

Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy

Sally Green, Rachele Buchbinder, Richard Glazier, Andrew Forbes

Department of
Epidemiology and
Preventive
Medicine, Monash
University,
Melbourne 3186,
Australia

Sally Green,
PhD scholar
Rachele
Buchbinder,
senior lecturer
Andrew Forbes,
senior lecturer

Family and
Community
Medicine,
Preventive Medicine
and Biostatistics,
University of
Toronto, Ontario,
Canada

Richard Glazier,
assistant professor

Correspondence to:
Dr Buchbinder
rachele.buchbinder@
med.monash.edu.au

BMJ 1998;316:354-60

Abstract

Objective: To review the efficacy of common interventions for shoulder pain.

Design: All randomised controlled trials of non-steroidal anti-inflammatory drugs, intra-articular and subacromial glucocorticosteroid injection, oral glucocorticosteroid treatment, physiotherapy, manipulation under anaesthesia, hydrodilatation, and surgery for shoulder pain that were identified by computerised and hand searches of the literature and had a blinded assessment of outcome were included.

Main outcome measures: Methodological quality (score out of 40), selection criteria, and outcome measures. Effect sizes were calculated and combined in a pooled analysis if study population, end point, and intervention were comparable.

Results: Thirty one trials met inclusion criteria. Mean methodological quality score was 16.8 (9.5-22). Selection criteria varied widely, even for the same diagnostic label. There was no uniformity in the outcome measures used, and their measurement properties were rarely reported. Effect sizes for individual trials were small (range - 1.4 to 3.0). The results of only three studies investigating "rotator cuff tendinitis" could be pooled. The only positive finding was that subacromial steroid injection is better than placebo in improving the range of abduction (weighted difference between means 35° (95% confidence interval 14 to 55)).

Conclusions: There is little evidence to support or refute the efficacy of common interventions for shoulder pain. As well as the need for further well designed clinical trials, more research is needed to establish a uniform method of defining shoulder disorders and developing outcome measures which are valid, reliable, and responsive in affected people.

Introduction

There are many accepted standard forms of conservative treatment for shoulder disorders, including non-steroidal anti-inflammatory drugs, corticosteroid injections, and physiotherapy, yet evidence of their efficacy is not well established. Furthermore, the interpretation of results of studies is often hampered by the fact that these disorders are labelled and defined

in diverse and often conflicting ways. To determine the efficacy of common interventions for shoulder pain, we performed a systematic review of randomised controlled trials investigating these treatments. To determine whether the results of the different studies could be compared or pooled, or both, we also undertook a methodological review of the selection criteria and outcome assessment used in these studies.

Methods

Identification and selection of studies

One of us (SG) searched computerised bibliographic databases (MEDLINE, EMBASE, CINAHL) without language restrictions from 1966 to September 1995 using the Cochrane Collaboration search strategy, which aims to identify all randomised controlled trials.¹ This was combined with the medical subject heading "shoulder" (exploded) and other keywords pertaining to shoulder disorders or their treatment. SG also hand searched relevant conference proceedings and reviewed textbooks and reference lists of all retrieved articles.

To determine whether a study should be included, two of us (RB and RG) reviewed the retyped methods sections of all identified trials independently according to predetermined criteria (that the trial be randomised, that the outcome assessment be blinded, and that the intervention was one of those under review). Randomised controlled trials which investigated common interventions for shoulder pain in adults (age greater than or equal to 18 years) were included provided that there was a blinded assessment of outcome. For the purposes of this review, interventions were broadly categorised as non-steroidal anti-inflammatory drugs, intra-articular and subacromial glucocorticosteroid injection, oral glucocorticosteroid treatment, physiotherapy, manipulation under anaesthesia, hydrodilatation (shoulder distension), and surgery. All studies which primarily concerned pain arising from the shoulder were included irrespective of diagnostic label. Studies that included various rheumatological disorders were considered if the results on shoulder pain were presented separately or if 90% or more of patients in the study had shoulder pain.

Assessment of validity

To determine the methodological quality, RB and RG independently reviewed the retyped methods and results sections of the included studies using predetermined criteria.² An overall score for methodological quality (out of a possible 40) was calculated to assess the overall standard of each trial, but these scores were not used to weight the pooled analysis.

Assessment of selection criteria and outcome

RB and SG reviewed the selection criteria and outcome measures. They determined whether distinct diagnoses were specified and whether definitions of these diagnoses were recorded. Inclusion and exclusion criteria were also examined. For outcome measures they recorded the types of measures used, the method of measurement, and whether the clinimetric properties had been considered. (Clinimetric or measurement properties refer to the reliability, validity, and responsiveness to change of an outcome measure, clinimetrics being the measurement of clinical phenomena.) The timing of outcome assessment was also noted.

Analysis of data

To assess efficacy, raw data (means and standard deviations) were extracted for reported outcomes when data were available in the published reports and entered into the Cochrane Collaboration's Review Manager software program.¹ For studies not reporting the required data, further details were requested of first authors, but no additional information was obtained. Range of motion scores were entered as degrees of movement, and all pain and overall effect scores were transformed to 100 point scales.

Effect sizes: individual trials

To determine the treatment effect for each end point assessed we calculated the effect size as the difference between the mean of the treatment group at follow up and the mean of the control group at follow up, divided by the pooled standard deviation:

$Effect\ size = (\bar{x}_t - \bar{x}_c) / PSD$ where \bar{x}_t = treatment group mean, \bar{x}_c = control group mean, and PSD = pooled standard deviation.

