dramatic changes in the ranking of the studies according to methodological quality. Exceptions are the study by Valtonen [31] and, in particular, the study by Yamamoto *et al.* [29], which would be ranked among the first 6 studies. Both publications report on rheumatic conditions affecting several joints, including only a limited amount of information on shoulder complaints.

Table 3 also presents the ranking of studies under the assumption that all methodological criteria are equally important. Calculation of unweighted methods scores resulted in some minor changes in the hierarchical order. However, the hierarchical order of the 5 studies with the highest methods scores remains the same when using a scoring system without weighting factors.

Comparison of NSAIDs with placebo intervention or other treatment modalities

Table 4 summarizes the design and results of five trials which have included a placebo treatment and/or other treatment modalities for shoulder complaints. The studies varied considerably with respect to the type of NSAID used, dosage schedule, diagnosis or indication, duration of complaints and the exact definition of success. These differences may account for some of the variation in success rates between studies, which ranged from 20% [14] to 86% [16] in groups taking NSAIDs.

Four of these five trials scored more than 50 points [14–16, 18]. The trial described by Petri *et al.* [14] was assigned the highest methods score (57 points). In this study, naproxen was compared with triamcinolone injections in 100 patients with a painful shoulder, using a double-dummy technique to ensure blinding of both patient and physician. All four intervention groups received an injection (triamcinolone or lidocaine) as well as tablets (naproxen or placebo). After 2 weeks, improvement in functional status and range of movement was significantly larger when using naproxen than placebo tablets (group i vs iv in Table 4). This difference was not statistically significant after conclusion of the intervention period at 4 weeks.

Mena *et al.* [16] conducted the only trial in this review which compared a NSAID (flurbiprofen)

with placebo tablets in patients with acute tendinitis or bursitis of the shoulder. The results show a significant difference in favour of flurbiprofen. This publication received a relatively high methods score (54 points). In all, only 4 trials compared the efficacy of NSAIDs with a placebo intervention, using placebo tablets alone [16] or in combination with local infiltration of an anaesthetic or corticosteroid [14, 15, 18]. Three of these studies, all of relatively high methodological quality, demonstrated superior short-term efficacy of NSAIDs [14–16].

In 2 out of 4 studies comparing NSAIDs with other treatment modalities, corticosteroid injections appeared to be significantly more effective than NSAIDs [14, 15].

Comparison between two types of NSAIDs

The majority of the studies included in this review were designed to study the differences between NSAIDs, with respect to efficacy or tolerability, and have not included a placebo intervention. Details of the design and results of 14 comparative studies are presented in Table 5.

Success rates in the intervention groups varied between 10% [26] and 79% [22, 29]. As mentioned earlier, this variation may partly be the result of differences with respect to intervention protocol, characteristics of the study population and the exact definition of success. Due to the intention-to-treat analysis, in some studies our calculations produced lower success rates than the results reported in the original publications [19, 23, 27, 31]. Menkes [19] presented success rates of 62% (etodolac) and 43% (piroxicam) for patients who did not deviate from the protocol, whereas we calculated success rates of 46 and 33%, respectively (Table 5).

The studies revealed few differences between the various types of NSAIDs. With the exception of the study by Valtonen [31], the conclusions of all trials were rated 'negative'. Valtonen demonstrated a significantly larger improvement of a combined symptom score in patients receiving diclofenac than in patients receiving naproxen. This difference was only statistically significant after 1 week, halfway through the intervention period. According to Yamamoto *et al.* [29], indomethacin was more effective than piroxicam with respect to relief of pain, but these differences were not reflected by differences in success rates. Smith *et al.* [28] reported a significantly larger reduction of pain at night for piroxicam, compared with naproxen. However,

Table 4. Details of trials studying the efficacy of NSAIDs for shoulder complaints: comparison with placebo treatment or other treatment modalities

