Behavioural treatment for chronic low-back pain (Review)

Henschke N, Ostelo RWJG, van Tulder MW, Vlaeyen JWS, Morley S, Assendelft WJJ, Main CJ

THE COCHRANE COLLABORATION®

This is a reprint of a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in *The Cochrane Library* 2011, Issue 2

http://www.thecochranelibrary.com

WILEY
Publishers Since 1807

TABLE OF CONTENTS

[Intervention Review]

Behavioural treatment for chronic low-back pain

Nicholas Henschke[1], Raymond WJG Ostelo[2], Maurits W van Tulder[3], Johan WS Vlaeyen[4], Stephen Morley[5], Willem JJ Assendelft [6], Chris J. Main[7]

[1]The George Institute for International Health, Sydney, Australia. [2]Department of Health Sciences, EMGO Institute for Health and Care Research, VU University, Amsterdam, Netherlands. [3]Department of Health Sciences, Faculty of Earth and Life Sciences, VU University, Amsterdam, Netherlands. [4]Department of Clinical Psychology, University of Maastricht, Maastricht, Netherlands. [5]Leeds Institute of Health Sciences, University of Leeds, Leeds, UK. [6]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands. [7]Primary Care Sciences, Keele University, Staffordshire, UK

Contact address: Nicholas Henschke, The George Institute for International Health, Level 7, 341 George Street, Sydney, NSW, 2000, Australia. nhenschke@georgeinstitute.org.au.

Editorial group: Cochrane Back Group.
Publication status and date: Edited (no change to conclusions), published in Issue 2, 2011.
Review content assessed as up-to-date: 31 July 2009.

Citation: Henschke N, Ostelo RWJG, van Tulder MW, Vlaeyen JWS, Morley S, Assendelft WJJ, Main CJ. Behavioural treatment for chronic low-back pain. *Cochrane Database of Systematic Reviews* 2010, Issue 7. Art. No.: CD002014. DOI: 10.1002/14651858.CD002014.pub3.

ABSTRACT

Background

Behavioural treatment is commonly used in the management of chronic low-back pain (CLBP) to reduce disability through modification of maladaptive pain behaviours and cognitive processes. Three behavioural approaches are generally distinguished: operant, cognitive, and respondent; but are often combined as a treatment package.

Objectives

To determine the effects of behavioural therapy for CLBP and the most effective behavioural approach.

Search strategy

The Cochrane Back Review Group Trials Register, CENTRAL, MEDLINE, EMBASE, and PsycINFO were searched up to February 2009. Reference lists and citations of identified trials and relevant systematic reviews were screened.

Selection criteria

Randomised trials on behavioural treatments for non-specific CLBP were included.

Data collection and analysis

Two review authors independently assessed the risk of bias in each study and extracted the data. If sufficient homogeneity existed among studies in the pre-defined comparisons, a meta-analysis was performed. We determined the quality of the evidence for each comparison with the GRADE approach.

Main results

We included 30 randomised trials (3438 participants) in this review, up 11 from the previous version. Fourteen trials (47%) had low risk of bias. For most comparisons, there was only low or very low quality evidence to support the results. There was moderate quality evidence that:

i) operant therapy was more effective than waiting list (SMD -0.43; 95%CI -0.75 to -0.11) for short-term pain relief;

ii) little or no difference exists between operant, cognitive, or combined behavioural therapy for short- to intermediate-term pain relief;

iii) behavioural treatment was more effective than usual care for short-term pain relief (MD -5.18; 95%CI -9.79 to -0.57), but there were no differences in the intermediate- to long-term, or on functional status;

iv) there was little or no difference between behavioural treatment and group exercise for pain relief or depressive symptoms over the intermediate- to long-term;

v) adding behavioural therapy to inpatient rehabilitation was no more effective than inpatient rehabilitation alone.

Authors' conclusions

For patients with CLBP, there is moderate quality evidence that in the short-term, operant therapy is more effective than waiting list and behavioural therapy is more effective than usual care for pain relief, but no specific type of behavioural therapy is more effective than another. In the intermediate- to long-term, there is little or no difference between behavioural therapy and group exercises for pain or depressive symptoms. Further research is likely to have an important impact on our confidence in the estimates of effect and may change the estimates.

PLAIN LANGUAGE SUMMARY

Behavioural treatment for chronic low-back pain

Low-back pain is a major health and economical problem that affects populations around the world. Chronic low-back pain, in particular, is a major cause of medical expenses, work absenteeism, and disability. Current management of chronic low-back pain includes a range of different treatments such as medication, exercise, and behavioural therapy. Research has shown that social roles and psychological factors have a role in the course of chronic low-back pain.

This review of 30 studies (3438 participants) evaluated three behavioural therapies for chronic low-back pain: (i) operant (which acknowledges that external factors associated with pain can reinforce it), (ii) cognitive (dealing with thoughts, feelings, beliefs, or a combination of the three, that trigger the pain), (iii) respondent (interrupts muscle tension with progressive relaxation techniques or biofeedback of muscle activity).

For pain relief, there was moderate quality evidence that:

(i) operant therapy was more effective than waiting list controls in the short-term,

(ii) there was little or no difference between operant therapy, cognitive therapy; or a combination of behavioural therapies in the short- or intermediate-term, and

(iii) behavioural treatment was more effective than usual care (which usually consists of physical therapy, back school and/or medical treatments) in the short-term.

Over a longer term, there was little or no difference between behavioural treatment and group exercise for pain relief or reduced depressive symptoms. The addition of behavioural therapy to inpatient rehabilitation did not appear to increase the effect of inpatient rehabilitation alone.

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON *[Explanation]*

Behavioural treatment compared with waiting list control for chronic low-back pain

Patient or population: adults with chronic low-back pain
Settings: primary and secondary health care centres
Intervention: behavioural treatment (respondent, cognitive, and operant therapy, or a combination)
Comparison: waiting list control

Outcomes	Illustrative means (95% CI)		No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Control group	Intervention group			
	Waiting list control	Respondent therapy (progressive relaxation)			
Pain intensity VAS scale (0-100) short-term follow-up	The mean pain intensity ranged across control groups from **44.4 to 77.0 points**	The mean pain intensity in the intervention groups was **19.77 points lower** (34.34 to 5.2 lower)	74 [3 studies]	++00 low[1,3]	
Functional status (generic) various scales short-term follow-up		The mean generic functional status in the intervention groups was **0.88 standard deviations lower** (1.36 to 0.39 lower)	74 [3 studies]	++00 low[1,3]	SMD -0.88 (-1.36 to -0.39)
Depression Beck Depression Inventory (0-63) short-term follow-up	The mean depression ranged across control groups from **7.2 to 22.4 points**	The mean depression in the intervention groups was **6.8 points lower** (19.73 lower to 6.12 higher)	58 [2 studies]	+000 very low[1,2,3]	
	Waiting list control	Respondent therapy (EMG biofeedback)			

Pain intensity various scales short-term follow-up		The mean pain intensity in the intervention groups was **0.8 standard deviations lower** (1.32 to 0.28 lower)	64 [3 studies]	++00 low[1,3]	SMD -0.8 (-1.32 to -0.28)
Functional status (generic) various scales short-term follow-up		The mean functional status in the intervention groups was **0.17 standard deviations lower** (1.56 lower to 1.22 higher)	44 [2 studies]	+000 very low[1,2,3]	SMD -0.17 (-1.56 to 1.22)
	Waiting list control	**Operant therapy**			
Pain intensity various scales short-term follow-up		The mean pain intensity in the intervention groups was **0.43 standard deviations lower** (0.75 to 0.11 lower)	153 [3 studies]	+++0 moderate[3]	SMD -0.43 (-0.75 to -0.11)
Functional status (generic) Sickness Impact Profile (0-136) short-term follow-up	The mean generic functional status ranged across control groups from **5.4 to 5.7 points**	The mean functional status in the intervention groups was **1.18 points lower** (3.53 lower to 1.18 higher)	87 [2 studies]	++00 low[1,3]	
Depression various scales short-term follow-up		The mean depression in the intervention groups was **0.11 standard deviations lower** (0.67 lower to 0.44 higher)	103 [2 studies]	++00 low[1,3]	SMD -0.11 (-0.67 to 0.44)
	Waiting list control	**Cognitive therapy**			

Pain intensity various scales short-term follow-up		The mean pain intensity in the intervention groups was **0.27 standard deviations lower** (0.75 lower to 0.22 higher)	68 [2 studies]	++00 low[1,3]	SMD -0.27 (-0.75 to 0.22)
Functional status (generic) various scales short-term follow-up		The mean functional status in the intervention groups was **0.15 standard deviations lower** (0.64 lower to 0.33 higher)	68 [2 studies]	++00 low[1,3]	SMD -0.15 (-0.64 to 0.33)
	Waiting list control	**Combined behavioural therapy**			
Pain intensity various scales short-term follow-up		The mean pain intensity in the intervention groups was **0.60 standard deviations lower** (0.97 to 0.22 lower)	239 [5 studies]	++00 low[1,3]	SMD -0.60 (-0.97 to -0.22)
Functional status (generic) various scales short-term follow-up		The mean functional status in the intervention groups was **0.37 standard deviations lower** (0.87 lower to 0.13 higher)	134 [4 studies]	++00 low[1,3]	SMD -0.37 (-0.87 to 0.13)
Depression Beck Depression Inventory (0-63) short-term follow-up	The mean depression ranged across control groups from **7.2 to 22.4 points**	The mean depression in the intervention groups was **1.92 points lower** (6.16 lower to 2.32 higher)	194 [4 studies]	+000 very low[1,2,3]	

CI: Confidence interval

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Serious limitations in study design (i.e. >25% of participants from studies with high risk of bias)
[2] Serious inconsistency of results (i.e. opposite direction of effects and/or significant statistical heterogeneity)
[3] Serious imprecision (i.e. total number of participants <300 for each outcome)

BACKGROUND

Low-back pain is a major health and economical problem which affects large populations around the world. In particular, chronic low-back pain (CLBP) is a major cause of medical expenses, work absenteeism, and disability (Koes 2006). Current management of CLBP includes a range of different intervention strategies such as medication, exercise, and behavioural therapy. The main assumption underlying a behavioural therapy approach is that pain and its resulting disability are not only influenced by somatic pathology, but by psychological and social factors as well. In this way, CLBP is not only a physical problem, but may also be influenced by the patient's attitudes and beliefs, psychologic distress, and illness behaviour (Waddell 2004). Consequently, the goal of behavioural treatment is to alter maladaptive thoughts, feelings and behaviours as well as dysfunctional sensory phenomena, and thereby the experience of pain. In general, three behavioural treatment approaches can be distinguished: operant, cognitive, and respondent (Turk 1984; Vlaeyen 1995). Each of these focuses on modifying one of the three response systems which characterize emotional experiences: behaviour, cognition, and physiological reactivity.

Operant treatments are based on the operant conditioning principles of Skinner (Skinner 1953) which have been applied to CLBP by Fordyce (Fordyce 1976). This model proposes that acute pain behaviours may be reinforced by external factors (such as attention of the spouse and the medical personnel, rest, or reduction of pain level by analgesic medication) and thus develop into a chronic pain problem. Therefore, operant treatment involves the removal of positive reinforcement of pain behaviours and the promotion of healthy behaviours (e.g. exercise, work). It often incorporates involvement of the spouse to help maintain these changes. Increased activity levels are promoted by establishing exercise quota and reinforcing exercise with positive feedback and verbal encouragement. The exercise quota are systematically increased for each treatment session, towards a pre-defined goal. Each successfully performed increment is positively reinforced by all treatment staff (Fordyce 1976).

Cognitive treatment aims to identify and modify harmful cognitions which patients may have regarding their pain and disability. Patients with CLBP often have maladaptive thoughts, feelings, and beliefs, which have an important role in their experience of low-back pain (Pincus 2006). It is proposed that beliefs about the meaning of pain and expectations regarding control over pain can be directly modified using cognitive restructuring techniques such as imagery and attention diversion (Turner 1993). Cognitions can also be indirectly altered through education and other treatments, so cognitive therapy is often used as part of a 'package' approach of behavioural treatment.

Respondent treatment aims to modify the physiological response system to pain, through reduction of muscular tension. The theoretical basis of this approach is the assumption of a pain-tension cycle, where pain is viewed as both a cause and a result of muscular tension. Respondent treatment attempts to interrupt this cycle by using a tension-incompatible reaction, such as relaxation. Electromyographic (EMG) biofeedback, progressive relaxation, and applied relaxation are frequently used to reduce the assumed muscular tension, relieve anxiety, and subsequently pain (Turk 1984; Vlaeyen 1995).

Behavioural treatments are often applied together, as part of a combined treatment approach, commonly referred to as cognitive-behavioural treatment. This combined behavioural treatment is based on a multidimensional model of pain that includes physical, affective, cognitive, and behavioural components. Treatment may therefore include education about a multidimensional view of pain, how to identify pain-eliciting and pain-aggravating situations, thoughts and behaviour, and use of coping strategies and applied relaxation. Goal setting and activity increase is encouraged, as the basis of a combined treatment approach is to reduce feelings of helplessness and assist the patient to gain control over the pain experience (Turk 1984; Rudy 1995). Furthermore, combined behavioural treatments are often applied in addition to other therapies such as physiotherapy, medication, or exercise.

Previous reviews have summarized the literature concerning behavioural therapy for chronic pain (Morley 1999) and CLBP (Pincus 2006; Hoffman 2007). Two previous versions of the current review have been published (van Tulder 2000; Ostelo 2005), systematically evaluating the effectiveness of behavioural therapy for CLBP, using the methods of The Cochrane Collaboration. In this second update, the most recent literature is included and the methods of data collection and analysis have been updated (Furlan 2009; Higgins 2009).

OBJECTIVES

The objective of this systematic review is to determine whether behavioural therapy is more effective than other treatments for non-specific CLBP, and to ascertain which type of behavioural therapy is most effective.

The following comparisons were investigated:

1. behavioural treatment versus placebo, no treatment, or waiting list controls

2. between different types of behavioural treatment

3. behavioural treatment versus other kinds of treatment

4. behavioural treatment in addition to another treatment (e.g. physiotherapy) versus the other treatment alone.

METHODS

Criteria for considering studies for this review

Types of studies

Only randomised controlled trials (RCTs) were included.

Types of participants

RCTs that investigated male and female subjects with non-specific CLBP, between 18 and 65 years of age, were included. CLBP was defined as back pain that had persisted for 12 weeks or more. If an RCT recruited LBP patients with a mixed duration of symptoms (i.e. including patients with < 12 weeks duration), it was only considered eligible if data for the CLBP patients were presented separately. Trials were excluded if they included subjects with specific low-back pain caused by pathological entities such as infection, neoplasm, metastasis, osteoporosis, rheumatoid arthritis, or fractures.

Types of interventions

RCTs that evaluated one or more types of behavioural treatment, that is, treatment based on the behavioural therapy principles outlined in the introduction, were included. As there exists some variability in the delivery and content of behavioural interventions, only RCTs in which the interventions were explicitly stated to follow the behavioural principles were considered eligible. Where this was not explicit and there remained uncertainty regarding the eligibility of an RCT (based on the intervention provided), a consensus was sought from the review co-authors. For examples of good descriptions of treatments following these behavioural principles, studies by Fordyce 1976 and Roberts 1986 for operant treatments, Vlaeyen 1991 for cognitive treatments, and Belar 1986 for respondent treatments were used as a point of reference. Where a combination of behavioural therapy components were evaluated as part of a treatment 'package' (e.g. cognitive-behavioural treatment), it was considered as a single intervention in the analyses. Because of the considerable overlap in the definitions of behavioural components and delivery of the interventions, it is difficult to disaggregate the effects of separate components in these combined treatments.

Types of outcome measures

RCTs were included that measured at least one of the outcome measures that the Cochrane Back Review Group and review authors considered to be important, i.e. overall improvement (either self-reported or observed), back pain-specific functional status (e.g. Roland-Morris Questionnaire, Oswestry Disability Index), generic functional status (e.g. Sickness Impact Profile), return-to-work (e.g. expressed as the number of days of sick leave or the proportion of patients returned to work), and pain intensity (e.g. expressed on a visual analogue (VAS) or numerical rating scale (NRS)) (Furlan 2009). In the behavioural domain, any type of behavioural outcome was considered relevant (e.g. observed pain behaviours, cognitive errors, perceived or observed levels of tension, anxiety, and depression).

Search methods for identification of studies

All relevant RCTs meeting our inclusion criteria were identified by:

- a computer-aided search of the Cochrane Back Review Group Trials Register (February 2009), CENTRAL (*The Cochrane Library* 2009, issue 2), MEDLINE (1966 to February 2009), EMBASE (1988 to February 2009), and PsycINFO (1974 to February 2009) databases, using the search strategy recommended by the Cochrane Back Review Group (Furlan 2009). A highly sensitive search strategy for retrieval of controlled trials was run in conjunction with a specific search for low-back pain and behavioural treatment. RCTs published in any language were considered (Appendix 2).
- screening references from relevant identified publications and reviews
- citation tracking of the RCTs identified by these first two steps, using Science Citation Index.

Data collection and analysis

The most recent version of the Cochrane review on this topic (Ostelo 2005) included 21 RCTs which were screened using the eligibility criteria of this update. The methods of data collection and analysis were updated to align with recent recommendations from the Cochrane Handbook (Higgins 2009) and the Cochrane Back Review Group (Furlan 2009).

Study selection

Two review authors (NH and RO) independently selected the trials to be included in this update of the systematic review. A consensus meeting was held to resolve disagreements concerning the inclusion of RCTs and co-authors were consulted if disagreements persisted.

Risk of bias assessment

The risk of bias of each included trial was independently assessed by two review authors (NH and RO), using the 12 criteria (Appendix 1) recommended by the Cochrane Back Review Group (Furlan 2009). This updated list contains two new assessment items (selective reporting and timing of outcomes) from the list used previously (Ostelo 2005). The articles were not blinded for authors, institution and journal, because the review authors who performed the risk of bias assessments were familiar with the literature. A consensus method was used to resolve disagreements and

a third review author was consulted if disagreements persisted. A total score was computed by adding the number of positive scores, and a low risk of bias was defined as studies fulfilling six or more of the 12 internal validity criteria. Empirical evidence seems to suggest that studies fulfilling fewer than six items report higher treatment effects than studies fulfilling six or more items (van Tulder 2009).

Clinical relevance

The clinical relevance of each included trial was also assessed by two review authors (NH and RO). A list of five questions has been recommended to facilitate decisions about the applicability of the results to other populations (Furlan 2009; Malmivaara 2006):

1. Are the patients described in detail so that you can decide whether they are comparable to those that you see in your practice?
2. Are the interventions and treatment settings described well enough so that you can provide the same for your patients?
3. Were all clinically relevant outcomes measured and reported?
4. Is the size of the effect clinically important?
5. Are the likely treatment benefits worth the potential adverse effects?

Data extraction

A standardized form was used to collect descriptive data on the study populations and the types of intervention, as well as quantitative data from the outcome measures. Two review authors (NH and RO) independently extracted the data, considering the following domains: i) behavioural outcomes, ii) overall improvement, iii) back-pain specific functional status, iv) generic functional status, v) return-to-work, and vi) pain intensity. Measures of effect and estimates of variability were extracted in the form of follow-up (post-intervention) measurements or change scores from baseline in all intervention and control groups. Where possible, only follow-up measures were entered in the meta-analyses.

Data analysis and the GRADE approach

If studies were clinically homogeneous regarding study population, types of treatment, outcomes and measurement instruments, a meta-analysis was performed. If possible, the mean difference (MD) was calculated, because this improves the interpretability of the results. If an MD was not possible, the standardized mean difference (SMD) was calculated. If trials reported outcomes as graphs, the mean scores and standard deviations were estimated from these graphs. If standard deviations (SD) for outcomes were not reported, they were calculated using the reported values of the confidence intervals, if possible. If an RCT evaluated multiple groups which were relevant for a comparison in the meta-analysis, all relevant experimental intervention groups of the study were combined to create a single group and all relevant control groups were combined to create a single control group (Higgins 2008).

GRADE (Grades of Recommendation, Assessment, Development and Evaluation) profiles were used to evaluate the overall quality of the evidence and the strength of the recommendations (Atkins 2004). Using this approach, as recommended in the recent update of *The Cochrane Handbook* (Higgins 2009), and the Cochrane Back Review Group method guidelines (Furlan 2009), the overall quality of evidence for a particular outcome is considered to be high when multiple RCTs with a low risk of bias provide consistent, generalisable, and precise data. The quality of the evidence was downgraded by one level for each of the five factors that was encountered: 1) limitations in design (i.e. >25% of participants from studies with high risk of bias), 2) inconsistency of results (i.e. opposite direction of effects and/or significant statistical heterogeneity), 3) indirectness (e.g. patients selected based on MRI results, surgical candidates), 4) imprecision (i.e. total number of participants <300 for each outcome), and 5) other considerations (e.g. reporting bias). The judgment of whether these factors were present for each outcome was made by two review authors (NH and RO) using the descriptions shown in brackets above. Single studies were considered inconsistent and imprecise (i.e. sparse data) and provided "low quality evidence". This could be further downgraded to "very low quality evidence" if there were also limitations in design or indirectness. The following definitions of quality of the evidence were applied (Guyatt 2008):

High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

Very low quality: We are very uncertain about the estimate.

To improve the readability of this review, a GRADE profile was only completed when we completed a meta-analysis. If only one study was present for a given comparison, the results are described in the text and in the Characteristics of included studies table.

RESULTS

Description of studies

See: Characteristics of included studies; Characteristics of excluded studies.

See Characteristics of included studies; Characteristics of excluded studies.

In the previous Cochrane review on this topic, which was published in 2005 (Ostelo 2005), 21 RCTs were included. After screening these RCTs, two studies were considered to be ineligible for this update, one for not reporting the number of patients with

CLBP (Bru 1994) and the other for including patients with subacute LBP (Lindström 1992).

The updated search strategy for the current review identified an additional 121 references from MEDLINE, 337 from EMBASE, 100 from CINAHL, and 20 from PsycINFO. However, 186 of these publications were duplicated in more than one database, resulting in a total of 392 unique titles from the updated search. After screening the titles and abstracts, full text copies of 75 trials and 16 reviews were retrieved. The reference lists of the reviews were checked but did not result in the identification of any further relevant studies. After reviewing the full text of the 75 selected trials, both review authors (NH, RO) agreed on the inclusion of nine trials and exclusion of 57 trials. There was uncertainty regarding the inclusion of the nine remaining trials, so the remaining review authors were consulted in order to arrive at a consensus. Two of these nine trials were eventually included in this review. Therefore, 11 additional trials were included in this update, giving a total of 30 RCTs included in this systematic review.

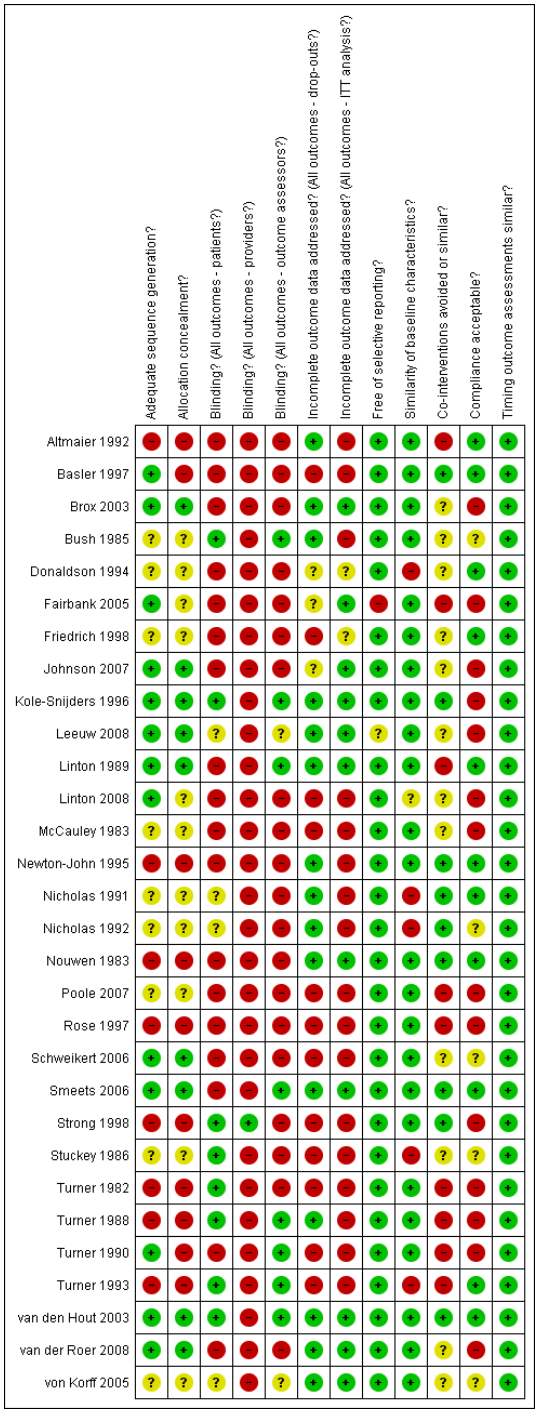
Twelve studies compared some type of behavioural treatment to waiting list controls (Bush 1985; Kole-Snijders 1996; Linton 1989; Linton 2008; Newton-John 1995; Nouwen 1983; Smeets 2006; Stuckey 1986; Turner 1982; Turner 1988; Turner 1990; Turner 1993); ten studies were identified that compared various types of behavioural treatment (Donaldson 1994; Kole-Snijders 1996; Leeuw 2008; Newton-John 1995; Nicholas 1991; Rose 1997; Turner 1982; Turner 1988; Turner 1993; van den Hout 2003); nine studies compared behavioural treatment with some other kind of treatment (Brox 2003; Donaldson 1994; Fairbank 2005; McCauley 1983; Poole 2007; Smeets 2006; Turner 1990; van der Roer 2008; von Korff 2005); and ten studies compared behavioural treatment in addition to another treatment to that treatment alone (Altmaier 1992; Basler 1997; Friedrich 1998; Johnson 2007; Nicholas 1991; Nicholas 1992; Schweikert 2006; Smeets 2006; Strong 1998; Turner 1990). Five studies did not report the results in a way that allowed us to include them in the statistical pooling (Bush 1985; Kole-Snijders 1996; Linton 1989; McCauley 1983; Strong 1998).

There were 20 RCTs that specifically mentioned the qualification of therapists and 10 RCTs where the qualification of therapists was not reported or reported insufficiently (Altmaier 1992; Brox 2003; Donaldson 1994; Fairbank 2005; Linton 2008; Newton-John 1995; Schweikert 2006; Stuckey 1986; van der Roer 2008; Von Korff 1998). An example of sufficient description of qualifications of therapists was "psychologist who has had five years of experience with chronic pain patients since completing his clinical qualifications" (Nicholas 1991).

Risk of bias in included studies

The final results of the risk of bias assessment are shown in Figure 1. Fourteen studies (47%) had a low risk of bias, meeting six or more of the criteria (Basler 1997; Brox 2003; Bush 1985; Johnson 2007; Kole-Snijders 1996; Leeuw 2008; Linton 1989; Newton-John 1995; Nouwen 1983; Smeets 2006; Strong 1998; Turner 1988; van den Hout 2003; van der Roer 2008). All studies were described as randomised, however only nine studies (30%) used a clearly described and adequate randomisation procedure in combination with an adequate concealment of treatment allocation. All 30 studies (100%) had similar timing of outcome measurements between groups and most studies were free of selective reporting (28 studies; 93%). Fifteen studies (50%) had an acceptable drop-out rate, 11 studies (37%) reported acceptable compliance, and in only nine studies (30%) were co-interventions avoided or similar between groups.

Figure 1. Summary of risks of bias for each included study



Clinical relevance

The clinical relevance scores for each trial are presented in Table 1. Twenty (67%) of the 30 RCTs were found to have moderate to high clinical relevance (a score of three or greater). The majority of studies could be easily assessed in terms of applicability to other populations because they provided sufficient descriptions of the included patients (26 trials; 87%), the interventions applied (27 trials; 90%), and measured appropriate outcome measures (26 trials; 87%). However, determining whether the study results were clinically relevant proved to be more difficult. Most studies did not provide enough information to ascertain whether the size of the effect was clinically important (16 trials, 53%) or whether treatment benefits outweighed the potential harms (29 trials; 97%). These findings indicate that while most studies were found to have moderate to high clinical relevance, the assessment is limited by poor reporting of outcome data and potential harms associated with treatment. However, It could be argued that the potential harms associated with behavioural therapy are negligible.