In this equation, effect size is expressed as a function of standard deviation. For example, an effect size of 0.43 reflects a difference between means of 0.43 of one standard deviation.

In the five studies in which standard deviation was not reported, effect size was calculated by conversion of the reported P value to a *t* statistic with appropriate degrees of freedom and use of the formula

$$Effect\ size = t \sqrt{\left(\frac{1}{n_t} + \frac{1}{n_c}\right)}$$

where n_t = number of subjects in treatment group, n_c = number of subjects in control group, and t = observed *t* statistic.

In the three studies which reported P values only as less than a value—for example, $P < 0.05$, $P < 0.01$ —a conservatively small effect size was calculated as though the P value equalled the quoted value—for example, $P < 0.05$ was taken as $P = 0.05$. When only the change in scores from baseline were reported (three studies), the within group standard deviation after intervention was computed assuming a (likely)

conservative correlation of 0.3 between measurements before and after intervention. One study provided only the range (maximum and minimum) of scores for each group, from which the standard deviation was estimated as one quarter of the range.

After reviewing the outcomes assessed in the individual trials, we examined three end points which together provided the most detailed data: pain (100 point scale), restriction of abduction (degrees), and overall effect as reported by the patient (100 point scale).

Pooled analysis

Only trials investigating the same intervention in similar populations and using the same outcome measure at the same or similar times of follow up were considered for pooling. Pooling was carried out using the Revman 3.0 software program¹ to calculate weighted difference between means, which is the difference between treatment and comparison group means at follow up, and (in contrast to effect size) is measured in the dimension of the outcome variable being combined. A random effects model was used for pooling to allow for heterogeneity between studies.³ Estimates and in particular confidence intervals from this model reduce to those from a fixed effects model when study results are homogeneous.

Results

Identification and selection of studies

Thirty one out of 58 identified studies met our selection criteria.⁴⁻³⁴ A list of excluded studies is available from us. Reasons for exclusion were lack of randomisation (23 studies), lack of blinding (12 studies), not the study population of interest (5 studies), and not the intervention of interest (6 studies). Despite blinding, 8 of the 58 trials were identified by RB or RG, or both.^{4 8 14 18 20 21 25 30}

The included studies assessed non-steroidal anti-inflammatory drugs (17 trials),^{4-8 10-13 16-19 22 26 29 33} intra-articular or subacromial glucocorticosteroid injection (10 trials),^{10 11 13 14 18 23-25 31 32} physiotherapy (5 trials),^{9 15 18 20 21} oral glucocorticosteroid treatment (2 trials),^{31 34} hydrodilatation (2 trials),^{24 32} and manipulation under anaesthesia (1 trial).²⁸ There were no randomised controlled trials of surgical interventions.

Selection criteria

Twenty four trials specified distinct diagnoses to characterise their study population. These included periarthritis (7 trials),^{4 5 16 27 29 30 31} adhesive capsulitis (6 trials),^{14 18 24-26 33} frozen shoulder (3 trials),^{28 29 31} rotator cuff tendinitis (7 trials),^{8 10 11 17 18 24 30} supraspinatus tendonitis (3 trials),^{29 30 32} bursitis (2 trials),^{8 14} infraspinatus tendonitis (1 trial),¹⁴ subscapularis tendonitis (1 trial),¹⁴ acromioclavicular joint sprain (1 trial),¹⁴ and rotator cuff lesion (1 trial).¹⁹ Definitions of distinct diagnoses were specified in 16 of the 24 trials. On the basis of review of the diagnostic labels or definitions of the study populations, or both, most trials could be broadly categorised as studying adhesive capsulitis (including periarthritis and frozen shoulder) (23 trials) or rotator cuff tendinitis (including supraspinatus, infraspinatus, and subscapularis tendonitis) (12 trials), or both. Six trials gave no selection criteria or population definition.^{6 11 21 28 30 35}

Selection criteria used in shoulder pain trials

Rotator cuff disease

- No definition given (2 trials)
- Shoulder pain and painful arc (1 trial)
- At least two of painful abduction, painful arc, and tenderness of supraspinatus insertion (1 trial)
- Painful arc between 40° and 120° abduction (1 trial)
- Pain on resisted abduction, tenderness over supraspinatus tendon, and normal passive motion (1 trial)
- Pain with resistance in abduction and external and internal rotation (1 trial)
- Pain with resisted movements of the shoulder and loss of passive abduction (1 trial)
- Pain on resisted abduction with or without resisted external rotation (1 trial)
- Pain exacerbated by resisted movement, passive range > active range, normal passive range (1 trial)
- Full range of passive motion and pain on resisted abduction (1 trial)

Adhesive capsulitis

- No definition given (8 trials)
- External rotation < 30° and abduction < 90° (1 trial)
- Total passive movement < 50% of normal with pain worse at night (1 trial)
- Generalised limitation of glenohumeral motion with pain at rest or on movement (5 trials)
- Pain, active and passive motion limited to > 20°, pain on resisted abduction or rotation, and impaired glenohumeral joint (1 trial)
- Loss of passive motion of glenohumeral joint (1 trial)
- Pain, restriction of movement, loss of full function, and pain at night with inability to lie on affected side (1 trial)
- Pain at night, inability to lie on affected side, restriction of active and passive motion, restriction of external rotation (2 trials)
- Restriction of abduction and external rotation (1 trial)
- Appreciable restriction of both active and passive motion (1 trial)
- Abduction and flexion < 90° and external rotation < 20° (1 trial)
- Abduction and flexion < 70% and external rotation < 20% (1 trial)

The box shows the selection criteria used to define the shoulder disorders studied. In general, adhesive capsulitis was defined as the presence of pain with restriction of active and passive glenohumeral joint movements, and rotator cuff tendinitis was defined by the presence of painful arc and pain with resisted movements or a normal passive range of motion. However, no standard definitions were used, and conflicting criteria often defined the same condition in different trials. Exclusion criteria were specified in 28 studies, although these also varied widely between studies.