First author [ref.]	Treatment with NSAID* (No. of patients)	Reference treatment(s) (No. of patients)	Methods score	Result† (p-values <0.05 were considered significant)	% Patients with adverse reactions
Petri [14]	i. Naproxen 1000 mg daily for 4 weeks plus 1 × intrabursal injection of lidocaine (25) ii. Naproxen 1000 mg daily for 4 weeks plus 1 × intrabursal injection of lidocaine and triamcinolone (25) <i>double-dummy</i>	iii. Naproxen matched placebo tablets plus 1 × intrabursal injection of triamcinolone and lidocaine (25) iv. Naproxen matched placebo tablets plus 1 × intrabursal injection of lidocaine (25)	57	Improvements of ROM and limitation of function were significantly larger when using naproxen than placebo at 2 weeks. At 4 weeks naproxen was not significantly better than placebo. Triamcinolone was superior to naproxen in relief of pain and global clinical index. SRs (perfect score of global index) after 4 weeks: 20% (i), 28% (ii), 28% (iii), 8% (iv). Difference i vs iv: 12%, 95% CI (-6, 31).	8% (i) 16% (ii) 12% (iii) 12% (iv)
Adebajo [15]	i. Diclofenac 150 mg daily for 4 weeks plus 1 × subacromial injection of lidocaine (20) <i>double-dummy</i>	ii. Diclofenac matched placebo tablets plus 1 × subacromial injection of lidocaine and triamcinolone (20) iii. Diclofenac placebo tablets plus 1 × subacromial injection of lidocaine (20)	56	Groups (i) and (ii) showed significant benefit over those taking placebo in improvement of pain, ROM and functional limitation. SRs (improvement in all three outcome measures) was significantly higher in group (ii) than in (i) and (iii) (SRs: 70%, 30%, 0%, respectively). Difference (i) vs (iii): 30%, 95% CI (10, 50).	?
Mena [16]	i. Flurbiprofen, dosage decreasing from 300 to 200 mg daily for a maximum period of 14 days (35)	ii. Flurbiprofen matched placebo tablets (34)	54	The pattern of dose reduction during treatment was evident in group (i), not in (ii). Group (i) showed a larger proportion of patients with improvements in pain, tenderness, swelling, and ROM. SRs (better according to investigator) were: 86% (i) and 56% (ii). Difference: 30%, 95% CI (10, 30).	23% (i) 9% (ii)
Berry [18]	i. Tolmetin sodium 1200 mg daily for 4 weeks plus 1 × intra-articular injection of prednisolone and lignocaine (12)	ii. Tolmetin matched placebo tablets plus 1 × injection of prednisolone + lignocaine (12) iii. Ultrasound therapy (12) iv. Tolmetin matched placebo tablets plus placebo ultrasound therapy (12) v. Acupuncture	52	There were no significant differences in pain score, ROM and SR between the five groups. SRs (success according to the investigator) were 42% (i), 50% (ii), 50% (iii), 75% (iv), 42% (v). Difference (i) vs (ii): -8%, 95% CI (-48, 32).	active tolmetin 33% (i) placebo: 25% (ii + iv)
White [25]	i. Indomethacin 100 mg daily for 25 days plus 1 × intrabursal injection of lidocaine (20) <i>double-dummy</i>	ii. Indomethacin placebo matched tablets plus 1 × intrabursal injection of lidocaine and triamcinolone (20)	40	In each group pain and ROM improved significantly, but there were no significant differences between the groups. SRs (global assessment score ≤4) were: 50% (i) and 45% (ii). Difference: 5%, 95% CI (-26, 36).	?

Abbreviations: SR = success rate; ROM = range of movement; 95% CI = 95% confidence interval.

*Some studies have used a double-dummy system in order to maintain the double-blind nature of the trial. Additional placebo tablets or injections are provided for each intervention group, matched to the reference medication.

†Success rates (SR) were calculated according to the intention-to-treat principle, using data presented in the original publications. Unless stated otherwise, results are presented for success rates according to the patient.

For each study the definition of success rate is given in the last column. Differences between success rates were calculated together with the 95% confidence intervals, using the following equation:

$$(p_1 - p_2) \pm z_{\alpha/2} \sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}$$

p_1 = success rate in group i (NSAIDs), p_2 = success rate in the reference group, n_1 = sample size in the reference group, n_2 = sample size in the reference group. $z_{\alpha} = 1.96$.

Table 5. Details of trials studying the efficacy of NSAIDs for shoulder complaints: comparison between two NSAIDs