Table 1. Results of clinical relevance assessment

Study	Patients	Interventions	Relevant outcomes	Size of effect	Benefit/Harms
Altmaier 1992	Y	N	Y	N	?
Basler 1997	Y	Y	Y	N	?
Brox 2003	Y	Y	Y	N	?
Bush 1985	Y	Y	Y	?	?
Donaldson 1994	Y	Y	N	?	?
Fairbank 2005	Y	Y	N	N	?
Friedrich 1998	Y	Y	Y	?	?
Johnson 2007	Y	Y	Y	Y	?
Kole-Snijders 1996	Y	Y	N	?	?
Leeuw 2008	Y	Y	Y	N	?
Linton 1989	N	Y	Y	?	?
Linton 2008	Y	N	Y	N	?
McCauley 1983	N	Y	Y	?	?
Newton-John 1995	Y	Y	Y	?	?
Nicholas 1991	Y	Y	Y	N	?
Nicholas 1992	Y	Y	Y	N	?
Nouwen 1983	Y	Y	N	?	?
Poole 2007	Y	Y	Y	N	?

Table 1. Results of clinical relevance assessment (*Continued*)

Rose 1997	N	N	Y	N	?
Schweikert 2006	Y	Y	Y	N	?
Smeets 2006	Y	Y	Y	Y	?
Strong 1998	Y	Y	Y	?	?
Stuckey 1986	N	Y	Y	?	?
Turner 1982	Y	Y	Y	?	?
Turner 1988	Y	Y	Y	?	?
Turner 1990	Y	Y	Y	?	?
Turner 1993	Y	Y	Y	?	?
van den Hout 2003	Y	Y	Y	?	?
van der Roer 2008	Y	Y	Y	?	Y
von Korff 2005	Y	Y	Y	N	?

Effects of interventions

See: **Summary of findings for the main comparison** Summary of findings: Behavioural treatment versus waiting list control; **Summary of findings 2** Summary of findings: Comparisons among behavioural treatments; **Summary of findings 3** Summary of findings: Behavioural treatments versus other treatments; **Summary of findings 4** Summary of findings: Behavioural treatments + other treatments versus other treatments alone

1. Behavioural treatment versus waiting list controls

Summary

As stated in the objectives, "no treatment" or "placebo" groups were considered as controls in this comparison, however, the included studies used only waiting list controls. Twelve studies, seven of which had a low risk of bias (Bush 1985; Kole-Snijders 1996; Linton 1989; Newton-John 1995; Nouwen 1983; Smeets 2006; Turner 1988), were identified as comparing some type of behavioural treatment (either cognitive, operant, respondent, or a combination) to waiting list controls. For this comparison, only post-treatment (short-term) data were available because after the treatment period, most studies allowed the waiting list controls to receive the intervention.

Using the GRADE approach, there is low quality evidence that respondent therapy (through progressive relaxation or EMG biofeedback) or a combination of behavioural therapies is more effective than a waiting list control for short-term pain relief. There is also low quality evidence that respondent therapy (through progressive relaxation) is more effective than a waiting list control for improving functional status in the short-term. Very low quality evidence exists that there is little or no difference between respondent therapy and a waiting list control on depression in the short-term.

We found moderate quality evidence that operant therapy is more effective than a waiting list control for pain relief in the short-term. There is low quality evidence that there is little or no difference between operant therapy, a combination of behavioural therapies, or a waiting list control for improved function or depression in the short-term. Low quality evidence exists that there is little or no difference between cognitive therapy and a waiting list control for pain relief or improved function in the short-term.

1.1 Respondent therapy (progressive relaxation) versus

waiting list controls

Three studies with high risk of bias were identified for this comparison and provided data which could be pooled (Stuckey 1986; Turner 1982; Turner 1993). All three studies measured pain intensity as an outcome measure using a 100-point scale. The Chi-square value for homogeneity was 4.69 (P > 0.05), indicating statistical homogeneity among these studies. The MD (95% CI) when these studies were pooled was -19.77 (-34.34 to -5.20) (Analysis 1.1). There is low quality evidence (three RCTs; N = 74; limitations in design, imprecision) that respondent therapy (progressive relaxation) is more effective than a waiting list control for pain relief in the short-term.

The three RCTs measured generic functional status using either a 7-point scale (Stuckey 1986) or the 136-item Sickness Impact Profile (Turner 1982; Turner 1993). The SMD (95% CI) between these studies was -0.88 (-1.36 to -0.39) (Analysis 1.2). There is low quality evidence (three RCTs; N = 74; limitations in design, imprecision) that respondent therapy (progressive relaxation) is more effective than a waiting list control at improving functional status over the short-term.

In the behavioural domain, two RCTs (Turner 1982; Turner 1993) measured symptoms of depression as an outcome, using the Beck Depression Inventory. The MD (95% CI) when these studies were pooled was -6.80 (-19.73 to 6.12) (Analysis 1.3). There is very low quality evidence (two RCTs; N = 58; limitations in design, inconsistency, imprecision) that there is no significant difference between respondent therapy (progressive relaxation) and a waiting list on symptoms of depression over a short-term.

1.2 Respondent therapy (EMG biofeedback) versus waiting list controls

Three RCTs provided data in a format that could be pooled for this comparison (Newton-John 1995; Nouwen 1983; Stuckey 1986). The RCT (N = 44) that was not included in the statistical pooling showed no differences between the two treatment arms (Bush 1985). The SMD (95% CI) between groups for pain intensity was -0.80 (-1.32 to -0.28) in favour of respondent therapy (Analysis 2.1). There is low quality evidence (three RCTs; N = 64; limitations in design, imprecision) that respondent therapy through EMG biofeedback is more effective than a waiting list control for pain relief in the short-term.

Two of the RCTs measured general functional status as an outcome measure (Newton-John 1995; Stuckey 1986). The SMD (95% CI) between groups was -0.17 (-1.56 to 1.22) (Analysis 2.2). There is very low quality evidence (two RCTs; N = 44; limitations in design, inconsistency, imprecision) that there is no significant difference between respondent therapy through EMG biofeedback and a waiting list control for improved function in the short-term.

1.3 Operant therapy versus waiting list controls

Of four RCTs which were identified, two RCTs with a low risk of bias (Linton 1989; Turner 1988) and one RCT with a high risk of bias (Turner 1990) could be pooled. The trial (N = 89) that was not included in the pooling found small effects in favour of operant therapy (Kole-Snijders 1996). Pain intensity was measured by the three pooled RCTs, and the pooled SMD (95% CI) between groups was -0.43 (-0.75 to -0.11) in favour of the operant therapy group (Analysis 3.1). There is moderate quality evidence (N = 153; imprecision) that operant therapy is more effective than waiting list control for pain relief in the short-term.

Two RCTs measured generic functional status using the SIP (Turner 1988; Turner 1990). The pooled MD (95% CI) between groups was -1.18 (-3.53 to 1.18) (Analysis 3.2). There is low quality evidence (two RCTs; N = 87; limitations in design, imprecision) that there is no significant difference between operant therapy and a waiting list control on improved functional status in the short-term.

Two RCTs measured symptoms of depression as an outcome measure, using different scales (Linton 1989; Turner 1990). The pooled SMD (95% CI) between groups was -0.11 (-0.67 to 0.44) (Analysis 3.3). There is low quality evidence (two RCTs; N = 103; limitations in design, imprecision) that there is no significant difference between operant therapy and a waiting list control on depressive symptoms in the short-term.

1.4 Cognitive therapy versus waiting list controls

Two RCTs with high risk of bias were identified for this comparison and could be pooled (Linton 2008; Turner 1993). With pain intensity as the outcome, the pooled SMD (95% CI) between groups was -0.27 (-0.75 to 0.22) (Analysis 4.1). There is low quality evidence (two RCTs; N = 68; limitations in design, imprecision) that there is no significant difference between cognitive therapy and a waiting list control for pain relief in the short-term.

The two RCTs also measured generic functional status as an outcome measure (Linton 2008; Turner 1993). The pooled SMD (95% CI) between groups was -0.15 (-0.64 to 0.33) (Analysis 4.2). There is low quality evidence (two RCTs, N = 68; limitations in design, imprecision) that there is no significant difference between cognitive therapy and a waiting list control on improved functional status in the short-term.

1.5 Combined behavioural therapy versus waiting list controls

There were five RCTs identified which compared a combination of behavioural therapy components (as a treatment 'package') to waiting list controls (Newton-John 1995; Smeets 2006; Turner 1982; Turner 1988; Turner 1993). Three RCTs with a low risk of bias (Newton-John 1995; Smeets 2006; Turner 1988) and two RCTs with a high risk of bias (Turner 1982; Turner 1993) measured pain intensity as an outcome of treatment. The pooled SMD

(95% CI) for pain intensity between groups was -0.60 (-0.97 to -0.22) in favour of combined behavioural therapy (Analysis 5.1). There is low quality evidence (five RCTs; N = 239; limitations in design, imprecision) that combined behavioural therapy is more effective than a waiting list control for pain relief in the short-term.

Four of the RCTs measured generic functional status as an outcome, using the Pain Disability Index (Newton-John 1995) or the SIP (Turner 1982; Turner 1988; Turner 1993). The pooled SMD (95% CI) between groups was -0.37 (-0.87 to 0.13) (Analysis 5.2). There is low quality evidence (four RCTs; N = 134; limitations in design, imprecision) that there is no significant difference between combined behavioural therapy and a waiting list control on improved functional status in the short-term.

Four RCTs used the BDI to evaluate symptoms of depression as an outcome measure (Newton-John 1995; Smeets 2006; Turner 1982; Turner 1993). The pooled MD (95% CI) between groups was -1.92 (-6.16 to 2.32) (Analysis 5.3). There is very low quality evidence (four RCTs; N = 194; limitations in design, inconsistency, imprecision) that there is no significant difference between combined behavioural therapy and a waiting list control on symptoms of depression in the short-term.

2. Comparison among different types of behavioural treatment

Summary

Ten studies, five of which had a low risk of bias (Kole-Snijders 1996; Leeuw 2008; Newton-John 1995; Turner 1988; van den Hout 2003), were identified as comparing one type of behavioural therapy (or a combination of behavioural therapies) to another behavioural therapy. For some of the analyses in this comparison, short-term (immediately post-treatment), intermediate-term (6-month follow-up), or long-term (12-month follow-up) data were available. Seven RCTs provided data in a form that was suitable for statistical pooling (Leeuw 2008; Newton-John 1995; Nicholas 1991; Turner 1982; Turner 1988; Turner 1993; van den Hout 2003).

Using the GRADE approach, we found moderate quality evidence that there is little or no difference between cognitive therapy and operant therapy, or operant therapy and a combination of behavioural therapies, on pain relief over a short- to intermediate-term follow-up. There is low quality evidence that respondent therapy is more effective than a combination of behavioural therapies for depression in the short-term. There is very low quality evidence that respondent therapy using EMG biofeedback is more effective than respondent therapy using progressive relaxation for pain relief in the short-term. For all of the other comparisons between different types of behavioural treatment, there is either no evidence, or only low to very low quality evidence that there is little or no difference in efficacy between one type or another.

2.1 Respondent therapy (EMG biofeedback) versus respondent therapy (progressive relaxation)

One RCT (N = 24) with a high risk of bias (Donaldson 1994) compared two types of respondent therapy, progressive relaxation and EMG biofeedback. The EMG biofeedback group improved significantly more than the relaxation group post-treatment, and after three months of follow-up. As this was the only study eligible for this comparison, there is only very low quality evidence (N = 24; limitations in design, inconsistency, imprecision) that EMG biofeedback is more effective than progressive relaxation for pain relief over the short-term.

2.2 Cognitive versus operant therapy

Two RCTs, one with a low risk of bias (Leeuw 2008) and one with high risk of bias (Nicholas 1991), compared cognitive therapy to operant therapy. All groups in the trial with high risk of bias also received physiotherapy, back-education and an exercise program (Nicholas 1991). Both studies evaluated pain intensity as an outcome measure over short- and intermediate-term follow-ups. The pooled SMD (95% CI) between groups was 0.41 (-0.63 to 1.45) (Analysis 6.1) in the short-term and 0.35 (-0.64 to 1.35) in the intermediate-term (Analysis 6.2). There is moderate quality evidence that there is no significant difference between cognitive therapy and operant therapy for pain relief over a short-term follow-up (two RCTs; N = 93; imprecision) or an intermediate-term follow-up (two RCTs; N = 82; imprecision).

2.3 Cognitive versus respondent therapy

One RCT with a high risk of bias compared cognitive therapy to respondent therapy with progressive muscle relaxation training (Turner 1993). This RCT (N = 47) found no significant differences between the groups post-treatment or after six and 12 months on pain intensity, a global measure of improvement, or functional status (Turner 1993). As this was the only study eligible for this comparison, there is only very low quality evidence (N = 47; limitations in design, inconsistency, imprecision) that there is no significant difference in effect between cognitive therapy and respondent therapy using progressive muscle relaxation over any length of follow-up.

2.4 Operant versus respondent therapy

There is no evidence, because no studies were identified that compared operant therapy to respondent therapy.

2.5 Combined behavioural therapy versus cognitive therapy

Two RCTs with high risk of bias compared a combination of behavioural treatments to cognitive therapy (Nicholas 1991; Turner 1993). The trial by Nicholas 1991 included two groups that underwent combined behavioural therapy and were included in the

analyses. One of these groups received a combination of operant and respondent therapy, while the other received a combination of cognitive and respondent therapy (Nicholas 1991). When these groups were pooled with the trial by Turner 1993, the SMD (95% CI) for pain intensity was -0.24 (-1.36 to 0.87) in the short-term (Analysis 7.1); -0.30 (-2.59 to 1.98) in the intermediate-term (Analysis 7.2); and -0.89 (-3.64 to 1.87) in the long-term (Analysis 7.3). There is very low quality evidence (two RCTs; N = 61; limitations in design, inconsistency, imprecision) that there is no significant difference between a combination of behavioural therapies and cognitive therapy alone for pain relief over the short-, intermediate-, or long-term.

The two RCTs also measured generic functional status using the SIP after short-, intermediate-, and long-term follow-ups (Nicholas 1991; Turner 1993). The MD (95% CI) between groups was -2.01 (-10.02 to 5.99) in the short-term (Analysis 7.4); -3.20 (-16.44 to 10.04) in the intermediate-term (Analysis 7.5); and -2.23 (-12.59 to 8.13) in the long-term (Analysis 7.6). There is low quality evidence over short-term follow-up (two RCTs; N = 61; limitations in design, imprecision) and very low quality evidence over intermediate- to long-term follow-up (two RCTs; N = 47; limitations in design, inconsistency, imprecision) that there is no significant difference between a combination of behavioural therapies and cognitive therapy for improved functional status.

Both RCTs evaluated symptoms of depression using the BDI (Nicholas 1991; Turner 1993). The pooled MD (95% CI) was -3.10 (-11.43 to 5.23) over the short-term (Analysis 7.7); -4.66 (-10.94 to 1.61) in the intermediate-term (Analysis 7.8); and -0.64 (-4.61 to 3.32) over a long-term follow-up (Analysis 7.9). There is very low quality evidence over short-term follow-up (two RCTs; N = 61; limitations in design, inconsistency, imprecision) and low quality evidence in the intermediate- to long-term (two RCTs; N = 47; limitations in design, imprecision) that there is no significant difference between a combination of behavioural therapies and cognitive therapy alone on symptoms of depression.

2.6 Combined behavioural therapy versus operant therapy

Three studies with a low risk of bias (Kole-Snijders 1996; Turner 1988; van den Hout 2003) and one study with a high risk of bias (Nicholas 1991) were identified which compared a combination of behavioural treatments to operant therapy alone. One RCT, comparing a combination of cognitive and operant therapy to operant therapy alone, did not provide data in a form that was suitable for statistical pooling (Kole-Snijders 1996). One of the pooled RCTs included two groups that underwent combined behavioural therapy and were included in the analyses (Nicholas 1991). One of these groups was a combination of operant and respondent therapy, while the other was a combination of cognitive and respondent therapy. When these groups were pooled with the other two studies (Turner 1988; van den Hout 2003) the SMD (95% CI) for pain intensity was -0.15 (-0.46 to 0.16) in the short-term (Analysis 8.1); -0.23 (-0.57 to 0.11) in the intermediate-term (Analysis 8.2); and -0.31 (-0.65 to 0.03) in the long-term (Analysis 8.3). There is moderate quality evidence (three RCTs; N = 161; imprecision) that there is no significant difference between combined behavioural treatment and operant therapy alone for pain relief in the short-, intermediate-, and long-term.

One RCT with a low risk of bias (Turner 1988) and one RCT with a high risk of bias (Nicholas 1991) measured generic functional status as an outcome. The pooled SMD (95% CI) was 0.21(-0.24 to 0.67) in the short-term (Analysis 8.4); -0.23 (-1.01 to 0.55) in the intermediate-term (Analysis 8.5); and -0.50 (-1.56 to 0.56) in the long-term (Analysis 8.6). There is low quality evidence (two RCTs; N = 77; limitations in design, imprecision) that there is no significant difference between combined behavioural therapy and operant therapy alone for improved functional status in the short-, intermediate-, and long-term.

2.7 Combined behavioural therapy versus respondent therapy

Three RCTs were identified which compared a combined behavioural treatment to respondent therapy alone (Newton-John 1995; Turner 1982; Turner 1993). Two had a high risk of bias (Turner 1982; Turner 1993) and one had a low risk of bias (Newton-John 1995). The pooled SMD (95% CI) for pain intensity was 0.09 (-0.31 to 0.50) in the short-term (Analysis 9.1) and 0.47 (-0.42 to 1.35) in the intermediate-term (Analysis 9.2). There is low quality evidence that there is no significant difference between combined behavioural therapy and respondent therapy alone for pain relief in the short-term (three RCTs; N = 97; limitations in design, imprecision), or in the intermediate-term (two RCTs; N = 62; limitations in design, imprecision).

The three RCTs measured generic functional status as an outcome, using the Pain Disability Index (Newton-John 1995) or the SIP (Turner 1982; Turner 1993). The pooled SMD (95% CI) was 0.38 (-0.02 to 0.78) in the short-term (Analysis 9.3) and 0.13 (-0.81 to 1.07) in the long-term (Analysis 9.4). There is low quality evidence that there is no significant difference between combined behavioural therapy and respondent therapy alone for improved functional status in the short-term (three RCTs; N = 97; limitations in design, imprecision), or in the intermediate-term (two RCTs; N = 62; limitations in design, imprecision).

Symptoms of depression, measured with the BDI, were also evaluated as an outcome measure in the three RCTs (Newton-John 1995; Turner 1982; Turner 1993). The pooled MD (95% CI) was 2.89 (0.55 to 5.24) in favour of respondent therapy in the short-term (Analysis 9.5) and 1.84 (-0.43 to 4.11) in the intermediate-term (Analysis 9.6). There is low quality evidence in the short-term (three RCTs; N = 97; limitations in design, imprecision) that respondent therapy is more effective than combined behavioural therapy; and low quality evidence in the intermediate-term (two RCTs; N = 62; limitations in design, imprecision) that there is no

significant difference between combined behavioural therapy and respondent therapy for symptoms of depression.
A separate RCT with a high risk of bias (Rose 1997) evaluated four groups of LBP patients following different programs of combined behavioural therapy, but found no significant differences between group and individual therapy (N = 50), or between a high and low intensity program (N = 60).

3. Behavioural treatment versus other kinds of treatment

Summary

Nine RCTs were identified which compared some form of behavioural treatment to another kind of treatment, two of which had a low risk of bias (Brox 2003; Smeets 2006). The treatments that behavioural therapy was compared to were: usual care (Poole 2007; von Korff 2005), group exercise (Smeets 2006; Turner 1990), spine stabilisation surgery (Brox 2003; Fairbank 2005), guideline-based care (van der Roer 2008), education (Donaldson 1994), and hypnosis (McCauley 1983).
Using the GRADE approach, we found moderate quality evidence that behavioural treatment is more effective than usual care for pain relief in the short-term, but there is no difference in the intermediate-term. There is also moderate quality evidence that there is no significant difference between behavioural treatment and usual care for improved back-specific functional status in the short- to intermediate-term. When compared to a group exercise program, there is moderate quality evidence that behavioural treatment is as effective for improving pain or depression in the intermediate- to long-term. There is also low quality evidence that there is no significant difference between behavioural therapy and group exercise for pain or depression in the short-term. We found low quality evidence that there is no significant difference between behavioural treatment and surgical stabilisation for back-specific functional status in the long-term. Only single RCTs were identified for comparisons between behavioural treatment and guideline-based care (van der Roer 2008), education (Donaldson 1994), or hypnosis (McCauley 1983). These studies are described in further detail in the Characteristics of included studies table.

3.1 Behavioural treatment versus usual care

Two RCTs with a high risk of bias compared behavioural treatment to usual care (Poole 2007; von Korff 2005) and were deemed to be clinically homogenous enough to allow for statistical pooling. For pain intensity measured on a 100-point VAS, the pooled MD (95% CI) was -5.18 (-9.79 to -0.57) in favour of behavioural treatment in the short-term (Analysis 10.1); and -4.29 (-9.28 to 0.69) in the intermediate-term (Analysis 10.2). There is moderate quality evidence (two RCTs; N = 330; limitations in design) that behavioural treatment is more effective than usual care for pain relief in the short-term, but there is little or no difference between the groups for pain relief in the intermediate-term.
The two RCTs also measured back-specific functional status as an outcome measure using the Roland-Morris Disability Questionnaire (RMDQ) (von Korff 2005) or the Oswestry Disability Index (ODI) (Poole 2007). The pooled SMD (95% CI) was -0.20 (-0.41 to 0.02) in the short-term (Analysis 10.3) and -0.12 (-0.34 to 0.10) in the intermediate-term (Analysis 10.4). There is moderate quality evidence (two RCTs; N = 330; limitations in design) that there is no significant difference between behavioural treatment and usual care for improved back-specific functional status in the short- to intermediate-term.

3.2 Behavioural treatment versus group exercise

Two RCTs, one with a low risk of bias (Smeets 2006) and one with a high risk of bias (Turner 1990), compared behavioural treatment to a group exercise program. Both trials measured pain intensity using the Pain Rating Index (PRI) as an outcome. The pooled MD (95% CI) between groups was -2.31 (-6.33 to 1.70) in the short-term (Analysis 11.1), 1.18 (-3.16 to 5.53) in the intermediate-term (Analysis 11.2), and 0.14 (-4.40 to 4.67) in the long-term (Analysis 11.3). There is low quality evidence (two RCTs; N = 146; limitations in design, imprecision) that there is no significant difference between behavioural treatment and a group exercise program for pain relief in the short-term. There is moderate quality evidence (two RCTs; N = 137; imprecision) that there is no significant difference between behavioural treatment and a group exercise program for pain relief in the intermediate- to long-term.
These two RCTs also measured symptoms of depression, using the BDI (Smeets 2006) or the CESD (Turner 1990). The pooled SMD (95% CI) between groups was 0.25 (-0.07 to 0.58) in the short-term (Analysis 11.4), 0.02 (-0.32 to 0.35) in the intermediate-term (Analysis 11.5), and 0.07 (-0.27 to 0.41) in the long-term (Analysis 11.6). There is low quality evidence (two RCTs; N = 146; limitations in design, imprecision) that there is no significant difference between behavioural treatment and a group exercise program for symptoms of depression in the short-term. There is moderate quality evidence (two RCTs; N = 137; imprecision) that there is no significant difference between behavioural treatment and a group exercise program for symptoms of depression in the intermediate- to long-term.

3.3 Behavioural treatment versus surgical stabilisation

Two RCTs, one with a low risk of bias (Brox 2003) and one with a high risk of bias (Fairbank 2005), compared behavioural treatment to surgical stabilisation in CLBP patients who were candidates for surgery. Both trials provided long-term follow-up data (>12 months) for back-specific functional status, measured by the ODI. The pooled MD (95% CI) between groups was 2.36 (-1.94

to 6.66) (Analysis 12.1). There is low quality evidence (two RCTs; N = 345; limitations in design, indirectness) that behavioural treatment is as effective as surgical stabilisation to improve back-specific functional status over long-term follow-up.

4. Behavioural treatment in addition to another treatment versus the other treatment alone

Summary

Ten RCTs, four of which had a low risk of bias (Basler 1997; Johnson 2007; Smeets 2006; Strong 1998), were identified in which behavioural treatment was added to another treatment and compared to the other treatment alone. Behavioural treatment was added to a number of different treatments including physiotherapy with back education (Nicholas 1991; Nicholas 1992), an educational booklet and audio cassette (Johnson 2007), an inpatient pain management program (Altmaier 1992; Schweikert 2006, Strong 1998), various forms of medical treatment (e.g., pain medication, nerve blocks, or physical therapy) (Basler 1997), and exercise therapy (Friedrich 1998; Smeets 2006; Turner 1990). The RCTs which evaluated the addition of behavioural treatment to exercise therapy were considered to be clinically heterogenous and so were not pooled (Friedrich 1998; Smeets 2006; Turner 1990). Using the GRADE approach, we found low quality evidence that adding behavioural treatment to physiotherapy and back education was no more effective for pain, depression, or function over the short- to intermediate-term than physiotherapy and back education alone. There was moderate quality evidence that adding behavioural treatment to inpatient pain rehabilitation was no more effective for short-term pain relief than inpatient rehabilitation alone.

4.1 Behavioural treatment in addition to physiotherapy

Two RCTs with a high risk of bias (Nicholas 1991; Nicholas 1992) evaluated the addition of a number of behavioural treatments to physiotherapy and back education. For pain intensity on a 6-point nominal scale, the pooled MD (95% CI) was -0.13 (-1.01 to 0.75) in the short-term (Analysis 13.1), and -0.11 (-0.67 to 0.44) in the intermediate-term (Analysis 13.2). There is low quality evidence (two RCTs; N = 59; limitations in design, imprecision) that the addition of behavioural treatment to physiotherapy and back education is no more effective than physiotherapy and back education alone for pain relief over the short- to intermediate-term.

The same two RCTs (Nicholas 1991; Nicholas 1992) also measured symptoms of depression using the BDI, and the pooled MD (95% CI) was 1.56 (-1.71 to 4.83) in the short-term (Analysis 13.3) and 0.17 (-6.85 to 7.19) in the intermediate-term (Analysis 13.4). There is low quality evidence (two RCTs; N = 59; limitations in design, imprecision) that the addition of behavioural treatment to physiotherapy and back education is no more effective than physiotherapy and back education alone for symptoms of depression over the short- to intermediate-term.

For generic functional status (measured with the SIP) in these two RCTs (Nicholas 1991; Nicholas 1992) the pooled MD (95% CI) was -6.26 (-12.71 to 0.19) in the short-term (Analysis 13.5), and -0.93 (-6.71 to 4.84) in the intermediate-term (Analysis 13.6). There is low quality evidence (two RCTs; N = 59; limitations in design, imprecision) that the addition of behavioural treatment to physiotherapy and back education is no more effective than physiotherapy and back education alone for improving functional status over the short- to intermediate-term.

4.1 Behavioural treatment in addition to inpatient pain rehabilitation

Two RCTs with a high risk of bias (Altmaier 1992; Schweikert 2006) compared the addition of behavioural treatment to an inpatient multidisciplinary pain program to the multidisciplinary program alone. One RCT with a low risk of bias did not provide data in a form suitable for pooling (Strong 1998). The SMD (95% CI) for pain intensity was -0.14 (-0.34 to 0.05) in the short-term (Analysis 14.1). There is moderate quality evidence (two RCTs; N = 405; limitations in design) that the addition of behavioural treatment to an inpatient multidisciplinary program is no more effective than the multidisciplinary program alone for pain intensity in the short-term.