Assessment of outcome

Table 1 summarises the outcomes assessed. Pain and range of motion were recorded in most trials (29 and 27 respectively). Only four studies included an adequate description of how range of motion was assessed (measurement tool and definition of end of range).^{16 21 22 35} Function was assessed in eight studies and was measured either by visual analogue scale^{14-19 26} or return to work.²⁰ No study included a disability index. Although two studies cited a reliability study for their method of range of motion assessment,^{21 35} no other reference was made to the reliability, validity, or responsiveness of the outcome measures used. Timing of assessment for the primary efficacy analysis varied between 1 and 24 weeks.

Validity of individual trials

Table 2 shows the overall scores for the quality of methods and for each category for each trial. The mean quality score of all trials combined was 16.8 out of a possible 40 points (42%). No trial scored greater than 22 out of a possible score of 40 (range 9.5-22.0).

Individual trials

Table 2 shows effect sizes according to intervention. In general, the effect sizes were small, suggesting a lack of clear benefit for any of the treatments investigated.

Pooled analysis

Most results of individual trials could not be pooled because of a lack of similarity of study population, outcome measures, timing of outcome assessment, or insufficient reported data. Pooling of results was performed for two studies of non-steroidal anti-inflammatory drugs and placebo in rotator cuff tendinitis (table 3).^{10 11} The weighted difference between means suggested that non-steroidal anti-inflammatory drugs may be superior to placebo in improving the degree of restriction of abduction (26 degrees (95% confidence interval -9 to 61)). The weighted difference between means for pain score was 3 (-19 to 25). Data from the same two trials were pooled to determine the efficacy of subacromial steroid injection and placebo in rotator cuff tendinitis (table 3).^{10 11} The weighted difference between means suggested injection was superior to placebo in improving range of abduction (35 (14 to 55)).

Pooling of results was also performed for two studies which compared non-steroidal anti-inflammatory drugs and steroid injection with non-steroidal anti-inflammatory drugs alone in rotator cuff tendinitis (table 3).^{10 18} The weighted difference between means showed no added benefit of injection with respect to both restriction of abduction (4 (-14 to 22)) and pain (-2 (-11 to 7)).

Discussion

This review has confirmed the lack of uniformity in the way shoulder disorders are labelled and defined. It has also highlighted the wide variation in assessment of

Table 1 Outcome measures used in 31 trials of intervention for shoulder pain and stiffness

Outcome measure	No of trials
Pain	29
Overall score	17
On activity	14
At night	11
At rest	4
During day	4
Range of motion	27
Active abduction	13
Passive abduction	14
Active external rotation	4
Passive external rotation	12
Active flexion	5
Passive flexion	5
Hand behind back (internal rotation)	5
Composite score	5
Function	8
Global improvement score (patient)	5
Global improvement score (blinded assessor)	7
Overall severity (patient)	6
Overall severity (blinded assessor)	5
Tenderness	3
Strength	1
Stiffness (patient)	4
Analgesia count	6
Adverse effects	9

Table 2 Summary of validity scores and effect sizes for overall pain, restriction in abduction, and overall efficacy for all included studies