First author [ref.]	Treatment with NSAID* (No. of patients)	Alternative treatment with NSAIDs (No. of patients)	Methods score	Results† (p-values <0.05 were considered significant)	% Patients with adverse reactions
Zuñen [17]	i. Diclofenac/misoprostol for 2 weeks 100–150 mg/ 400–600 µg daily (185)	ii. Diclofenac 100–150 mg daily for 2 weeks (187)	54	Both groups showed steady improvements in pain, limitation of movement and tenderness, but there were no significant differences between the groups. SRs (maximal or moderate improvement from baseline) are 48% (i) and 47% (ii). Difference: 1%, 95% CI (–4, 6).	38% (i) 25% (ii)
Menkes [19]	i. Etodolac 400 mg daily for 2 weeks (52) <i>double-dummy</i>	ii. Piroxicam 40 mg daily for 2 days, followed by 20 mg daily for 12 days (52)	49	Both groups improved significantly within 1 week, but there were no significant differences with regard to spontaneous pain, pain on movement, functional score, and use of acetaminophen. SRs (better compared to initial state): 46% (i) and 33% (ii). Difference: 13%, 95% CI (–6, 32).	27% (i) 23% (ii)
Friis [20]	i. Sustained release ibuprofen 2400 mg daily: 2 × daily 2 tablets for three weeks (73) <i>double-dummy</i>	ii. Conventional ibuprofen 2400 mg daily: 2 × daily 1 tablet for 3 weeks (74)	49	A significantly larger proportion showed complete relief of pain in group (i): 21% vs 7%. SR (better or complete relief) was not significantly different between the groups: 66% (i) and 76% (ii). Difference: –10%, 95% CI (–25, 5).	51% (i) 38% (ii)
Duke [21]	i. Naproxen 825 mg daily for 4 weeks (30) <i>double-dummy</i>	ii. Indomethacin 100 mg daily for 4 weeks (29)	45	There were no significant differences with regard to pain, ROM, stiffness, restriction or limitation of activity between the groups. SRs (cured or better): 50% (i) and 52% (ii). Difference: –2%, 95% CI (–28, 24).	33% (i) 28% (ii)
Thumb [22]	i. Fentiazac 400 mg daily for 3 weeks (19)	ii. Diclofenac 50 mg daily for 3 weeks (19)	43	There were no differences between the groups with respect to pain, ROM, global benefit or side-effects. SRs (well or very well) were 79% (i) and 68% (ii) for global benefit. Difference: 11%, 95% CI (–17, 39).	26% (i) 21% (ii)
Ginsberg [23]	i. Slow release fentiazac 300 mg daily: 1 × daily 1 tablet for 2 weeks (17) <i>double-dummy</i>	ii. Conventional fentiazac 400 mg daily: 4 × daily 1 tablet for 2 weeks (15)	42	There was significant improvement in tenderness, pain and ROM, but there were no significant differences between the groups. SRs (complete relief according to the investigator) were 59% (i) and 53% (ii). Difference: 6%, 95% CI (–28, 40).	13% (i) 20% (ii)
Ginsberg [23]	i. Slow release fentiazac 300 mg daily: 1 × daily 1 tablet for 2 weeks (17) <i>double-dummy</i>	ii. Conventional fentiazac 200 mg daily: 2 × daily 1 tablet for 2 weeks (16)	42	There were no significant differences between the two formulations. Improvements in pain, ROM, and tenderness were significant in both groups. SRs were 47% (i) and 44% (ii). Difference: 3%, 95% CI (–31, 37).	20% (i) 27% (ii)

Famaey [24]	i. Ibuprofen 1600 mg daily for 2 weeks (25)	ii. Diclofenac 100 mg daily for 2 weeks (25)	41	There were no significant differences in perceived benefit and improvements of pain and ROM between the groups. SRs (efficacy good or very good) were 52% (i) and 48% (ii). Difference: 4%, 95% CI (– 24, 32).	20% (i) 24% (ii)
Huskinson [26]	i. Diclofenac 150 mg daily for 2 weeks (20) <i>double-dummy</i>	ii. Ibuprofen 1200 mg daily for 2 weeks (20)	40	There were no significant differences with regard to pain score or ROM. Diclofenac showed a faster improvement of pain scores (results presented in graphs). SRs (efficacy excellent or good) were: 15% (i) and 10% (ii). Difference: 5%, 95% CI (– 15, 25).	35% (i) 20% (ii)
Wielandt [27]	i. Fentiazac 400 mg daily for 1 week (14) (one patient with tendinitis of the elbow was included in the study)	ii. Phenylbutazone 400 mg daily (12)	38	There were no significant differences with respect to tenderness, pain, function loss or ROM. SRs (improved compared to initial state) were 50% (i) and 42% (ii). Difference: 8%, 95% CI (– 30, 46).	?
Smith [28]	i. Piroxicam 20 mg daily for 21 days (20) <i>double-dummy</i>	ii. Naproxen 250 mg daily for 21 days (20)	36	Piroxicam was significantly more successful in relieving pain at night. No drug was superior for treatment of pain on movement or ROM. No data have been given on any measure of global benefit.	40% (i) 50% (ii)
Yamamoto [29]	i. Indomethacin 25 mg daily for 2 weeks (148, 67 with PHS) <i>double-dummy</i>	ii. Piroxicam 20 mg daily for 2 weeks (147, 79 with PHS)	33	Patients' perceptions of improvement slightly favoured indomethacin. SRs (improvement according to the physician) were 79% (i) and 73% (ii). Difference: 6%, 95% CI (– 8, 19).	18% (i) 14% (ii) (all patients)
Rhind [30]	i. Naproxen 750 mg daily for 4 weeks (20) <i>double-dummy</i>	ii. Indomethacin 50 mg daily for 4 weeks (21)	32	There were significant improvements of stiffness, ROM, and pain in both groups. 40% of patients in group (i) and 48% in (ii) reported being cured or better. Difference: – 8%, 95% CI (– 38, 22).	70% (i) 76% (ii)
Valtonen [31]	i. Diclofenac sodium 25 mg daily for 2 weeks (60, 16 with shoulder complaints) <i>double-dummy</i>	ii. Naproxen 250 mg daily for 2 weeks (60, 15 with shoulder complaints)	31	Results of total symptom score (pain, swelling, tenderness, functional impairment, ROM, disturbed sleep) were significantly better in group (i) than in group (ii) after the first week. SRs (improvement according to the physician) in the total study population after 2 weeks were 72% (i) and 57% (ii). Difference 15%, 95% CI (– 2, 32). (No separate presentation of these data for shoulder patients.)	26% (i) 25% (ii) (all patients)

