Review Article

Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache

Stephen Morley[a,*], Christopher Eccleston[b], Amanda Williams[c]

[a]*Division of Psychiatry and Behavioural Sciences, School of Medicine, University of Leeds, Leeds, LS2 9JT, UK*
[b]*Department of Psychology, University of Bath, Bath, UK*
[c]*United Dental and Medical School, University of London, London, UK*

Abstract

A computer and a hand search of the literature recovered 33 papers from which 25 trials suitable for meta-analysis were identified. We compared the effectiveness of cognitive-behavioural treatments with the waiting list control and alternative treatment control conditions. There was a great diversity of measurements which we grouped into domains representing major facets of pain. Effect sizes, corrected for measurement unreliability, were estimated for each domain. When compared with the waiting list control conditions cognitive-behavioural treatments were associated with significant effect sizes on all domains of measurement (median effect size across domains = 0.5). Comparison with alternative active treatments revealed that cognitive-behavioural treatments produced significantly greater changes for the domains of pain experience, cognitive coping and appraisal (positive coping measures), and reduced behavioural expression of pain. Differences on the following domains were not significant; mood/affect (depression and other, non-depression, measures), cognitive coping and appraisal (negative, e.g. catastrophization), and social role functioning. We conclude that active psychological treatments based on the principle of cognitive behavioural therapy are effective. We discuss the results with reference to the complexity and quality of the trials. © 1999 International Association for the Study of Pain. Published by Elsevier Science B.V.

Keywords: Systematic review; Meta-analysis; Chronic pain; Cognitive behavioural therapy

1. Introduction

Behavioural and cognitive treatments for chronic pain have become established in the 30 years since their exposition (Fordyce et al., 1968, 1973; Turk et al., 1983). There are many published open trials of treatment but fewer use control groups in which patients are randomized to treatments. Reviews, however, conclude that there is strong, if not overwhelming evidence for the efficacy of cognitive behavioural therapy (CBT) in restoring function and mood and in reducing pain and disability-related behaviour. Recently, one reviewer regretted that CBT is not provided routinely for chronic pain sufferers rather than medical and physical interventions for which there is less evidence of efficacy (Loeser, 1991). Other overviews of pain management are more critical (Ashburn, 1996). However, to date there has been no systematic review and meta-analysis of randomized controlled trials.

Of the three extant meta-analyses of CBT for chronic pain, one (Malone and Strube, 1988) combined physical and psychological treatment for chronic pain including headache and dental pain; a second (Flor et al., 1992) restricted its scope to psychological treatments and excluded headache; the most recent (Turner, 1996) selected a small sample of randomized controlled trials (RCT) of educational, behavioural and cognitive interventions for chronic low-back pain in the setting of primary care. Both meta-analyses, which included uncontrolled studies, found

* Corresponding author. Tel.: +44-113-233-2733; fax: +44-113-243-3719; e-mail: s.j.morley@leeds.ac.uk

the largest effect sizes for treatment in outcome measures of mood, behaviour and pain ratings, and somewhat smaller ones for drug and health care use. Flor et al. (1992) concluded that: 'overall the results of this meta-analysis provide support for the conclusion that multidisciplinary pain clinics are efficacious. Even at long-term follow-up, patients who are treated in such a setting are functioning better than 75% of a sample that is either untreated or that has been treated by conventional, unimodal treatment approaches' (p. 226). Turner's (1996) findings were consistent with this excepting that the change in mood, in this case depression, was not replicated. This finding may be attributable to a floor effect as patients in her trials were mostly community volunteers and scored low on depression instruments at intake.

In this paper we report a systematic review and meta-analysis of published RCTs of CBT for chronic pain excluding headache. We sought to answer two broad questions: (1) is cognitive behavioural therapy (including behaviour therapy and biofeedback) an effective treatment for chronic pain, i.e. is it 'better' than no treatment? (2) Is cognitive behavioural therapy more effective than alternative active treatments? We chose to exclude headache due to the different emphasis in treatment, both in treatment provision and in outcomes, where pain relief is a much more realistic result of treatment than in other chronic pain. Otherwise, chronic pain was accepted as a label for a heterogeneous group of pain problems in which neither diagnosis, nor site of pain, nor medical findings are an apparent major source of variance in any of the targets of treatment (Turk, 1996; van Tulder et al., 1997). The variety of control conditions found in trials reflects the difficulties in designing suitable controls, e.g. 'inert' controls such as a waiting list can, on ethical and practical grounds, be only short-term and 'active' controls contain an unknown mixture of therapeutic ingredients (O'Leary and Borkovec, 1978; Turner et al., 1994; Schwartz et al., 1997). The comparative treatment groups were, therefore, similarly heterogeneous.

2. Methods

2.1. Search strategy

A search was conducted for published reports of randomized controlled trials of BT and CBT for adults presenting with chronic pain. A priori decisions were made to search only for studies published in full, in peer reviewed journals between 1974 and 1996. Although previous systematic reviews in pain have relied upon Medline (McQuay et al., 1996) it was recognized that the sensitivity of searches using Medline alone has been reported to be low (Adams et al., 1994; Dickersin et al., 1994). Only relevant computer based abstracting services were searched. In order to capture efficiently the maximum number of published trials a three stage plan was chosen (Jadad-Bechara, 1994; NHSCRD, 1996).

A high yield, imprecise, search-term strategy was used. The search strategy contained the word 'pain' and 22 relevant phrases (copy available from authors). Relevant Medline MeSH terms were used (e.g. behaviour therapy). This search strategy had low precision of 0.243% yielding a total of 13 598 articles. Of these 21 were relevant randomized controlled trials and 15 were relevant unrandomized trials.

Four computer abstracting services were selected and their yields compared; Medline, Psychlit, Embase and Social Science Citation Index (SSCI)

Reference lists and bibliographies were searched from all retrieved articles and relevant published reviews. The final list was cross-checked with the PARED database (Jadad et al., 1996a). Twelve additional papers were recovered from searching reference lists. This gave a total sample of 33 papers. Of the 12 that were not found by the search, ten were abstracted on Medline, five on PsychLit, ten on Embase and 12 on SSCI. Of the full set of 33 papers, searching for the specific paper by author and title, Medline has 25 abstracted, Psychlit has 24, Embase has 30 and SSCI also has 30. PARED had recorded 17. Each paper appeared in at least two databases.

SSCI and Embase covered the largest number of journals. Searching for RCTs of psychological therapy in Medline or PsychLit alone did not recover all relevant research reports due, largely, to the omission of specific journals. The 33 papers appeared in 12 journals. Of these four are not regularly abstracted for Medline and three are not regularly abstracted for PsychLit. A three step searching strategy, as employed for this study, is recommended for systematic reviews of psychological therapy.

Papers were read by each of the authors and a consensus decision was taken as to whether the paper contained data suitable for meta-analysis, i.e. contained post treatment means and variances or contrast statistics between two groups (t or F). Where this was not the case we attempted to contact the authors requesting further information about the trial and access to unpublished data. The 33 papers contained data from 30 trials, some papers reported additional or follow-up data. Five trials were excluded from the statistical analysis as the authors were unable to provide data suitable for computing effect size statistics. This left 25 controlled trials for analysis.