Trial (regimen)	Sample size (total)	Length of follow up	Methodology scores				Total (out of 40)	Effect size		
			A	B	C	D		Overall pain	Restriction in abduction	Overall efficacy
NSAID v placebo										
Cohen and Cohen ¹² (25 mg indomethacin 4xday for 7 days)	36	1, 2, 3, 7 days	3.5	1	7.5	1	13	0.6	-0.1	0.7
Ward et al ²⁶ (50 mg diclofenac 3xday for 7 days)*	25	7 days	4	1	6	1	12	NA	NM	NM
Adebajo et al ¹¹ (50 mg diclofenac 3xday for 28 days)	40	4 weeks	7	1	7	5	20	1.5	1.5	NM
Petri et al ¹⁰ (1000 mg naproxen 1xday for 30 days)	50	2 and 4 weeks	7.5	1	8.5	4.5	21.5	-0.4	0.2	NM
One type of NSAID v another										
Yamamoto et al ⁷ (25 mg indomethacin 3xday v piroxicam 20 mg 1xday)	139	2 weeks	3	2	7	3	15	NM	NM	0.2
Friis et al ⁶ (1200 mg slow release ibuprofen 2xday v 600 mg ibuprofen 4xday for 3 weeks)	145	3 weeks	6	4	8	4	22	NM	NM	-0.3
Ginsberg and Famaey ¹⁹ (30 mg slow release fentiazac 1xday v standard fentiazac 4xday for 2 weeks)†	60	1 and 2 weeks	3	1	8	3.5	15.5	-0.3	NM	NM
Wielandts and Dequeker ¹⁶ (100 mg phenylbutazone 4xday v 100 mg fentiazac 4xday)	25	7 days	3.5	1	8.5	2.5	15.5	0.5	NM	NM
Smith et al ²² (20 mg piroxicam 1xday v 250 mg naproxen 2xday)†	40	1, 2, 3 weeks	7	1	7	4	19	0.1	0.0	NM
Duke et al ¹⁷ (825 mg naproxen 1xday v 100 mg indomethacin 1xday)	59	2 and 4 weeks	5	1	8	3.5	17.5	-0.4	-0.5	-0.3
Rhind et al ³³ (750 mg naproxen 1xday v 100 mg indomethacin 1xday for 4 weeks)‡	41	4 weeks	5.5	1	8	3	17.5	-0.4	0.2	-0.4
Famaey and Ginsberg ²⁹ (400 mg ibuprofen 4xday v 25 mg diclofenac 4xday)*	50	1 and 2 weeks	4.5	1	8	3.5	17	NA	NA	NA
Huskisson and Bryans ⁵ (400 mg ibuprofen 3xday v 50 mg diclofenac 3xday for 2 weeks)	40	2 weeks	6.5	1.5	7.5	3	18.5	NA	NA	-0.2
Thumb et al ⁴ (200 mg fentiazac 2xday v 50 mg diclofenac 2xday for 3 weeks)	19	1, 2, 3 weeks	5.5	1.5	7	3.5	17.5	0.2	NA	0.3
NSAID plus injection v NSAID alone										
Petri et al ¹⁰ (100 mg naproxen 1xday v 3 ml lignocaine and 1 ml triamcinolone 40 mg/ml)	50	2 and 4 weeks	7.5	1	8.5	4.5	21.5	-0.1	0.7	NM
Berry et al ¹⁸ (400 mg tolmetin sodium 3xday plus anterior injection of methylprednisolone 40 mg and 2 ml of 2% lignocaine v 400 mg tolmetin sodium 3xday alone)	60	2 and 4 weeks	5	0.5	7	4.5	17	-0.1	-0.2	NA
NSAID plus steroid injection v placebo										
Petri et al ¹⁰ (subacromial bursa injection of 4 ml of 1% lignocaine plus 500 mg naproxen 2xday for 30 days v placebo)	50	2 and 4 weeks	7.5	1	8.5	4.5	21.5	-0.9	0.9	NM
NSAID v steroid injection										
Adebajo et al ¹¹ (50 mg diclofenac 3xday for 28 days v subacromial injection of 2 ml of 0.5% lignocaine plus 1 ml triamcinolone hexacetone 80 mg/ml)	40	4 weeks	7	1	7	5	20	1.4	0.1	NM
White et al ⁸ (100 mg indomethacin 1xday for 3 or 6 weeks v 1 or 2 injections of 40 mg triamcinolone acetone)	40	6 weeks	7	2	6.5	3	18.5	0.0	-0.5	NM
Steroid injection										
Injection v placebo:										
Richardson ²³ (2 injections 2 weeks apart of prednisolone acetate into subdeltoid bursae and glenohumeral joint v injections of saline)†	101	2, 4, 6 weeks	7.5	2	4.5	2.5	16.5	NA	-0.5	NA
Rizk et al ²⁵ (intra-articular methylprednisolone and lignocaine v subacromial methylprednisolone and lignocaine v identical injections of saline)*	48	4, 8, 11, 24 weeks	5	1.5	5.5	2.5	14.5	NA	NA	NM
Withrington et al ¹⁴ (subacromial injection of 80 mg methylprednisolone and 2 ml of 2% lignocaine v identical injection of saline)	25	2 and 8 weeks	3.5	2	4	1	10.5	NM	NM	NA
Petri et al ¹⁰ (subacromial bursa injection of 4 ml of 1% lignocaine v identical injection of saline)	50	2 and 4 weeks	7.5	1	8.5	4.5	21.5	-0.7	0.6	NM
Adebajo et al ¹¹ (subacromial injection of 2ml of 0.5% lignocaine plus 1 ml of triamcinolone hexacetone 80 mg/ml v identical injection of saline)	40	4 weeks	7	1	7	5	20	3.0	1.4	NM
Bulgen et al ³¹ (subacromial and intra-articular injection of 20 mg methylprednisolone acetate and 1% lignocaine v no treatment)*	42	6 weeks	4.5	0.5	4.5	0	9.5	NA	NA	NA
Berry et al ¹⁸ (anterior injection of 40 mg methylprednisolone and 2 ml of 2% lignocaine v identical injection of placebo)	60	2 and 4 weeks	5	0.5	7	4.5	17	-0.2	-0.6	NA
Trigger point injection v anatomical injection:										
Hollingworth et al ¹³ (trigger point injection of 2 ml of 40 mg methylprednisolone acetate and 1% lignocaine v subacromial or intra-articular injection of 2 ml of 40 mg methylprednisolone acetate and 1% lignocaine)	77	1, 4, 8 weeks	4.5	2	4	4	14.5	NA	NA	-5.9§
Injection v physiotherapy:										
Bulgen et al ³¹ (subacromial and intra-articular injection of 20 mg methylprednisolone acetate and 1% lignocaine v Maitland's mobilisations 3xweek for 6 weeks)*	42	6 weeks	4.5	0.5	4.5	0	9.5	NA	NA	NA
Berry et al ¹⁸ (anterior injection 40 mg methylprednisolone and 2 ml of 2% lignocaine v 8x10 minute ultrasound sessions)	60	2 and 4 weeks	5	0.5	7	4.5	17	0.5	-0.5	NA