Abbreviations: SR = success rate; ROM = range of movement; CI = 95% confidence interval; PHS = peri arthritis humero-scapularis.
*Some studies have used a double-dummy system in order to maintain the double-blind nature of the trial. Additional placebo tablets or injections are provided for each intervention group, matched to the reference medication.
†Success rates (SR) were calculated according to the intention-to-treat principle, using data presented in the original publications. Unless stated otherwise, results are presented for success rates according to the patient.
For each study the definition of success rate is given in the last column. Differences between success rates were calculated together with the 95% confidence intervals, using the following equation:

$$(p_i - p_k) \pm z_{\alpha} \sqrt{p_i(1 - p_i)/n_i + p_k(1 - p_k)/n_k}$$

p_i = success rate in group i (NSAIDs), p_k = success rate in the reference group, n_i = sample size in group i (NSAIDs), n_k = sample size in the reference group, z_{α} = 1.96.

it must be noted that the intervention groups were not comparable with respect to severity of nocturnal pain at baseline. There was no adjustment for this difference in the analysis of the results. Methods scores of these 3 studies [28, 29, 31] were low (31, 33 and 36 points, respectively).

Only one of the studies listed in Table 5 was assigned more than 50 points [17]. In this trial the efficacy and tolerability of diclofenac (187 patients) was compared with a combination of diclofenac and misoprostol (185 patients). The two formulations proved to be of equal efficacy, showing similar success rates and improvement in tenderness, pain and range of movement.

In Table 6 the results of power calculations are presented for each trial. Only three studies were designed with sufficient power ($1 - \beta > 0.80$) to demonstrate a difference in success rate of 25% [17, 20, 29].

Adverse reactions

Adverse reactions were reported by 8% [14] to 76% [30] of the patients receiving NSAIDs (Tables 4 and 5). Side-effects were mainly of

gastrointestinal nature or included skin rash, headache or dizziness. Adverse reactions were usually considered to be of a moderate or mild nature, and were the reason for withdrawal in less than 10% of all participants in most trials. However, Rhind *et al.* [30], Friis *et al.* [20] and Famaey and Ginsberg [24] reported higher drop-out rates due to side-effects: 20, 19 and 12%, respectively. In studies comparing two different types of NSAIDs, the incidence of adverse reactions was approximately similar, except for the trials conducted by Friis *et al.* [20], Huskisson *et al.* [26] and Zuinen [17] (Table 5). Notwithstanding a higher incidence of gastrointestinal side-effects in his study, Zuinen recommends use of the combined formulation diclofenac/misoprostol, referring to previous studies in rheumatoid arthritis and osteoarthritis in which misoprostol was ascribed a gastro-protective benefit.

The issue of benefit-risk ratio was only addressed by Menkes [19], who calculated the ratio of opinions on efficacy and safety of both patient and investigator, with very similar results for the two NSAIDs under study, etodolac and piroxicam.

Table 6. Power calculations for each study: the ability of the trials to detect a clinically relevant difference* between intervention groups

First author [ref.]	Mean number of patients per group	Success rate in least successful group	Power ($1 - \beta$)	Conclusion
Petri [14]	25	8%	0.59	Positive at 2 weeks Negative at 4 weeks
Adebajo [15]	20	0%	0.68	Positive
Mena [16]	35	59%	0.64	Positive
Zuinen [17]	186	47%	0.99	Negative
Berry [18]	12	42%	0.23	Negative
Menkes [19]	52	43%	0.74	Negative
Friis [20]	74	67%	0.97	Negative
Duke [21]	30	50%	0.52	Negative
Thumb [22]	19	68%	0.49	Negative
Ginsberg [23]	16	53%	0.31	Negative
Ginsberg [23]	17	47%	0.31	Negative
Famaey [24]	25	48%	0.44	Negative
White [25]	20	45%	0.36	Negative
Huskisson [26]	20	11%	0.46	Negative
Wielandts [27]	13	50%	0.25	Negative
Smith [28]†	20	—	—	Negative
Yamamoto [29]	73	73%	0.99	Negative
Rhind [30]	21	40%	0.36	Negative
Valtonen [31]	16	63%	0.37	Positive at 1 week Negative at 2 weeks

*A difference in success rate of 25% between intervention groups was considered to be clinically relevant. Power ($1 - \beta$) was calculated according to the formula described by Friedman *et al.* [13]:

$$z_{\beta} = \frac{-z_{\alpha}\sqrt{2p(1-p)} + \sqrt{N(p_c - p_i)}}{\sqrt{p_c(1-p_c) + p_i(1-p_i)}}$$

p_i = success rate in group i ($p_c + 25\%$), p_c = success rate in the reference group, p = mean success rate, N = mean sample size. Values of z_{β} were converted to β using statistical tables of standard normal distributions.