ADDITIONAL SUMMARY OF FINDINGS *[Explanation]*

Comparisons among behavioural treatments for chronic low-back pain

Patient or population: adults with chronic low-back pain
Settings: primary and secondary health care centres
Intervention: behavioural treatments (respondent, cognitive, and operant therapy, or a combination)
Comparison: behavioural treatments (respondent, cognitive, and operant therapy, or a combination)

Outcomes	Illustrative means (95% CI)		No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Control group	Intervention group			
	Operant therapy	Cognitive therapy			
Pain intensity various scales short-term follow-up		The mean pain intensity in the intervention groups was **0.41 standard deviations higher** (0.63 lower to 1.45 higher)	93 [2 studies]	+++O moderate[3]	SMD 0.41 (-0.63 to 1.45)
Pain intensity various scales intermediate-term follow-up		The mean pain intensity in the intervention groups was **0.35 standard deviations higher** (0.64 lower to 1.35 higher)	82 [2 studies]	+++O moderate[3]	SMD 0.35 (-0.64 to 1.35)
	Cognitive therapy	**Combined behavioural therapy**			
Pain intensity various scales short-term follow-up		The mean pain intensity in the intervention groups was **0.24 standard deviations lower** (1.36 lower to 0.87 higher)	61 [2 studies]	+OOO very low[1,2,3]	SMD -0.24 (-1.36 to 0.87)

Pain intensity various scales intermediate-term follow-up		The mean pain intensity in the intervention groups was **0.3 standard deviations lower** (2.59 lower to 1.98 higher)	44 [2 studies]	+000 very low[1,2,3]	SMD -0.3 (-2.59 to 1.98)
Pain intensity various scales long-term follow-up		The mean pain intensity in the intervention groups was **0.89 standard deviations lower** (3.64 lower to 1.87 higher)	48 [2 studies]	+000 very low[1,2,3]	SMD -0.89 (-3.64 to 1.87)
Functional status (generic) Sickness Impact Profile (0-136) short-term follow-up	The mean generic functional status ranged across control groups from **8.0 to 24.3 points**	The mean functional status in the intervention groups was **2.01 points lower** (10.02 lower to 5.99 higher)	61 [2 studies]	++00 low[1,3]	
Functional status (generic) Sickness Impact Profile (0-136) intermediate-term follow-up	The mean generic functional status ranged across control groups from **5.9 to 25.7 points**	The mean functional status in the intervention groups was **3.2 points lower** (16.44 lower to 10.04 higher)	47 [2 studies]	+000 very low[1,2,3]	
Functional status (generic) Sickness Impact Profile (0-136) long-term follow-up	The mean generic functional status ranged across control groups from **5.3 to 20.8 points**	The mean functional status in the intervention groups was **2.23 points lower** (12.59 lower to 8.13 higher)	51 [2 studies]	+000 very low[1,2,3]	
Depression Beck Depression Inventory (0-63) short-term follow-up	The mean depression ranged across control groups from **8.8 to 18.4 points**	The mean depression in the intervention groups was **3.1 points lower** (11.43 lower to 5.23 higher)	61 [2 studies]	+000 very low[1,2,3]	
Depression Beck Depression Inventory (0-63) intermediate-term follow-up	The mean depression ranged across control groups from **9.4 to 16.1 points**	The mean depression in the intervention groups was **4.66 points lower** (10.94 lower to 1.61 higher)	47 [2 studies]	++00 low[1,3]	

Depression Beck Depression Inventory (0-63) long-term follow-up	The mean depression ranged across control groups from **6.5 to 12.8 points**	The mean depression in the intervention groups was **0.64 points lower** (4.61 lower to 3.32 higher)	51 [2 studies]	++00 low[1,3]	
	Operant therapy	**Combined behavioural therapy**			
Pain intensity various scales short-term follow-up		The mean pain intensity in the intervention groups was **0.15 standard deviations lower** (0.46 lower to 0.16 higher)	161 [3 studies]	+++0 moderate[3]	SMD -0.15 (-0.46 to 0.16)
Pain intensity various scales intermediate-term follow-up		The mean pain intensity in the intervention groups was **0.23 standard deviations lower** (0.57 lower to 0.11 higher)	139 [3 studies]	+++0 moderate[3]	SMD -0.23 (-0.57 to 0.11)
Pain intensity various scales long-term follow-up		The mean pain intensity in the intervention groups was **0.31 standard deviations lower** (0.65 lower to 0.03 higher)	140 [3 studies]	+++0 moderate[3]	SMD -0.31 (-0.65 to 0.03)
Functional status (generic) various scales short-term follow-up		The mean functional status in the intervention groups was **0.21 standard deviations higher** (0.24 lower to 0.67 higher)	77 [2 studies]	++00 low[1,3]	SMD 0.21 (-0.24 to 0.67)
Functional status (generic) various scales intermediate-term follow-up		The mean functional status in the intervention groups was **0.23 standard deviations lower** (1.01 lower to 0.55 higher)	61 [2 studies]	++00 low[1,3]	SMD -0.23 (-1.01 to 0.55)

Functional status (generic) various scales long-term follow-up		The mean functional status in the intervention groups was **0.50 standard deviations lower** (1.56 lower to 0.56 higher)	66 [2 studies]	++00 low[1,3]	SMD -0.50 (-1.56 to 0.56)
	Respondent therapy	**Combined behavioural therapy**			
Pain intensity various scales short-term follow-up		The mean pain intensity in the intervention groups was **0.09 standard deviations higher** (0.31 lower to 0.5 higher)	97 [3 studies]	++00 low[1,3]	SMD 0.09 (-0.31 to 0.5)
Pain intensity various scales intermediate-term follow-up		The mean pain intensity in the intervention groups was **0.47 standard deviations higher** (0.42 lower to 1.35 higher)	62 [2 studies]	++00 low[1,3]	SMD 0.47 (-0.42 to 1.35)
Functional status (generic) various scales short-term follow-up		The mean functional status in the intervention groups was **0.38 standard deviations higher** (0.02 lower to 0.78 higher)	97 [3 studies]	++00 low[1,3]	SMD 0.38 (-0.02 to 0.78)
Functional status (generic) various scales intermediate-term follow-up		The mean functional status in the intervention groups was **0.13 standard deviations higher** (0.81 lower to 1.07 higher)	62 [2 studies]	++00 low[1,3]	SMD 0.13 (-0.81 to 1.07)
Depression Beck Depression Inventory (0-63) short-term follow-up	The mean depression ranged across control groups from **6.2 to 8.1 points**	The mean depression in the intervention groups was **2.89 points higher** (0.55 to 5.24 higher)	97 [3 studies]	++00 low[1,3]	

Depression Beck Depression Inventory (0-63) intermediate-term follow-up	The mean depression ranged across control groups from **5.3 to 7.4 points**	The mean depression in the intervention groups was **1.84 points lower** (0.43 lower to 4.11 higher)	62 [2 studies]	++00 low[1,3]

CI: Confidence interval

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Serious limitations in study design (i.e. >25% of participants from studies with high risk of bias)
[2] Serious inconsistency of results (i.e. opposite direction of effects and/or significant statistical heterogeneity)
[3] Serious imprecision (i.e. total number of participants <300 for each outcome)

Behavioural treatment compared with other treatments for chronic low-back pain

Patient or population: adults with chronic low-back pain
Settings: primary or secondary health care settings
Intervention: behavioural treatment (respondent, cognitive, and operant therapy, or a combination)
Comparison: other chronic low-back pain treatments (i.e. usual care, exercise, surgery)

Outcomes	Illustrative means (95% CI)		No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Control group	Intervention group			
	Usual care	Behavioural treatment			
Pain intensity VAS (0-100) short-term follow-up	The mean pain intensity ranged across control groups from **48.9 to 53.0 points**	The mean pain intensity in the intervention groups was **5.18 points lower** (9.79 to 0.57 lower)	330 [2 studies]	+++O moderate[1]	
Pain intensity VAS (0-100) intermediate-term follow-up	The mean pain intensity ranged across control groups from **42.7 to 47.0 points**	The mean pain intensity in the intervention groups was **4.29 points lower** (9.28 lower to 0.69 higher)	319 [2 studies]	+++O moderate[1]	
Functional status (back-specific) various scales short-term follow-up		The mean back-specific functional status in the intervention groups was **0.2 standard deviations lower** (0.41 lower to 0.02 higher)	330 [2 studies]	+++O moderate[1]	SMD -0.2 (-0.41 to 0.02)
Functional status (back-specific) various scales intermediate-term follow-up		The mean back-specific functional status in the intervention groups was **0.12 standard deviations lower** (0.34 lower to 0.1 higher)	319 [2 studies]	+++O moderate[1]	SMD -0.12 (-0.34 to 0.1)

	Exercise	**Behavioural treatment**			
Pain intensity Pain Rating Index (0-45) short-term follow-up	The mean pain intensity ranged across control groups from **17.5 to 17.8 points**	The mean pain intensity in the intervention groups was **2.31 points lower** (6.33 lower to 1.7 higher)	146 [2 studies]	++00 low[1,2]	
Pain intensity Pain Rating Index (0-45) intermediate-term follow-up	The mean pain intensity ranged across control groups from **15.2 to 15.7 points**	The mean pain intensity in the intervention groups was **1.18 points higher** (3.16 lower to 5.53 higher)	137 [2 studies]	+++0 moderate[3]	
Pain intensity Pain Rating Index (0-45) long-term follow-up	The mean pain intensity ranged across control groups from **14.9 to 16.6 points**	The mean pain intensity in the intervention groups was **0.14 points higher** (4.4 lower to 4.67 higher)	136 [2 studies]	+++0 moderate[3]	
Depression various scales short-term follow-up		The mean depression in the intervention groups was **0.25 standard deviations higher** (0.07 lower to 0.58 higher)	146 [2 studies]	++00 low[1,3]	SMD 0.25 (-0.07 to 0.58)
Depression various scales intermediate-term follow-up		The mean depression in the intervention groups was **0.02 standard deviations higher** (0.32 lower to 0.35 higher)	137 [2 studies]	+++0 moderate[3]	SMD 0.02 (-0.32 to 0.35)
Depression various scales long-term follow-up		The mean depression in the intervention groups was **0.07 standard deviations higher** (0.27 lower to 0.41 higher)	136 [2 studies]	+++0 moderate[3]	SMD 0.07 (-0.27 to 0.41)
	Surgery	**Behavioural treatment**			

Functional status (back specific) Oswestry Disability Index (0-100) long-term follow-up	The mean back-specific functional status ranged across control groups from **26.4 to 34.0 points**	The mean back-specific functional status in the intervention groups was **2.36 points higher** (1.94 lower to 6.66 higher)	345 [2 studies]	++00 low[1,4]

CI: Confidence interval

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Serious limitations in study design (i.e. >25% of participants from studies with high risk of bias)
[2] Serious inconsistency of results (i.e. opposite direction of effects and/or significant statistical heterogeneity)
[3] Serious imprecision (i.e. total number of participants <300 for each outcome)
[4] Serious indirectness (i.e. not directly applicable to all patients with chronic low-back pain)

Behavioural treatment in addition to other treatments for chronic low-back pain

Patient or population: adults with chronic low-back pain
Settings: primary, secondary, or tertiary health care settings
Intervention: behavioural treatment (respondent, cognitive, and operant therapy, or a combination) in addition to another treatment
Comparison: the other treatment alone

Outcomes	Illustrative means (95% CI)		No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Control group	Intervention group			
	Physiotherapy	Behavioural treatment + physiotherapy			
Pain intensity 5-point scale short-term follow-up	The mean pain intensity ranged across control groups from **2.7 to 3.0 points**	The mean pain intensity in the intervention groups was **0.13 points lower** (1.01 lower to 0.75 higher)	59 [2 studies]	++00 low[1,2]	
Pain intensity 5-point scale intermediate-term follow-up	The mean pain intensity ranged across control groups from **2.6 to 2.8 points**	The mean pain intensity in the intervention groups was **0.11 points lower** (0.67 lower to 0.44 higher)	45 [2 studies]	++00 low[1,2]	
Depression Beck Depression Inventory (0-63) short-term follow-up	The mean depression ranged across control groups from **12.1 to 16.4 points**	The mean depression in the intervention groups was **1.56 points higher** (1.71 lower to 4.83 higher)	59 [2 studies]	++00 low[1,2]	
Depression Beck Depression Inventory (0-63) intermediate-term follow-up	The mean depression ranged across control groups from **9.9 to 18.5 points**	The mean depression in the intervention groups was **0.17 points higher** (6.85 lower to 7.19 higher)	50 [2 studies]	++00 low[1,2]	

Functional status (generic) Sickness Impact Profile (0-136) short-term follow-up	The mean generic functional status ranged across control groups from **25.3 to 26.1 points**	The mean generic functional status in the intervention groups was **6.26 points lower** (12.71 to 0.19 lower)	59 [2 studies]	++00 low[1,2]	
Functional status (generic) Sickness Impact Profile (0-136) intermediate-term follow-up	The mean generic functional status ranged across control groups from **19.4 to 25.3 points**	The mean generic functional status in the intervention groups was **0.93 points lower** (6.71 lower to 4.84 higher)	51 [2 studies]	++00 low[1,2]	
	Inpatient rehabilitation	**Behavioural treatment + in-patient rehabilitation**			
Pain intensity various scales short-term follow-up		The mean pain intensity in the intervention groups was **0.14 standard deviations lower** (0.34 lower to 0.05 higher)	405 [2 studies]	+++0 moderate[1]	SMD -0.14 (-0.34 to 0.05)

CI: Confidence interval

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Serious limitations in study design (i.e. >25% of participants from studies with high risk of bias)
[2] Serious imprecision (i.e. total number of participants <300 for each outcome)

DISCUSSION

Thirty RCTs were included in this updated systematic review. In addition to eleven new trials identified since the previous review, an updated risk of bias assessment was performed, and the GRADE approach was used to judge the quality of the evidence. Generally, the included trials were small in size, did not provide long-term outcomes, and suffered from poor reporting of a number of methodological quality items. None of the comparisons made in this systematic review provided high quality evidence, either for or against behavioural treatment. For most of the comparisons made in this review, there was only low or very low quality evidence to support the results. Moderate quality evidence was found for a small number of comparisons evaluated in this updated systematic review. Operant therapy was found to be more effective than a waiting list control for short-term pain relief. No significant differences were found for pain relief when operant therapy was compared to cognitive therapy or to a combined behavioural treatment. There was also moderate quality evidence that behavioural treatment is more effective than usual care for pain relief in the short-term, but there was no difference in the intermediate-term. There were no differences between behavioural treatment and a usual care package on back-pain related disability over the short- to intermediate-term. When behavioural treatment was compared to a group exercise program, no significant difference was found for pain relief or symptoms of depression over the intermediate- to long-term. There was moderate quality evidence that the addition of behavioural treatment to an inpatient pain management program is no more effective than an inpatient program alone for short-term pain relief.

The rationale upon which behavioural treatments for CLBP are based recognises that the aim is not to treat the pain directly, but to attempt to modify one of the three response systems: behaviour, cognition, or physiological reactivity (Vlaeyen 1995). Therefore, if a behavioural treatment is effective, we would expect to see significant changes in outcomes which measure one of these three responses, but not necessarily a reduction in pain intensity. A large number and variety of behavioural outcomes were measured by the included trials in this review, which served to express the diversity of behavioural constructs which are considered part of the three response systems. Behavioural outcomes measuring the effect of treatment on coping skills, pain behaviours, mood, or social role functioning were identified, in most cases however, these were only measured in individual studies. This variation in outcome measures hampered the comparability between studies in this domain. Apart from pain intensity and functional status (generic and back specific), depressive symptoms (measured by the Beck Depression Inventory) were the only other outcomes to be pooled in the meta-analyses. Despite general recommendations that RCTs on interventions for CLBP measure pain intensity and pain-related disability to evaluate the efficacy of treatment (Bombardier 2000), there is still uncertainty regarding the appropriate behavioural outcomes to use. A recent study compared 20 psychological constructs in predicting LBP outcome in primary care and found that four factors (personal control, acute/chronic timeline, illness identify, and pain self-efficacy) were strongly predictive of outcome after six months of follow-up (Foster 2010). These appeared to be better predictors of disability than more common targets for behavioural treatment such as fear avoidance, catastrophising, and depression. Future research is needed to translate this information about the most predictive psychological factors to recovery into targeted interventions and improved outcomes for patients.

While the evidence from the current review shows a promising effect on pain relief in favour of behavioural therapy, it is important to consider how the interventions are applied in practice and the clinical importance of the effect size before it is recommended as part of CLBP management (Malmivaara 2006). The applicability and clinical relevance of the results deal with the questions of whether and how to use the evidence in practice. It is difficult to provide firm definitions of the therapeutic components of behavioural treatment, such as cognitive or operant methods, as there are often considerable differences in the descriptions of the applied interventions between trials. However, in this review, it was found that most of the included RCTs provided adequate descriptions of the interventions (90%), the patient characteristics (87%), and outcome measures (87%) to assess whether the results are applicable to other populations. A clinically significant effect size was defined as a 30% difference between groups in pain intensity, as measured with a visual analogue scale (VAS); and an improvement of 8% to12% in functional ability measured with the Roland-Morris Disability Questionnaire (RMDQ) (Furlan 2009). However, because many of the included trials did not measure pain and functional ability with a VAS or the RMDQ (respectively), there was usually insufficient information to determine if the size of the effect was clinically significant. While these outcomes are recommended for use in most CLBP trials (Bombardier 2000), further research is needed to expand the definition of a clinically significant effect size, especially when measured with other commonly used outcome measures. Only one of the included RCTs (van der Roer 2008) reported on whether any adverse effects related to the intervention were observed. This made it difficult to determine whether the benefits gained from behavioural treatment are worth the potential harms. From the results of the meta-analyses, it can be seen that behavioural treatment for CLBP generally results in small effect sizes. However, as behavioural treatment can be considered inherently harmless, small benefits could be considered worthwhile if they prove to be cost-effective.

Risk of Bias

The risk of bias of the trials included in this review was generally high (Figure 1). Considering the nature of behavioural therapy, blinding of patients and care providers is difficult, if not impossible in some cases. Many of the other criteria used to assess risk

of bias were not reported, especially details about the randomisation procedure and concealment, compliance, and tracking of co-interventions. In order for an RCT to be considered to have a low risk of bias, adequate reporting of essential methodological quality criteria is required. There was a trend for more recent studies (e.g. Smeets 2006, van der Roer 2008) to fulfil more of these criteria, which can be attributed to an increasing awareness of methodological issues and reporting guidelines (Plint 2006). The limitations found in the design and reporting of the included RCTs contributed to the overall judgment of the quality of evidence, using the GRADE approach, and served to downgrade the quality for most of the comparisons.

Data-analysis

A number of important changes in data collection and analysis have been made to the review since the previous version, the most significant of which was the method of grading the quality of the evidence for each comparison. Using a "levels of evidence" approach, the previous version of this review (Ostelo 2005) found moderate evidence that respondent therapy (using progressive relaxation) has a large positive effect on pain intensity and behavioural outcomes in the short-term. In the same comparison performed in the current review, there was only low quality evidence to support a positive effect of respondent therapy (using progressive relaxation) on pain relief. This change in the strength of evidence was not a result of the addition of new trials to the meta-analysis, but of applying the GRADE approach, which considers limitations in the design of trials as well as factors such as inconsistency, imprecision, and indirectness. We can see from the case of respondent therapy with progressive relaxation that there were serious limitations in the design of the pooled trials but also sparse data. The small sample sizes of the pooled trials meant there was serious imprecision in the estimate and thus the evidence was downgraded from moderate to low quality.

In an effort to remain consistent when applying the GRADE approach to a meta-analysis, provisional cut-offs were made for each factor. For example, when coming to a decision about the extent of limitations in the design of pooled trials, we downgraded the quality of evidence if more than 25% of the pooled sample came from studies with a high risk of bias. The evidence for a comparison was considered inconsistent if significant statistical heterogeneity existed, and imprecise if the pooled sample size was less than 300. While these cut-offs have not been empirically tested, they can be considered to have face validity and their intention was to ensure the consensus process remained consistent and manageable. It is important to note that small variations in the decision making process when applying the GRADE approach can lead to changes in the reported quality of evidence.

AUTHORS' CONCLUSIONS

Implications for practice

Operant therapy was found to be an effective treatment modality for short-term pain reduction in patients with chronic low-back pain when compared to a waiting list. However, no significant differences were detected when operant therapy was compared to other types of behavioural treatment. Behavioural treatment was found to be more effective than usual care for pain relief in the short-term, but no differences were found over the long-term or on functional status measures. No differences were found between behavioural treatment and a group exercise program on pain relief or for symptoms of depression. It is still unknown what type of patients benefit most from what type of behavioral treatment. Whether clinicians should refer chronic low-back pain patients to behavioural treatments or to other active conservative treatments cannot be concluded from this review.

Implications for research

More fundamental or basic research is warranted to identify which psychological factors have the strongest influence on a patient's experience of LBP and which of these factors can be utilised as appropriate outcome measures. Only after these factors are better understood can possible mechanisms of behavioural therapy on pain relief be determined and subsequent improvements made to the interventions. In future trials, we advocate the use of valid and reliable outcome measures in the low-back pain field and also a move to determine the most reliable and valid outcome measures in the behavioural domain. Finally, in future studies, behavioural treatment should be compared to other active treatments for CLBP, and a cost-effectiveness analysis should be included.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions of Steven Linton, Rob Scholten, and Dirk Knol in the previous versions of this systematic review.

REFERENCES

References to studies included in this review

Altmaier 1992 *{published data only}*

Altmaier EM, Lehmann TR, Russell DW, Weinstein JN, Feng Kao C. The effectiveness of psychological interventions for the rehabilitation of low back pain: a randomized controlled trial evaluation. *Pain* 1992;**49**:329–35.

Basler 1997 *{published data only}*

Basler HD, Jäkle C, Kröner-Herwig B. Incorporation of cognitive-behavioral treatment into the medical care of chronic low back patients: a controlled randomized study in German pain treatment centers. *Patient Education & Counseling* 1997;**31**:113–24.

Brox 2003 *{published data only}*

Brox JI, Sørensen R, Friis A, Nygaard Ø, Indahl A, Keller A, et al. Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine* 2003;**28**(17):1913–21.

Bush 1985 *{published data only}*

Bush C, Ditto B, Feuerstein M. A controlled evaluation of paraspinal EMG biofeedback in the treatment of chronic low back pain. *Health Psychology* 1985;**4**:307–21.

Donaldson 1994 *{published data only}*

Donaldson S, Romney D, Donaldson M, Skubick D. Randomized study of the application of single motor unit biofeedback training to chronic low back pain. *J Occup Rehabil* 1994;**4**:23–37.

Fairbank 2005 *{published data only}*

Fairbank J, Frost H, Wilson-MacDonald J, Yu LM, Barker K, Collins R, Spine Stabilisation Trial Group. Randomised controlled trial to compare surgical stabilisation of the lumbar spine with an intensive rehabilitation programme for patients with chronic low back pain: the MRC spine stabilisation trial. *British Medical Journal* 2005;**330**(7505):1233–9.

Friedrich 1998 *{published data only}*

Friedrich M, Gittler G, Arendasy M, Friedrich KM. Long-term effect of a combined exercise and motivational program on the level of disability of patients with chronic low back pain. *Spine* 2005;**30**(9):995–1000.

* Friedrich M, Gittler G, Halberstadt Y, Cermak T, Heiller I. Combined exercise and motivation program: effect on the compliance and level of disability of patients with chronic low back pain: a randomized controlled trial. *Arch Phys Med Rehabil* 1998;**79**:475–87.

Johnson 2007 *{published data only}*

Johnson RE, Jones GT, Wiles NJ, Chaddock C, Potter RG, Roberts C, et al. Active exercise, education, and cognitive behavioral therapy for persistent disabling low back pain: a randomized controlled trial. *Spine* 2007;**32**(15):1578–85.

Kole-Snijders 1996 *{published and unpublished data}*

* Goossens MEJB, Rutten-van Mölken MPMH, Kole-Snijders AMJ, Vlaeyen JWS, van Breukelen G, Leidl R. Health economic assessment of behavioural rehabilitation in chronic low back pain: a randomised clinical trial. *Health Economics* 1998;**7**:39–51.

Kole-Snijders AMJ, Vlaeyen JWS, Goossens MEJB. Chronic low back pain: what does cognitive coping skills training add to operant behavioral treatment? Results of a randomized clinical trial. *J Consult Clin Psychol* 1999;**67**:931–44.

Vlaeyen JWS, Kole-Snijders AMJ, Goossens MEJB, Rutten-van Mölken MPMH. Clinical treatment of chronic low back pain: the effectiveness of a cognitive treatment and relaxation therapy in addition to an operant treatment. Report, Institute for Rehabilitation Research, Hoensbroek, The Netherlands: 1996.

Leeuw 2008 *{published data only}*

Leeuw M, Goossens MEJB, van Breukelen GJP, de Jong JR, Heuts PHTG, Smeets RJEM, et al. Exposure in vivo versus operant graded activity in chronic low back pain patients: results of a randomized controlled trial. *Pain* 2008;**138**:192–207.

Linton 1989 *{published data only}*

Linton SJ, Bradley LA. An 18-month follow-up of a secondary prevention program for back pain: help and hindrance factors related to outcome maintenance. *Clin J Pain* 1992;**8**:227–36.

* Linton SJ, Bradley LA, Jensen I, Spangfort E, Sundell L. The secondary prevention of low back pain: a controlled study with follow-up. *Pain* 1989;**36**:197–207.

Linton 2008 *{published data only}*

Linton SJ, Boersma K, Jansson M, Overmeer T, Lindblom K, Vlaeyen JWS. A randomized controlled trial of exposure in vivo for patients with spinal pain reporting fear of work-related activities. *European Journal of Pain* 2008;**12**:722–30.

McCauley 1983 *{published data only}*

McCauley JD, Thelen MH, Frank RG, Willard RR, Callen E. Hypnosis compared to relaxation in the outpatient management of chronic low back pain. *Arch Phys Med Rehabil* 1983;**64**:548–52.

Newton-John 1995 *{published data only}*

Newton-John TR, Spence SH, Schotte D. Cognitive behavioural therapy versus EMG biofeedback in the treatment of chronic low back pain. *Behav Res Ther* 1995;**33**:691–7.

Nicholas 1991 *{published data only}*

Nicholas MK, Wilson PH, Goyen J. Operant-behavioural and cognitive behavioural treatment for chronic low back pain. *Behav Res Ther* 1991;**29**:225–38.

Nicholas 1992 *{published data only}*

Nicholas MK, Wilson PH, Goyen J. Comparison of cognitive behavioral group treatment and an alternative non-psychological treatment for chronic low back pain. *Pain* 1992;**48**:339–47.

Nouwen 1983 *{published data only}*

Nouwen A. EMG biofeedback used to reduce standing levels of paraspinal muscle tension in chronic low back pain. *Pain* 1983;**17**:353–60.

Poole 2007 *{published data only}*

Poole H, Glenn S, Murphy P. A randomised controlled study of reflexology for the management of chronic low back pain. *European Journal of Pain* 2007;**11**:878–887.

Rose 1997 *{published data only}*

Rose MJ, Reilly JP, Pennie B, Bowen-Jones K, Stanley IM, Slade PD. Chronic low back pain rehabilitation programs: a study of the optimum duration of treatment and a comparison of group and individual therapy. *Spine* 1997;**22**:2246–53.

Schweikert 2006 *{published data only}*

Schweikert B, Jacobi E, Seitz R, Cziske R, Ehlert A, Knab J, et al.Effectiveness and cost-effectiveness of adding a cognitive behavioral treatment to the rehabilitation of chronic low back pain. *Journal of Rheumatology* 2006;**33**:2519–26.

Smeets 2006 *{published data only}*

Smeets RJ, Severens JL, Beelen S, Vlaeyen JW, Knottnerus JA. More is not always better: Cost-effectiveness analysis of combined, single behavioral and single physical rehabilitation programs for chronic low back pain. *European Journal of Pain* 2010;**13**:71–81.

Smeets RJEM, Vlaeyen JWS, Hidding A, Kester ADM, van der Heijden GJMG, Knottnerus JA. Chronic low back pain: physical training, graded activity with problem solving training, or both? The one-year post-treatment results of a randomized controlled trial. *Pain* 2008;**134**(3):263–76.

* Smeets RJEM, Vlaeyen JWS, Hidding A, Kester ADM, van der Heijden GJMG, van Geel ACM, et al.Active rehabilitation for chronic low back pain: cognitive-behavioral, physical, or both? First direct post-treatment results from a randomized controlled trial. *BMC Musculoskeletal Disorders* 2006;**7**:5.

Smeets RJEM, Vlaeyen JWS, Kester ADM, Knottnerus JA. Reduction of pain catastrophizing mediates the outcome of both physical and cognitive-behavioral treatment in chronic low back pain. *The Journal of Pain* 2006;**7**(4):261–71.