Continued over

Table 2 (continued)

Trial (regimen)	Sample size (total)	Length of follow up	Methodology scores				Total (out of 40)	Effect size		
			A	B	C	D		Overall pain	Restriction in abduction	Overall efficacy
Injection v hydrodilatation:										
Jacobs et al ³² (intra-articular injection of 40 mg triamcinolone acetonide (total volume 1 ml) v 40 mg triamcinolone acetonide plus 6 ml of 0.25% bupivacaine plus 3 ml of air (total volume 10 ml)‡	47	6 and 12 weeks	6	1.5	7.5	3.5	18.5	NA	0.0	NA
Corbell et al ²⁴ (intra-articular injection of 40 mg triamcinolone acetonide v 40 mg acetonide plus 20 ml lignocaine to cause distension of capsule to 20 cm ³)*	40	1, 2, 3 months	3.5	1.5	4.5	0.5	10	NA	NA	NM
Physiotherapy										
Physiotherapy v placebo:										
Leclaire and Bourgooin ¹⁵ (routine care (hot pack, exercises, and manual treatment) plus 17 sessions of magnetotherapy v routine care alone)	47	4, 8, 12 weeks	4	2.5	8	4.5	19	-0.2	-0.4	NM
Binder et al ³⁴ (magnetotherapy active coils for 5-9 hours per day for 4 weeks v identical regimen of placebo coils for 4 weeks, then active coils)†	29	2, 4, 6, 8, 12, 16 weeks	4	2.5	7	2.5	16	-0.9	-1.1	NM
England et al ²¹ (active infra red laser treatment at 904 nm 3x/week for 2 weeks v placebo)†	30	2 weeks	3.5	1.5	6.5	2.5	14	-1.4	-1.4	NM
Berry et al ¹⁸ (8x10 minute ultrasound sessions v placebo)	60	2 and 4 weeks	5	0.5	7	4.5	17	-0.6	NA	NA
Mobilisation plus exercise v exercise alone:										
Nicholson ⁹ (mobilisation and active exercise 2-3x/week for 4 weeks v active exercise alone)	20	1, 2, 3, 4 weeks	4	1	2	1	8	0.0	0.6	NM
Manipulation under anaesthesia										
Thomas et al ²⁸ (manipulation under anaesthesia plus intra-articular cortisone injection v intra-articular injection alone)*	30	4 and 12 weeks	3	1	1.5	1	6.5	NA	NA	NM
Oral steroid treatment										
Blockey et al ³⁰ (cortisone acetate suspension in syrup (concentration 12.5 mg/ml) daily v placebo)*	32	1, 4, 18 weeks	6	2	5	2	15	NA	NA	NA
Binder et al ³⁴ (10 mg oral steroid daily for 4 weeks v no treatment)¶	40	6 weeks and 8 months	3	1	6	5	15	-0.6**	-0.2	NA
Average validity scores										
All included studies			4.9	1.5	6.5	2.7	16.8			

NSAID=non-steroidal anti-inflammatory drug. A=validity score out of 10 for study population (selection criteria, randomisation, comparable baseline characteristics, and complete follow up). B=validity score out of 9 for description of therapeutic regimen. C=validity score out of 14 for effect measurement (blinding of patient and assessor, as well as outcome measures used). D=validity score out of 7 for data analysis (complete presentation of results and power of study). NA=not able to calculate effect size from data given in trial. NM=not measured. Effect size= $(\bar{x}_t - \bar{x}_c) / PSD$ where \bar{x}_t =treatment group mean, \bar{x}_c =control group mean, and PSD=pooled standard deviation. *Unable to calculate effect size from data available. †Effect size calculated from P value and t statistic. ‡Effect size calculated from change score using correlation of 0.3 before and after treatment. §Odds ratio, 95% CI -15 to -2. ¶Standard deviation calculated from range (SD=one quarter of range). **Odds ratio, 95% CI -7 to 0.

outcome in clinical trials investigating the efficacy of interventions for painful shoulder. These factors limit the degree to which the results of different trials can be compared and pooled. In addition, the heterogeneity of the interventions studied, the timing of outcome assessment, the overall poor methodological quality, inadequate reporting of results, and small sample sizes preclude the drawing of firm conclusions about the efficacy of any of the interventions studied.

On the basis of our review, the only conclusions that may be drawn about efficacy are that non-steroidal anti-inflammatory drugs and subacromial glucocorticosteroid injection may be superior to placebo in improving range of abduction in rotator cuff tendinitis and that the addition of corticosteroid injection to non-steroidal anti-inflammatory drugs does not seem to confer further benefit. These pooled results should be interpreted cautiously. In addition to the above concerns, statistical manipulation of inadequately presented results was necessary to perform these analyses. No conclusions can be drawn about the efficacy of the interventions studied for adhesive capsulitis.