†Smith *et al.* have not presented data on success rate or any other measure of perceived benefit. Therefore, it was not possible to make a power calculation for this study.

DISCUSSION

Studies of low methodological quality may introduce bias and may consequently result in different effect sizes. Therefore, we assessed the efficacy of NSAIDs for shoulder complaints after evaluation of the methodological quality of the randomized clinical trials, in order to avoid conclusions based on trials of inferior methodological quality. Blinding of the publications enabled us to evaluate the methods of the studies objectively. Each methodological criterion was assigned a weighting factor which reflected our assessment of the relative importance of the criteria. These weighted criteria were summated into a final score, reflecting the methodological quality of a study. The resulting hierarchical order of studies, according to methodological quality, was not severely affected when the methods scores were recalculated under different assumptions. The 6 studies with the highest methods scores remained in the first 6 positions when points were assigned for inadequately described methodological criteria or when weighting factors were totally discarded (Table 3). This means that our scoring system could discriminate between studies of higher and lower methodological quality, regardless of weighting factors. Conclusions on the efficacy of NSAIDs for shoulder complaints were therefore based on trials which are listed in the top of our ranking. When weighting factors are used, the quality of the trials is measured on a numerical scale, rather than a ranking scale. This may have consequences for concluding statements if final conclusions are based on studies with, for example, at least 50 points or a minimum of 60 points for methodological quality.

The methodological quality of the majority of trials was disappointingly low; only 5 studies received a methods score exceeding 50 points. Furthermore, a considerable amount of information, relevant for the evaluation of the methodological quality, seemed to be missing. An adequate description of the randomization procedure is very important, as valid randomization is the only way to prevent selection bias. A brief description of the randomization procedure was only provided by two studies [14, 15]. Even when an adequate randomization procedure has been carried out, it does not guarantee equal distribution of prognostic factors and confounding variables among the study groups, particularly when a study size is small. Documentation of relevant prognostic indicators and baseline rates of outcome measures gives some indication

whether the randomization procedure has been successful for these variables. However, in a considerable number of studies, baseline characteristics were not presented, or were only partly presented (Table 2). In addition, only 4 studies received points for description of co-interventions [20, 23, 26, 27] or control of compliance [16, 20, 28]. Interpretation of the efficacy of NSAIDs is difficult when information on the use of co-interventions is limited and no information is presented with respect to the adherence of the participants to the treatment regimen. A complete and more informative report may have revealed additional flaws in the design or conduct of the studies included in this review.

In our scoring system no category for the quality of outcome measure is included, although this is a very relevant aspect of a clinical trial. Apart from the frequently applied visual analogue scales for pain, a wide variety of outcome measures was used for the assessment of global improvement, functional limitation, swelling and tenderness. Most of these measures have probably been designed on the basis of face validity, and may have proved useful in clinical practice. A thorough assessment of the quality of outcome measures for shoulder complaints requires a detailed study of their applicability, reliability, validity and responsiveness. Unfortunately, this was not feasible within the scope of this review.

The efficacy of NSAIDs was compared with a placebo intervention in 4 trials [14–16, 18]. Three of these studies, all highly positioned in our methodological hierarchy, demonstrated the superior short-term efficacy of NSAIDs. Corticosteroid injections showed even better results in two of these trials [14, 15], but NSAIDs may be considered useful in the short-term management of shoulder complaints. We hoped to find out whether NSAIDs are useful for shoulder disorders in general and, if possible, for which type of patient. In order to detect these differences, trials should have a relatively homogeneous study population with respect to prognosis and susceptibility to treatment. However, selection criteria varied over the studies and were often inadequately described, or a wide variety of patients were included within one study. Consequently, there was no strong evidence for any differences in efficacy of NSAIDs between the various indications and diagnoses in this review.

The intention-to-treat analysis for the calculation of success rates may have underestimated

true effect rates, but it might also be possible that the review has overestimated effect rates. Publication bias, for example, may have resulted in an overrepresentation in the literature of studies with positive results. Definite answers on the efficacy of NSAIDs can only be provided after conducting trials of adequate methodological quality and power.