Strong 1998 *{published data only}*

Strong J. Incorporating cognitive-behavioural therapy with occupational therapy: a comparative study with patients with low back pain. *J Occup Rehabil* 1998;**8**:61–71.

Stuckey 1986 *{published data only}*

Stuckey SJ, Jacobs A, Goldfarb J. EMG biofeedback training, relaxation training, and placebo for the relief of chronic back pain. *Percept Motor Skills* 1986;**63**:1023–36.

Turner 1982 *{published data only}*

Turner JA. Comparison of group progressive-relaxation training and cognitive-behavioral group therapy for chronic low back pain. *J Consult Clin Psychol* 1982;**50**:757–65.

Turner 1988 *{published data only}*

* Turner JA, Clancy S. Comparison of operant behavioral and cognitive-behavioral group treatment for chronic low back pain. *J Consult Clin Psychol* 1988;**56**:261–66.

Turner JA, Clancy S. Strategies for coping with chronic low back pain: Relationship to pain and disability. *Pain* 1986;**24**:355–64.

Turner 1990 *{published data only}*

Turner JA, Clancy S, McQuade KJ, Cardenas DD. Effectiveness of behavioral therapy for chronic low back pain: a component analysis. *J Consult Clin Psychol* 1990;**58**:573–9.

Turner 1993 *{published data only}*

Turner JA, Jensen MP. Efficacy of cognitive therapy for chronic low back pain. *Pain* 1993;**52**:169–77.

van den Hout 2003 *{published and unpublished data}*

van den Hout JHC. *To solve or not to solve? Effects of problem solving and graded activity in non-specific low back pain*. Thesis Maastricht Univeristy, The Netherlands, 2002.

* van den Hout JHC, Vlaeyen JWS, Heuts PH, Zijlema JH, Wijnen JA. Secondary prevention of work-related disability in nonspecific low back pain: does problem solving therapy help?. *Clin J Pain* 2003;**19**(2):87–96.

van der Roer 2008 *{published data only}*

* van der Roer N, van Tulder MW, Barendse JM, Knol D, van Mechelen W, de Vet HCW. Intensive group training protocol versus guideline physiotherapy for patients with chronic low back pain: a randomised controlled trial. *European Spine Journal* 2008; **17**:1193–200.

van der Roer N, van Tulder MW, Barendse JM, van Mechelen W, Franken WK, Ooms AC, et al.Cost-effectiveness of an intensive group training protocol compared to physiotherapy guideline care for sub-acute and chronic low back pain: design of a randomised controlled trial with an economic evaluation. *BMC Musculoskeletal Disorders* 2004;**5**:45.

van der Roer N, van Tulder MW, van Mechelen W, de Vet HCW. Economic evaluation of an intensive group training protocol compared with usual care physiotherapy in patients with chronic low back pain. *Spine* 2008;**33**(4):445–51.

von Korff 2005 *{published data only}*

Von Korff M, Balderson BHK, Saunders K, Miglioretti DL, Lin EHB, Berry S, et al. A trial of an activating intervention for chronic back pain in primary care and physical therapy settings. *Pain* 2005;**113**:323–30.

References to studies excluded from this review

Asfour 1990 *{published data only}*

Asfour SS, Khalil TM, Waly SM, Goldberg ML, Rosomoff RS, Rosomoff HL. Biofeedback in back muscle strengthening. *Spine* 1990;**15**:510–3.

Brox 2006 *{published data only}*

Brox JI, Reikeras O, Nygaard Ø, Sørensen R, Indahl A, Holm I, et al.Lumbar instrumented fusion compared with cognitive intervention and exercises in patients with chronic back pain after previous surgery for disc herniation: A prospective randomized controlled study. *Pain* 2006;**122**:145–55.

Bru 1994 *{published data only}*

Bru E, Mykletun RJ, Terjesen Berge W, Svebak S. Effects of different psychological interventions on neck, shoulder and low back pain in female hospital staff. *Psychology and Health* 1994;**9**: 371–82.

Carson 2005 *{published data only}*

Carson JW, Keefe FJ, Lynch TR, Carson KM, Goli V, Fras AM, et al. Loving-kindness meditation for chronic low back pain: results from a pilot trial. *Journal of Holistic Nursing* 2005;**23**:287–304.

Ferrari 2006 *{published data only}*

Ferrari R, Fipaldini E, Birbaumer N. Individual characteristics and results of biofeedback training and operant treatment in patients with chronic pain. *Psicoterapia Cognitiva e Comportamentale* 2006; **12**:161–79.

Field 2007 *{published data only}*

Field T, Hernandez-Reif M, Diego M, Fraser M. Lower back pain and sleep disturbance are reduced following massage therapy. *Journal of Bodywork and Movement Therapies* 2007;**11**:141–5.

Hernandez-Reif 2001 *{published data only}*
Hernandez-Reif M, Field T, Krasnegor J, Theakston H. Lower back pain is reduced and range of motion increased after massage therapy. *Intern J Neuroscience* 2001;**106**:131–45.

Hlobil 2005 *{published data only}*
Hlobil H, Staal JB, Twisk J, Köke A, Ariëns G, Smid T, et al.The effects of a graded activity intervention for low back pain in occupational health on sick leave, functional status and pain: 12-month results of a randomized controlled trial. *Journal of Occupational Rehabilitation* 2005;**15**:569–80.

Jenssen 1997 *{published data only}*
Jenssen IB, Dahlquist C, Nygren A, Royen E, Stenberg M. Treatment for "Helpless" women suffering from chronic spinal pain: a randomized controlled 18-month follow-up study. *J Occup Rehab* 1007;**7**(4):225–238.

Jenssen 2001 *{published data only}*
Jenssen IB, Bergstrom G, Ljungquist T, Bodin L, Nygren AL. A randomized controlled component analysis of a behavioral medicine rehabilitation program for chronic spinal pain: are the effects dependent on gender?. *Pain* 2001;**91**:65–78.

Jousset 2004 *{published data only}*
Jousset N, Fanello S, Bontoux L, Dubus V, Billabert C, Vielle B, et al. Effects of functional restoration versus 3 hours per week physical therapy: a randomized controlled study. *Spine* 2004;**29**(5): 487–93.

Kool 2005 *{published data only}*
Kool JP, Oesch PR, Bachmann S, Knuesel O, Dierkes JG, Russo M, et al.Increasing days at work using function-centered rehabilitation in non-acute nonspecific low back pain: a randomized controlled trial. *Archives of Physical Medicine & Rehabilitation* 2005;**86**: 857–64.

Kääpä 2006 *{published data only}*
Kääpä EH, Frantsi K, Sarna S, Malmivaara A. Multidisciplinary group rehabilitation versus individual physiotherapy for chronic nonspecific low back pain: a randomized trial. *Spine* 2006;**31**: 371–6.

Lamb 2010 *{published data only}*
Lamb SE, Hansen Z, Lall R, Castelnuovo E, Withers EJ, Nichols V, et al.Group cognitive behavioural treatment for low-back pain in primary care: a randomised controlled trial and cost-effectiveness analysis. *Lancet* 2010;**375**:916–23.

Lindell 2009 *{published data only}*
Lindell O, Johansson SE, Strender LE. Subacute and chronic, non-specific back and neck pain: cognitive-behavioural rehabilitation versus primary care. A randomized controlled trial. *BMC Musculoskeletal Disorders* 2009;**9**:172.

Lindström 1992 *{published data only}*
Lindström I, Öhlund C, Eek C, Wallin L, Peterson LE, Fordyce WE, et al.The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. *Physical Therapy* 1992; **72**(4):279–90.

Linton 2000 *{published data only}*
Linton SJ, Andersson T. Can chronic disability be prevented? A randomized trial of a cognitive-behavioral intervention and two forms of information for patients with spinal pain. *Spine* 2000;**25** (21):2825–31.

Linton 2001 *{published data only}*
Linton SJ, Ryberg M. A cognitive-behavioral group intervention as prevention for persistent neck and back pain in a non-population: a randomized controlled trial. *Pain* 2001;**90**:83–90.

Linton 2005 *{published data only}*
Linton SJ, Boersma K, Jansson M, Svärd L, Botvalde M. The effects of cognitive-behavioral and physical therapy preventive interventions on pain-related sick leave: a randomized controlled trial. *Clinical Journal of Pain* 2005;**21**:109–19.

Magnussen 2005 *{published data only}*
Magnussen L, Rognsvåg T, Tveito TH, Eriksen HR. Effect of a brief cognitive training programme in patients with long-lasting back pain evaluated as unfit for surgery. *Journal of Health Psychology* 2005;**10**:233–43.

Mangels 2009 *{published data only}*
Mangels M, Schwarz S, Worringen U, Holme M, Rief W. Evaluation of a behavioral-medical inpatient rehabilitation treatment including booster sessions: A randomized controlled study. *Clinical Journal of Pain* 2009;**25**:356–64.

Mehling 2005 *{published data only}*
Mehling WE, Hamel KA, Acree M, Byl N, Hecht FM. Randomized, controlled trial of breath therapy for patients with chronic low-back pain. *Alternative Therapies in Health and Medicine* 2005;**11**:44–52.

Menzel 2006 *{published data only}*
Menzel NN, Robinson ME. Back pain in direct patient care providers: early intervention with cognitive behavioral therapy. *Pain Management Nursing* 2006;**7**:53–63.

Mitchell 1994 *{published data only}*
Mitchell RI, Carmen GM. The functional restoration approach to the treatment of chronic pain in patients with soft tissue and back injuries. *Spine* 1994;**19**:633–42.

Moore 2000 *{published data only}*
Moore JE, Von Korff M, Cherkin D, Saunders K, Lorig K. A randomized trial of a cognitive-behavioral program for enhancing back pain self care in a primary care setting. *Pain* 2000;**88**:145–53.

Steenstra 2006 *{published data only}*
Steenstra IA, Anema JR, Bongers PM, de Vet HCW, Knol DL, van Mechelen W. The effectiveness of graded activity for low back pain in occupational healthcare. *Occupational and Environmental Medicine* 2006;**63**:718–725.

Von Korff 1998 *{published data only}*
Von Korff M, Moore JE, Lorig K, Cherkin DC, Saunders K, Gonzalez VM, et al.A randomized trial of a lay person-led self-management group intervention for back pain patients in primary care. *Spine* 1998;**23**(23):2608–15.

Vowles 2007 *{published data only}*
Vowles KE, McNeil DW, Gross RT, McDaniel ML, Mouse A, Bates M, et al. Effects of pain acceptance and pain control strategies on physical impairment in individuals with chronic low back pain. *Behavior Therapy* 2007;**38**:412–25.

Woods 2008 *{published data only}*

Woods MP, Asmundson GJG. Evaluating the efficacy of graded in vivo exposure for the treatment of fear in patients with chronic back pain: A randomized controlled clinical trial. *Pain* 2008;**136**: 271–80.

Additional references

Atkins 2004

Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, GRADE Working Group. Grading quality of evidence and strength of recommendations. *British Medical Journal* 2004;**328**:1490.

Belar 1986

Belar CD, Kibrick SA. Biofeedback in the treatment of chronic back pain. In: AD Holzman, DC Turk editor(s). *Pain management: a handbook of psychological treatment approaches*. New York: Pergamon Press, Inc, 1986:131–50.

Bombardier 2000

Bombardier C. Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 2000;**25**(24):3100–3.

Fordyce 1976

Fordyce WE. *Behavioral methods for chronic pain and illness*. St. Louis: Mosby, 1976.

Foster 2010

Foster NE, Thomas E, Bishop A, Dunn KM, Main CJ. Distinctiveness of psychological obstacles to recovery in low back pain patients in primary care. *Pain* 2010;**148**:398–406.

Furlan 2009

Furlan AD, Pennick V, Bombardier C, van Tulder MW, Editorial Board, Cochrane Back Review Group. 2009 Updated Method Guidelines for Systematic Reviews in the Cochrane Back Review Group. *Spine* 2009;**34**(18):1929–41.

Guyatt 2008

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; **336**(7650):924–6.

Higgins 2008

Higgins JPT, Deeks JJ (editors). Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S editor(s). *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updated September 2008]*. The Cochrane Collaboration, 2008, Available from www.cochrane–handbook.org.

Higgins 2009

Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.2 [updated September 2009]*. The Cochrane Collaboration, 2009, Available from www.cochrane–handbook.org.

Hoffman 2007

Hoffman BM, Papas RK, Chatkoff DK, Kerns RD. Meta-analysis of psychological interventions for chronic low back pain. *Health Psychology* 2007;**26**(1):1–9.

Koes 2006

Koes BW, van Tulder MW, Thomas S. Diagnosis and treatment of low back pain. *British Medical Journal* 2006;**332**(7555):1430–4.

Malmivaara 2006

Malmivaara AMC, Koes BW, Bouter LM, van Tulder MW. Applicability and clinical relevance of results in randomized controlled trials: the Cochrane review on exercise therapy for low back pain as an example. *Spine* 2006;**31**(13):1405–9.

Morley 1999

Morley S, Eccleston C, Williams A. Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. *Pain* 1999;**80**:1–13.

Pincus 2006

Pincus T, Vogel S, Burton AK, Santos R, Field AP. Fear avoidance and prognosis in back pain: A systematic review and synthesis of current evidence. *Arthritis & Rheumatism* 2006;**54**(12):3999–4010.

Plint 2006

Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al.Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Medical Journal of Australia* 2006;**185**(5):263–7.

Roberts 1986

Roberts AH. The operant approach to the management of pain and excess disability. In: AD Holzman, DC Turk editor(s). *Pain management: a handbook of psychological treatment approaches*. New York: Pergamon Press, Inc, 1986.

Rudy 1995

Rudy TE, Turk DC. Cognitive-behavioral treatment of the patient with chronic pain. In: AH White, JA Schofferman editor(s). *Spine Care Volume 1: Diagnosis and Conservative Treatment*. St Louis: Mosby-Year Book, Inc, 1995.

Skinner 1953

Skinner BF. *Science and human behaviour*. New York: McMillan, 1953.

Turk 1984

Turk DC, Flor H. Etiological theories and treatments for chronic back pain. II. Psychological models and interventions. *Pain* 1984; **19**:209–33.

van Tulder 2009

van Tulder MW, Suttorp M, Morton S, Bouter L, Shekelle P. Empirical evidence of an association between internal validity and effect size in randomized controlled trials of low-back pain. *Spine* 2009;**34**(16):1685–92.

Vlaeyen 1991

Vlaeyen JWS. *Chronic low back pain. Assessment and treatment from a behavioral rehabilitation perspective*. Lisse: Swets & Zeitlinger, 1991.

Vlaeyen 1995

Vlaeyen JWS, Haazen IWC, Schuerman JA, Kole-Snijders AMJ, 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* 1995;**34**:95–118.

Waddell 2004

Waddell G. *The Back Pain Revolution*. 2nd Edition. London: Churchill Livingstone, 2004.

References to other published versions of this review

Ostelo 2005

Ostelo RWJG, van Tulder MW, Vlaeyen JWS, Linton SJ, Morley SJ, Assendelft WJJ. Behavioural treatment for chronic low-back pain. *Cochrane Database of Systematic Reviews* 2005, Issue 1. [DOI: Art. No.: CD002014. DOI:10.1002/14651858.CD002014.pub2]

van Tulder 2000

van Tulder MW, Ostelo RWJG, Vlaeyen JWS, Linton SJ, Morley SJ, Assendelft WJJ. Behavioural treatment for chronic low-back pain. *Cochrane Database of Systematic Reviews* 2000, Issue 2. [Art. No.: CD002014. DOI: 10.1002/14651858.CD002014]

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Altmaier 1992

Methods	RCT; 'randomly assigned'
Participants	47 consecutive patients with CLBP, aged 18-63 years, who were disabled and not working for at least 3 months. All patients were admitted to an inpatient LBP rehabilitation program, but excluded from the study if they were involved with litigation, or had pain relating to pregnancy or severe vertebral fracture. Two patients failed to complete the program and were not included in analysis.
Interventions	Behavioural treatment (I): standard 3-week rehabilitation program with additional operant conditioning components, relaxation training, biofeedback, and cognitive-behavioural coping skills training (N = 24). Reference treatment (R): standard 3-week inpatient rehabilitation program: twice daily physical therapy, aerobic exercises, education, and vocational rehabilitation (N = 21).
Outcomes	Mean scores on McGill Pain Questionnaire pretreatment, at short-term follow-up (post-treatment), and in the intermediate-term (6 months): (I) 24.24, 23.76, and 22.66 vs. (R) 20.33, 18.05, and 18.19. No significant differences between groups on any outcome measures (pain intensity, pain interference, return to work, disability).
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	High risk	'randomly assigned'
Allocation concealment?	High risk	Inadequate
Blinding? All outcomes - patients?	High risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	

Altmaier 1992 (*Continued*)

Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	High risk	
Compliance acceptable?	Low risk	
Timing outcome assessments similar?	Low risk	

Basler 1997

Methods	RCT; 'assignment of random numbers'.
Participants	94 patients with CLBP from three pain treatment centres who had not previously received behavioural treatment. 18 patients dropped out and were not included in analysis.
Interventions	Behavioural treatment (I): cognitive-behavioural therapy (education, progressive muscle relaxation (Jacobson), modifying thoughts and feelings, pleasant activity scheduling, training of posture) plus medical treatment, groups of 5-8 patients, 12 sessions of 150 min at weekly intervals (N = 36). Reference treatment (R): medical treatment (pain medication, nerve blocks, TENS, physical therapy) (N = 40).
Outcomes	Mean (SD) scores on pain intensity (NRS) pre- and post-treatment (I) 4.6 (1.8), 4.1 (2.1) vs. (R) 4.0 (1.0), 4.2 (1.4); control over pain (I) 2.6 (1.9), 3.8 (2.5) vs. (R) 2.7 (1.7) , 2.8 (1.8); (I) significantly more improved than (R). In behavioural domain: avoidance behaviour, pleasant activities, and catastrophising significantly more improved in (I) than (R); cognitive strategies, social support, philosophical beliefs, and active coping had no differences between groups. In functional domain: social roles, physical functions, mental performance significantly more improved in (I); social relations and physical performance no differences between groups.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'assignment of random numbers'.
Allocation concealment?	High risk	Inadequate
Blinding? All outcomes - patients?	High risk	

Basler 1997 (*Continued*)

Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing outcome assessments similar?	Low risk	

Brox 2003

Methods	RCT; 'computer-generated random list'
Participants	64 patients, aged 25-60 years, with CLBP referred for spine surgery. All had pain lasting longer than 1 year, a score of at least 30 on the ODI, and evidence of disc degeneration at L4-L5 and/or L5-S1 on radiographic examination. Exclusion criteria were widespread pain, pelvic pain, spinal stenosis, recurrent herniation, generalised disc degeneration, previous surgery or fracture, and other serious disease. 3 patients were lost to follow-up at 1 year and were not included in the analysis.
Interventions	Behavioural treatment (I): cognitive intervention consisting of a lecture; reinforced by three daily physical exercise sessions (~25 hours/week) for 3 weeks (N = 26). Reference treatment (R): instrumented lumbar fusion with posterior transpedicular screws followed by post-operative rehabilitation at the choice of the surgeon (N = 35).
Outcomes	Overall improvements in ODI at the 1-year follow-up visit did not differ significantly between (I) and (R). The mean difference (95%CI) in change between groups was 2.3 (-6.8 to 11.4) in favour of (R). Fear avoidance beliefs and fingertip-floor distance were significantly improved more after (I) than (R), and lower limb pain was significantly reduced more after (R) than (I). The success rate according to an independent observer was 70% after (R) and 76% after (I). The early complication rate in the (R) group was 18%.
Notes	

Brox 2003 (*Continued*)

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'computer-generated random list'
Allocation concealment?	Low risk	Adequate
Blinding? All outcomes - patients?	High risk	not reported, unable to blind due to nature of intervention
Blinding? All outcomes - providers?	High risk	as above
Blinding? All outcomes - outcome assessors?	High risk	"physical therapist and a specialist in physical medicine and rehabilitation carried out blind follow-up measurements"
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	Figure 1
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	"The results were analysed according to the method of intention-to-treat"
Free of selective reporting?	Low risk	all pre-specified outcomes are reported
Similarity of baseline characteristics?	Low risk	Table 1, 2
Co-interventions avoided or similar?	Unclear risk	Unclear from text
Compliance acceptable?	High risk	"Four patients randomised to lumbar fusion did not have surgery. Two patients randomised to cognitive intervention and exercises did not attend treatment. Additionally, 1 patient dropped out after the first treatment period and had surgery."
Timing outcome assessments similar?	Low risk	Figure 1

Bush 1985

Methods	RCT; 'randomly assigned...matched for age and sex'.
Participants	72 patients with CLBP (minimum of 2 years history, twice weekly LBP) who responded to media announcements and were aged 20-65 years. Patients were excluded if receiving compensation, had clinical depression, psychosis, or previous low-back surgery. 6 patients

Bush 1985 (*Continued*)

	withdrew before completion of the study and were not included in the analysis.
Interventions	Behavioural treatment (I): auditory EMG biofeedback training in sitting position until decrease and increase of 2 μv without feedback was reached, max. 8 sessions (N = 22). Reference treatment (R1): placebo feedback of back temperature (N = 22). Reference treatment (R2): waiting list control (N = 22).
Outcomes	No significant differences between groups on pain intensity, functional status, or psychosocial status at short-term (post-treatment) or intermediate-term (3 months) follow-up. No data presented.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Unclear risk	'randomly assigned...matched for age and sex'
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	Low risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	Low risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	Unclear risk	Unclear from text
Compliance acceptable?	Unclear risk	Unclear from text
Timing outcome assessments similar?	Low risk	

Donaldson 1994

Methods	RCT; 'randomly assigned'.
Participants	36 volunteers, recruited through advertisements, with CLBP in the area T8-S1 for at least 1 year, aged 18-55 years and experiencing daily pain. Subjects were excluded for a history of back surgery, positive straight leg raise test, loss of reflexes, lower limb weakness, scoliosis, or other serious disease.
Interventions	Behavioural treatment (I1): progressive relaxation training (Lehrer & Woolfolk), 10 35-min. sessions (N = 12). Behavioural treatment (I2): single motor unit biofeedback training (Johnson, Mulder), 10 x 35-min. sessions (N = 12). Reference treatment (R): education on anatomy, exercise, depression, and stress management, 10 x 35-min. sessions (N = 12).
Outcomes	Mean scores on McGill Pain Questionnaire (MPQ) and pain intensity (VAS) pretreatment (I1) 31.08, 2.51, (I2) 28.75, 2.23, (R) 34.50, 3.48; posttreatment (I1) 27.67, 1.90, (I2) 16.08, 1.26, (R) 28.58, 2.47; and after 3 months (I1) 32.33, 1.78, (I2) 15.33, 0.72, (R) 20.08, 0.87. (I2) significantly more improved after 3 months than (I1). No significant differences on pain intensity between groups.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Unclear risk	'randomly assigned'
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	High risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Unclear risk	Unclear from text
Incomplete outcome data addressed? All outcomes - ITT analysis?	Unclear risk	Unclear from text
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	High risk	

Donaldson 1994 (*Continued*)

Co-interventions avoided or similar?	Unclear risk	Unclear from text
Compliance acceptable?	Low risk	
Timing outcome assessments similar?	Low risk	

Fairbank 2005

Methods	RCT; 'randomisation was generated centrally by computer program'
Participants	349 participants who were considered candidates for spinal fusion, aged 18-55 years, with CLBP of at least one year duration. Patients were excluded if they had an infection or other comorbidity which contraindicated treatment, psychiatric disease, or were pregnant.
Interventions	Behavioural treatment (I): intensive rehabilitation programme of education and exercise based on principles of cognitive-behaviour therapy (identify and overcome fears and unhelpful beliefs), 5 days per week for 3 weeks (N = 173) Reference treatment (R): spinal stabilisation surgery at discretion of the surgeon (N = 176)
Outcomes	The mean (SD) ODI changed from 46.5 (14.6) to 34.0 (21.1) in the (R) group and from 44.8 (14.8) to 36.1 (20.6) in the (I) group after 24 months follow-up. The estimated difference in change (from baseline to 24 months) between the groups was -4.1 (95%CI -8.1 to -0.1, P=0.045) in favour of (R). No significant differences between groups were observed in any of the other outcome measures. Intra-operative complications occurred in 19 surgical cases.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'randomisation was generated centrally by computer program'
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	High risk	not reported
Blinding? All outcomes - providers?	High risk	"We were not able to blind the trial research therapists to patient allocation after the baseline assessment."
Blinding? All outcomes - outcome assessors?	High risk	not reported

Fairbank 2005 (*Continued*)

Incomplete outcome data addressed? All outcomes - drop-outs?	Unclear risk	Borderline acceptable drop-out rate, reasons not described
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	"We carried out an intention to treat analysis"
Free of selective reporting?	High risk	did not report DRAM at follow up
Similarity of baseline characteristics?	Low risk	Table 1
Co-interventions avoided or similar?	High risk	not similar, Table 2
Compliance acceptable?	High risk	"Forty eight (28%) patients randomised to rehabilitation had surgery by two years. Seven (4%) patients randomised to surgery had rehabilitation instead of surgery."
Timing outcome assessments similar?	Low risk	Figure, flow of participants

Friedrich 1998

Methods	RCT; 'randomly assigned'	
Participants	98 patients, aged between 20 and 60 years, with back pain of at least 4 months duration or three episodes of LBP in the last 6 months. Criteria for exclusion were cardiovascular diseases, acute lumbar radicular lesions, previous low-back surgery, other lumbar spine pathologies, spinal stenosis, high degree of instability, psychiatric disorder, pregnancy, and patients involved in litigation.	
Interventions	Behavioural treatment (I): Individual exercise program + motivational program (counselling and information, reinforcement techniques, treatment contract, exercise diary), 10 25-min sessions (N = 49) Reference treatment (R): Individual exercise program, 10 25-min sessions (N = 49)	
Outcomes	Patients in the (I) group were significantly more likely to attend their exercise therapy appointments. Mean disability (low-back outcome score) and pain intensity (VAS) pretreatment (I): 42.5, 50.2; (R): 42.8, 54.5; after four months (I): 57.2, 32.7; (R): 51.0, 39.8; and after 12 months (I): 58.9, 26.4; (R): 50.9, 41.9. There was a significant improvement in favour of (I). No significant differences were found in motivation scores, self-reported compliance with long-term exercise, or modified Waddell score.	
Notes		
Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**

Friedrich 1998 (*Continued*)

Adequate sequence generation?	Unclear risk	'randomly assigned'
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	High risk	not reported
Blinding? All outcomes - providers?	High risk	not reported
Blinding? All outcomes - outcome assessors?	High risk	not reported
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	Figure 1
Incomplete outcome data addressed? All outcomes - ITT analysis?	Unclear risk	Unclear from text
Free of selective reporting?	Low risk	pre-specified outcomes reported
Similarity of baseline characteristics?	Low risk	"There were no significant differences between the motivational and control groups on socio-demographic variables, pain histories, disability scores, and working ability data at study entry."
Co-interventions avoided or similar?	Unclear risk	Unclear from text
Compliance acceptable?	Low risk	Table 1, same in both groups
Timing outcome assessments similar?	Low risk	Figure 1

Johnson 2007

Methods	RCT; 'randomised by means of computer generated code'
Participants	234 patients, 18 to 65 years of age, still reporting LBP 3 months after an initial GP consultation. Criteria for exclusion included having a consultation for LBP in past 6 months, presence of '"red flags", pregnancy or recent childbirth, major rheumatologic, neurologic, neoplastic, or other conditions, previous spinal surgery, major psychiatric illness, or a history of drug or alcohol abuse.
Interventions	Behavioural treatment (I): community-based treatment program using a cognitive behavioral approach (problem solving, pacing and regulation of activity, challenging distorted cognitions, identifying helpful and unhelpful thoughts), eight 2-hour group sessions over a 6-week period + educational pack (booklet + audio cassette) (N = 116) Reference treatment (R): educational pack (booklet + audio cassette) (N = 118)

Johnson 2007 (*Continued*)