Table 3 Pooled analyses. Values are weighted differences between means (95% confidence intervals)

Study	Range of abduction (degrees)*	Pain (100 point scale)
NSAID v placebo		
Petri et al ¹⁰	7 (-15 to 30)	-9 (-21 to 3)
Adebajo et al ¹¹	43 (25 to 61)	13 (8 to 19)
Pooled	26 (-9 to 61)	3 (-19 to 25)
Subacromial steroid injection v placebo		
Petri et al ¹⁰	24 (3 to 44)	26 (21 to 32)
Adebajo et al ¹¹	45 (25 to 64)	-14 (-29 to 0)
Pooled	35 (14 to 55)	7 (-33 to 47)
NSAID plus injection v NSAID alone		
Berry et al ¹⁸	-7 (-33 to 18)	-3 (-21 to 16)
Petri et al ¹⁰	12 (-8 to 31)	-2 (-13 to 8)
Pooled	4 (-14 to 22)	-2 (-11 to 7)

NSAID=non-steroidal anti-inflammatory drug. *Mean in treatment group minus mean in control group, positive values representing improved abduction and more pain.

Previous reviews

Previous separate reviews of non-steroidal anti-inflammatory drugs, steroid injections, and physiotherapy have been performed by investigators in the Netherlands.³⁵⁻³⁷ Although our study confirms their conclusions on the poor overall methodological quality of reviewed trials, our study differs in several important respects.

Firstly, we tried to differentiate studies on the basis of the nature of the populations being studied, recognising that the benefits of treatment may vary for different underlying causes of shoulder pain.

Secondly, we calculated effect sizes for the same reported outcome measures in different trials. This enables a direct comparison between studies using the

same outcome measurement, although it is important to note that if one effect size is larger than another it may be because in the different studies the numerator (the treatment effect) is larger, the denominator (the variability between subjects in each group) is smaller, or there is some combination of the two. In the previous reviews the overall efficacy of interventions was described by calculation of success rates for each intervention group. These were determined by dividing the number of documented successes (defined as recovery or substantial improvement from baseline, according to the patient) at the end of the intervention period by the number allocated to the intervention by randomisation. The exact definition of success therefore differed between papers and is, in essence, subjective.³⁶

Finally, we performed a detailed assessment of the outcome measures used in previous trials. While most studies incorporated a measure of pain and range of motion, the accompanying disability, which may be of utmost importance to the patient and therefore a more appropriate end point, was neglected in most studies. Furthermore, the reliability and responsiveness of the outcome measures used in most trials has not been established. To improve our ability to interpret and compare the results of different studies, further work is needed to address these issues.

Another recent systematic review of corticosteroid injection for rotator cuff tendinitis concluded that steroid injection is effective in treating it.³⁸ However, this conclusion cannot be verified by our results. The other review included non-randomised studies, the results of primary studies were reported only as significant or not significant, and no attempt was made to measure effect sizes or pool results.

Further research

Systematic reviews of randomised controlled trials serve to identify parts of clinical practice where further research is required. We could not draw firm conclusions about the efficacy of any of the common interventions currently being used to treat painful shoulders from this review. Therefore further clinical trials are needed to justify or censure current treatment strategies. These trials should carefully consider their study population and measures of outcome. Adoption of a uniform method of labelling and defining shoulder disorders and incorporation of a standard set of outcome measures may greatly enhance these efforts. This review will be updated through the Musculoskeletal Review Group of the Cochrane Collaboration as further trials are performed.

We thank the Australasian Cochrane Centre and the Musculoskeletal Cochrane Review Group for their methodological support.

Contributors: RB initiated and lodged the protocol for the project with the Cochrane Collaboration. SG coordinated the project under RB's supervision. SG searched the literature, retrieved data, and performed blinding. RB and RG performed the methodological review and determined inclusion. SG entered and synthesised the data and analysed and interpreted the results. AF, RG, and RB also analysed and interpreted the results. All authors wrote the paper. RB is the guarantor for the study.

Funding: None.

Conflict of interest: None.

- Oxman A, ed. Preparing and maintaining systematic reviews. In *Cochrane Collaboration handbook*. Oxford: Cochrane Collaboration, 1994.
- Glazier R, Buchbinder R, Bell M. Critical appraisal of continuing medical education in the rheumatic diseases for primary care physicians. *Arthritis Rheum* 1995;38:533-8.

Key messages

- This systematic review found little evidence to support the use of any of the common interventions in managing shoulder pain
- There is currently no uniformity in the way shoulder disorders are labelled or defined
- Measurement of outcome varies widely between clinical trials and, in general, the reliability, validity, and responsiveness of these outcome measures are not established
- Further clinical trials are needed to determine the optimal treatment strategies for shoulder pain