2.2. Coding

Development work on the first 20 papers retrieved, which included two papers not entered into the final analysis (Linton and Gotestam, 1984; Linton et al., 1985), generated coding schemes to extract information about the following features of the studies: (1) source of paper (2) the design of study (3) the participants (4) the treatments and (5) the measures and their associated effect sizes (Stock, 1994). Each paper was read to extract data for each coding scheme, i.e. a paper was read five times by each coder during the course of data extraction. Data were extracted by two or

three coders and the reliability of coding was assessed by computing Kappa or percentage agreement for categorical data, and the intraclass correlation for continuous measures. Differences between coders were resolved by consensus. As a large number of features were coded we report coding reliability data, only where necessary. Overall reliability was high.

2.3. Extracting reliability data for study measures

Our choice of meta-analytic strategy (Hunter and Schmidt, 1990) required estimates of the reliability of outcome measures for computing effect size estimates. We generated a list of all the outcome measures used in the studies and sought information about the reliability of each measure. We obtained information from a variety of sources; the study paper, references to measures contained therein, published test manuals, and unpublished data were obtained by contacting authors. In preference we used measures of test stability (test–retest) as the reliability estimate. Where this was not available we used measures of internal consistency (Cronbach's alpha) or inter-rater reliability (Kappa).

2.4. Effect size (ES) computations

We estimated the effect size using Hedges's g (Hedges and Olkin, 1985). The sign of the result was adjusted so that improvements on every measure were denoted as positive. Where g could not be computed directly from means and standard deviations given in the source paper we computed it indirectly from the available test statistics, e.g. t, using the formula of Rosenthal (1994). The estimates of g were corrected for small sample bias (Hedges and Olkin, 1985; p.79, Eq. 7) prior to further analysis. In one study (Kerns et al., 1986) outcomes were presented as z scores standardized on pre- and post-treatment data. Rather than eliminate the study from consideration we computed a proxy estimate of the effect size by calculating the difference between the z scores for the treatment and control groups. This was not corrected for bias as the distribution is not known.

2.5. Analytic rational and methods

We used the meta-analysis psychometric method of Hunter and Schmidt (1990) which assumes that the computed ES is an estimate with an associated error from which a confidence interval can be estimated. Hunter and Schmidt (1990) have provided a series of algorithms for estimating the ES and its associated variance including corrections for variations in the reliability in the dependent variable; which if uncorrected will cause variation in the ES estimate beyond the variation due to sampling error. The analytic strategy was therefore: (1) to estimate ESs correcting for measurement artefact (2) to estimate mean ES over the domain of interest and test the hypothesis that variation is

due to statistical artefacts (3). Finally, if the hypothesis that $ES > 0$ cannot be rejected, to investigate the influence of study characteristics (other than those involved in measurement artefact) by disaggregation of the sample into subsamples with shared characteristics. This step is not without problems because as a sample is disaggregated the sizes of the sub-samples may become too small to yield robust estimates.

2.5.1. Comparisons: decisions concerning the multivariate nature of the data

For statistical purposes the ideal meta-analysis would be conducted on a single common measure of interest, e.g. pain intensity, or behavioural activity, extracted from every relevant study. Furthermore, each study would contribute only one effect size derived from a comparison between a single well specified treatment and a control. The current data set met neither of these criteria as most studies had both multiple measures and more than one treatment arm.

2.5.2. Multiple measures

Conducting an analysis in which all outcome measures with computed effect sizes are entered presents problems of bias and independence of measures. A composite effect size may be generated by estimating a mean ES for each study and methods for multivariate solutions of this problem are available (Raudenbusch et al., 1988; Gleser and Olkin, 1994). These methods require information about the correlations between measures and moderately large sample sizes. As these conditions were not uniformly met in the current data set we considered another, statistically simpler, solution.

On the basis of previous reviews and papers (Malone and Strube, 1988; Flor et al., 1992; Gatchel and Turk, 1996) we hypothesized that treatment outcomes would be differentially effective over different measurement domains and conducted separate analysis several domains of measurement. We identified the following domains from our knowledge of the literature and detailed cataloguing of all the measures used in the trials: pain experience; mood/affect; cognitive-coping and appraisal; pain behaviour; social role performance; biological and physical fitness measures; use of health care services; miscellaneous. Definitions of each of these are given in Table 1. The data extraction protocol enabled the assessment of the interrater agreement for assigning measures to domains. This is also given in Table 1. Although data were extracted on use of health care, biological and miscellaneous domains there were too few ESs to merit analysis. These were, therefore, excluded from the comparative analysis of treatment.

As many studies used more than one outcome measure in a given domain. We chose an analytic strategy which selected one measure from each study using the following criteria: select the most frequently occurring measure across studies, e.g. the Beck depression inventory (BDI) in preference to other measures of depression; select multi-item

Table 1

Measurement domains, inter rater agreement for allocating measures to domain and example measures

Domain	% Agree	Example measure
Pain experience: Measures of subjective pain experience captured by ratings of intensity, sensation and unpleasantness	97.5%	McGill pain questionnaire; visual analogue scales of intensity; composite diary measure of numerical ratings
Mood/affect: Primary measure of mood or affective state, but not a trait assessment (these measures were subdivided into assessment of depression and measures of other affective states)	94%	Beck depression inventory; CES-D; STAI-S
Cognitive coping and appraisal: Reports of cognitive strategies and appraisals used in attempts to manage pain (these measures were subdivided into: negative coping, measures known to be correlated with poor adjustment; and positive coping, measures associated with good adjustment)	91%	Coping strategies questionnaire and subscales, e.g. catastrophization, passive coping, active coping
Pain behaviour: Overt behavioural acts associated with pain. There were two subcategories: pain behaviour, referring to behaviour which apparently signals the presence of pain; and activity level, such as distance walked	60%	Pain behaviour, direct observational system; grimacing, guarding, bracing; activity level, distance walked
Biology/physical fitness: Assessment of biological function and physical fitness, but not including measures of behavioural activity as in previous category	85%	Vo_{max}, joint flexibility
Social role functioning: Assessments of the impact of pain on the ability of the person to function in a variety of social roles: work, leisure, marital and family	83%	Sickness impact profile, MPI-Interference, ratings of interference (VAS)
Use of health care system: Use of health care facilities, including clinic visits and drug use	100%	Outpatient medical visits, drug use
Miscellaneous: All other measures not categorized in previous categories	67%	Pain drawings, personality measures, repertory grids

measures in preference to single item (e.g. McGill pain questionnaire (MPQ) in preference to a Visual analogue scale (VAS)), since they are likely to be more reliable; select measures with known reliability coefficients wherever possible.

2.5.3. Multi-armed trials

Many (21 out of 25) studies compared more than one treatment with a control, and there were a variety of control groups used. This presented two issues to be considered: classification and combination of treatment groups and choice of comparison group for estimating ESs.

We considered pooling treatment effect sizes within studies to estimate a study effect size, but rejected this option because there is an expectation in the literature that different treatments may produce different outcomes. We, therefore, estimated overall treatment impact by including all treatment arms within a trial; acknowledging that the ES estimates in this comparison are not independent, as those drawn from a single trial will have a common control condition. We anticipated that further analyses might be possible by estimating the mean ES for treatments with common ingredients. Coding the details of treatments reported in the papers revealed wide variation between treatments described with a generic term, e.g. cognitive therapy, but there was marked variability between studies in the detail provided. We categorized the treatments into three primary types based on a consensus judgement of therapy derived from the source paper. Subcategories for several types were also coded. Details of these are given in Table 2.