If studies do not recruit a sufficiently large study population, they are *a priori* only able to detect very large, improbable differences between groups. An adequate sample size enhances the precision of an estimated effect size and enables the detection of more subtle though clinically relevant differences. However, the results of power calculations vary depending on a number of factors, including the goal of the study, the choice of outcome measure, the variability of the results of the outcome measure, the size of a clinically relevant difference and the level of significance. We considered a 25% difference between groups in success rate (on a dichotomous scale) to be clinically relevant. Power calculations showed that only three trials had sufficient power to detect this rather large difference. These three studies [17, 20, 29] compared different types of NSAIDs and detected only minor differences between intervention groups. The methodological quality of two of these powerful studies [20, 29] was disappointing. Nevertheless, other trials studying the efficacy of NSAIDs in rheumatoid arthritis, osteoarthritis, soft tissue rheumatism and sports injuries have also demonstrated that not many differences should be expected between the various types of NSAIDs [10, 32–34]. Any preference for a particular NSAID will probably be influenced by differences in tolerability, cost or convenience of dosage schedule. The less frequent administration of slow-release formulations may offer advantages in terms of patient compliance, although this issue could not be studied by Friis *et al.* [20] or Ginsberg and Famaey [23]. The use of a double-dummy system in these trials hindered the assessment of differences in compliance, as all participants received the same number of tablets.

Despite the lack of power of most studies, the results of the trials were not pooled in order to avoid pooling data from studies of widely varying methodological quality. As mentioned earlier, the outcome of studies of low quality may be biased. Furthermore, pooling was impeded by the fact that most trials differed considerably with respect to the intervention under study (type

of NSAID and dosage schedule), the selection criteria for the study population and the choice of outcome measures. The possibilities of pooling were further restricted by incomplete reporting in some publications with respect to selection criteria and outcome measures.

The tolerability of NSAIDs varies considerably between patients, but the use of NSAIDs is quite frequently accompanied by adverse reactions, mostly of a gastrointestinal nature. More serious adverse reactions, such as bronchospasm, renal insufficiency, gastrointestinal bleeding, hepatitis and allergic reactions are less often reported and may be prevented by careful prescription [35]. The trials in this review reported few differences between NSAIDs with respect to the incidence of adverse reactions. Fries *et al.* [36] studied the comparative toxicity of NSAIDs by constructing a summary index of symptoms, laboratory abnormalities and hospitalization in a large number of patients with rheumatoid arthritis. The most toxic drugs appeared to be indomethacin, tolmetin and meclofenamate, whereas coated aspirin, salsalate and ibuprofen were the least toxic drugs.

Although side-effects were often considered to be moderate or mild in the evaluated studies, a substantial number of participants discontinued medication because of adverse reactions. Moreover, it is uncertain whether the NSAIDs were merely effective as an analgesic or also because of their anti-inflammatory properties. It would be interesting to know whether simple analgesics, such as paracetamol, are as effective as NSAIDs without producing the accompanying adverse reactions. Bradley *et al.* [37] demonstrated a similar short-term efficacy of acetaminophen and ibuprofen in patients with osteoarthritis, whether ibuprofen was administered in an analgesic or in a higher, anti-inflammatory dose. Unfortunately, we could not find any reports of trials comparing the efficacy of NSAIDs with simple analgesics for soft-tissue rheumatic conditions of the shoulder. Studies comparing the benefit–risk ratios of NSAIDs and analgesics should indicate whether the benefit of NSAIDs outweighs the risk of adverse reactions, which may be avoided by prescribing more simple analgesics. This is particularly relevant for the treatment of musculoskeletal disorders in elderly people, who are at an increased risk of NSAID-induced toxicity [38].

In conclusion, the trials with the highest methods scores seemed to demonstrate a superior short-term efficacy of NSAIDs, com-

pared with a placebo intervention. However, the methodological quality of the majority of clinical trials studying the efficacy of NSAIDs for shoulder complaints was rather disappointing. When designing and conducting trials studying the efficacy of NSAIDs for shoulder complaints, more attention should be given to recruiting a sufficient sample size, limitation of the number of drop-outs, adequate data analysis and the writing of a more informative report of the study. There was no conclusive evidence in favour of any particular type of NSAID for the treatment of shoulder complaints. Future studies should compare the benefit-risk ratios of NSAIDs and analgesics for shoulder complaints, in order to establish whether the use of NSAIDs is more favourable than analgesics, despite the higher risk of adverse reactions from NSAIDs.