Outcomes	The (I) group showed only a small and non-significant effect at reducing pain (-3.6 mm; 95%CI -8.5 to 1.2mm) and improving back pain specific function using the RMDQ (-0.6 points; 95%CI, -1.6 to 0.4). The mean incremental cost-effectiveness ratio was £5000 (U.S. $8650) per QALY.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'randomised by means of computer generated code'
Allocation concealment?	Low risk	Adequate
Blinding? All outcomes - patients?	High risk	"Given the nature of the interventions, it was not possible to blind subjects or staff to treatment allocation"
Blinding? All outcomes - providers?	High risk	as above
Blinding? All outcomes - outcome assessors?	High risk	as above
Incomplete outcome data addressed? All outcomes - drop-outs?	Unclear risk	Acceptable drop-out rate, no reasons given
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	"All analyses were conducted according to "intention to treat" (ITT) principles"
Free of selective reporting?	Low risk	all outcome measures reported
Similarity of baseline characteristics?	Low risk	Table 1
Co-interventions avoided or similar?	Unclear risk	Unclear from text
Compliance acceptable?	High risk	compliance unacceptable, 63% attended at least half of the sessions
Timing outcome assessments similar?	Low risk	Figure 1

Kole-Snijders 1996

Methods	RCT; 'before first pretreatment measurement each patient was given a number written down on a card and folded up. An independent researcher blindly drew a card and assigned it to one of the three treatments.'
Participants	148 patients with CLBP for at least 6 months, aged 18-65 years, observable pain behaviour, partner willing to participate, referred by general practitioner or specialist. Criteria to exclude patients included: illiteracy, pregnancy, involvement in litigation, alcohol or drug abuse, serious psychopathology, and specific medical pathology.
Interventions	Behavioural treatment (I1): operant treatment (graded aerobic exercises, partner involvement) and cognitive treatment plus relaxation (decreasing distorted pain cognitions, increasing self-expectations, imagery, applied relaxation) (N = 59). Behavioural treatment (I2): operant treatment plus group discussion, groups of 5 patients (N = 58). Reference treatment (R): waiting list control group (N = 31).
Outcomes	Post-treatment (I1) and (I2) had significantly less negative affect, higher activity tolerance, less pain behaviour, higher pain coping and higher pain control than (R). Post-treatment (I1) had better pain coping and pain control than (I2). No significant differences at follow-up on any outcome measure including costs and quality of life. Data in graphs.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'before first pretreatment measurement each patient was given a number written down on a card and folded up. An independent researcher blindly drew a card and assigned it to one of the three treatments.'
Allocation concealment?	Low risk	Adequate
Blinding? All outcomes - patients?	Low risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	Low risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	

Kole-Snijders 1996 (*Continued*)

Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	High risk	
Timing outcome assessments similar?	Low risk	

Leeuw 2008

Methods	RCT; 'predetermined and computer-generated randomisation schedule'
Participants	85 patients, 18-65 years of age, with CLBP reporting some disability (RMDQ > 3) and moderate pain-related fear (Tampa Scale for Kinesiophobia > 33), referred by physicians or responding to newspaper advertisements. Exclusion criteria were illiteracy, pregnancy, substance abuse, involvement in litigation, specific medical disorders preventing participation in physical exercise, and serious psychopathology.
Interventions	Behavioural treatment (I1): exposure in vivo (cognitive therapy, education, engaging in fear-provoking activities) for approximately 16 sessions (N = 42). Behavioural treatment (I2): operant graded activity (positive reinforcement of healthy behaviours, education, activity quotas) for approximately 26 sessions (N = 43).
Outcomes	No significant difference was found between (I1) and (I2) in functional disability (QBPDS) and patient specific complaints post-treatment and after an intermediate-term (6 months) follow-up. (I1) had significantly lower PHODA-SeV and PCS scores immediately and 6 months post-treatment. No difference between groups was found for daily activity level or pain intensity (MPQ) immediately post-treatment or after 6 months follow-up.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'computer-generated randomisation schedule'
Allocation concealment?	Low risk	Adequate

Leeuw 2008 (*Continued*)

Blinding? All outcomes - patients?	Unclear risk	treatment names not revealed to participants, credibility of treatments not evaluated
Blinding? All outcomes - providers?	High risk	"all teams provided both treatment arms"
Blinding? All outcomes - outcome assessors?	Unclear risk	self-report primary outcome measures, patient blinding unclear
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	reported in Figure 1. Flow of participants
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	"Intention-to-treat analyses were employed, including all patients that were originally enrolled in the study, irrespective of their completing therapy or assessments"
Free of selective reporting?	Unclear risk	various questionnaires were administered and not reported (considered irrelevant within the scope of the study)
Similarity of baseline characteristics?	Low risk	Table 1 and 2
Co-interventions avoided or similar?	Unclear risk	advised to avoid co-interventions, not checked
Compliance acceptable?	High risk	"treatment was prematurely terminated either by patient or the therapist in 12 patients (29%) of the EXP and 14 patients (33%) of the GA"
Timing outcome assessments similar?	Low risk	"Assessments occurred ... directly after termination of treatment (Post-treatment), and 6 months (Follow-up 1), and 12 months (Follow-up 2) afterwards."

Linton 1989

Methods	RCT; 'randomly assigned using a table of random numbers'.
Participants	66 female nurses, aged 20-59 years, current episode of LBP, sick listed some time during previous 2 years, currently working.
Interventions	Behavioural treatment (I): 5 week-period in a back clinic; 8 hours / day mostly in groups of 6 patients; exercise activities (walking, swimming, jogging, cycling) 4 hours / day; ergonomic education, individual physical therapy programs, behaviour therapy techniques (N = 36).

Linton 1989 (*Continued*)

	Reference treatment (R): waiting-list control: no additional active treatment (N = 30).	
Outcomes	Pain intensity (VAS) significantly better in (I) than (R) after 6-weeks and 6-months (data in graphs). Psychological outcome measures (fatigue, anxiety, sleep quality, observed pain behaviour, mood, helplessness) similar results.	
Notes		

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'randomly assigned using a table of random numbers'
Allocation concealment?	Low risk	Adequate
Blinding? All outcomes - patients?	High risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	Low risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	High risk	
Compliance acceptable?	Low risk	
Timing outcome assessments similar?	Low risk	

Linton 2008

Methods	RCT; crossover design; 'randomised according to a computer generated list'
Participants	46 patients, 18-60 years old, with chronic LBP who were deemed fearful (TSK score > 35) and had no red flags, recruited through referrals via local primary care facilities, advertisements in local newspapers, and National Insurance Authority offices.
Interventions	Behavioural treatment (I): Exposure in vivo (13 sessions of psycho-education, goal setting, graded exposure to fearful activities) and treatment as usual (N = 21). Reference treatment (R): Waiting list (median 14 weeks) and treatment as usual (N = 25).
Outcomes	Post-treatment, a significant interaction was found for disability as measured by the ADL Scale where (I) improved more than (R). Mean (SD) scores post-treatment were: (I) 31.0 (11.8); (R) 29.6 (11.4). There was no significant difference between the groups on pain intensity or pain related fear. There were 12 dropouts (8 in (I) and 4 in (R)) during the first treatment phase and an additional 4 when (R) crossed over to (I).
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'randomised according to a computer generated list'
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	High risk	not reported
Blinding? All outcomes - providers?	High risk	not reported
Blinding? All outcomes - outcome assessors?	High risk	not reported
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	not acceptable, high drop out rate ~30%
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	drop out rate too high
Free of selective reporting?	Low risk	all specified outcomes reported
Similarity of baseline characteristics?	Unclear risk	prognostic indicators different, no outcome scores

Linton 2008 (*Continued*)

Co-interventions avoided or similar?	Unclear risk	both groups had treatment as usual, uncontrolled
Compliance acceptable?	High risk	Figure 2
Timing outcome assessments similar?	Low risk	Figure 2

McCauley 1983

Methods	RCT; 'randomly assigned'.
Participants	17 patients with CLBP referred by local physicians, aged 17-73 years, who had previous unsuccessful medical treatment and had no pending litigation regarding their LBP.
Interventions	Behavioural treatment (I): progressive muscle relaxation training (Bernstein & Borkovec) and differential relaxation; 50 min / week, 8 weeks (N = 8). Reference treatment (R): self-hypnosis (Barber) and hypno-analgesic techniques; 50 min / week, 8 weeks (N = 9).
Outcomes	Mean scores on pain (VAS) pre-treatment, post-treatment, and after 3 months (I) 56.9, 39.1, 35.9 vs (R) 63.1, 43.6, 42.2. No significant differences between groups on pain intensity or depression.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Unclear risk	'randomly assigned'
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	High risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	

McCauley 1983 (*Continued*)

Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	Unclear risk	Unclear from text
Compliance acceptable?	High risk	
Timing outcome assessments similar?	Low risk	

Newton-John 1995

Methods	RCT; 'randomly assigned on the basis of alternate allocation'. Waiting list controls not randomised.
Participants	44 patients with chronic LBP, aged 18-65 years, referred by medical practitioners or self-referred through media publicity. Patients were excluded if they had other chronic pain conditions, alcohol abuse, or a history of psychosis.
Interventions	Behavioural treatment (I1): cognitive behavior therapy (education, goal setting, autogenic relaxation, cognitive pain control and restructuring techniques, homework tasks); groups of 4 subjects, 8 sessions of 1 hr, twice weekly (N = 16). Behavioural treatment (I2): electromyographic biofeedback (multiple, short criterion oriented feedback trials, sitting position, plus psycho-educational session, diaphragmatic breathing exercises), 1 hr, twice weekly, 8 sessions (N = 16). Reference treatment (R): waiting list controls (N = 12).
Outcomes	Mean (SD) pretreatment, posttreatment and after 6 months for pain: (I1) 15.72 (13.97) , 10.38 (11.37), 8.68 (10.54), (I2) 16.81 (11.66), 8.42 (6.05), 8.40 (7.31), (R) 16.37 (11.47), 17.56 (9.05), for disability: (I1) 27.25 (19.71), 18.00 (15.19), 16.38 (14.02), (I2) 22.56 (9.93), 15.12 (8.38), 23.06 (23.28), (R) 25.17 (13.80), 26.33 (17.09). No significant differences between (I1) and (I2) for pain, functional status and behavioral outcomes (depression, anxiety, coping, pain beliefs) posttreatment and after 6 months.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	High risk	'randomly assigned on the basis of alternate allocation'. Waiting list controls not randomised.
Allocation concealment?	High risk	Inadequate

Newton-John 1995 (*Continued*)

Blinding? All outcomes - patients?	High risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing outcome assessments similar?	Low risk	

Nicholas 1991

Methods	RCT; 'randomly assigned'.
Participants	58 patients with chronic LBP, aged 21-63 years, selected from patients referred from a pain clinic, and by specialists and general medical practitioners. Data available on 48 patients post-treatment and 39 at follow-up.
Interventions	Behavioural treatment (I1): operant conditioning (Fordyce) and physiotherapy; one 2 hour and one 1.5 hour session / week / 5 weeks (N = 10). Behavioural treatment (I2): behavioural and physiotherapy and progressive muscle relaxation training; one 2 hour and one 1.5 hour session / week / 5 weeks (N = 9). Behavioural treatment (I3): cognitive treatment (coping strategies) and physiotherapy; one 2 hour and one 1.5 hour session / week / 5 weeks (N = 10). Behavioural treatment (I4): cognitive treatment and physiotherapy and progressive muscle relaxation training; one 2 hour and one 1.5 hour session / week / 5 weeks (N = 8). Reference treatment (R1): physiotherapy: information, exercises and handouts (one 2 hour and one 1.5 hour session / week / 5 weeks) (N = 11). Reference treatment (R2): physiotherapy (one 2 hour and one 1.5 hour session / week / 5 weeks) and attention (5 sessions) (N = 10).

Nicholas 1991 (*Continued*)

Outcomes	Posttreatment (I1), (I2), (I3) and (I4) significantly more improved on pain intensity (6-point nominal scale), self-rated functional status (SIP) and pain beliefs than (R1) and (R2), but no significant differences after 6 and 12 months. (I1) and (I2) significantly more improved post-treatment on self-rated SIP than (I3) and (I4). No other differences between behavioural treatments after 6 and 12 months on any of the outcome measures.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Unclear risk	'randomly assigned'
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	Unclear risk	Unclear from text
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	High risk	
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing outcome assessments similar?	Low risk	

Nicholas 1992

Methods	RCT; 'randomly assigned'.
Participants	20 patients with chronic LBP, aged 20-60 years, selected from patients referred from a pain clinic, and by specialists and general medical practitioners. 2 patients dropped out from each group and were not included in analysis.
Interventions	Behavioural treatment (I): cognitive-behavioural approach, including progressive muscle relaxation training (Bernstein & Borkovec), and physiotherapy; one 2 hour and one 1.5 hour session/week, 5 weeks (N = 10). Reference treatment (R): physiotherapy: information, exercises and handouts (one 2 hour and one 1.5 hour session/week./5 weeks) and attention (5 sessions) (N = 10).
Outcomes	Mean (SD) scores of pain intensity (6-point nominal scale) and functional status (SIP) pretreatment, posttreatment and after 6 months: (I) 3.13 (0.88-), 3.07 (0.79), 2.89 (0.64) and 30.87 (12.17-), 18.81 (10.9-7), 18.30 (11.1-8) vs (R) 2.84 (0.85), 2.72 (0.77) , 2.75 (1.1-1) and 32.10 (13.4-5), 26.08 (16.40), 25.31 (14.34). Not significant. (I) significantly better posttreatment than (R) on coping strategies, pain self-efficacy and medication use. After 6 months (I) significantly better coping strategies.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Unclear risk	'randomly assigned'
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	Unclear risk	Unclear from text
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	High risk	
Co-interventions avoided or similar?	Low risk	

Nicholas 1992 (*Continued*)

Compliance acceptable?	Unclear risk	Unclear from text
Timing outcome assessments similar?	Low risk	

Nouwen 1983

Methods	RCT; 'randomly divided', 'alternation method'.
Participants	20 patients with chronic LBP, aged 20-55 years, volunteers recruited through newspaper article.
Interventions	Behavioural treatment (I): auditory and visual EMG biofeedback training in standing position, 15 sessions in 3 weeks (N = 10). Reference treatment (R): waiting list control, no treatment (N = 10).
Outcomes	Mean (SD) pain level (duration x intensity) pre- and post-treatment (I) 15.8 (9.4), 14.3 (8.6) vs (R) 18.4 (11.8), 19.1 (15.6). Not significant.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	High risk	'randomly divided', 'alternation method'
Allocation concealment?	High risk	Inadequate
Blinding? All outcomes - patients?	High risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	

Nouwen 1983 (*Continued*)

Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing outcome assessments similar?	Low risk	

Poole 2007

Methods	pragmatic RCT; randomised using a 'minimisation technique'
Participants	234 patients with chronic LBP, 18 to 65 years of age, referred from general practitioners. Patients were excluded for the following reasons: pregnancy; co-existing major medical illness; psychiatric disorder; in litigation; previous use of reflexology and contraindication to reflexology including: recent surgery and circulatory disorders of the lower limb.
Interventions	Behavioural treatment (I): Progressive muscle relaxation for six 1 hour sessions at weekly intervals for six weeks in groups of 1-4 (N = 69). Reference treatment (R): Reflexology for six treatments of approximately 1 hour duration, over a period of 6-8 weeks (N = 68). Reference treatment (R2): Usual care (N = 54).
Outcomes	There were no significant differences between groups on the primary outcome variables of pain (SF-36) and functioning (ODI) or any of the secondary outcomes over the period of the trial. Mean (SD) scores at baseline, post-treatment, and six month follow-up on the primary outcome measure of pain intensity (SF-36) were: (I) 40.7 (28.6), 37.9 (27.0) , 41.3 (28.5); (R) 44.5 (24.8), 35.0 (25.9), 41.3 (28.5); (R2) 40.6 (26.7), 48.9 (29.3), 42.7 (28.4).
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Unclear risk	Method of randomisation unclear
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	High risk	not reported, pragmatic trial
Blinding? All outcomes - providers?	High risk	as above
Blinding? All outcomes - outcome assessors?	High risk	as above

Poole 2007 (*Continued*)

Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	Figure 1, no reasons for drop-out described
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	number of patients in analysis was not reported
Free of selective reporting?	Low risk	all specified outcomes are reported
Similarity of baseline characteristics?	Low risk	Table 1, 3
Co-interventions avoided or similar?	High risk	Table 2 describes usual care throughout trial
Compliance acceptable?	High risk	Figure 1
Timing outcome assessments similar?	Low risk	Figure 1, Table 3

Rose 1997

Methods	Two part RCT: part A (individual vs. group therapy), part B (short duration vs. long duration); 'allocated randomly'.
Participants	Part A : 120 patients with CLBP, aged 18-65 years, referred from orthopaedic departments and primary care physicians. Part B : 161 patients with chronic LBP, aged 18-65 years, referred from primary care physicians. Patients who were awaiting surgery or were receiving treatment elsewhere were excluded.
Interventions	All groups received a multimodal behavioural treatment program: education, cognitive therapy, graded aerobic exercise, and relaxation therapy. In addition, the 30- and 60-hour groups of Part B swam daily. Part A: (I1) group treatment, 8-10 patients (N = 26) vs. (I2) individual treatment (N = 24). Part B (group treatment): (I3) 15-hour program (N = 22) vs. (I4) 30-hour program (N = 22) vs. (I5) 60-hour program (N = 16).
Outcomes	Part A: No significant differences between groups on pain, functional status, and psychological domain (somatic perception, depression, locus of control, self-efficacy) post-treatment or after 6 months. Data in graphs. Part B: No significant differences between groups on pain, functional status, and psychological domain (somatic perception, depression, locus of control, self-efficacy) post-treatment or after 6 months. Data in graphs.
Notes	

Risk of bias

Rose 1997 (*Continued*)

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	High risk	'allocated randomly'
Allocation concealment?	High risk	Inadequate
Blinding? All outcomes - patients?	High risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	High risk	
Compliance acceptable?	High risk	
Timing outcome assessments similar?	Low risk	

Schweikert 2006

Methods	RCT; 'randomisation performed... using Rancode Professional 3.6'
Participants	409 patients with non-specific LBP for at least 6 months, pre-screened by pension insurance administration. Exclusion criteria were severe co-morbidities, an indication of severe spinal pathology such as rheumatoid arthritis, osteoporosis, fibromyalgia, oncologic diseases, or radiologically proven intervertebral disc rupture. Further, patients were excluded if they had filed an application for early retirement or if they were unemployed for more than 12 months.
Interventions	Behavioural treatment (I): cognitive-behavioural pain management program (relaxation, distraction of attention, cognitive reappraisal of pain and stress, coping strategies) in 6 group sessions, in addition to standard inpatient rehabilitation (N = 200). Reference treatment (R): standard inpatient rehabilitation (conventional 3-week inpatient rehabilitation program consisting of physiotherapy in small groups, massage, elec-

Schweikert 2006 (*Continued*)

	trotherapeutic measures, education, twice-daily exercise program) (N = 209).
Outcomes	Post-treatment, none of the behavioral, pain, or disability outcome measures displayed significant differences between (I) and (R). After 6 months, there were no significant differences between groups in quality-adjusted life-years gained or in direct medical or non-medical costs.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'randomisation performed... using Rancode Professional 3.6'
Allocation concealment?	Low risk	Adequate
Blinding? All outcomes - patients?	High risk	"Blinding of patients, clinic physicians, and the psychologists administering the treatment was not possible due to the nature of the intervention."
Blinding? All outcomes - providers?	High risk	as above
Blinding? All outcomes - outcome assessors?	High risk	as above
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	Figure 1
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	Table 2, only analysed patients who completed treatment
Free of selective reporting?	Low risk	all specified outcomes are reported
Similarity of baseline characteristics?	Low risk	Table 1
Co-interventions avoided or similar?	Unclear risk	Unclear from text
Compliance acceptable?	Unclear risk	Unclear from text
Timing outcome assessments similar?	Low risk	Figure 1

Smeets 2006

Methods	RCT; 'randomised using permuted blocks generated by computer'
Participants	227 patients, aged 18-65 years old, with non-specific chronic LBP were recruited by general practitioners and medical specialists. 4 patients were not included in the analysis. Exclusion criteria were vertebral fracture, spinal inflammatory disease, spinal infections or malignancy, current nerve root pathology, spondylolysis or spondylolisthesis, lumbar spondylodesis, cardiovascular or metabolic disease which contraindicates intensive exercise, or a clear treatment preference.
Interventions	Behavioural treatment (I1): Cognitive-behavioural treatment (graded activity training consisting of 20 individual sessions, and problem solving training of 10 sessions) 3 times / week for 10 weeks (N = 60). Behavioural treatment (I2): Combined therapy (active physical training plus cognitive-behavioural treatment) for 19 sessions (N = 62). Reference treatment (R1): Active physical training (aerobic training, strengthening exercises) 3 times / week for 10 weeks (N = 54) Reference treatment (R2): Waiting list for 10 weeks (N = 51).
Outcomes	Post-treatment, the observed change on the RMDQ was +0.04 ± 2.90 for (R2), -2.25 ± 4.51 for (R1), -2.65 ± 4.66 for (I1) and -2.27 ± 4.19 for (I2). Significant reductions were observed in functional limitations, patient's main complaints, and pain intensity for (I1), (I2), and (R1) compared to (R2). No clinically relevant differences were found between (I2) and (R1), or between (I2) and (I1). R1, followed by I2 showed, although not significant, higher total costs than I1. Reduction of disability and gain in QALY did not differ significantly between I2 and the single treatment modalities.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'randomised using permuted blocks generated by computer'
Allocation concealment?	Low risk	Adequate
Blinding? All outcomes - patients?	High risk	"it was not possible to keep the therapists blinded. The patients could not be blinded because of ethical reasons."
Blinding? All outcomes - providers?	High risk	as above
Blinding? All outcomes - outcome assessors?	Low risk	"Assessments were carried out by blinded research assistants at baseline and immediately post-treatment"

Smeets 2006 (*Continued*)

Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	Figure 1
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	"Statistical analyses were carried out according to the intention-to-treat principle"
Free of selective reporting?	Low risk	all specified outcomes in the protocol are reported
Similarity of baseline characteristics?	Low risk	Table 1
Co-interventions avoided or similar?	Low risk	Table 2
Compliance acceptable?	Low risk	"In APT, 83% of all patients met the criterion of at least 2/3 attendance. Of all CBT-patients, 78% and 76% had a sufficient number of sessions of GA and PST respectively."
Timing outcome assessments similar?	Low risk	Figure 1

Strong 1998

Methods	RCT; 'randomly allocated', 'alternation'.
Participants	30 consecutive patients with CLBP admitted to a pain clinic. 12 patients dropped out and were not included in analysis.
Interventions	Behavioural treatment (I): inpatient pain management program (anaesthesia, psychiatry, occupational therapy, physiotherapy) plus four 2-hour psycho-educational individual treatment sessions (video, information on pain and anatomy, importance of behaviours, cognitions and emotions, pain management strategies) (N = 15). Reference treatment (R): inpatient pain management program plus 8-hour non-specific program (N = 15).
Outcomes	Mean (SD) illness behavior pretreatment, posttreatment and after 3 months: (I) 0.528 (0.665), 0.311 (0.687), 0.031 (0.913), (R) 0.231 (0.871), 0.035 (1.130), -0.038 (0.713) . Mean (SD) depressed and negative cognitions: (I) -0.33 (0.792), -3.55 (0.890), -0.033 (0.840), (R) 0.304 (0.738), 0.663 (0.762), 0.197 (1.219). Mean (SD) using acute pain strategies: (I) 0.441 (1.678), 0.278 (1.027), 0.279 (1.327), (R) -0.316 (0.890), -0.325 (0.960), 0.484 (0.894). (I) significantly better posttreatment reduction than (R) on depressed and negative cognitions.
Notes	
Risk of bias	

Strong 1998 (*Continued*)

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	High risk	'randomly allocated', 'alternation'
Allocation concealment?	High risk	Inadequate
Blinding? All outcomes - patients?	Low risk	
Blinding? All outcomes - providers?	Low risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	High risk	
Timing outcome assessments similar?	Low risk	

Stuckey 1986

Methods	RCT; 'randomly but equally assigned'.
Participants	30 patients with CLBP volunteering from an orthopaedic clinic. Patients were excluded if they were waiting for surgery, were unable to ambulate, or were not willing to practice daily. 6 patients left the study and were not included in analysis.
Interventions	Behavioural treatment (I): relaxation training: progressive relaxation, breathing techniques, autogenic training, visual imagery; 8 sessions of 45 min. (N = 8). Behavioural treatment (I2): EMG-biofeedback training; 8 sessions of 45 min. (N = 8). Reference treatment (R): Placebo EMG: no feedback, no relaxation instructions; 8 sessions of 45 min. (N = 8).