- Dersimmon R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;7:177-88.
- Thumb N, Kolarz G, Scherak O, Mayrhofer F. The efficacy and safety of fentiazac and diclofenac sodium in periarthritis of the shoulder: a multicenter, double-blind comparison. *J Int Med Res* 1987;15:327-34.
- Huskisson EC, Bryans R. Diclofenac sodium in the treatment of painful stiff shoulder. *Curr Med Res Opin* 1983;8:350-3.
- Friis J, Jarner D, Toft B, Christense K, Christophersen J, Ibfelt H, et al. Comparison of two ibuprofen formulations in the treatment of shoulder tendonitis. *Clin Rheumatol* 1992;11:105-8.
- Yamamoto M, Sugano T, Kashiwazaki T, Kageyama T, Mizushima Y, Kameyama M. Double blind comparison of piroxicam and indomethacin in the treatment of cervicobrachial syndrome and periarthritis scapulo-humeralis (stiff shoulder). *Eur J Rheum Inflamm* 1983;6:266-73.
- White RH, Paull DM, Fleming KW. Rotator cuff tendonitis: comparison of subacromial injection of a long acting corticosteroid versus oral indomethacin therapy. *J Rheumatol* 1986;13:608-13.
- Nicholson G. The effects of passive joint mobilisation on pain and hypomobility associated with adhesive capsulitis of the shoulder. *J Orthop Sports Physiother* 1985;6:238-46.
- Petri M, Dobrow R, Neiman R, Whiting-O'Keefe Q, Seamen WE. Randomised, double-blind, placebo-controlled study of the treatment of the painful shoulder. *Arthritis Rheum* 1987;30:1040-5.
- Adebajo A, 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 tendonitis. *J Rheumatol* 1990;7:1207-10.
- Cohen A, Cohen RW. Short term treatment of acute bursitis of the shoulder. *Pa Med* 1968;1:66-70.
- Hollingsworth GR, Ellis RM, Hattersley TS. Comparison of injection techniques for shoulder pain: results of a double blind, randomised study. *BMJ* 1983;287:1339-41.
- Withrington RH, Girgis FL, Seifert MH. A placebo-controlled trial of steroid injections in the treatment of supraspinatus tendonitis. *Scand J Rheumatol* 1985;14:76-8.
- Leclaire R, Bourgouin J. Electromagnetic treatment of shoulder periarthritis: a randomised controlled trial of the efficiency and tolerance of magnotherapy. *Arch Phys Med Rehabil* 1991;72:284-7.
- Wielandts L, Dequeker J. Double-blind trial comparing fentiazac with phenylbutazone in patients with tendinitis. *Curr Med Res Opin* 1979;6(suppl 2):85-9.
- Duke O, Zecler E, Grahame R. Anti-inflammatory drugs in periarthritis of the shoulder: a double-blind, between-patient, study of naproxen versus indomethacin. *Rheum Rehabil* 1981;20:54-9.
- 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;7:121-6.
- Ginsberg F, Famaey JP. A double-blind comparison of slow-release and standard tablet formulations of fentiazac in the treatment of patients with tendonitis and bursitis. *Curr Med Res Opin* 1985;9:442-8.
- Binder A, Parr G, Hazleman B, Fitton-Jackson S. Pulsed electromagnetic field therapy of persistent rotator cuff tendonitis. *Lancet* 1984;323:695-8.
- England S, Farrell AJ, Coppock JS, Struthers G, Bacon PA. Low power laser therapy of shoulder tendonitis. *Scand J Rheumatol* 1989;18:427-31.
- Smith MD, Thomas D, McCredie M, Brooks PM. Piroxicam versus naproxen in the treatment of painful shoulder. *Pharm* 1986;4:585-9.
- Richardson AT. The painful shoulder. *Proc R Soc Med* 1975;68:11-6.
- Corbeil V, Dussault RG, Leduc BE, Fleury J. Capsulite retractile de l'épaule: étude comparative de l'arthrographie avec corticothérapie intra-articulaire avec ou sans distension capsulaire. *J Can Assoc Radiol* 1992;43:127-30.
- Rizk TE, Pinals RS, Talaiverm AS. Corticosteroid injections in adhesive capsulitis: investigation of their value and site. *Arch Phys Med Rehabil* 1991;72:20-2.
- Ward MC, Kirwan JR, Norris P, Murray N. Paracetamol and diclofenac in the painful shoulder syndrome. *Br J Rheumatol* 1986;25:412-20.
- Gotter G. Comparative evaluation of tenoxicam and piroxicam in the treatment of humeroscapular periarthritis. *Eur J Rheum Inflamm* 1987;9:95-7.
- Thomas D, Williams RA, Smith DS. The frozen shoulder: a review of manipulative treatment. *Rheum Rehabil* 1980;19:173-9.
- Famaey JP, Ginsberg F. Treatment of periarthritis of the shoulder: a comparison of ibuprofen and diclofenac. *J Int Med Res* 1984;12:238-43.
- Blokey NJ, Wright JK, Kellgren JH. Oral cortisone therapy in periarthritis of the shoulder. *BMJ* 1954;1:1455-7.

- 31 Bulgen DY, Binder AI, Hazleman BL, Dutton J, Roberts S. Frozen shoulder: prospective clinical study with an evaluation of three treatment regimes. *Ann Rheum Dis* 1984;43:353-60.
- 32 Jacobs LGH, Barton MAJ, Wallace WA, Ferrous J, Dunn NA, Bossingham DH. Intra-articular distension and steroids in the management of capsulitis of the shoulder. *BMJ* 1991;302:1498-501.
- 33 Rhind V, Downie WW, Bird HA, Wright V, Engler C. Naproxen and indomethacin in periarthritis of the shoulder. *Rheum Rehabil* 1982;21:51-3.
- 34 Binder A, Hazleman BL, Parr G, Roberts S. A controlled study of oral prednisolone in frozen shoulder. *Br J Rheumatol* 1986;25:288-92.
- 35 Van der Windt DAWM, van der Heijden GJMG, Scholten RJPM, Koes BW, Bouter LM. The efficacy of non-steroidal anti-inflammatory drugs (NSAIDs) for shoulder complaints. A systematic review. *J Clin Epidemiol* 1995;48:691-704.
- 36 Van der Heijden GJMG, van der Windt DAWM, de Winter SF. Physiotherapy for patients with soft tissue shoulder disorders: a systematic review of randomised clinical trials. *BMJ* 1997;315:25-30.
- 37 Van der Heijden GJMG, van der Windt DAWM, Kleijnen J, Koes B, Bouter L. Steroid injections for shoulder disorders: a systematic review of randomised clinical trials. *Br J Gen Pract* 1996;46:309-16.
- 38 Goupille P, Sibillia J. Local corticosteroid injections in the treatment of rotator cuff tendinitis (except for frozen shoulder and calcific tendinitis). *Clin Exp Rheumatol* 1996;14:561-6.