REFERENCES

1. Bjelle A. Epidemiology of shoulder problems. *Baillieres Clin Rheumatol* 1989; 3: 437-451.
2. Bergenudd H, Lindgärde F, Nilsson B, Petersson CJ. Shoulder pain in middle age. A study of prevalence and relation to occupational work load and psychosocial factors. *Clin Orthop* 1988; 231: 234-238.
3. Chakravarty KK, Webley M. Disorders of the shoulder: an often unrecognised cause of disability in elderly people. *Br Med J* 1990; 300: 848-849.
4. Miedema HS. *Reuma-onderzoek Meerdere Echelons (ROME): Basisrapport*. Leiden: Nederlands Instituut voor Praeventieve Gezondheidszorg TNO; 1994.
5. Reeves B. The natural history of the frozen shoulder syndrome. *Scand J Rheumatol* 1975; 4: 193-196.
6. Binder AI, Bulgen DY, Hazleman BL, Roberts S. Frozen shoulder: A long-term prospective study. *Ann Rheum Dis* 1984; 43: 361-364.
7. Chard MD, Satelle LM, Hazleman BL. The long-term outcome of rotator cuff tendinitis—A review study. *Br J Rheumatol* 1988; 27: 385-389.
8. Chandnani V, Ho C, Gerhart J, Neumann C, Kursunoglu-Brahme S, Sartoris DJ, Resnick D. MR findings in asymptomatic shoulders: A blind analysis using symptomatic shoulders as controls. *Clin Imaging* 1992; 16: 25-30.
9. Stiles RG, Otte MT. Imaging of the shoulder. *Radiology* 1993; 188: 603-613.
10. Todd PA, Clissold SP. Tenoxicam. An update of its pharmacology and therapeutic efficacy in rheumatic diseases. *Drugs* 1991; 41: 625-646.
11. Koes BW, Bouter LM, Beckerman H, Van der Heijden GJMG, Knipschild PG. Physiotherapy exercises and back pain: a blinded review. *Br Med J* 1991; 302: 1572-1576.
12. Koes BW, Assendelft WJJ, Van der Heijden GJMG, Bouter LM, Knipschild PG. Spinal manipulation and mobilization for back and neck pain: a blinded review. *Br Med J* 1991; 303: 1298-1303.
13. Friedman LM, Furberg CD, Demets DL. *Fundamentals of Clinical Trials*. Littleton MA: PSG Publishing Company Inc; 1985.
14. Petri M, Dobrow R, Neiman R, Whiting-O'Keefe Q, Seaman WE. Randomized, double-blind, placebo-controlled study of the treatment of the painful shoulder. *Arthritis Rheum* 1987; 30: 1040-1045.
15. Adebajo AO, Nash P, Hazleman BL. A prospective double blind dummy placebo controlled study comparing triamcinolone hexacetonide injection with oral diclofenac 50 mg TDS in patients with rotator cuff tendinitis. *J Rheumatol* 1990; 17: 1207-1210.
16. Mena HR, Lomen PL, Turner LF, Lamborn KR, Brinn EL. Treatment of acute shoulder syndrome with flurbiprofen. *Am J Med* 1986; 80: 141-144.
17. Zuinen C. Diclofenac/misoprostol vs diclofenac/placebo in treating acute episodes of tendinitis/bursitis of the shoulder. *Drugs* 1993; 45 (Suppl. 1): 17-23.
18. Berry H, Fernandes L, Bloom B, Clark RJ, Hamilton EBD. Clinical study comparing acupuncture, physiotherapy, injection and oral anti-inflammatory therapy in shoulder-cuff lesions. *Curr Med Res Opin* 1980; 7: 121-126.
19. Menkes CJ. Périarthrite scapulo-humérale: efficacité, tolérance et bénéfice thérapeutique compare de l'étodolac (600 mg/j) et du piroxicam (40/20 mg/j). *Rhumatologie* 1990; 42: 195-200.
20. Friis J, Jarner D, Toft B, Christensen K, Christophersen J, Ibfeldt HH, Korsgaard J, Skinhoj A, Andersen F. Comparison of two ibuprofen formulations in the treatment of shoulder tendonitis. *Clin Rheumatol* 1992; 11: 105-108.
21. Duke O, Zecler E, Grahame R. Anti-inflammatory drugs in periarthritis of the shoulder: a double-blind, between-patient study of naproxen versus indomethacin. *Rheumatol Rehabil* 1981; 20: 54-59.
22. Thumb N, Kolarz G, Scherak O, Mayrhofer F. The efficacy and safety of fentiazac and diclofenac sodium in periarthritis of the shoulder: a multi-centre, double-blind comparison. *J Int Med Res* 1987; 15: 327-334.
23. Ginsberg F, Famaey JP. A double-blind comparison of slow-release and standard tablet formulations of fentiazac in the treatment of patients with tendinitis and bursitis. *Curr Med Res Opin* 1985; 9: 442-448.
24. Famaey JP, Ginsberg F. Treatment of periarthritis of the shoulder: a comparison of ibuprofen and diclofenac. *J Int Med Res* 1984; 12: 238-243.
25. White RH, Paull DM, Fleming KW. Rotator cuff tendinitis: Comparison of subacromial injection of a long acting corticosteroid versus oral indomethacin therapy. *J Rheumatol* 1986; 13: 608-613.
26. Huskisson EC, Bryans R. Diclofenac sodium in the treatment of painful stiff shoulder. *Curr Med Res Opin* 1983; 8: 350-353.
27. Wielands L, Dequecker D. Double-blind trial comparing fentiazac with phenylbutazone in patients with tendinitis. *Curr Med Res Opin* 1979; 6 (Suppl. 2): 85-89.
28. Smith MD, Thomas D, McCredie M, Brooks PM. Piroxicam versus naproxen in the treatment of the painful shoulder. *Pharmatherapeutica* 1986; 4: 585-589.
29. Yamamoto M, Sugano T, Kashiwazaki S, Kageyama T, Mizushima Y, Kameyama M. Double-blind comparison of piroxicam and indomethacin in the treatment of cervico-brachial syndrome and periarthritis scapulo-humeralis (stiff shoulder). *Eur J Rheumatol Inflamm* 1983; 6: 266-273.
30. Rhind V, Downie WW, Bird HA, Wright V, Engler C. Naproxen and indomethacin in periarthritis of the shoulder. *Rheumatol Rehabil* 1982; 21: 51-53.
31. Valtonen EJ. A comparative short-term trial with Voltaren (diclofenac sodium) and naproxen in soft-tissue rheumatism. *Scand J Rheumatol* 1978; (Suppl. 22): 69-73.
32. Moser U, Walburger H, Schwarz HA, Gobelet CA. A double-blind randomised multicentre study with tenoxicam, piroxicam and diclofenac sodium retard in