We identified two classes of control group. (1) Waiting list control (WLC), where no 'new' treatment was prescribed, although the possibility that WLC patients obtained 'some' treatment, e.g. continued medication, private visits

Table 2

Treatment types

Type	Definition
Biofeedback and relaxation	Use of biofeedback and/or a form of relaxation
Behaviour therapy	Managed approach to behavioural change using the basic concepts and principles of operant psychology
Cognitive behaviour therapy	Primary focus on changing cognitive activity to achieve changes in behaviour, thought and emotion. We identified two broad groupings; coping skills training (CST) and cognitive therapy (CT)
	CST focuses on inculcating improved cognitively mediated coping skills
	CT contains CST but with additional component of Beck's cognitive therapy

to other therapists, cannot be excluded. (2) Treatment control (TC), in which a participant was allocated to a 'new' treatment for the duration of the trial. The TC conditions comprised an heterogeneous collection of treatments, including access to regular treatment provided in a pain clinic, physiotherapy, occupational therapy, and the provision of a standard educational and advice package, particularly associated with rheumatoid arthritis (Lorig, 1982). We conducted two main analyses; first a comparison of CBT and BT treatments with WLC, and a second comparison with TC. Studies also used more than one potential control group and as a result of this some studies contributed data to both sets of comparisons, i.e. treatment versus WLC and TC. Table 3 lists the studies, the treatments and the codings we allocated on the basis of which the comparisons were made.

3. Results

We report summary statistics describing the trials entered into the meta-analysis before reporting details of the effect sizes.

3.1. Trial design

Of the 25 trials suitable for meta-analysis only four (16%) provided explicit and replicable information about the method of randomization (Altman, 1996). In the remaining 21 trials the fact that randomization had occurred was simply stated in the methods section or elsewhere in the text or title. Information about the randomization procedure did not include details of whether the randomization was independent of the trialists. Nineteen trials appeared to be true randomized trials and six trials used some form of pseudo-randomization, e.g. by time period. We rated the explicitness of exclusion and inclusion criteria; seven trials gave explicit replicable exclusion criteria while 16 gave explicit inclusion criteria. Only two trials reported a priori power calculations, and four reported post hoc calculations. Eighteen trials used samples of convenience from a specified source, e.g. rehabilitation and pain clinics; two trials recruited consecutive referrals to a clinic, and information was not reported in five trials.

3.2. Participants

Only nine trials reported details of the sample size from which patients were selected, i.e. number of referrals to the trial prior to selection. When all 25 trials are considered 1672 patients were entered into the trials, 38% male, 62% female. The average age (unweighted by number in the trials) was $M = 48.35$ (SD between trials = 7.19), and the mean chronicity of the samples was 12.27 years (SD between trials = 7.47). The average number of patients entered into a trial was 67 (SD = 30.21, range = 18 to 131); the average number of subjects at the end of treatment = 57 (SD = 25.38, range = 18 to 112), giving a crude estimate of drop out rate of 14%. The primary diagnostic labels reported for the patient groups were: chronic low-back pain (36%); rheumatoid arthritis (20%); mixed, predominantly back pain (16%); osteo-arthritis (8%); upper limb pain (8%); fibromyalgia (4%); unspecified (8%).

3.3. Treatments

The modal and median number of treatment arms in the trials was 3 (14/25 trials). There were five trials each with two and four treatment arms, and one trial had six treatments. Treatment was typically delivered in groups (76%), with 20% of treatments delivered as a combination of group and individual therapy. Treatment mode was unspecified in 4% of trials. The mean treatment duration was 6.74 weeks (SD = 2.32, range = 3 to 10 weeks), and the median number of hours in treatment was 16 (range = 6 to 90; interquartile range 10 to 18 h). Sixty percent of therapists were either specifically trained for the trial or were reported as having a general training in CBT and pain. Details were not available in 24% of trials, and the remaining 16% used therapists with general training (i.e. CBT not mentioned). Twenty percent of trials used graduate students (clinical psychologist in training) as therapists; 32% used professionals qualified for more than 5 years; 20% of trials used experienced therapists drawn from several disciplines; and no details were provided in 28% of trials. Eight trials (32%) reported providing regular or some supervision given to therapists during the course of the trial (68% no details given). Only 40% reported making checks on adherence to treatment protocols. Nine trials (36%) reported that treatment was fully manualised; four (16%) referred to partial manualization; and the remaining trials (48%) reported no manualization or a general reference to a text such as Turk et al. (1983). Patients' pre-treatment expectations and the credibility of treatments were assessed in ten trials (40%) but not reported in the remaining trials. In 16 trials (64%) the therapists and treatment delivery were not confounded, i.e. each therapist delivered all treatments. Information on therapists' allegiance to therapeutic mode was not provided.

3.4. Measures

The 25 trials reported total of 221 outcome measures for which effect sizes were computable, an average of 9.21 per trial (SD = 3.59, range = 4 to 16). The majority of these outcomes were patient self-ratings (77.4%); 11% were observations made by a researcher blind to the treatment condition; 6% were made by a non-blinded researcher or therapist; and 5% were made by a spouse or family member. The outcomes were not equally distributed amongst the domains of measurement. The frequencies and percentages are shown in Table 4. The assessment of the use of health care system, biological and fitness indices were relatively

Table 3

Details of studies entered into the meta-analysis. Treatment names as given by the authors and coding used in the study