(Accepted 23 September 1997)

Understanding controlled trials

What is a patient preference trial?

David J Torgerson, Bonnie Sibbald

This is the third of an occasional series on the methods of randomised controlled trials

National Primary Care Research and Development Centre, Centre for Health Economics, University of York, York YO1 5DD

David Torgerson, senior research fellow

National Primary Care Research and Development Centre, University of Manchester, Manchester M13 9PL

Bonnie Sibbald, reader in health services research

Correspondence to: Dr Torgerson.

BMJ 1998;316:360

A common problem in randomised controlled trials arises when patients (or their clinicians) have such strong treatment preferences that they refuse randomisation.¹ The absence of these patients from trials may restrict generalisation of the results, as participants may not be representative. A further potential source of bias exists when patients with strong treatment preferences are recruited and randomised. When it is not possible to blind patients to their treatment allocation they may suffer resentful demoralisation² if they do not receive their preferred treatment and may comply poorly. On the other hand, patients receiving their preferred treatment may comply better than average. There may therefore be a treatment effect which results from patient preferences and not from therapeutic efficacy. The effects of resentful demoralisation are so far a theoretical concern which have yet to be shown in practice, in part because they are difficult to evaluate.

Patients may be placed in one of three groups according to preference and willingness to be randomised: (a) patients who have no strong preferences and therefore consent to randomisation; (b) patients with a preference who still consent to randomisation; and (c) patients who refuse randomisation and opt for their treatment of choice

To cope with patient preferences the use of a comprehensive cohort design³ or the patient preference trial has been suggested.⁴ Patients with treatment preferences are allowed their desired treatment; those who do not have strong views are randomised conventionally. Hence, in a trial of two interventions, A and B, we end up with four groups: randomised to A; prefer A; randomised to B; prefer B. The analysis of such a trial is uncertain. Any comparison using the non-randomised groups is unreliable because of the presence of unknown and uncontrolled confounders.⁵ At least one analysis should therefore be a comparison between the two randomised arms alone. Analyses which include the unrandomised groups should be treated as observational studies with known confounding factors adjusted for in the analysis. Olschewski and Scheurlen have suggested that an analysis using randomisation status as a covariate might be helpful.³ A further limitation of the patient preference approach is that it may increase the size and cost of trials.

An alternative to the partially randomised approach has been proposed whereby the strength and direction

of patient preferences are elicited before randomisation, with all consenting patients randomised.⁶ This approach combines the advantage of the partially randomised design—that is, gathering information on the effect of preference on outcome—but retains the rigour of a full randomised design.⁶ The design has been used in a randomised trial of physiotherapy treatment for back pain and, despite most patients expressing a preference, no patient refused randomisation.⁶ The practical advantages of establishing and including patient preferences in trials has not been fully established. However, using a patient preference design, Henshaw et al in a comparison of medical versus surgical abortion produced important additional information on the acceptability of the two treatments in different preference groups which would not have been available in the usual trial.⁷ In addition, a recent preference trial of early amniocentesis versus chorionic villus sampling for diagnosing fetal abnormalities showed that rate of pregnancy loss did not differ between the preference group and their randomised equivalent.⁸ This trial is important in that only 38% of patients accepted randomisation. Thus, including the unrandomised patients in the trial offered some reassurance that the results could be extrapolated to a wider group of patients.

Patient preference designs complement, but do not replace randomised trials. However, measuring patient preferences within a fully randomised design deserves further use as this conserves all the advantages of a fully randomised design with the additional benefit of allowing for the interaction between preference and outcome to be assessed.

- 1 Fairhurst K, Dowrick C. Problems with recruitment in a randomized controlled trial of counselling in general practice: causes and implications. *J Health Serv Res Policy* 1996;1:77-80.
- 2 Bradley C. Designing medical and educational intervention studies. *Diabetes Care* 1993;16:509-18.
- 3 Olschewski M, Scheurlen H. Comprehensive cohort study: An alternative to randomised consent design in a breast preservation trial. *Meth Inform Med* 1985;24:131-4.
- 4 Brewin CR, Bradley C. Patient preferences and randomised clinical trials. *BMJ* 1989;299:313-5.
- 5 Silverman WA, Altman DG. Patient preferences and randomised trials. *Lancet* 1996;347:171-4.
- 6 Torgerson DJ, Klaber-Moffett J, Russell IT. Patient preferences in randomised trials: threat or opportunity? *J Health Serv Res Policy* 1996;1:194-7.
- 7 Henshaw RC, Naji SA, Russell IT, Templeton AA. Comparison of medical abortion with surgical vacuum aspiration: women's preferences and acceptability of treatment. *BMJ* 1993;307:714-7.
- 8 Nicolaidis K, de Lourdes Brizot M, Patel F, Snijders R. Comparison of chorionic villus sampling and amniocentesis for fetal karyotyping at 10-13 weeks gestation. *Lancet* 1994;344:435-9.