- the treatment of ambulant patients with osteoarthritis and extra-articular rheumatism. *Scand J Rheumatol* 1989; (Suppl. 80): 71–80.
33. Pena M. Etodolac: analgesic effects in musculoskeletal and postoperative pain. *Rheumatol Int* 1990; 10(Suppl.): 9–16.
 34. Weiler JM. Medical modifiers of sports injury. The use of non-steroidal anti-inflammatory drugs (NSAIDs) in sports soft-tissue injury. *Clin Sports Med* 1992; 11: 625–644.
 35. Avouac B. Adverse effects of long-term treatment with non-steroid anti-inflammatory drugs. In: Munderthaler M, Van Zwieten PA, Farcot JM, Eds. **Treatment of Chronic Pain Possibilities**. Harvard Academic Publications; 1990: 47–62.
 36. Fries JF, Williams CA, Bloch DA. The relative toxicity of non-steroidal anti-inflammatory drugs. *Arthritis Rheum* 1991; 34: 1353–1360.
 37. Bradley JD, Brandt KD, Katz BP, Kalasinski LA, Ryan SI. Comparison of an anti-inflammatory dose of ibuprofen, an analgesic dose of ibuprofen, and acetaminophen in the treatment of patients with osteoarthritis of the knee. *N Engl J Med* 1991; 325: 87–91.
 38. Afable RF, Ettinger WH. Musculoskeletal disease in the aged. Diagnosis and management. *Drugs Aging* 1993; 3: 49–59.
- of drop-outs (1 pt) and the reason for withdrawal (1 pt) is presented for each study group.
- E. Loss to follow-up: $100 - \{(\text{number of patients at main moment of effect measurement} / \text{number of patients at randomization}) \times 100\}\%$.
 - F. Smallest study group immediately after randomization.
 - G. Description of type of intervention (4 pts), dosage schedule or intensity of treatment (2 pts), duration of treatment or number of injections (2 pts), control of compliance (2 pts).
 - H. A comparison is made with a study group receiving a placebo treatment only.
 - I. A comparison is made between two or more existing interventions, medication or otherwise. (No addition of any placebo to the other study group.)
 - J. All other interventions are avoided (5 pts) or comparable co-interventions between study groups (3 pts).
 - K. Double-dummy system (5 pts) or some other attempt at blinding of the patient (2 pts) and the physician (1 pt). Blinding of patient (1 pt) or physician (1 pt) is evaluated and successful.
 - L. Results are presented of: pain (2 pts), global measure of improvement (2 pts), functional status (1 pt), range of movement (1 pt), adverse reactions (2 pts).
 - M. Blinded assessment of the outcome measures mentioned under L (1 pt each).
 - N. Timing of effect measurement is identical for all study groups (3 pts). Final effect measurement at least 3 months after randomization (2 pts).
 - O. When loss to follow-up $\leq 10\%$: analysis according to intention-to-treat principle: results on all randomized patients are presented for most outcome measures, and on the most important moments of effect measurement, irrespective of drop-outs or missing values. When loss to follow-up $> 10\%$: intention-to-treat-analysis plus alternative analysis that accounts for drop-outs and missing values.
 - P. Presentation of frequencies or means and SD of the most important outcome measures for each study group at most important moments of effect measurement.
 - Q. Adjustment of the results for confounding variables (e.g. differences of prognostic indicators at baseline, compliance, additional medication) by using multivariate analysis.

APPENDIX

More detailed description of the methodological criteria

- A. Clearly described criteria for selection (1 pt). Restriction to a homogenous population with respect to diagnosis, duration of complaint, previous treatments, contra-indications for treatment with NSAIDs (3 pts).
- B. Randomization procedure is described and is a procedure which excludes bias (e.g. sealed envelopes, allocation by telephone, pre-coded packaged medication (3 pts)).
- C. Comparability for duration of complaints (2 pts), baseline scores for outcome measures (2 pts), age (1 pt), recurrence status (1 pt), previous treatment of complaint (1 pt).
- D. No drop-outs (3 pts). If there are drop-outs: the number