Authors	Randomization quality	Treatments and code	Patient group, location,sample size and randomization quality
Altmaier et al. (1992)	1	3.4 Psychological treatment + routine care	Low-back pain: inpatient, $n = 45$
		4.7 Control: routine care	
Appelbaum et al. (1988)	2	3.4 Cognitive behavioural treatment	Rheumatoid arthritis: outpatients, $n = 18$
		4.7 Control: symptom monitoring	
Bradley et al. (1987)	1	1.3 Cognitive behavioural treatment	Rheumatoid arthritis: outpatient, $n = 68$
		(BFB + relax + goal setting)	
		4.7 Structured group social support therapy	
		4.7 Control: no adjunct treatment	
Flor and Birbaumer (1993)	1	1.1 EMG biofeedback	Chronic musculoskeletal pain (low-back,
		3.4 Cognitive behaviour therapy	temporomandibular): outpatients, $n = 78$
		4.7 Medical	
Keefe et al. (1990a,b)	1	3.5 Pain coping skills	Osteoarthritis (knee): outpatients, $n = 99$
		4.6 Education	
		4.7 Standard routine care	
Keefe et al. (1996)	1	3.5 Spouse assisted coping skills	Osteoarthritis (knee): outpatients, $n = 88$
		3.5 Coping skills training	
		4.6 Education-spouse support	
Kerns et al. (1986)	1	3.4 Cognitive behaviour therapy	Chronic pain – mixed (low back, neck, RSD,
		2;- Behavioural therapy	PHN, rheumatic disease, musculoskeletal):
		4.8 Wait list control	outpatient, $n = 28$
Kraaimaat et al. (1995)	2	3.4 Cognitive behaviour therapy	Rheumatoid arthritis: outpatients, $n = 77$
		4.7 Occupational therapy	
		4.8 Wait list control	
Moore and Chaney (1985)	2	3.4 Cognitive behaviour therapy (couples)	Chronic pain – mixed (low-back, arm,
		3.4 CBT (individuals)	knee, phantom limb): outpatients, $n = 43$
		4.8 Wait list control	
Newton-John et al. (1995)	3	1.1 EMG biofeedback	Low-back: outpatient, $n = 44$
		3.5 Cognitive behaviour therapy	
		4.8 Wait list control	
Nicholas et al. (1992)	1	3.5 Cognitive behavioural + physiotherapy	Low-back; outpatients, $n = 20$
		4.7 Physiotherapy	
Nicholas et al. (1991)	1	3.5 Cognitive treatment	Low-back: outpatient, $n = 58$
		3.5 Cognitive treatment + relaxation	
		2.- Behavioural treatment	
		2.- Behavioural treatment + relaxation	
		4.7 Attention control + physiotherapy	
		4.7 No attention control + physiotherapy	
O'Leary et al. (1988)	2	3.4 Cognitive behavioural therapy + bibliotherapy	Rheumatoid arthritis: outpatient, $n = 33$
		4.7 Bibliotherapy	
Peters and Large (1990)	1	3.5 Inpatient pain management	Chronic pain – mixed (back, arms, head,
		3.4 Outpatient pain management	legs, chest): in- and outpatient, $n = 68$
		4.7 Control (standard care allowed)	
Puder (1988)	1	3.4 Cognitive behavioural treatment (SIT)	Chronic pain – mixed: outpatient, $n = 69$
		4.8 Waiting list control	
Radojevic et al. (1992)	1	3.4 Behavioural treatment (CBT) with family support	Rheumatoid arthritis: outpatient, $n = 59$
		3.4 Behavioural treatment (CBT)	
		4.7 Education family support	
		4.7 No treatment control	
Spence (1989,1991)	1	3.5 Individual cognitive behavior therapy	Upper limb (work related): outpatient,
		3.5 Group CBT	$n = 45$
		4.8 Waiting list control	
Spence et al. (1995)	1	1.1 EMG biofeedback	Musculoskeletal – cervicobrachial
		1.2 Applied relaxation	(work related): outpatients, $n = 48$
		1.3 Combined EMG + relaxation	
		4.8 Waiting list control	
Turner (1982)	2	3.4 Cognitive behaviour therapy	Low-back pain: outpatients, $n = 36$
		1.2 Progressive relaxation	
		4.8 Waiting list control	

Table 3 *continued*

Details of studies entered into the meta-analysis. Treatment names as given by the authors and coding used in the study

Authors	Randomization quality	Treatments and code	Patient group, location, sample size and randomization quality
Turner and Clancy (1988)	1	2.- Operant behavioural 3.4 Cognitive behavioural 4.8 Waiting list control	Low-back pain: outpatient, $n = 81$
Turner and Jensen (1993)	1	1.2 Relaxation 3.5 Cognitive therapy 3.5 Cognitive therapy + relaxation 4.8 Wait list control	Low-back pain: outpatient, $n = 102$
Turner et al. (1990)	1	2.- Behavioural therapy + exercise 2.- Behavioural therapy 4.7 Exercise 4.8 Wait list control	Low-back pain: outpatient: $n = 96$
Vlaeyen et al. (1995)	3	2.- Operant treatment 3.4 Cognitive treatment 1.1 Respondent treatment 4.8 Wait list control	Low-back pain: outpatients, $n = 71$
Vlaeyen et al. (1996)	1	3.4 Combined cognitive/educational 4.6 Attention control (education) 4.8 Waiting list control	Fibromyalgia: outpatients, $n = 131$
Williams et al. (1996)	1	3.5 Inpatient cognitive behavioural 3.5 Outpatient cognitive behavioural 4.8 Waiting list control	Chronic pain – mixed: in- and outpatients, $n = 121$
Linton and Gotestam, (1984)[a]	–	–	–
Linton et al, (1985)[a]	–	–	–
Parker et al., (1988)[a]	–	–	–
Peters et al., (1992)[a]	–	–	–
Pilowsky et al., (1995)[a]	–	–	–
Strauss et al, (1986)[a]	–	–	–

[a]References that were retrieved in the literature search but did not contain useable data.
Randomization quality: 1, random assignment; 2, random assignment of matched pairs or counter balanced by explicit criterion; 3, random assignment compromised. *Treatment*: *first digit*; 1, biofeedback and relaxation (Bfb); 2, behaviour therapy; 3, cognitive therapy; 4, control group. *Second digit*: 1, biofeedback; 2, relaxation; 3, 1 and 2; 4, coping skills training (after Turk et al., 1983); 5, cognitive restructuring (after Beck et al., 1979); 6, education/bibliotherapy; 7, active treatment (treatment as usual – TAU); 8, waiting list control. *Patient group*: n, total number treated in the trial.

under-sampled. Table 4 also shows the numbers of trials which sampled each domain and the average number of measures taken per trial.

3.5. Effect sizes

Inspection of Table 4 reveals that three domains, biological, use of health care system, and miscellaneous, were sampled by very few trials. We, therefore, did not compute ESs for these domains. Inspection of the measures used in other domains led us to consider subdividing three of them on the basis of the measures within them. In domain 2 (mood/affect) there was a clear division between measures of depression (BDI, CES-D) and measures of other affective states, predominantly anxiety (STAI-S). Domain 3 (cognitive appraisal and coping) contained measures which might broadly be defined as 'negative', i.e. related to poor adjustment (catastrophizing, passive coping), and 'positive', i.e. related to good adjustment (active coping). This basic conceptual division has been substantiated in the coping literature and we, therefore, conducted separate analyses on the two components. Finally, we noted that in domain 4 (beha-

vioural) the measures broadly tap two components of pain behaviour: (1) the behavioural expression of pain, as indicated by postural adjustments and para-vocalisations, and assessed by measures such as Keefe's pain behaviour observation system (Keefe and Block, 1982) (2) increasing activity levels, usually measured by self report, e.g. MPI-Activity (Kerns et al., 1985). Successful treatment is expected to decrease the overt expression of pain and increase behavioural activity. We therefore, divided the pain behaviour domain to reflect these differences.

3.5.1. Treatment versus waiting list control

Table 5 displays the results for the comparisons between the all treatments and the waiting list control conditions. The left side of the table shows the number of comparisons (*n*) contributing to the estimated ES and the estimates of mean effect size, weighted by the sample sizes of each contributing comparison, and corrected for unreliability in the measurements for each measurement domain. The homogeneity of each set of ESs was also computed with one or two exceptions the application of the Hunter and Schmidt (1990) '75% rule' indicated that the samples

Table 4

Distribution of outcome measures in the 25 trials entered into the meta-analysis and the percentage of trials contributing to each domain. The mean number of measures per trial (column 3) is calculated only for those trials which contributed a measure in the relevant domain

Domain	Number of trials sampling the domain	Mean number of measures per trial	SD	Range
Pain experience	25 (100%)	1.64	1.20	1–7
Mood/affect	22 (88%)	1.64	0.77	1–3
Cognitive coping and appraisal	17 (68%)	2.34	1.16	1–5
Behavioural activity	17 (68%)	1.88	1.23	1–4
Biological	9 (36%)%	2.89	2.28	1–9
Social role functioning	19 (76%)	2.00	1.12	1–4
Use of health care system	3 (12%)	1.00	0.00	0
Miscellaneous	5 (20%)	1.40	0.49	1–2

were heterogeneous. We, therefore, decided to report all the data on the assumption that the ESs are heterogeneous[1]. The reported 95% confidence intervals in the both Tables 5 and 6 were calculated on this assumption as are the z values: z values ≥ 1.96 indicate that the mean ES is significantly greater than 0 at the conventional 5% (two-tailed) level, i.e. the null hypothesis that treatment is no more efficacious than the WLC condition is rejected. Without exception this hypothesis was rejected for all measurement domains. The median value of the ES for the measurement domains shown in Table 5 is 0.5, i.e. patients in receipt of treatment are, on average, improved by half a standard deviation relative to those assigned to WLC conditions.