Stuckey 1986 (*Continued*)

Outcomes	Mean scores of pain intensity during function test (range 0-100) at first and last treatment session (I1) 36.8, 28.0, (I2) 26.2, 31.6, (R) 42.4, 44.4 and ADL (range 1-7); (I1) 2.4, 2.9, (I2) 2.6, 2.5, (R) 2.2, 2.4. (I1) significantly more improved on pain intensity than (I2) and (R), and significantly more improved on ADL than (I2).
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Unclear risk	'randomly but equally assigned'
Allocation concealment?	Unclear risk	Unclear
Blinding? All outcomes - patients?	Low risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	High risk	
Co-interventions avoided or similar?	Unclear risk	Unclear from text
Compliance acceptable?	Unclear risk	Unclear from text
Timing outcome assessments similar?	Low risk	

Turner 1982

Methods	RCT; 'randomly assigned'.
Participants	46 patients with chronic LBP, aged 20-63 years, referred by orthopaedic surgeons, no need for further medical or surgical treatment. 6 dropped out before first treatment session, 4 were excluded from analysis because of complicating medical and psychiatric

Turner 1982 (*Continued*)

	factors.
Interventions	Behavioural treatment (I1): progressive muscle relaxation training (Bernstein & Borkovec) (N = 14 post-treatment; N = 18 follow-up). Behavioural treatment (I2): cognitive behavioural therapy, relaxation, coping, imagery (N = 13 post-treatment; N = 16 follow-up). Reference treatment (R): waiting-list control group (N = 9).
Outcomes	Mean (SD) score on self-rated functional impairment (SIP) and pain (VAS) pretreatment (I1) 14.6 (8.2), 57.9 (21.6), (I2) 18.6 (7.9), 55.2 (24.8), (R) 20.2 (11.1), 54.0 (32.0) and posttreatment (I1) 9.1 (8.3), 42.3 (20.2), (I2) 10.2 (6.9), 36.5 (22.7) and (R) 20.2 (8.2), 77.0 (21.6). (I1) and (I2) significantly better posttreatment than (R). Pain score (I1) after 1 month significantly better than (I2), no other differences between (I1) and (I2) posttreatment, after 1 month and 1.5 year on pain, depression and functional status.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	High risk	'randomly assigned'
Allocation concealment?	High risk	Inadequate
Blinding? All outcomes - patients?	Low risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	High risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	High risk	
Compliance acceptable?	High risk	

Turner 1982 (*Continued*)

Timing outcome assessments similar?	Low risk	

Turner 1988

Methods	RCT; 'randomly assigned'.
Participants	81 patients with chronic LBP, aged 20-65 years, referred by community and pain clinic physicians or self-referred following media publicity, current marriage or cohabitation.
Interventions	Behavioural treatment (I1): aerobic exercises and operant conditioning (Fordyce), participation of spouses; 2 hours / week, 8 weeks (N = 30). Behavioural treatment (I2): systematic progressive muscle relaxation (Bernstein & Borkovec) and imagery; 2 hours / week, 8 weeks (N = 26). Reference treatment (R): waiting-list control group (N = 25).
Outcomes	Mean (SD) scores on McGill Pain Questionnaire and SIP pre-treatment, post-treatment, and after 6 and 12 months: (I1) 23.07 (12.37), 18.50 (12.43), 19.57 (15.31), 15.07 (11.62) vs (I2) 18.30 (10.43), 15.91 (11.63), 12.70 (12.75), 10.80 (6.38). Not significant. (I1) significantly better post-treatment than (R) on pain, and physical and psychosocial functioning.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	High risk	'randomly assigned'
Allocation concealment?	High risk	Inadequate
Blinding? All outcomes - patients?	Low risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	Low risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	

Turner 1988 (*Continued*)

Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	High risk	
Compliance acceptable?	High risk	
Timing outcome assessments similar?	Low risk	

Turner 1990

Methods	RCT; 'randomly assigned using a centralized random number assignment scheme'.
Participants	96 patients with chronic LBP, aged 20-65 years, referred by community and pain clinic physicians or self-referred following media publicity; current marriage or cohabitation. Exclusion criteria included current infection, cardiovascular disease, spine fracture or dislocation, spondylolisthesis, spine instability, ankylosis spondylitis, rheumatoid arthritis, connective tissue disease, history of cancer, surgery in the past year, and leg pain with sciatic tension signs.
Interventions	Behavioural treatment (I1): operant conditioning (Fordyce), participation of spouses, group discussion, role playing, feedback; 2 hour / week, 8 weeks (N = 25). Behavioural treatment (I2): operant conditioning (Fordyce), participation of spouses, group discussion, role playing, feedback; 2 hour / week, 8 weeks, plus aerobic exercise, 10-20 min., 5 times / week, 8 weeks (N = 24). Reference treatment (R1): aerobic exercise 10-20 min., 5 times / week, 8 weeks (N = 24) . Reference treatment (R2): waiting-list control group (N = 23).
Outcomes	Mean scores on McGill Pain Questionnaire, SIP, and depression pre-treatment (I1) 20.96, 7.90, 10.40; (I2) 25.54, 8.50, 12.38; (R1) 19.42, 8.42, 11.95; and (R2) 21.17, 6.24, 10.48; and post-treatment (I1) 17.71, 4.72, 8.08; (I2) 12.41, 4.59, 7.31; (R1) 17.52, 5.49, 7.38; and (R2) 20.95, 5.37, 7.03. (I2) significantly more improved than (R1) and (R2). No significant differences after 6 and 12 months between (I1), (I2) and (R1).
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'randomly assigned using a centralized random number assignment scheme'
Allocation concealment?	High risk	Inadequate
Blinding? All outcomes - patients?	High risk	

Turner 1990 (*Continued*)

Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	Low risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	High risk	
Compliance acceptable?	High risk	
Timing outcome assessments similar?	Low risk	

Turner 1993

Methods	RCT; 'randomly assigned'.
Participants	102 patients with chronic LBP, aged 20-65 years, referred by community and pain clinic physicians or self-referred following media publicity. Subjects were excluded if they had evidence of current infectious disease or cancer, rheumatoid arthritis, connective tissue disease, or indications for surgical treatment.
Interventions	Behavioural treatment (I1): cognitive therapy and relaxation training (N = 25). Behavioural treatment (I2): cognitive therapy (Beck) (N = 23). Behavioural treatment (I3): progressive muscle relaxation training (Bernstein & Borkovec) and imagery (N = 24). Reference treatment (R): waiting-list control (N = 30).
Outcomes	Mean (SD) pain score (VAS) pre- vs post-treatment (I1) 60.68 (22.04) vs 44.33 (28-.45), (I2) 56.91 (18.47) vs 36.88 (20.45), (I3) 51.29 (21.68) vs 37.88 (20.07) and (R) 50.07 (21.14) vs 48.06 (20.97). (I1), (I2) and (I3) significantly more improved than (R) . No significant differences between (I1), (I2) and (I3) post-treatment and after 6 and 12 months on pain, global measure of improvement or functional status (SIP).
Notes	
Risk of bias	

Turner 1993 (*Continued*)

Bias	Authors' judgement	Support for judgement
Adequate sequence generation?	High risk	'randomly assigned'
Allocation concealment?	High risk	Inadequate
Blinding? All outcomes - patients?	Low risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	Low risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	High risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	High risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	High risk	
Co-interventions avoided or similar?	High risk	
Compliance acceptable?	Low risk	
Timing outcome assessments similar?	Low risk	

van den Hout 2003

Methods	RCT; 'randomly assigned using a computer generated scheme only known to the logistics planner of the rehabilitation centre'
Participants	115 patients (all employers) with LBP for at least 6 weeks (68% of patients had more than 12 weeks LBP), aged 18-65 years, on sick leave with LBP but no longer than 20 weeks, no more than 120 days of sick leave in last year. Exclusion criteria were: vertebral fracture, infectious disease, rheumatoid arthritis, ankylosing spondylitis, herniated disc; predominant psychopathology; and pregnancy.
Interventions	Both treatment groups received: 19 x ½-day sessions in 8 weeks plus 3 educational sessions on LBP. Behavioural treatment (I1): operant therapy (graded activity (Fordyce)) 18 one-hour session with PT and OT training for personal-relevant activities for 30 min a week + cognitive problem-solving therapy (Nezu '86, '89) in groups in 10 x 90-min sessions +

	group education 10 x 90 min lessons (N = 58) Behavioural treatment (I2): operant therapy + group education 10 x 90 min lessons (N = 57)
Outcomes	Mean (SD) scores pretreatment, post-treatment and after 6 and 12 months: on McGill Pain Questionnaire (I1) 17.6 (10.1) 13.4 (9.6) 12.6 (10.4) 11.3 (11.0) vs. (I2) 18.1 (8.4) 15.2 (10.5) 13.6 (9.8) 14.0 (11.5) No significant differences. On RMDQ (I1) 13.7 (5.2) 9.1 (6.3) 6.2 (6.0) 6.5 (6.7) vs. (I2) 12.4 (4.8) 8.5 (5.6) 6.8 (6.3) 7.7 (6.5) Only at 12 mos (I1) statistically significant better. On Tampa Scale for Kinesiophobia (I1) 36.7 (8.7) 33.5 (7.9) 32.3 (8.2) 33.1 (9.2) vs. (I2) 37.1 (6.8) 33.3 (8.3) 33.3 (8.3) 32.8 (8.7) No significant differences, also no differences on Pain Catastrophizing Scale.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'randomly assigned using a computer generated scheme only known to the logistics planner of the rehabilitation centre'
Allocation concealment?	Low risk	Adequate
Blinding? All outcomes - patients?	Low risk	
Blinding? All outcomes - providers?	High risk	
Blinding? All outcomes - outcome assessors?	Low risk	
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	
Free of selective reporting?	Low risk	
Similarity of baseline characteristics?	Low risk	
Co-interventions avoided or similar?	Low risk	
Compliance acceptable?	Low risk	
Timing outcome assessments similar?	Low risk	

van der Roer 2008

Methods	RCT; 'randomization lists independently generated using random number tables'
Participants	114 patients, aged 18-65 years, a new episode of non-specific CLBP lasting > 12 weeks recruited by physiotherapists. Exclusion criteria were: specific LBP, e.g. infection, tumour, osteoporosis, rheumatoid arthritis, fracture, inflammatory process, radicular syndrome, or cauda equina syndrome; advice by medical practitioner not to perform physically straining activities; pregnancy or pelvic girdle pain; and legal involvement related to LBP or work disability.
Interventions	Behavioural treatment (I): Intensive group training (exercise therapy, back school, behavioural principles) 10 individual sessions + 20 group sessions (N = 60) Reference treatment (R): Guideline based treatment (treated individually and the number of treatment sessions was at the discretion of the physiotherapists) approximately 13 sessions (N = 54)
Outcomes	No statistically significant differences were found for functional status. At 26 weeks, (I) reported significantly lower pain intensity, however this difference was absent after 1 year follow-up. The cost-effectiveness planes indicated no significant differences in cost-effectiveness between the two groups.
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Low risk	'randomisation lists independently generated using random number tables'
Allocation concealment?	Low risk	Adequate
Blinding? All outcomes - patients?	High risk	"Due to the pragmatic design both patients and physiotherapists could not be blinded for the interventions."
Blinding? All outcomes - providers?	High risk	as above
Blinding? All outcomes - outcome assessors?	High risk	self-reported outcomes, patients not blinded
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	Figure 1 described and acceptable drop out rate
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	"An intention-to-treat analysis was conducted for each follow-up moment using multilevel modelling."

van der Roer 2008 (*Continued*)

Free of selective reporting?	Low risk	physical measures were not reported, but are only a secondary outcome
Similarity of baseline characteristics?	Low risk	Table 1
Co-interventions avoided or similar?	Unclear risk	co-interventions discouraged but not checked
Compliance acceptable?	High risk	"in 18% of the patients the protocol was not adequately followed"
Timing outcome assessments similar?	Low risk	Figure 1

von Korff 2005

Methods	RCT; 'patients were randomly assigned'
Participants	240 CLBP patients aged 25-64 who were mailed a screening questionnaire 8-10 weeks after a primary care back pain visit, those scoring greater than 7 on the RMDQ were eligible. Patients being considered for back surgery were ineligible.
Interventions	Behavioural treatment (I): Activating intervention (addressing fears and encouraging normal activities and physical exercise, education on self-management) in 4 individual sessions (N = 119) Reference treatment (R): Usual care (use of prescription and non-prescription pain medications, infrequent primary care visits for back pain, and use of ancillary services such as physical therapy) (N = 121)
Outcomes	The mean difference in RMDQ score between (I) and (R) adjusted for baseline RMDQ score was 2.0 at 6 months, 1.7 at 12 months, and 2.0 at 24 months (P < 0.01). The mean difference in average pain intensity ratings between (I) and (R) adjusted for baseline pain intensity rating was 0.47 at 6 months, 0.67 at 12 months, and 0.34 at 24 months (P < 0.05). (I) showed significantly greater reductions in worry and fear-avoidance beliefs than (R) at each follow-up (P < 0.01). SF-36 measures of psychological distress and social functioning did not differ between (I) and (R).
Notes	

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Adequate sequence generation?	Unclear risk	'randomly assigned'
Allocation concealment?	Unclear risk	Unclear

von Korff 2005 (*Continued*)

Blinding? All outcomes - patients?	Unclear risk	variable treatments, unsure of what patients were told
Blinding? All outcomes - providers?	High risk	not reported
Blinding? All outcomes - outcome assessors?	Unclear risk	self-report outcomes, unsure of patient blinding
Incomplete outcome data addressed? All outcomes - drop-outs?	Low risk	Figure 1
Incomplete outcome data addressed? All outcomes - ITT analysis?	Low risk	"Intent to treat analyses included all randomised participants for whom follow-up data were available"
Free of selective reporting?	Low risk	data reported on all outcome measures
Similarity of baseline characteristics?	Low risk	Table 1
Co-interventions avoided or similar?	Unclear risk	Unclear from text
Compliance acceptable?	Unclear risk	Unclear from text
Timing outcome assessments similar?	Low risk	Figure 1

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Asfour 1990	This study evaluated EMG biofeedback treatment given in addition to a standard pain and rehabilitation program. The aim of the biofeedback treatment was to increase the strength of trunk extensors. The intervention did not have a behavioural component.
Brox 2006	Included chronic LBP patients after previous surgery for disc herniation.
Bru 1994	Included patients with neck, shoulder, and low-back pain. Did not report the number of patients with CLBP.
Carson 2005	This study compared 'loving-kindness meditation' to usual care, but the reported intervention was not considered a true behavioural treatment.
Ferrari 2006	Mixed patient population with chronic neck and low-back pain. Unable to distinguish results for LBP patients separately.

(*Continued*)

Field 2007	Compared the effect of soft tissue massage to muscle relaxation therapy. Not considered true behavioural interventions.
Hernandez-Reif 2001	This study evaluated massage therapy and the relaxation therapy was included to control for potential placebo effects. Therefore there was only was instruction session. The relaxation therapy evaluated in this study was not considered a true relaxation therapy
Hlobil 2005	Inclusion criterion was non-specific LBP for at least 4 weeks prior to inclusion in the study, i.e. included sub-acute LBP patients
Jenssen 1997	In this study only women were included with spinal pain, including neck pain, or a combination of neck and back pain
Jenssen 2001	In this study patients were included with spinal pain, including neck pain, or a combination of neck and back pain
Jousset 2004	This study reported sampling 'chronic' patients, but did not explicitly state the duration of symptoms prior to entry into the study.
Kool 2005	Patients were included if they had had at least 6 weeks of sick leave in the previous 6 months, i.e. included sub-acute LBP patients
Kääpä 2006	Evaluated a multidisciplinary rehabilitation program which incorporated psychological therapy sessions. Not strictly a behavioural treatment.
Lamb 2010	Included patients with LBP > 6 weeks, i.e. sub-acute LBP patients.
Lindell 2009	Included sub-acute and chronic neck and back pain patients. Data for CLBP patients was not analysed separately.
Lindström 1992	Included patients with LBP > 6 weeks, i.e. sub-acute LBP patients.
Linton 2000	This study included patients with acute and subacute spinal pain
Linton 2001	This study included a non-patient population
Linton 2005	Included patients with non-specific neck and back pain.
Magnussen 2005	LBP patients who were considered "unfit for surgery", no further details given.
Mangels 2009	No details provided regarding the duration of LBP in the sample
Mehling 2005	Evaluated "breath therapy" versus physiotherapy, not considered to be a proper behavioural treatment.
Menzel 2006	Duration of LBP not reported.

(*Continued*)

Mitchell 1994	The functional restoration treatment evaluated in this study was not considered a true behavioural treatment. Some type of behavioural or psychosocial support was provided, but the amount of time given to each of the components of the treatment program varied in each clinic and was tailored to suit the individual patient. It is unclear if all patients received behavioural therapy and by whom.
Moore 2000	This study included patients with subacute and chronic pain
Steenstra 2006	This study included patients with subacute and chronic pain
Von Korff 1998	This study included patients with subacute and chronic pain
Vowles 2007	This study was not considered to evaluate an explicit behavioural treatment.
Woods 2008	Duration of LBP in patients prior to enrolment was not reported.

DATA AND ANALYSES

Comparison 1. Respondent therapy (progressive relaxation) versus waiting list control

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short-term)	3	74	Mean Difference (IV, Random, 95% CI)	-19.77 [-34.34, -5.20]
2 Functional status (short term)	3	74	Std. Mean Difference (IV, Fixed, 95% CI)	-0.88 [-1.36, -0.39]
3 Depression (short term)	2	58	Mean Difference (IV, Random, 95% CI)	-6.80 [-19.73, 6.12]

Comparison 2. Respondent therapy (EMG biofeedback) versus waiting list control

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short term)	3	64	Std. Mean Difference (IV, Fixed, 95% CI)	-0.80 [-1.32, -0.28]
2 Functional status (short term)	2	44	Std. Mean Difference (IV, Random, 95% CI)	-0.17 [-1.56, 1.22]

Comparison 3. Operant therapy versus waiting list control

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short term)	3	153	Std. Mean Difference (IV, Fixed, 95% CI)	-0.43 [-0.75, -0.11]
2 Functional status (short term)	2	87	Mean Difference (IV, Fixed, 95% CI)	-1.18 [-3.53, 1.18]
3 Depression (short term)	2	103	Std. Mean Difference (IV, Random, 95% CI)	-0.11 [-0.67, 0.44]

Comparison 4. Cognitive therapy versus waiting list control

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short term)	2	68	Std. Mean Difference (IV, Fixed, 95% CI)	-0.27 [-0.75, 0.22]
2 Functional status (short term)	2	68	Std. Mean Difference (IV, Fixed, 95% CI)	-0.15 [-0.64, 0.33]

Comparison 5. Cognitive-behavioural therapy versus waiting list control

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short term)	5	239	Std. Mean Difference (IV, Random, 95% CI)	-0.60 [-0.97, -0.22]
2 Functional status (short term)	4	134	Std. Mean Difference (IV, Random, 95% CI)	-0.37 [-0.87, 0.13]
3 Depression (short term)	4	194	Mean Difference (IV, Random, 95% CI)	-1.92 [-6.16, 2.32]

Comparison 6. Cognitive therapy versus operant therapy

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short term)	2	93	Std. Mean Difference (IV, Random, 95% CI)	0.41 [-0.63, 1.45]
2 Pain (intermediate term)	2	82	Std. Mean Difference (IV, Random, 95% CI)	0.35 [-0.64, 1.35]

Comparison 7. Cognitive-behavioural therapy versus cognitive therapy

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short-term)	2	61	Std. Mean Difference (IV, Random, 95% CI)	-0.24 [-1.36, 0.87]
2 Pain (intermediate-term)	2	44	Std. Mean Difference (IV, Random, 95% CI)	-0.30 [-2.59, 1.98]
3 Pain (long-term)	2	48	Std. Mean Difference (IV, Random, 95% CI)	-0.89 [-3.64, 1.87]
4 Functional status (short-term)	2	61	Mean Difference (IV, Random, 95% CI)	-2.01 [-10.02, 5.99]
5 Functional status (intermediate-term)	2	47	Mean Difference (IV, Random, 95% CI)	-3.20 [-16.44, 10.04]
6 Functional status (long-term)	2	51	Mean Difference (IV, Random, 95% CI)	-2.23 [-12.59, 8.13]
7 Depression (short-term)	2	61	Mean Difference (IV, Random, 95% CI)	-3.10 [-11.43, 5.23]
8 Depression (intermediate-term)	2	47	Mean Difference (IV, Random, 95% CI)	-4.66 [-10.94, 1.61]
9 Depression (long-term)	2	51	Mean Difference (IV, Fixed, 95% CI)	-0.64 [-4.61, 3.32]

Comparison 8. Cognitive-behavioural therapy versus operant therapy

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short-term)	3	161	Std. Mean Difference (IV, Fixed, 95% CI)	-0.15 [-0.46, 0.16]
2 Pain (intermediate-term)	3	139	Std. Mean Difference (IV, Fixed, 95% CI)	-0.23 [-0.57, 0.11]
3 Pain (long-term)	3	140	Std. Mean Difference (IV, Fixed, 95% CI)	-0.31 [-0.65, 0.03]
4 Functional status (short-term)	2	77	Std. Mean Difference (IV, Fixed, 95% CI)	0.21 [-0.24, 0.67]
5 Functional status (intermediate-term)	2	61	Std. Mean Difference (IV, Random, 95% CI)	-0.23 [-1.01, 0.55]

6 Functional status (long-term)	2	66	Std. Mean Difference (IV, Random, 95% CI)	-0.50 [-1.56, 0.56]

Comparison 9. Cognitive-behavioural therapy versus respondent therapy

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short term)	3	97	Std. Mean Difference (IV, Fixed, 95% CI)	0.09 [-0.31, 0.50]
2 Pain (intermediate term)	2	62	Std. Mean Difference (IV, Random, 95% CI)	0.47 [-0.42, 1.35]
3 Functional status (short term)	3	97	Std. Mean Difference (IV, Fixed, 95% CI)	0.38 [-0.02, 0.78]
4 Functional status (intermediate term)	2	62	Std. Mean Difference (IV, Random, 95% CI)	0.13 [-0.81, 1.07]
5 Depression (short term)	3	97	Mean Difference (IV, Fixed, 95% CI)	2.89 [0.55, 5.24]
6 Depression (intermediate term)	2	62	Mean Difference (IV, Fixed, 95% CI)	1.84 [-0.43, 4.11]

Comparison 10. Behavioural treatment versus other treatments (usual care)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short term)	2	330	Mean Difference (IV, Fixed, 95% CI)	-5.18 [-9.79, -0.57]
2 Pain (intermediate term)	2	319	Mean Difference (IV, Fixed, 95% CI)	-4.29 [-9.28, 0.69]
3 Back specific functional status (short term)	2	330	Std. Mean Difference (IV, Fixed, 95% CI)	-0.20 [-0.41, 0.02]
4 Back specific functional status (intermediate term)	2	319	Std. Mean Difference (IV, Fixed, 95% CI)	-0.12 [-0.34, 0.10]

Comparison 11. Behavioural treatment versus other treatments (group exercise)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain (short term)	2	146	Mean Difference (IV, Fixed, 95% CI)	-2.31 [-6.33, 1.70]
2 Pain (intermediate term)	2	137	Mean Difference (IV, Fixed, 95% CI)	1.18 [-3.16, 5.53]
3 Pain (long term)	2	136	Mean Difference (IV, Fixed, 95% CI)	0.14 [-4.40, 4.67]
4 Depression (short term)	2	146	Std. Mean Difference (IV, Fixed, 95% CI)	0.25 [-0.07, 0.58]
5 Depression (intermediate term)	2	137	Std. Mean Difference (IV, Fixed, 95% CI)	0.02 [-0.32, 0.35]
6 Depression (long term)	2	136	Std. Mean Difference (IV, Fixed, 95% CI)	0.07 [-0.27, 0.41]

Comparison 12. Behavioural treatment versus other treatments (surgery)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Back specific functional status (long term)	2	345	Mean Difference (IV, Fixed, 95% CI)	2.36 [-1.94, 6.66]

Comparison 13. Behavioural treatment in addition to physiotherapy versus physiotherapy

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain intensity (short term)	2	59	Mean Difference (IV, Random, 95% CI)	-0.13 [-1.01, 0.75]
2 Pain Intensity (intermediate term)	2	45	Mean Difference (IV, Fixed, 95% CI)	-0.11 [-0.67, 0.44]
3 Depression (short term)	2	59	Mean Difference (IV, Fixed, 95% CI)	1.56 [-1.71, 4.83]
4 Depression (intermediate term)	2	50	Mean Difference (IV, Random, 95% CI)	0.17 [-6.85, 7.19]
5 Functional status (short term)	2	59	Mean Difference (IV, Fixed, 95% CI)	-6.26 [-12.71, 0.19]
6 Functional status (intermediate term)	2	51	Mean Difference (IV, Fixed, 95% CI)	-0.93 [-6.71, 4.84]

Comparison 14. Behavioural treatment in addition to inpatient rehabilitation versus inpatient rehabilitation

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain intensity (short term)	2	405	Std. Mean Difference (IV, Fixed, 95% CI)	-0.14 [-0.34, 0.05]

Analysis 1.1. Comparison 1 Respondent therapy (progressive relaxation) versus waiting list control, Outcome 1 Pain (short-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 1 Respondent therapy (progressive relaxation) versus waiting list control

Outcome: 1 Pain (short-term)

Study or subgroup	Respondent therapy N	Mean(SD)	Waiting list control N	Mean(SD)	Mean Difference IV,Random,95% CI	Weight	Mean Difference IV,Random,95% CI
Turner 1982	14	42.3 (20.2)	9	77 (21.6)		31.4 %	-34.70 [-52.34, -17.06]
Stuckey 1986	8	28 (20.2)	8	44.4 (17.1)		30.3 %	-16.40 [-34.74, 1.94]
Turner 1993	17	37.9 (20.5)	18	48.1 (21)		38.3 %	-10.20 [-23.95, 3.55]
Total (95% CI)	**39**		**35**			**100.0 %**	**-19.77 [-34.34, -5.20]**

Heterogeneity: $Tau^2 = 95.01$; $Chi^2 = 4.69$, df = 2 (P = 0.10); $I^2 = 57\%$

Test for overall effect: Z = 2.66 (P = 0.0078)

Analysis 1.2. Comparison 1 Respondent therapy (progressive relaxation) versus waiting list control, Outcome 2 Functional status (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 1 Respondent therapy (progressive relaxation) versus waiting list control

Outcome: 2 Functional status (short term)

Study or subgroup	Respondent therapy N	Mean(SD)	Waiting list control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Stuckey 1986	8	-2.9 (1)	8	-2.4 (0.6)		23.4 %	-0.57 [-1.58, 0.43]
Turner 1982	14	9.1 (8.3)	9	20.2 (8.2)		27.2 %	-1.29 [-2.23, -0.36]
Turner 1993	17	4.8 (4)	18	9.6 (7.3)		49.5 %	-0.79 [-1.48, -0.10]
Total (95% CI)	**39**		**35**			**100.0 %**	**-0.88 [-1.36, -0.39]**

Heterogeneity: $Chi^2 = 1.18$, df = 2 (P = 0.55); $I^2 = 0.0\%$

Test for overall effect: Z = 3.53 (P = 0.00041)

Analysis 1.3. Comparison 1 Respondent therapy (progressive relaxation) versus waiting list control, Outcome 3 Depression (short term).

Review: Behavioural treatment for chronic low-back pain

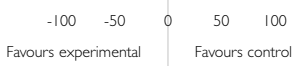
Comparison: 1 Respondent therapy (progressive relaxation) versus waiting list control

Outcome: 3 Depression (short term)

Study or subgroup	Respondent therapy N	Mean(SD)	Waiting list control N	Mean(SD)	Mean Difference IV,Random,95% CI	Weight	Mean Difference IV,Random,95% CI
Turner 1982	14	8.1 (6.2)	9	22.4 (14)		43.6 %	-14.30 [-24.01, -4.59]
Turner 1993	17	6.2 (3.4)	18	7.2 (4.9)		56.4 %	-1.00 [-3.78, 1.78]
Total (95% CI)	**31**		**27**			**100.0 %**	**-6.80 [-19.73, 6.12]**

Heterogeneity: $Tau^2 = 75.18$; $Chi^2 = 6.67$, df = 1 (P = 0.01); $I^2 = 85\%$

Test for overall effect: Z = 1.03 (P = 0.30)



Analysis 2.1. Comparison 2 Respondent therapy (EMG biofeedback) versus waiting list control, Outcome 1 Pain (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 2 Respondent therapy (EMG biofeedback) versus waiting list control

Outcome: 1 Pain (short term)

Study or subgroup	Respondent therapy N	Mean(SD)	Waiting list control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Newton-John 1995	16	8.4 (6.1)	12	17.6 (9.1)		40.1 %	-1.19 [-2.01, -0.37]
Nouwen 1983	10	14.3 (8.6)	10	19.1 (15.6)		34.4 %	-0.36 [-1.25, 0.52]
Stuckey 1986	8	31.6 (13.5)	8	44.4 (17.1)		25.5 %	-0.79 [-1.81, 0.24]
Total (95% CI)	**34**		**30**			**100.0 %**	**-0.80 [-1.32, -0.28]**

Heterogeneity: $Chi^2 = 1.79$, df = 2 (P = 0.41); $I^2 = 0.0\%$

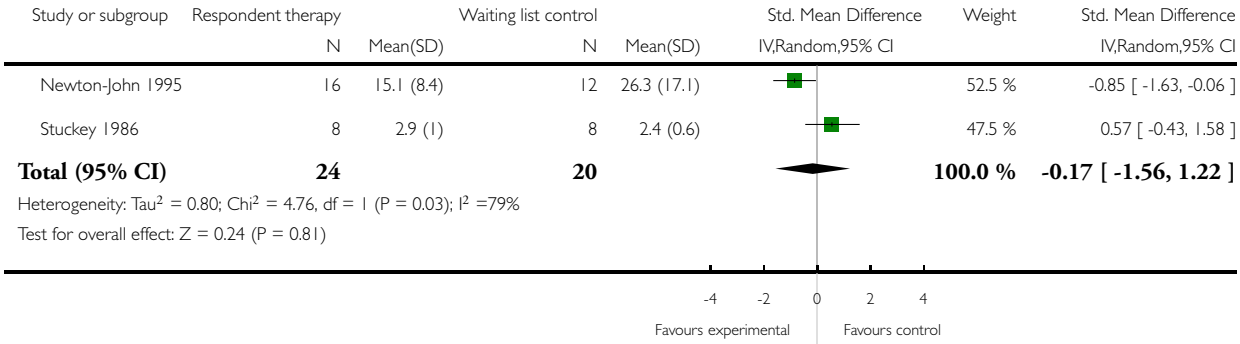
Test for overall effect: Z = 3.03 (P = 0.0025)

Analysis 2.2. Comparison 2 Respondent therapy (EMG biofeedback) versus waiting list control, Outcome 2 Functional status (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 2 Respondent therapy (EMG biofeedback) versus waiting list control

Outcome: 2 Functional status (short term)

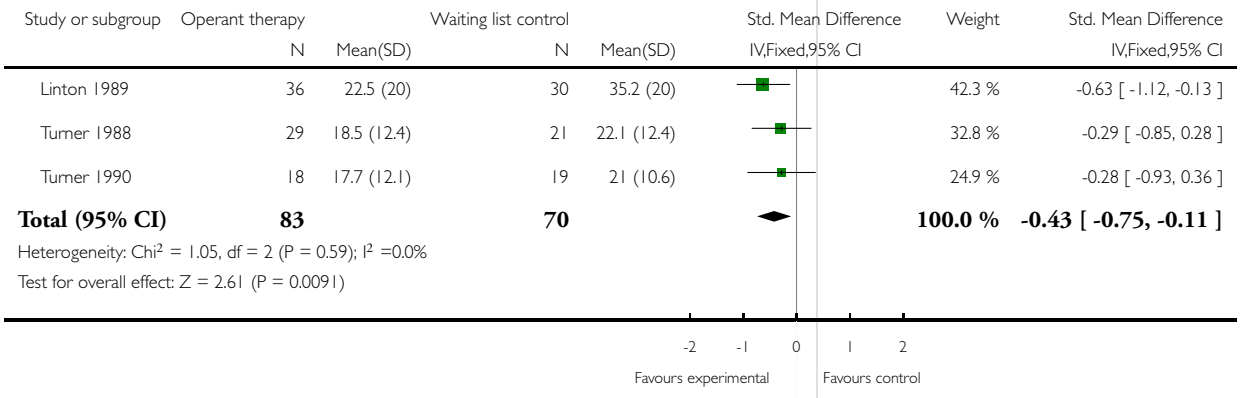


Analysis 3.1. Comparison 3 Operant therapy versus waiting list control, Outcome 1 Pain (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 3 Operant therapy versus waiting list control