The right side of Table 5 shows the same statistics for sub-groups of treatment types. The null hypothesis tested is that the particular treatment is no more efficacious than the WLC condition: no comparisons between treatment type are made. The number of treatment versus control comparisons, on which each of the ESs is made, is variable with most data being due to the CBT subgroup. CBT is more efficacious than the WLC control conditions for all measurement domains except the expression of pain behaviour. There were relatively fewer comparisons between behaviour therapy and WLC conditions and the estimates of mean ES are based on smaller samples. There were ES > 0 for the domains of pain experience, mood/affect (other than depression), social role functioning (reduced interference) and most markedly for the expression of pain behaviour. The number of comparisons between biofeedback and relaxation treatments and WLC conditions was also small. There were ES > 0 for pain experience, mood/affect (depression), positive and negative coping, and social role functioning.

All three types of treatment are effective in changing pain experience, i.e. reducing pain intensity, improving social role functioning, and (accepting the single behaviour therapy comparison available) in reducing negative appraisal and coping (predominantly catastrophization).

3.5.2. Treatment versus treatment control

Summary statistics for the comparisons between treatments and active treatment controls are shown in Table 6, which has the same format as Table 5. Altogether there were fewer comparisons between treatments and ATC conditions and the majority of these comprised CBT treatments. When the overall (left side of Table 6) mean ESs are estimated, treatments are reliably more effective (ES > 0) than ATC conditions for the domains of pain experience, cognitive coping and appraisal (increasing positive coping), and pain behaviour (reducing expression of pain). There was no effect of treatment on the other domains, although it should be noted that no data were available to estimate an ES for the increasing activity component of pain behaviour.

When treatment subtypes are considered the results for the largest group, CBT correspond to the findings for the overall estimate. This is not surprising given that CBT contributes most to the overall estimate. The ES estimates for the small number of behaviour therapy comparisons are generally not >0, with two notable exceptions, a reduction in the expression of pain behaviour, and an improvement in social role functioning. The latter is notable since the estimate of the overall ES $= 0$.

4. Discussion

4.1. Resume

In answer to the two questions addressed by this study, we conclude that active psychological treatments based on the principles of cognitive-behavioural therapy (including behaviour therapy and biofeedback) are effective relative to waiting list control conditions. CBT produced significant changes in measures of pain experience, mood/affect, cog-

---

[1] The decision to regard the ESs as heterogeneous seemed prudent given that most of the analyses indicated heterogeneity, and that where homogeneity was indicated it might have been attributable to the fact that estimates were based on samples in which individual ESs were drawn from the same study. The effect of the assumption is to increase the confidence interval which is tantamount to increasing the probability of a Type II error. We note that in no case, where homogeneity was indicated, did the assumption of heterogeneity change the significance of the result.

Table 5

Effect Sizes for treatments versus waiting list controls. See text for explanation of table. Figures in parenthesis in the penultimate column (95%CI) are the standard errors of a single effect size. CBT, cognitive behavioural therapy, BT, behavior therapy, BFB, biofeedback

Domain	Overall					Sub groups			
	n	Mean ES	95% CI	z			n	Mean ES	95% CI	z
Pain experience	28	0.40	0.22–0.58	4.28	CBT	16	0.33	0.09–0.57	2.78
					BT	5	0.32	−0.09–0.55	2.73
					BFB	7	0.74	0.28–1.20	3.17
Mood/affect depression	24	0.36	0.13–0.59	3.11	CBT	13	0.38	0.07–0.69	2.43
					BT	4	−0.03	−0.21–0.15	−0.33
					BFB	7	0.74	0.28–1.20	3.17
Mood/affect other	16	0.52	0.19–0.84	3.10	CBT	9	0.41	0.00–0.82	1.96
					BT	2	0.74	0.41–1.08	4.34
					BFB	5	0.71	−0.01–1.43	1.94
Cognitive coping and appraisal negative	16	0.50	0.27–0.73	4.20	CBT	8	0.41	0.08–0.73	2.44
					BT	1	1.41	[(±0.41)]	(3.79)
					BFB	7	0.52	0.29–0.76	4.47
Cognitive coping and appraisal positive	11	0.53	0.28–0.78	4.20	CBT	8	0.58	0.28–0.89	3.72
					BT	1	0.56	[(±0.37)]	(1.51)
					BFB	2	0.17	0.03–0.32	2.41
Behaviour expression	12	0.50	0.22–0.78	3.49	CBT	5	0.49	−0.08–1.05	1.68
					BT	5	0.45	0.31–0.59	6.28
					BFB	2	0.71	−0.03–1.45	1.89
Behaviour activity	14	0.46	0.25–0.72	4.34	CBT	7	0.48	0.20–0.77	3.31
					BT	2	0.54	0.28–0.79	4.12
					BFB	5	0.39	−0.03–0.80	1.81
Social role functioning (social role interference)	25	0.60	0.44–0.76	7.28	CBT	15	0.61	0.39–0.84	5.35
					BT	4	0.34	0.17–0.51	3.90
					BFB	6	0.85	0.58–1.31	6.05

nitive coping and appraisal (reduction of negative coping and increase in positive coping), pain behaviour and activity level, and social role function. When compared across the same range of outcomes with other treatments or control conditions, the efficacy of CBT is of a smaller size and limited to the outcomes of pain experience, positive coping and social role function. The overall effect sizes in the order of 0.5, is concordant with those from larger meta-analyses for psychological treatments for a variety of disorders (Shadish et al., 1997), and our conclusion is similar to that reached by Compas et al. (1998) in a narrative review of selected studies.

How might our conclusion be affected by our methods? We identify three areas where we made clear decisions about the treatment of the data: exclusion of unpublished trials, treatment of study measures, and not weighting studies by quality. (1) *Published trials*: the use of only published trials assumes that no unpublished trials would qualify for inclusion, but given the liberality of inclusion criteria in this review, that may well be unfounded. However, there would need to be many such trials, or ones with large samples and effects, to make a significant difference (Chalmers, 1991). Like reviewers in many other fields, we judged this to be unlikely, but the 'amnesty' for trials is wholeheartedly welcomed as they will provide a more satisfactory basis for such judgements. (2) *Measurement*: We attempted to gain control of the variability in the measure-

ments by two means: we corrected for unreliability in the measures, and we grouped measures into reliably defined domains. We recognise that neither of these procedures is perfect. The correction for reliability was dependent on the availability of published coefficients and the degree to which these coefficients may be generalized to the samples is not always known. Our decision to analyze the outcome measures by treatment domains was pragmatic. Clearly investigators expect changes in conceptually distinct areas of measurement, which we believe are reflected in the domains used in this analysis. (3) *Weighting trials*: It is unusual to reject the trial weighting approach in the pain field since the quality of medical trials in pain is known to be associated with the likelihood of finding a positive effect (Jadad et al., 1996b). However, the practice is by no means universally endorsed (Egger and Davey Smith, 1997): the judgements of quality are necessarily subjective and the weightings arbitrary (Thompson and Pocock, 1991; Egger et al., 1997). Although there are excellent guidelines (Altman, 1996), there is no 'gold standard' (Chalmers, 1991). We decided instead to be catholic in our criteria for study inclusion and conservative in the statistical treatment of their results.