Outcome: 1 Pain (short term)

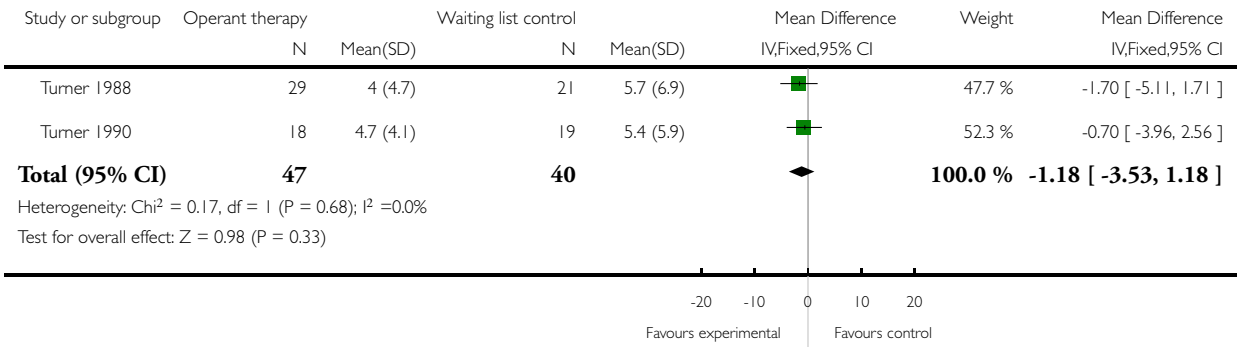


Analysis 3.2. Comparison 3 Operant therapy versus waiting list control, Outcome 2 Functional status (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 3 Operant therapy versus waiting list control

Outcome: 2 Functional status (short term)

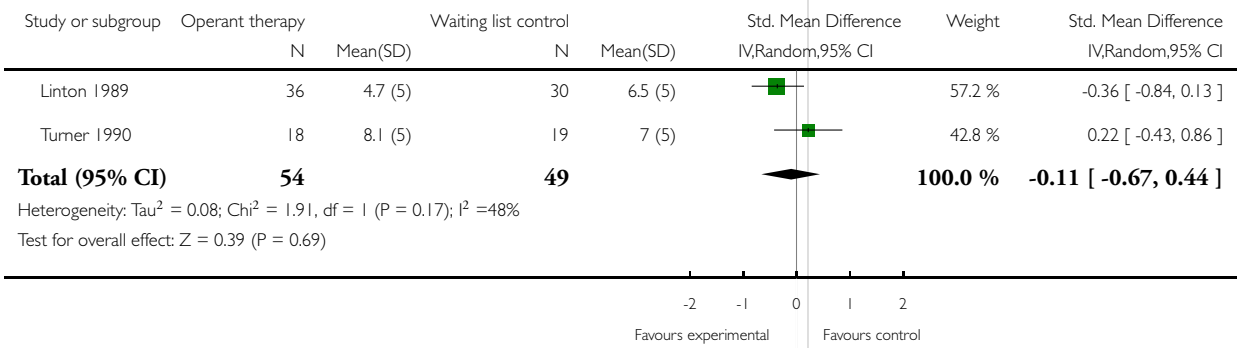


Analysis 3.3. Comparison 3 Operant therapy versus waiting list control, Outcome 3 Depression (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 3 Operant therapy versus waiting list control

Outcome: 3 Depression (short term)



Analysis 4.1. Comparison 4 Cognitive therapy versus waiting list control, Outcome 1 Pain (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 4 Cognitive therapy versus waiting list control

Outcome: 1 Pain (short term)

Study or subgroup	Cognitive therapy N	Mean(SD)	Waiting list control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Linton 2008	13	5 (2.1)	21	5 (1.9)		49.6 %	0.0 [-0.69, 0.69]
Turner 1993	16	36.9 (20.5)	18	48.1 (21)		50.4 %	-0.53 [-1.21, 0.16]
Total (95% CI)	**29**		**39**			**100.0 %**	**-0.27 [-0.75, 0.22]**

Heterogeneity: Chi² = 1.12, df = 1 (P = 0.29); I² =11%

Test for overall effect: Z = 1.07 (P = 0.29)

Analysis 4.2. Comparison 4 Cognitive therapy versus waiting list control, Outcome 2 Functional status (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 4 Cognitive therapy versus waiting list control

Outcome: 2 Functional status (short term)

Study or subgroup	Cognitive therapy N	Mean(SD)	Waiting list control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Linton 2008	13	-31 (11.8)	21	-29.6 (11.4)		48.7 %	-0.12 [-0.81, 0.57]
Turner 1993	16	8 (9.3)	18	9.6 (7.3)		51.3 %	-0.19 [-0.86, 0.49]
Total (95% CI)	**29**		**39**			**100.0 %**	**-0.15 [-0.64, 0.33]**

Heterogeneity: Chi² = 0.02, df = 1 (P = 0.89); I² =0.0%

Test for overall effect: Z = 0.63 (P = 0.53)

Analysis 5.1. Comparison 5 Cognitive-behavioural therapy versus waiting list control, Outcome 1 Pain (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 5 Cognitive-behavioural therapy versus waiting list control

Outcome: 1 Pain (short term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Waiting list control N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
Turner 1982	13	36.5 (22.7)	9	77 (21.6)		10.4 %	-1.75 [-2.77, -0.73]
Newton-John 1995	16	10.4 (11.4)	12	17.6 (9.1)		15.8 %	-0.67 [-1.44, 0.11]
Smeets 2006	55	38.6 (31.2)	50	53.4 (22.6)		31.7 %	-0.54 [-0.93, -0.15]
Turner 1988	24	15.9 (11.6)	21	22.1 (12.4)		21.7 %	-0.51 [-1.10, 0.09]
Turner 1993	21	44.3 (28.5)	18	48.1 (21)		20.4 %	-0.15 [-0.78, 0.48]
Total (95% CI)	**129**		**110**			**100.0 %**	**-0.60 [-0.97, -0.22]**

Heterogeneity: $Tau^2 = 0.07$; $Chi^2 = 6.95$, df = 4 (P = 0.14); $I^2 = 42\%$

Test for overall effect: Z = 3.14 (P = 0.0017)

Analysis 5.2. Comparison 5 Cognitive-behavioural therapy versus waiting list control, Outcome 2 Functional status (short term).

Review: Behavioural treatment for chronic low-back pain

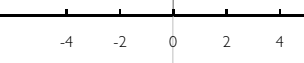
Comparison: 5 Cognitive-behavioural therapy versus waiting list control

Outcome: 2 Functional status (short term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Waiting list control N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
Turner 1982	13	10.2 (6.9)	9	20.2 (8.2)		18.1 %	-1.29 [-2.24, -0.34]
Newton-John 1995	16	18 (15.2)	12	26.3 (17.1)		23.5 %	-0.50 [-1.26, 0.26]
Turner 1988	24	5.4 (3.9)	21	5.7 (6.9)		30.1 %	-0.05 [-0.64, 0.53]
Turner 1993	21	9.5 (8.2)	18	9.6 (7.3)		28.3 %	-0.01 [-0.64, 0.62]
Total (95% CI)	**74**		**60**			**100.0 %**	**-0.37 [-0.87, 0.13]**

Heterogeneity: $Tau^2 = 0.13$; $Chi^2 = 5.95$, df = 3 (P = 0.11); $I^2 = 50\%$

Test for overall effect: Z = 1.45 (P = 0.15)

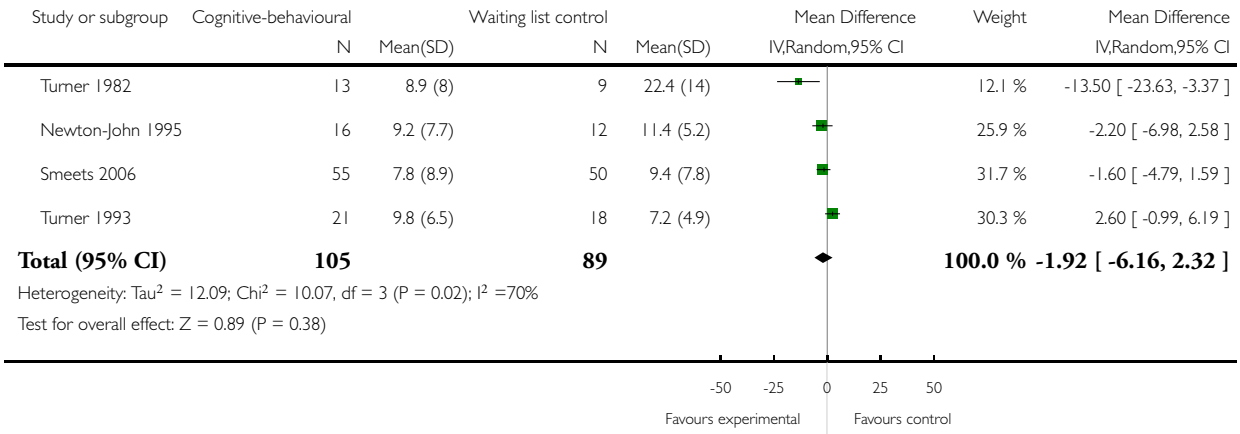


Analysis 5.3. Comparison 5 Cognitive-behavioural therapy versus waiting list control, Outcome 3 Depression (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 5 Cognitive-behavioural therapy versus waiting list control

Outcome: 3 Depression (short term)

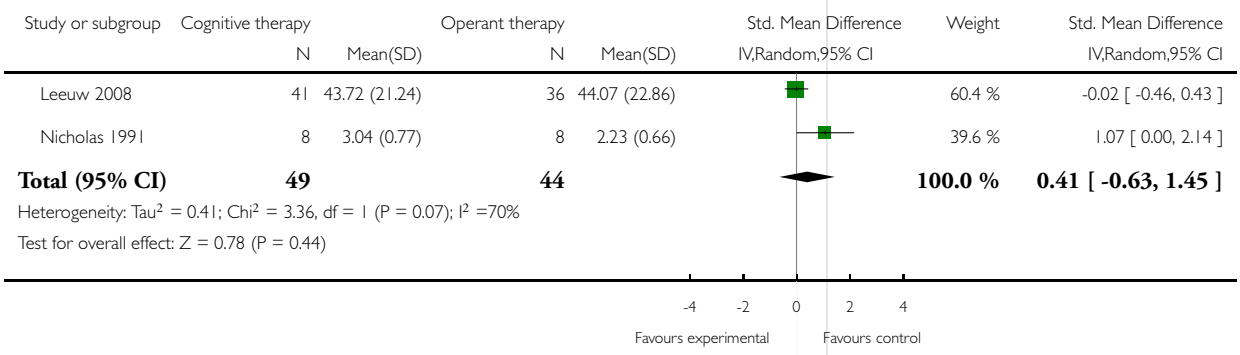


Analysis 6.1. Comparison 6 Cognitive therapy versus operant therapy, Outcome 1 Pain (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 6 Cognitive therapy versus operant therapy

Outcome: 1 Pain (short term)

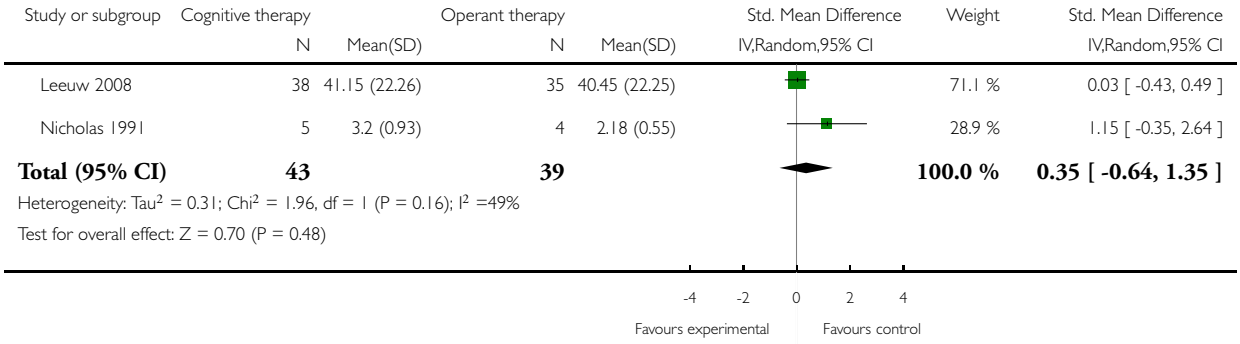


Analysis 6.2. Comparison 6 Cognitive therapy versus operant therapy, Outcome 2 Pain (intermediate term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 6 Cognitive therapy versus operant therapy

Outcome: 2 Pain (intermediate term)

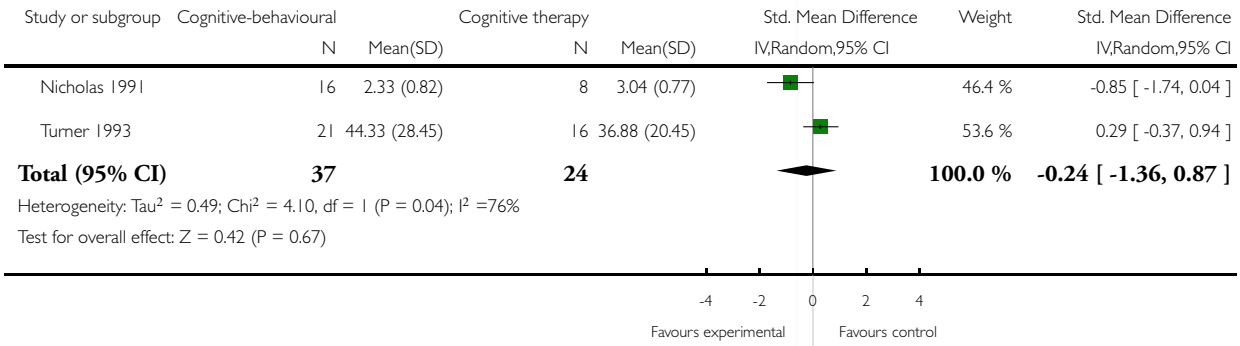


Analysis 7.1. Comparison 7 Cognitive-behavioural therapy versus cognitive therapy, Outcome 1 Pain (short-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 7 Cognitive-behavioural therapy versus cognitive therapy

Outcome: 1 Pain (short-term)



Analysis 7.2. Comparison 7 Cognitive-behavioural therapy versus cognitive therapy, Outcome 2 Pain (intermediate-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 7 Cognitive-behavioural therapy versus cognitive therapy

Outcome: 2 Pain (intermediate-term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Cognitive therapy N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
Nicholas 1991	11	2.03 (0.63)	5	3.2 (0.93)		47.9 %	-1.52 [-2.74, -0.30]
Turner 1993	14	45.86 (26.7)	14	25.99 (20.37)		52.1 %	0.81 [0.04, 1.59]
Total (95% CI)	**25**		**19**			**100.0 %**	**-0.30 [-2.59, 1.98]**

Heterogeneity: $Tau^2 = 2.45$; $Chi^2 = 10.00$, df = 1 (P = 0.002); I^2 =90%

Test for overall effect: Z = 0.26 (P = 0.79)

Analysis 7.3. Comparison 7 Cognitive-behavioural therapy versus cognitive therapy, Outcome 3 Pain (long-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 7 Cognitive-behavioural therapy versus cognitive therapy

Outcome: 3 Pain (long-term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Cognitive therapy N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
Nicholas 1991	12	1.175 (0.852)	4	3.3 (0.83)		47.2 %	-2.37 [-3.85, -0.90]
Turner 1993	18	41.94 (21.01)	14	31.54 (25.1)		52.8 %	0.44 [-0.26, 1.15]
Total (95% CI)	**30**		**18**			**100.0 %**	**-0.89 [-3.64, 1.87]**

Heterogeneity: $Tau^2 = 3.61$; $Chi^2 = 11.37$, df = 1 (P = 0.00075); I^2 =91%

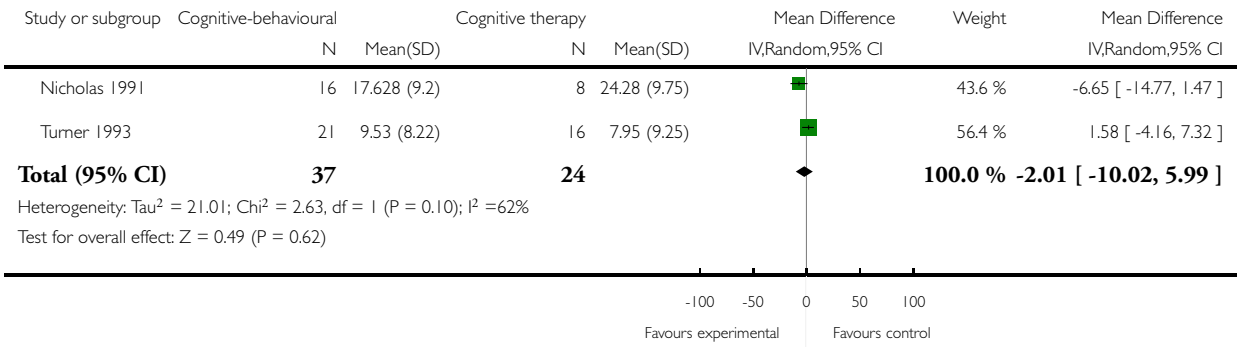
Test for overall effect: Z = 0.63 (P = 0.53)

Analysis 7.4. Comparison 7 Cognitive-behavioural therapy versus cognitive therapy, Outcome 4 Functional status (short-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 7 Cognitive-behavioural therapy versus cognitive therapy

Outcome: 4 Functional status (short-term)

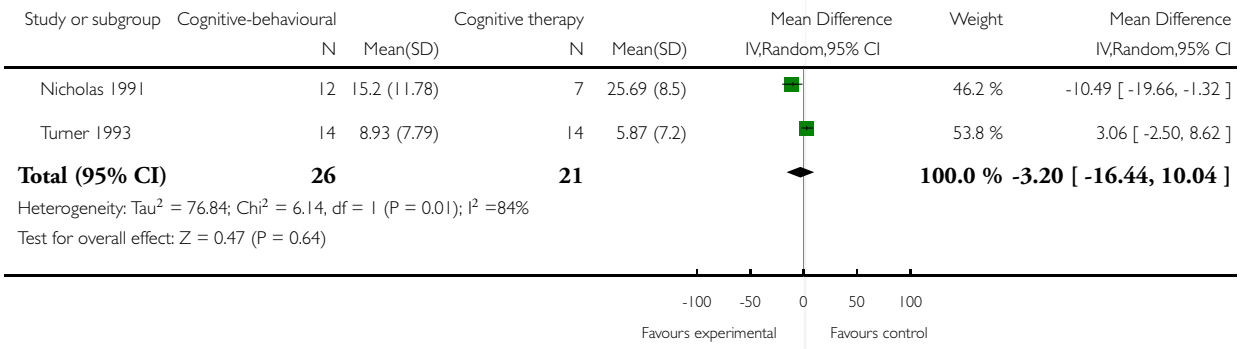


Analysis 7.5. Comparison 7 Cognitive-behavioural therapy versus cognitive therapy, Outcome 5 Functional status (intermediate-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 7 Cognitive-behavioural therapy versus cognitive therapy

Outcome: 5 Functional status (intermediate-term)

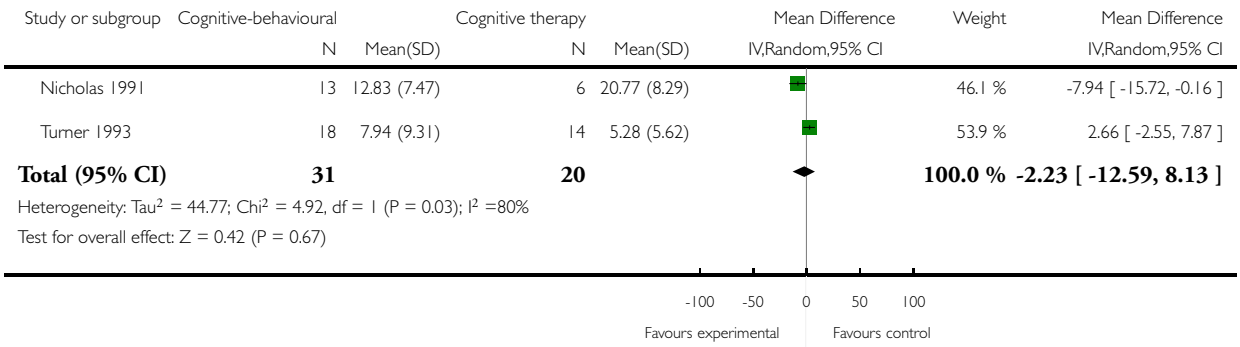


Analysis 7.6. Comparison 7 Cognitive-behavioural therapy versus cognitive therapy, Outcome 6 Functional status (long-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 7 Cognitive-behavioural therapy versus cognitive therapy

Outcome: 6 Functional status (long-term)

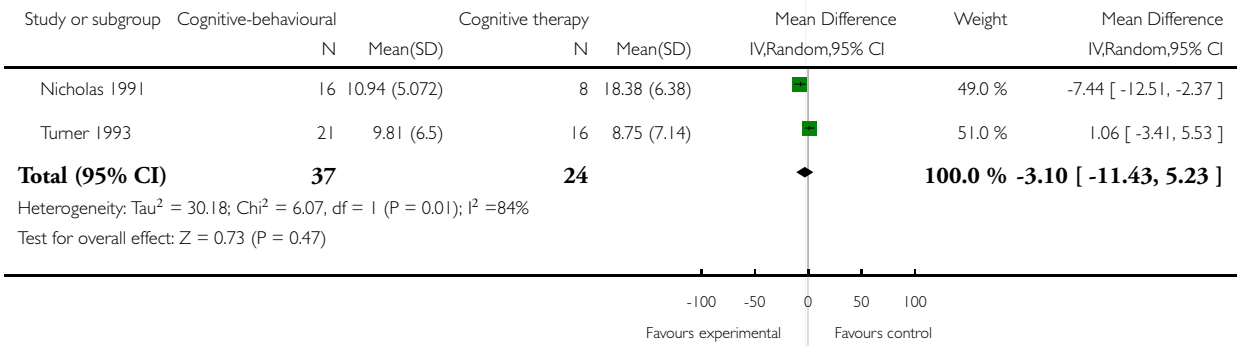


Analysis 7.7. Comparison 7 Cognitive-behavioural therapy versus cognitive therapy, Outcome 7 Depression (short-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 7 Cognitive-behavioural therapy versus cognitive therapy

Outcome: 7 Depression (short-term)



Analysis 7.8. Comparison 7 Cognitive-behavioural therapy versus cognitive therapy, Outcome 8 Depression (intermediate-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 7 Cognitive-behavioural therapy versus cognitive therapy

Outcome: 8 Depression (intermediate-term)

Study or subgroup	Cognitive-behavioural		Cognitive therapy		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Random,95% CI		IV,Random,95% CI
Nicholas 1991	12	8.498 (5.635)	7	16.14 (3.8)		53.6 %	-7.64 [-11.90, -3.39]
Turner 1993	14	8.14 (4.35)	14	9.36 (9.42)		46.4 %	-1.22 [-6.66, 4.22]
Total (95% CI)	**26**		**21**			**100.0 %**	**-4.66 [-10.94, 1.61]**

Heterogeneity: $Tau^2 = 14.42$; $Chi^2 = 3.33$, df = 1 (P = 0.07); $I^2 = 70\%$

Test for overall effect: Z = 1.46 (P = 0.15)

Analysis 7.9. Comparison 7 Cognitive-behavioural therapy versus cognitive therapy, Outcome 9 Depression (long-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 7 Cognitive-behavioural therapy versus cognitive therapy

Outcome: 9 Depression (long-term)

Study or subgroup	Cognitive-behavioural		Cognitive therapy		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Fixed,95% CI		IV,Fixed,95% CI
Nicholas 1991	13	10.386 (7.413)	6	12.83 (6.69)		35.0 %	-2.44 [-9.14, 4.26]
Turner 1993	18	6.83 (6.44)	14	6.5 (7.48)		65.0 %	0.33 [-4.59, 5.25]
Total (95% CI)	**31**		**20**			**100.0 %**	**-0.64 [-4.61, 3.32]**

Heterogeneity: $Chi^2 = 0.43$, df = 1 (P = 0.51); $I^2 = 0.0\%$

Test for overall effect: Z = 0.32 (P = 0.75)

Analysis 8.1. Comparison 8 Cognitive-behavioural therapy versus operant therapy, Outcome 1 Pain (short-term).

Review: Behavioural treatment for chronic low-back pain

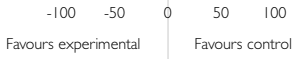
Comparison: 8 Cognitive-behavioural therapy versus operant therapy

Outcome: 1 Pain (short-term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Operant therapy N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Nicholas 1991	16	2.3275 (0.8186)	8	2.23 (0.66)		13.6 %	0.12 [-0.73, 0.97]
Turner 1988	24	15.91 (11.63)	29	18.5 (12.43)		33.3 %	-0.21 [-0.75, 0.33]
van den Hout 2003	45	13.4 (9.6)	39	15.2 (10.5)		53.1 %	-0.18 [-0.61, 0.25]
Total (95% CI)	**85**		**76**			**100.0 %**	**-0.15 [-0.46, 0.16]**

Heterogeneity: Chi^2 = 0.46, df = 2 (P = 0.79); I^2 =0.0%

Test for overall effect: Z = 0.93 (P = 0.35)



Analysis 8.2. Comparison 8 Cognitive-behavioural therapy versus operant therapy, Outcome 2 Pain (intermediate-term).

Review: Behavioural treatment for chronic low-back pain

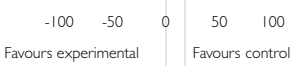
Comparison: 8 Cognitive-behavioural therapy versus operant therapy

Outcome: 2 Pain (intermediate-term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Operant therapy N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Nicholas 1991	11	2.03 (0.626)	4	2.18 (0.55)		8.7 %	-0.23 [-1.38, 0.92]
Turner 1988	20	12.7 (12.75)	23	19.57 (15.31)		31.1 %	-0.48 [-1.08, 0.13]
van den Hout 2003	44	12.6 (10.4)	37	13.6 (9.8)		60.2 %	-0.10 [-0.54, 0.34]
Total (95% CI)	**75**		**64**			**100.0 %**	**-0.23 [-0.57, 0.11]**

Heterogeneity: Chi^2 = 0.98, df = 2 (P = 0.61); I^2 =0.0%

Test for overall effect: Z = 1.31 (P = 0.19)



Analysis 8.3. Comparison 8 Cognitive-behavioural therapy versus operant therapy, Outcome 3 Pain (long-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 8 Cognitive-behavioural therapy versus operant therapy

Outcome: 3 Pain (long-term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Operant therapy N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Nicholas 1991	12	2.27 (0.8687)	5	2.56 (0.97)		10.4 %	-0.31 [-1.36, 0.74]
Turner 1988	20	10.8 (6.38)	27	15.07 (11.62)		33.5 %	-0.43 [-1.02, 0.16]
van den Hout 2003	41	11.3 (11)	35	14 (11.5)		56.0 %	-0.24 [-0.69, 0.21]
Total (95% CI)	**73**		**67**			**100.0 %**	**-0.31 [-0.65, 0.03]**

Heterogeneity: $Chi^2 = 0.26$, df = 2 (P = 0.88); $I^2 = 0.0\%$

Test for overall effect: Z = 1.79 (P = 0.073)

Analysis 8.4. Comparison 8 Cognitive-behavioural therapy versus operant therapy, Outcome 4 Functional status (short-term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 8 Cognitive-behavioural therapy versus operant therapy

Outcome: 4 Functional status (short-term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Operant therapy N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Nicholas 1991	16	17.63 (9.2)	8	18.14 (11.46)		29.2 %	-0.05 [-0.90, 0.80]
Turner 1988	24	5.39 (3.91)	29	3.96 (4.7)		70.8 %	0.32 [-0.22, 0.87]
Total (95% CI)	**40**		**37**			**100.0 %**	**0.21 [-0.24, 0.67]**

Heterogeneity: $Chi^2 = 0.52$, df = 1 (P = 0.47); $I^2 = 0.0\%$

Test for overall effect: Z = 0.92 (P = 0.36)

Analysis 8.5. Comparison 8 Cognitive-behavioural therapy versus operant therapy, Outcome 5 Functional status (intermediate-term).