4.2. Comment on the quality of trials

Arguably most trials were statistically under powered.

Table 6

Effect sizes for treatments versus treatment controls. Details of the table are given in the text

Domain	Overall					Sub groups			
	n	Mean ES	95% CI	z			n	Mean ES	95% CI	z
Pain experience	22	0.29	0.11–0.46	3.21	CBT	17	0.26	0.05–0.47	2.45
					BT	4	0.33	−0.04–0.71	1.70
					BFB	1	0.52	[(±0.28)]	(1.83)
Mood/affect depression	15	−0.14	−0.32–0.04	−1.52	CBT	11	−0.14	−0.36–0.08	−1.22
					BT	4	−0.14	−0.38–0.11	−1.07
					BFB	0	–	–	–
Mood/affect other	16	0.05	−0.27–0.37	−0.30	CBT	13	0.01	−0.34–0.36	0.05
					BT	2	0.62	−0.55–1.79	1.03
					BFB	1	0.15	[(±0.28)]	(0.54)
Cognitive coping and appraisal (negative)	14	0.17	−0.08–0.42	1.35	CBT	11	0.09	−0.18–0.35	0.63
					BT	2	0.53	−0.16–1.22	1.53
					BFB	1	0.54	[(±0.28)]	(1.92)
Cognitive coping and appraisal (positive)	15	0.40	0.21–0.62	3.60	CBT	12	0.55	0.38–0.72	6.57
					BT	2	−0.13	−0.55–0.29	−0.61
					BFB	1	−0.40	[(±0.28)]	(−1.42)
Behaviour expression	11	0.27	0.08–0.47	2.76	CBT	8	0.31	0.05–0.56	2.36
					BT	2	0.06	0.02–0.10	3.11
					BFB	1	0.34	(±0.28)	(1.20)
Behaviour activity (no data)	–	–	–	–	–	–	–	–	–
Social role functioning (social role interference)	14	0.17	−0.08–0.34	1.62	CBT	10	0.10	−0.15–0.35	0.76
					BT	4	0.37	0.17–0.58	3.61
					BFB	0	–	–	–

This is unsurprising given the demands of delivering a complex multicomponent treatment with sufficient consistency for large numbers of patients over a prolonged period of time. On the other hand, some trials might be regarded as over-complex with multiple treatment and control groups. In future trialists might consider the merits of simple two armed trials with sufficient numbers comparing a treatment with suitable control. The content of and differentiation of control groups from treatment requires careful consideration (Schwartz et al., 1997). Patients assigned to a waiting list in one trial may continue to receive existing treatments (such as physical therapy or pharmacotherapy) which may be equivalent to the treatment control in another trial. Unless expectations of efficacy are monitored, it would be invidious to make assumptions about equivalence in terms of patients' experience. There was also variability within the class of treatment controls from continuing previously ineffective treatments to starting a treatment with demonstrated benefit, such as arthritis education (e.g. Lorig, 1982; Keefe et al., 1996). The distinction between the content of active treatment and control condition can also be a fine one. We surmize that being allocated to a control condition would have different psychological consequences to being allocated to an active treatment, even though that treatment is based on predominantly non-psychological principles, e.g. physical therapy. In addition long-term comparison of treatment and control groups was rendered difficult by the use of waiting list controls. Patients in these groups were

commonly entered into an active treatment group or drop out of the trial.

When treatments were considered, across both comparisons (WLC and TC), there was considerable variability in quality and quantity of treatment as reported in the results. While some authors gave explicit accounts of the treatment procedures with reference to manualized interventions which were appropriately monitored, this was not universally so. It is possible that expediency and economy of reporting is a product of external pressures (e.g. editorial demands), but this does not account for what appear to be rather brief interventions delivered by relatively inexperienced therapists to chronically distressed patients for any realistic expectation of change to take place. In addition we note that the measurement of process variables, such as patients' expectations of change, adherence to treatment methods, and therapist competence, were generally lacking. In comparison with best practice in the psychotherapy outcome literature the design and implementation of psychological treatment trials for chronic pain has considerable scope for development (Kazdin, 1994; Lambert and Hill, 1994).

Our analysis of outcome measures revealed a lack in some domains which have economic importance and are of concern to health service managers, third party payers and patients themselves. Data were notably sparse on health service use, drug intake, uptake of additional treatment, and change in work and occupational status as a consequence of

treatment. The over reliance on self report measures is also notable. While many of these measures are psychometrically reliable the extent to which they are influenced by measurement reactivity is often unknown. We note that this feature is not confined to the field of pain (Smith et al., 1980) and while many psychological states can only be measured through self report the development of robust measures of direct observation or independent blind assessors would be beneficial. Kaplan (1990) has argued persuasively for the desirability of behavioural outcomes in health care trials.

4.3. Conclusion

Published randomized controlled trials provide good evidence for the effectiveness of cognitive behavioural therapy and behaviour therapy for chronic pain in adults. This systematic review raised methodological issues which should be considered in the design of future trials. Psychological treatment of chronic pain is complex, lengthy and variable, outcomes cannot be easily dichotomized, and it is rarely possible to blind patients and therapists to treatment conditions. We see the comments, criticisms and questions which arise from our review as a cause for optimism and we hope provide material for debate.

Acknowledgements

References

Adams, C.E., Power, A., Frederick, K. and Lefebvre, C., An investigation of the adequacy of medline searches for randomized controlled trials (RCTs) of the effects of mental-health-care, Psychol. Med., 24 (1994) 741–748.

Altman, D.G., Better reporting of randomized controlled trials: the CONSORT statement, Br. Med. J., 313 (1996) 570–571.

Altmaier, E.M., Lehmann, T.R., Russell, D.W., Weinstein, J.N. and Kao, C.F., The effectiveness of psychological interventions for the rehabilitation of low-back pain: a randomized controlled trial evaluation, Pain, 49 (1992) 329–335.

Appelbaum, K.A., Blanchard, E.B., Hickling, E.J. and Alfonso, M., Cognitive behavioral treatment of a veteran population with moderate to severe rheumatoid arthritis, Behav. Ther., 19 (1988) 489–502.

Ashburn, M.A., Interdisciplinary pain management: center of excellence or home of the dinosaur?, Clin. J. Pain, 12 (1996) 258–259.

Beck, A.T., Rush, A.J., Shaw, B.F. and Emery, G., Cognitive Therapy of Depression, Guildford Press, New York, 1979.

Chalmers, T.C., Problems induced by meta-analyses, Stat. Med., 10 (1991) 971–980.

Bradley, L.A., Young, L.D., Anderson, K.O., Turner, R.A., Agudelo, C.A., McDaniel, L.K., Pisko, E.J., Semble, E.L. and Morgan, T.M., Effects of psychological therapy on pain behavior of rheumatoid arthritis patients, Treatment outcome and 6-month follow-up, Arth. Rheum., 30 (1987) 1105–1114.

Compas, B.E., Haaga, D.A.F., Keefe, F.J., Leitenberg, H. and Williams, D.A., A sampling of empirically supported treatments from health psychology: smoking, chronic pain, cancer and bulimia nervosa. J. Consult. Clin. Psychol., 66 (1999) 89–112.

Dickersin, K., Scherer, R. and Lefebvre, C., Identifying relevant studies for systematic reviews, Br. Med. J., 309 (1994) 1286–1291.

Egger, M. and Davey Smith, G., Meta-analysis: potentials and promise, Br. Med. J., 315 (1997) 1371–1374.

Egger, M., Davey Smith, G. and Phillips, A., Meta-analysis: principles and procedures, Br. Med. J., 315 (1997) 1533–1537.

Flor, H., Fydrich, T. and Turk, D.C., Efficacy of multidisciplinary pain treatment centers: a meta-analytic review, Pain, 49 (1992) 221–230.

Flor, H. and Birbaumer, N., Comparison of the efficacy of electromyographic biofeedback, cognitive-behavioral therapy, and conservative medical interventions in the treatment of chronic musculoskeletal pain, J. Consult. Clin. Psychol., 61 (1993) 653–658.

Fordyce, W.E., Fowler, R.S., Lehmann, J.F. and DeLateur, B.J., Some implications of learning on problems of chronic pain, J. Chronic Dis., 21 (1968) 179–190.

Fordyce, W.E., Fowler, R.S., Lehmann, J.F., Delateur, B.J., Sand, P.L. and Treischmann, R.B., Operant conditioning in the treatment of chronic pain, Arch. Phys. Med. Rehab., 54 (1973) 399–408.

Gatchel, R.J. and Turk, D.C., Psychological Approaches to Pain Management: A Practitioner's Handbook, Guilford Press, New York, 1996.

Gleser, L.J. and Olkin, I., Stochastically dependent effect sizes. In: H. Cooper and L.V. Hedges (Eds.), The Handbook of Research Synthesis. Russell Sage Foundation, New York, 1994, pp. 339–355.

Hedges, L.V. and Olkin, I. Statistical Methods for Meta-Analysis. Academic Press, Orlando, FL., 1985.

Hunter, J.E. and Schmidt, F.L. Methods of Meta-Analysis. Sage, Beverly Hills, CA., 1990.

Jadad, A.R., Carroll, D., Moore, A., McQuay, H., Developing a database of published reports of randomized clinical trials in pain research, Pain, 66 (1996a) 239–246.

Jadad, A.R., Moore, R.A., Carroll, D., Jenkinson, C., Reynolds, D.J.M., Gavaghan, D.J., McQuay, H.J., Assessing the quality of randomized clinical trials: is blinding necessary? Control. Clin. Trials, 17 (1996b) 1–12.

Jadad-Bechara, A.R., Meta-Analysis of Randomized Clinical Trials in Pain relief, Balliol College, University of Oxford, Oxford, 1994.

Kaplan, R.M., Behavior as the central outcome in health care, Am. Psychol., 45 (1990) 1211–1220.

Kazdin, A.E., Methodology design and evaluation in psychotherapy research. In: A.E. Bergin and S.L. Garfield (Eds.), Handbook of Psychotherapy and Behavior Change (4th edn.), Wiley, New York, 1994, pp. 19–71.

Keefe, F.J. and Block, A.R., Development of an observational method for assessing pain behavior in chronic low-back pain patients, Behav. Ther., 13 (1982) 363–375.

Keefe, F.J., Caldwell, D.S., Williams, D.A., Gil, K.M., Mitchell, D., Robertson, C., Martinez, S., Nunley, J., Beckham, J.C., Crisson, J.E. and Helms, M., Pain coping skills training in the management of osteoarthritic knee pain: a comparative study, Behav. Ther., 21 (1990a) 49–63.

Keefe, F.J., Caldwell, D.S., Williams, D.A., Gil, K.M., Mitchell, D., Robertson, C., Martinez, S., Nunley, J., Beckham, J.C., Crisson, J.E. and Helms, M., Pain coping skills training in the management of

osteoarthritic knee pain – II: follow-up results, Behav. Ther., 21 (1990b) 435–447.

Keefe, F.J., Caldwell, D.S., Baucom, D., Salley, A., Robinson, E., Timmons, K., Beaupre, P., Weisberg, J. and Helms, M., Spouse - assisted coping skills training in the management of osteoarthritic knee pain, Arth. Care Res., 9 (1996) 279–291.

Kerns, R.T., Turk, D.C. and Rudy, T.E., The West Haven–Yale multi-dimensional pain inventory, Pain, 23 (1985) 345–356.

Kerns, R.D., Turk, D.C., Holzman, A.D. and Rudy, T.E., Comparison of cognitive behavioral and behavioral approaches to the outpatient treatment of chronic pain, Clin. J. Pain, 1 (1986) 195–203.

Kraaimaat, F.W., Brons, M.R., Greenen, R. and Bijlsma, J.W.J., The effect of cognitive behavior therapy in patients with rheumatoid arthritis, Behav. Res. Ther., 33 (1995) 487–495.

Lambert, M.J. and Hill, C.E. Assessing psychotherapy outcomes and process. In: A.E. Bergin and S.L. Garfield (Eds.), Handbook of Psychotherapy and Behavior Change (4th edn.), Wiley, New York, 1994, pp. 72–113.

Linton, S.J. and Gotestam, K.G., A controlled study of the effects of applied relaxation plus operant procedures in the regulation of chronic pain, Br. J. Clin. Psychol., 23 (1984) 291–299.

Linton, S.J., Melin, L. and Stjernlof, K., The effects of applied relaxation and operant activity on chronic pain, Behav. Psychother., 13 (1985) 87–100.

Loeser, J.D., The role of chronic pain clinics in managing back pain. In: J.W. Frymoyer (Ed.), The Adult Spine: Principles and Practice, Raven Press, New York, 1991, pp. 221–229.

Lorig, K., Arthritis Self Management; Leader's Manual, Arthritis Foundation, Atlanta, GA., 1982.

Malone, M.D. and Strube, M.J., Meta-analysis of non-medical treatments for chronic pain, Pain, 34 (1988) 231–244.

McQuay, H., Nye, B.A., Carroll, D., Wiffen, P., Tramer, M. and Moore, A., A systematic review of antidepressants in neuropathic pain, Pain, 68 (1996) 217–227.

Moore, J.E. and Chaney, E.F., Cognitive behavioural treatment of chronic pain: effects of spouse involvement, J.Consult. Clin. Psychol., 53 (1985) 326–334.

Newton-John, T.R.O., Spence, S.H. and Schotte, D., Cognitive-behavioural therapy versus EMG biofeedback in the treatment of chronic low-back pain, Behav. Res. Ther., 33 (1995) 691–697.

NHSCRD, Undertaking Systematic Reviews of Research on Effectiveness: CDR Guidelines for Those Carrying Out or Commissioning Reviews, 4, York, 1996.

Nicholas, M.K., Wilson, P.H. and Goyen, J., Operant-behavioural and cognitive-behavioural treatment for chronic low-back pain, Behav. Res. Ther., 29 (1991) 238–255.

Nicholas, M.K., Wilson, P.H. and Goyen, J., Comparison of cognitive-behavioral group treatment and an alternative non-psychological treatment for chronic low-back pain, Pain, 48 (1992) 339–347.

O'Leary, K.D. and Borkovec, T.D., Conceptual, methodological, and ethical problems of placebo groups in psychotherapy research, Am. Psychol., 33 (1978) 821–830.

O'Leary, A., Shoor, S., Lorig, K. and Holman, H.R., A cognitive behavioral treatment for rheumatoid arthritis, Health Psychol., 7 (1988) 537–544.

Parker, J.C., Frank, R.G., Beck, N.C., Smarr, K.L., Buesher, K.L., Smith, E.I., Anderson, S.K. and Walker, S.E., Pain management in rheumatoid arthritis patients: a cognitive behavioral approach, Arthritis Rheum., 31 (1988) 593–600.

Peters, J.L. and Large, R.G., A randomised control trial evaluating in- and outpatient pain management programmes, Pain, 41 (1990) 283–293.

Peters, J.L., Large, R.G. and Elkind, G., Follow-up results from a randomised controlled trial evaluating in- and outpatient pain management programmes, Pain, 50 (1992) 41–50.

Pilowsky, I., Spence, N., Rounsefell, B., Forsten, C. and Soda, J., Out-patient cognitive-behavioural therapy with amitriptyline for chronic non-malignant pain: a comparative study with 6-month follow-up, Pain, 60 (1995) 49–54.

Puder, R.S., Age analysis of cognitive-behavioral group therapy for chronic pain outpatients, Psychol. Aging, 3 (1988) 204–207.

Radojevic, V., Nicassio, P.M. and Weisman, M.H., Behavioral intervention with and without family support for rheumatoid arthritis, Behav. Ther., 23 (1992) 13–30.

Raudenbusch, S.W., Becker, B.J. and Kalaian, H., Modeling multivariate effect sizes, Psychol. Bull., 103 (1988) 111–120.

Rosenthal, R., Parametric measures of effect size. In: H. Cooper and L.V. Hedges (Eds.), The Handbook of Research Synthesis, Russell Sage Foundation, New York, 1994, pp. 231–244.

Schwartz, C.E., Chesney, M.A., Irvine, M.J. and Keefe, F.J., The control groups dilemma in clinical research: applications for psychosocial and behavioral medicine trials, Psychosomat. Res., 59 (1997) 362–371.

Shadish, W.R., Matt, G.E., Navarro, A.M., Siegle, G., Crits-Cristoph, P., Hazelrigg, M.D., Jorm, A.F., Lyons, L.C., Nietzel, M.T., Prout, H.T., Robinson, L., Smith, M.L., Svartberg, M. and Weiss, B., Evidence that therapy works in clinically representative conditions, J. Consult. Clin. Psychol., 65 (1997) 355–365.

Smith, M.L., Glass, G.V. and Miller, T.I., The Benefits of Psychotherapy. Johns Hopkins University Press, Baltimore, 1980.

Spence, S.H., Cognitive-behavior therapy in the management of chronic, occupational pain of the upper limbs, Behav. Res. Ther., 27 (1989) 435–466.

Spence, S.H., Cognitive-behavior therapy in the treatment of chronic, occupational pain of the upper limbs: a 2 year follow-up, Behav. Res. Ther., 29 (1991) 503–509.

Spence, S.H., Sharpe, L., Newton-John, T. and Champion, D., Effect of EMG biofeedback compared to applied relaxation training with chronic, upper extremity cumulative trauma disorders, Pain, 63 (1995) 199–203.

Stock, W.A., Systematic coding for research synthesis. In: H. Cooper and L.V. Hedges (Eds.), The Handbook of Research Synthesis, Russell Sage Foundation, New York, 1994, pp. 125–138.

Strauss, G.D., Spiegel, J.S., Daniels, M., Spiegel, T., Landsverk, J., Roy-Byrne, P., Edelstein, C., Ehlhardt, J., Falke, R., Hinden, L. and Zackler, L., Group therapies for rheumatoid arthritis: a controlled study of two approaches, Arthritis Rheum., 29 (1986) 1203–1209.

Thompson, S.G. and Pocock, S.J., Can meta-analyses be trusted?, Lancet, 338 (1991) 1127–1130.

Turk, D.C., Meichenbaum, D. and Genest, M., Pain and Behavioral Medicine. Guilford Press, New York, 1983.

Turk, D.C., Biopsychosocial perspective on chronic pain. In: R.J. Gatchel and D.C. Turk (Eds.), Psychological Approaches to Pain Management: A Practitioner's Handbook. Guilford Press, New York, 1996.

Turner, J.A., Comparison of group progressive-relaxation training and cognitive-behavioral group therapy for chronic low-back pain, J. Consult. Clin. Psychol., 50 (1982) 757–765.

Turner, J.A. and Clancy, S., Comparison of operant behavioral and cognitive behavioral group treatment for chronic low-back pain, J. Consult. Clin. Psychol., 56 (1988) 261–266.

Turner, J.A., Clancy, S., McQuade, K.J. and Cardenas, D.D., Effectiveness of behavioral therapy for chronic low-back pain: a component analysis, J. Consult. Clin. Psychol., 58 (1990) 573–579.

Turner, J.A. and Jensen, M.P., Efficacy of cognitive therapy for chronic low-back pain, Pain, 52 (1993) 169–177.

Turner, J.A., Deyo, R.A., Loeser, J.D., Von Korff, M. and Fordyce, W.E., The importance of placebo effects in pain treatment and research, J. Am. Med. Assoc., 271 (1994) 1609–1614.

Turner, J.A., Educational and behavioral interventions for back pain in primary care, Spine, 21 (1996) 2851–2859.

van Tulder, M.W., Assendelft, W.J.J., Koes, B.W. and Bouter, L.M., Spinal radiographic findings and non-specific low-back pain, Spine, 22 (1997) 427–434.

Vlaeyen, J.W.S., Haazen, I.W.C.J., Schuerman, J.A., Kole-Snijders, A.M.J. and van Eek, H., Behavioural rehabilitation of chronic low-back pain: comparison of an operant treatment, an operant cognitive

treatment and an operant-respondent treatment, Br. J. Clin. Psychol., 43 (1995) 95–118.

Vlaeyen, J.W.S., Teeken-Gruben, N.J.G., Goossens, M.E.J.B., Rutten-van Molken, M.M.P.H., Pelt, R.A.G.B., van Eek, H. and Heuts, P.H.T.G., Cognitive-educational treatment of fibromyalgia: a randomised clinical trial. I. Clinical effects, J. Rheumatol., 23 (1996) 1237–1245.

Williams, A.Cde.C., Richardson, P.H., Nicholas, M.K., Harding, V.R., Ridout, K.L., Ralphs, J.A., Richardson, I.H., Justins, D.M. and Chamberlain, J.H., Inpatient versus outpatient pain management: results of a randomised controlled trial, Pain, 66 (1996) 13–22.