Review: Behavioural treatment for chronic low-back pain

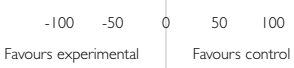
Comparison: 8 Cognitive-behavioural therapy versus operant therapy

Outcome: 5 Functional status (intermediate-term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Operant therapy N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
Nicholas 1991	12	15.198 (11.78)	6	24.42 (11.78)		37.0 %	-0.75 [-1.76, 0.27]
Turner 1988	20	3.57 (3.9)	23	3.27 (3.7)		63.0 %	0.08 [-0.52, 0.68]
Total (95% CI)	**32**		**29**			**100.0 %**	**-0.23 [-1.01, 0.55]**

Heterogeneity: $Tau^2 = 0.16$; $Chi^2 = 1.87$, df = 1 (P = 0.17); $I^2 = 46\%$

Test for overall effect: Z = 0.57 (P = 0.57)



Analysis 8.6. Comparison 8 Cognitive-behavioural therapy versus operant therapy, Outcome 6 Functional status (long-term).

Review: Behavioural treatment for chronic low-back pain

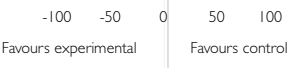
Comparison: 8 Cognitive-behavioural therapy versus operant therapy

Outcome: 6 Functional status (long-term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Operant therapy N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
Nicholas 1991	13	12.832 (7.466)	6	23.85 (12.5)		41.7 %	-1.14 [-2.19, -0.09]
Turner 1988	20	2.91 (2.87)	27	3.07 (4.49)		58.3 %	-0.04 [-0.62, 0.54]
Total (95% CI)	**33**		**33**			**100.0 %**	**-0.50 [-1.56, 0.56]**

Heterogeneity: $Tau^2 = 0.42$; $Chi^2 = 3.23$, df = 1 (P = 0.07); $I^2 = 69\%$

Test for overall effect: Z = 0.92 (P = 0.36)

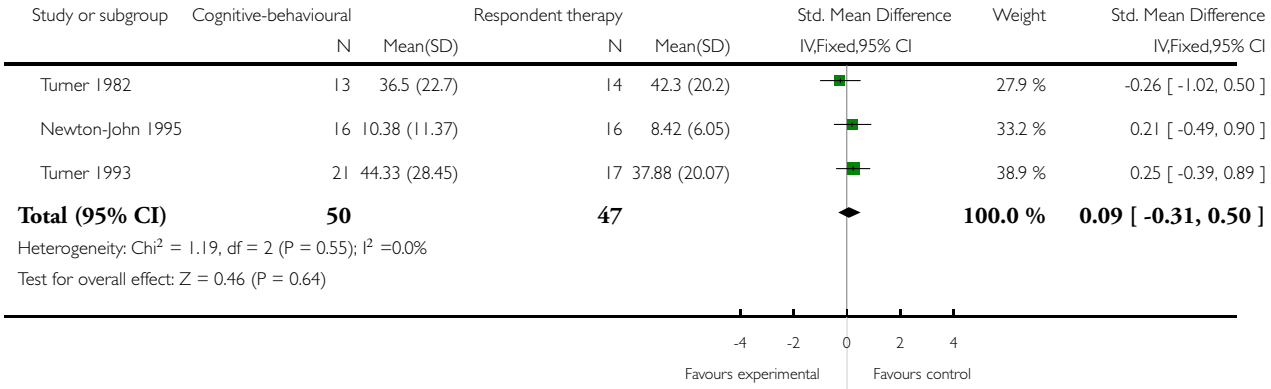


Analysis 9.1. Comparison 9 Cognitive-behavioural therapy versus respondent therapy, Outcome 1 Pain (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 9 Cognitive-behavioural therapy versus respondent therapy

Outcome: 1 Pain (short term)

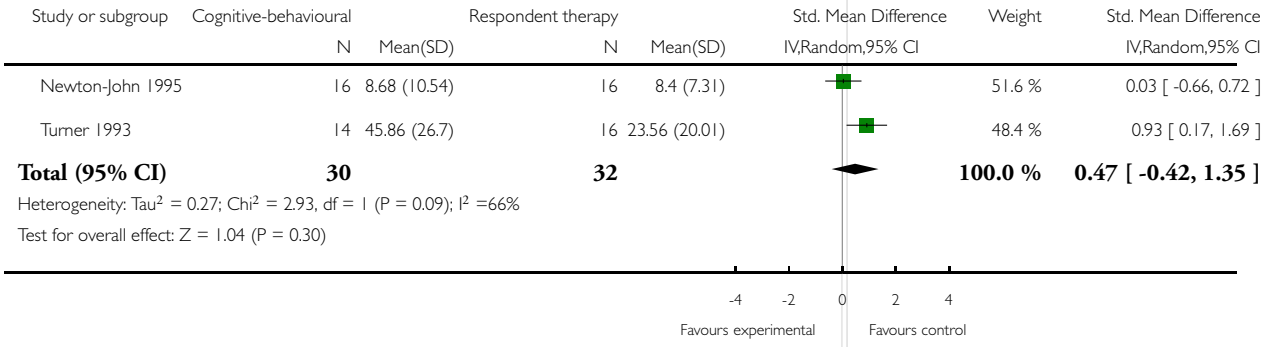


Analysis 9.2. Comparison 9 Cognitive-behavioural therapy versus respondent therapy, Outcome 2 Pain (intermediate term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 9 Cognitive-behavioural therapy versus respondent therapy

Outcome: 2 Pain (intermediate term)

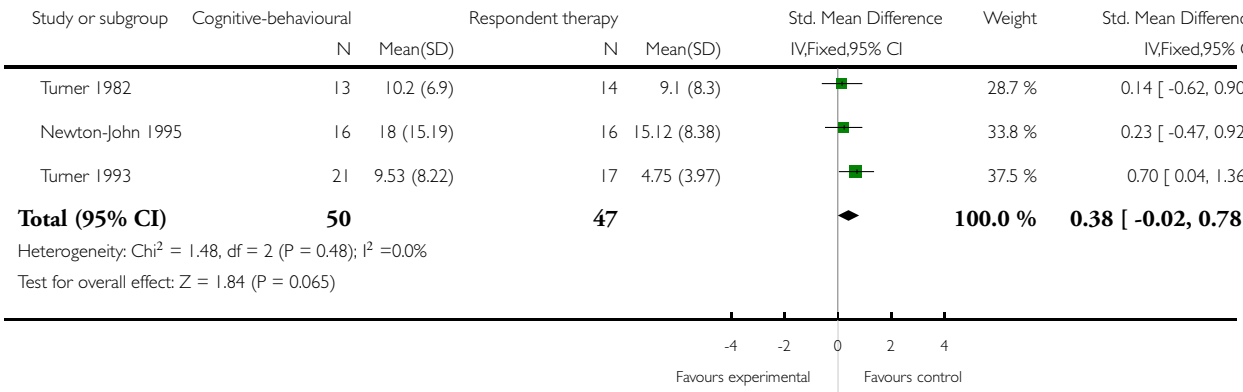


Analysis 9.3. Comparison 9 Cognitive-behavioural therapy versus respondent therapy, Outcome 3 Functional status (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 9 Cognitive-behavioural therapy versus respondent therapy

Outcome: 3 Functional status (short term)

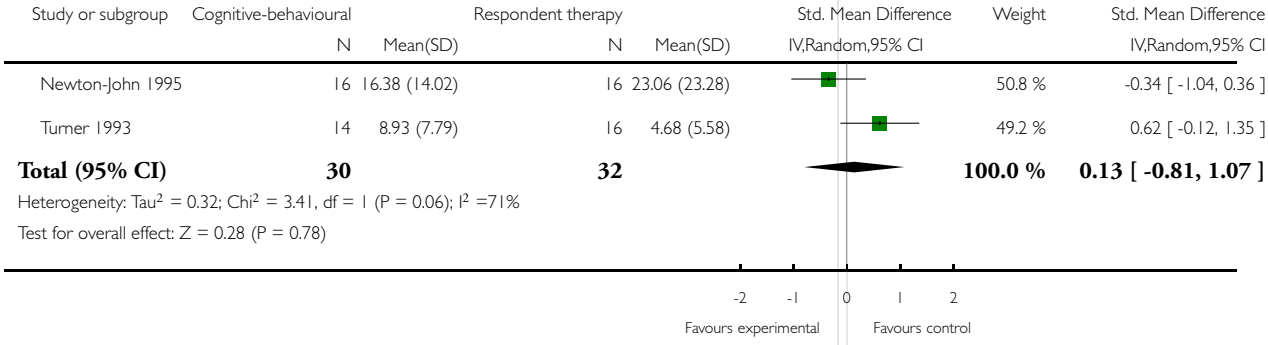


Analysis 9.4. Comparison 9 Cognitive-behavioural therapy versus respondent therapy, Outcome 4 Functional status (intermediate term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 9 Cognitive-behavioural therapy versus respondent therapy

Outcome: 4 Functional status (intermediate term)



Analysis 9.5. Comparison 9 Cognitive-behavioural therapy versus respondent therapy, Outcome 5 Depression (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 9 Cognitive-behavioural therapy versus respondent therapy

Outcome: 5 Depression (short term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Respondent therapy N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Turner 1982	13	8.9 (8)	14	8.1 (6.2)		18.7 %	0.80 [-4.63, 6.23]
Newton-John 1995	16	9.18 (7.67)	16	6.18 (4.87)		27.8 %	3.00 [-1.45, 7.45]
Turner 1993	21	9.81 (6.5)	17	6.24 (3.36)		53.5 %	3.57 [0.36, 6.78]
Total (95% CI)	**50**		**47**			**100.0 %**	**2.89 [0.55, 5.24]**

Heterogeneity: $Chi^2 = 0.74$, df = 2 (P = 0.69); $I^2 = 0.0\%$

Test for overall effect: Z = 2.42 (P = 0.016)

Analysis 9.6. Comparison 9 Cognitive-behavioural therapy versus respondent therapy, Outcome 6 Depression (intermediate term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 9 Cognitive-behavioural therapy versus respondent therapy

Outcome: 6 Depression (intermediate term)

Study or subgroup	Cognitive-behavioural N	Mean(SD)	Respondent therapy N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Newton-John 1995	16	7.43 (5.17)	16	7.37 (5.57)		37.1 %	0.06 [-3.66, 3.78]
Turner 1993	14	8.14 (4.35)	16	5.25 (3.53)		62.9 %	2.89 [0.03, 5.75]
Total (95% CI)	**30**		**32**			**100.0 %**	**1.84 [-0.43, 4.11]**

Heterogeneity: $Chi^2 = 1.40$, df = 1 (P = 0.24); $I^2 = 28\%$

Test for overall effect: Z = 1.59 (P = 0.11)

Analysis 10.1. Comparison 10 Behavioural treatment versus other treatments (usual care), Outcome 1 Pain (short term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 10 Behavioural treatment versus other treatments (usual care)

Outcome: 1 Pain (short term)

Study or subgroup	Behavioural N	Mean(SD)	Usual care N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Poole 2007	57	37.9 (27)	43	48.9 (29.3)		16.9 %	-11.00 [-22.22, 0.22]
von Korff 2005	110	49 (20)	120	53 (19)		83.1 %	-4.00 [-9.05, 1.05]
Total (95% CI)	**167**		**163**			**100.0 %**	**-5.18 [-9.79, -0.57]**

Heterogeneity: Chi^2 = 1.24, df = 1 (P = 0.26); I^2 =20%

Test for overall effect: Z = 2.20 (P = 0.028)

Analysis 10.2. Comparison 10 Behavioural treatment versus other treatments (usual care), Outcome 2 Pain (intermediate term).

Review: Behavioural treatment for chronic low-back pain

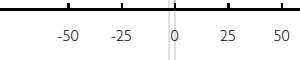
Comparison: 10 Behavioural treatment versus other treatments (usual care)

Outcome: 2 Pain (intermediate term)

Study or subgroup	Behavioural N	Mean(SD)	Usual care N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
von Korff 2005	110	42 (20)	110	47 (22)		80.4 %	-5.00 [-10.56, 0.56]
Poole 2007	54	41.3 (28.5)	45	42.7 (28.4)		19.6 %	-1.40 [-12.65, 9.85]
Total (95% CI)	**164**		**155**			**100.0 %**	**-4.29 [-9.28, 0.69]**

Heterogeneity: Chi^2 = 0.32, df = 1 (P = 0.57); I^2 =0.0%

Test for overall effect: Z = 1.69 (P = 0.091)



Analysis 10.3. Comparison 10 Behavioural treatment versus other treatments (usual care), Outcome 3 Back specific functional status (short term).

Review: Behavioural treatment for chronic low-back pain

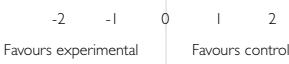
Comparison: 10 Behavioural treatment versus other treatments (usual care)

Outcome: 3 Back specific functional status (short term)

Study or subgroup	Behavioural N	Mean(SD)	Usual care N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
von Korff 2005	110	10.2 (6.3)	120	11.5 (5.8)		70.0 %	-0.21 [-0.47, 0.05]
Poole 2007	57	33.4 (22.3)	43	36.7 (19.9)		30.0 %	-0.15 [-0.55, 0.24]
Total (95% CI)	**167**		**163**			**100.0 %**	**-0.20 [-0.41, 0.02]**

Heterogeneity: Chi^2 = 0.06, df = 1 (P = 0.80); I^2 =0.0%

Test for overall effect: Z = 1.77 (P = 0.077)



Analysis 10.4. Comparison 10 Behavioural treatment versus other treatments (usual care), Outcome 4 Back specific functional status (intermediate term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 10 Behavioural treatment versus other treatments (usual care)

Outcome: 4 Back specific functional status (intermediate term)

Study or subgroup	Behavioural N	Mean(SD)	Usual care N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
von Korff 2005	110	9.2 (6.6)	110	10.1 (6.4)		69.1 %	-0.14 [-0.40, 0.13]
Poole 2007	54	31.3 (21.1)	45	32.9 (17.6)		30.9 %	-0.08 [-0.48, 0.31]
Total (95% CI)	**164**		**155**			**100.0 %**	**-0.12 [-0.34, 0.10]**

Heterogeneity: Chi^2 = 0.05, df = 1 (P = 0.81); I^2 =0.0%

Test for overall effect: Z = 1.07 (P = 0.28)

Analysis 11.1. Comparison 11 Behavioural treatment versus other treatments (group exercise), Outcome 1 Pain (short term).

Review: Behavioural treatment for chronic low-back pain

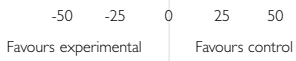
Comparison: 11 Behavioural treatment versus other treatments (group exercise)

Outcome: 1 Pain (short term)

Study or subgroup	Behavioural N	Mean(SD)	Exercise N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Smeets 2006	55	14.34 (12.23)	52	17.84 (13.44)		67.8 %	-3.50 [-8.38, 1.38]
Turner 1990	18	17.71 (12.08)	21	17.52 (10.2)		32.2 %	0.19 [-6.89, 7.27]
Total (95% CI)	**73**		**73**			**100.0 %**	**-2.31 [-6.33, 1.70]**

Heterogeneity: Chi² = 0.71, df = 1 (P = 0.40); I² =0.0%

Test for overall effect: Z = 1.13 (P = 0.26)



Analysis 11.2. Comparison 11 Behavioural treatment versus other treatments (group exercise), Outcome 2 Pain (intermediate term).

Review: Behavioural treatment for chronic low-back pain

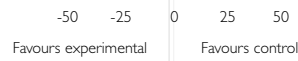
Comparison: 11 Behavioural treatment versus other treatments (group exercise)

Outcome: 2 Pain (intermediate term)

Study or subgroup	Behavioural N	Mean(SD)	Exercise N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Smeets 2006	55	15.65 (12.24)	51	15.21 (13.46)		78.2 %	0.44 [-4.47, 5.35]
Turner 1990	14	19.5 (15.72)	17	15.65 (9.15)		21.8 %	3.85 [-5.46, 13.16]
Total (95% CI)	**69**		**68**			**100.0 %**	**1.18 [-3.16, 5.53]**

Heterogeneity: Chi² = 0.40, df = 1 (P = 0.53); I² =0.0%

Test for overall effect: Z = 0.53 (P = 0.59)



Analysis 11.3. Comparison 11 Behavioural treatment versus other treatments (group exercise), Outcome 3 Pain (long term).

Review: Behavioural treatment for chronic low-back pain

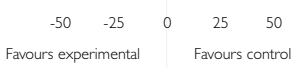
Comparison: 11 Behavioural treatment versus other treatments (group exercise)

Outcome: 3 Pain (long term)

Study or subgroup	Behavioural N	Mean(SD)	Exercise N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Smeets 2006	52	16.02 (14.2)	51	16.64 (15.14)		63.9 %	-0.62 [-6.29, 5.05]
Turner 1990	17	16.41 (13.63)	16	14.94 (7.86)		36.1 %	1.47 [-6.07, 9.01]
Total (95% CI)	**69**		**67**			**100.0 %**	**0.14 [-4.40, 4.67]**

Heterogeneity: Chi² = 0.19, df = 1 (P = 0.66); I² =0.0%

Test for overall effect: Z = 0.06 (P = 0.95)



Analysis 11.4. Comparison 11 Behavioural treatment versus other treatments (group exercise), Outcome 4 Depression (short term).

Review: Behavioural treatment for chronic low-back pain

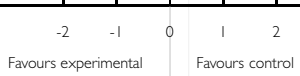
Comparison: 11 Behavioural treatment versus other treatments (group exercise)

Outcome: 4 Depression (short term)

Study or subgroup	Behavioural N	Mean(SD)	Exercise N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Turner 1990	18	8.08 (4.95)	21	7.38 (4.57)		26.8 %	0.14 [-0.49, 0.77]
Smeets 2006	55	9.07 (6.53)	52	6.94 (7.91)		73.2 %	0.29 [-0.09, 0.67]
Total (95% CI)	**73**		**73**			**100.0 %**	**0.25 [-0.07, 0.58]**

Heterogeneity: Chi² = 0.15, df = 1 (P = 0.69); I² =0.0%

Test for overall effect: Z = 1.52 (P = 0.13)



Analysis 11.5. Comparison 11 Behavioural treatment versus other treatments (group exercise), Outcome 5 Depression (intermediate term).

Review: Behavioural treatment for chronic low-back pain

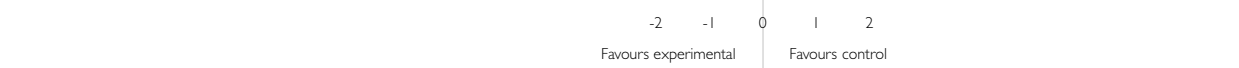
Comparison: 11 Behavioural treatment versus other treatments (group exercise)

Outcome: 5 Depression (intermediate term)

Study or subgroup	Behavioural N	Mean(SD)	Exercise N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Smeets 2006	55	7.34 (7.63)	51	7.75 (9.26)		77.7 %	-0.05 [-0.43, 0.33]
Turner 1990	14	11.36 (8.3)	17	9.29 (8.3)		22.3 %	0.24 [-0.47, 0.95]
Total (95% CI)	**69**		**68**			**100.0 %**	**0.02 [-0.32, 0.35]**

Heterogeneity: Chi² = 0.50, df = 1 (P = 0.48); I² =0.0%

Test for overall effect: Z = 0.10 (P = 0.92)



Analysis 11.6. Comparison 11 Behavioural treatment versus other treatments (group exercise), Outcome 6 Depression (long term).

Review: Behavioural treatment for chronic low-back pain

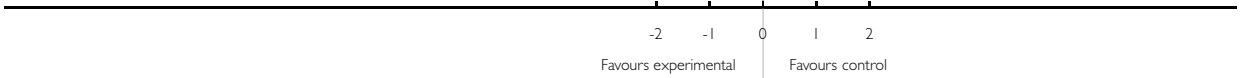
Comparison: 11 Behavioural treatment versus other treatments (group exercise)

Outcome: 6 Depression (long term)

Study or subgroup	Behavioural N	Mean(SD)	Exercise N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Turner 1990	17	8.29 (7.74)	16	9.31 (7.73)		24.2 %	-0.13 [-0.81, 0.55]
Smeets 2006	52	8.37 (8.82)	51	7.15 (9.26)		75.8 %	0.13 [-0.25, 0.52]
Total (95% CI)	**69**		**67**			**100.0 %**	**0.07 [-0.27, 0.41]**

Heterogeneity: Chi² = 0.43, df = 1 (P = 0.51); I² =0.0%

Test for overall effect: Z = 0.41 (P = 0.68)



Analysis 12.1. Comparison 12 Behavioural treatment versus other treatments (surgery), Outcome 1 Back specific functional status (long term).

Review: Behavioural treatment for chronic low-back pain

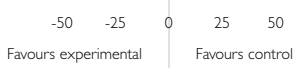
Comparison: 12 Behavioural treatment versus other treatments (surgery)

Outcome: 1 Back specific functional status (long term)

Study or subgroup	Behavioural N	Mean(SD)	Surgery N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Fairbank 2005	146	36.1 (20.6)	138	34 (21.1)		78.6 %	2.10 [-2.75, 6.95]
Brox 2003	26	29.7 (19.6)	35	26.4 (16.4)		21.4 %	3.30 [-5.99, 12.59]
Total (95% CI)	**172**		**173**			**100.0 %**	**2.36 [-1.94, 6.66]**

Heterogeneity: Chi^2 = 0.05, df = 1 (P = 0.82); I^2 =0.0%

Test for overall effect: Z = 1.07 (P = 0.28)



Analysis 13.1. Comparison 13 Behavioural treatment in addition to physiotherapy versus physiotherapy, Outcome 1 Pain intensity (short term).

Review: Behavioural treatment for chronic low-back pain

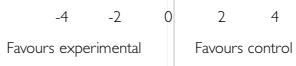
Comparison: 13 Behavioural treatment in addition to physiotherapy versus physiotherapy

Outcome: 1 Pain intensity (short term)

Study or subgroup	Behavioural N	Mean(SD)	Control N	Mean(SD)	Mean Difference IV,Random,95% CI	Weight	Mean Difference IV,Random,95% CI
Nicholas 1992	9	3.07 (0.79)	9	2.72 (0.77)		46.8 %	0.35 [-0.37, 1.07]
Nicholas 1991	32	2.48 (0.816)	9	3.03 (0.75)		53.2 %	-0.55 [-1.12, 0.02]
Total (95% CI)	**41**		**18**			**100.0 %**	**-0.13 [-1.01, 0.75]**

Heterogeneity: Tau^2 = 0.30; Chi^2 = 3.71, df = 1 (P = 0.05); I^2 =73%

Test for overall effect: Z = 0.29 (P = 0.77)



Analysis 13.2. Comparison 13 Behavioural treatment in addition to physiotherapy versus physiotherapy, Outcome 2 Pain Intensity (intermediate term).

Review: Behavioural treatment for chronic low-back pain

Comparison: 13 Behavioural treatment in addition to physiotherapy versus physiotherapy

Outcome: 2 Pain Intensity (intermediate term)

Study or subgroup	Behavioural N	Mean(SD)	Control N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Nicholas 1992	9	2.89 (0.64)	8	2.75 (1.11)		40.6 %	0.14 [-0.74, 1.02]
Nicholas 1991	20	2.353 (0.838)	8	2.64 (0.9)		59.4 %	-0.29 [-1.01, 0.44]
Total (95% CI)	**29**		**16**			**100.0 %**	**-0.11 [-0.67, 0.44]**

Heterogeneity: $Chi^2 = 0.54$, df = 1 (P = 0.46); $I^2 = 0.0\%$

Test for overall effect: Z = 0.40 (P = 0.69)

Analysis 13.3. Comparison 13 Behavioural treatment in addition to physiotherapy versus physiotherapy, Outcome 3 Depression (short term).

Review: Behavioural treatment for chronic low-back pain

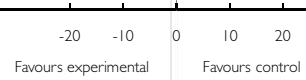
Comparison: 13 Behavioural treatment in addition to physiotherapy versus physiotherapy

Outcome: 3 Depression (short term)

Study or subgroup	Behavioural N	Mean(SD)	Control N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Nicholas 1991	32	14.35 (7.6)	9	12.11 (3.73)		82.9 %	2.24 [-1.35, 5.83]
Nicholas 1992	9	14.69 (6.2)	9	16.44 (10.39)		17.1 %	-1.75 [-9.65, 6.15]
Total (95% CI)	**41**		**18**			**100.0 %**	**1.56 [-1.71, 4.83]**

Heterogeneity: $Chi^2 = 0.81$, df = 1 (P = 0.37); $I^2 = 0.0\%$

Test for overall effect: Z = 0.93 (P = 0.35)



Analysis 13.4. Comparison 13 Behavioural treatment in addition to physiotherapy versus physiotherapy, Outcome 4 Depression (intermediate term).

Review: Behavioural treatment for chronic low-back pain

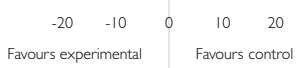
Comparison: 13 Behavioural treatment in addition to physiotherapy versus physiotherapy

Outcome: 4 Depression (intermediate term)

Study or subgroup	Behavioural N	Mean(SD)	Control N	Mean(SD)	Mean Difference IV,Random,95% CI	Weight	Mean Difference IV,Random,95% CI
Nicholas 1992	9	14.44 (5.98)	8	18.5 (9.26)		41.7 %	-4.06 [-11.57, 3.45]
Nicholas 1991	25	13.08 (7.49)	8	9.88 (5.46)		58.3 %	3.20 [-1.59, 7.99]
Total (95% CI)	**34**		**16**			**100.0 %**	**0.17 [-6.85, 7.19]**

Heterogeneity: $Tau^2 = 16.02$; $Chi^2 = 2.55$, df = 1 (P = 0.11); $I^2 = 61\%$

Test for overall effect: Z = 0.05 (P = 0.96)



Analysis 13.5. Comparison 13 Behavioural treatment in addition to physiotherapy versus physiotherapy, Outcome 5 Functional status (short term).

Review: Behavioural treatment for chronic low-back pain

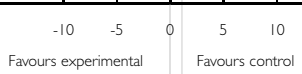
Comparison: 13 Behavioural treatment in addition to physiotherapy versus physiotherapy

Outcome: 5 Functional status (short term)

Study or subgroup	Behavioural N	Mean(SD)	Control N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Nicholas 1992	9	18.81 (10.97)	9	26.08 (16.4)		25.0 %	-7.27 [-20.16, 5.62]
Nicholas 1991	32	19.42 (10.02)	9	25.34 (10.09)		75.0 %	-5.92 [-13.37, 1.53]
Total (95% CI)	**41**		**18**			**100.0 %**	**-6.26 [-12.71, 0.19]**

Heterogeneity: $Chi^2 = 0.03$, df = 1 (P = 0.86); $I^2 = 0.0\%$

Test for overall effect: Z = 1.90 (P = 0.057)



Analysis 13.6. Comparison 13 Behavioural treatment in addition to physiotherapy versus physiotherapy, Outcome 6 Functional status (intermediate term).

Review: Behavioural treatment for chronic low-back pain

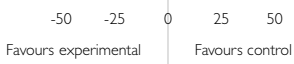
Comparison: 13 Behavioural treatment in addition to physiotherapy versus physiotherapy

Outcome: 6 Functional status (intermediate term)

Study or subgroup	Behavioural N	Mean(SD)	Control N	Mean(SD)	Mean Difference IV,Fixed,95% CI	Weight	Mean Difference IV,Fixed,95% CI
Nicholas 1992	9	18.3 (11.18)	9	25.31 (14.34)		23.7 %	-7.01 [-18.89, 4.87]
Nicholas 1991	25	20.35 (11.67)	8	19.4 (6.89)		76.3 %	0.95 [-5.66, 7.56]
Total (95% CI)	**34**		**17**			**100.0 %**	**-0.93 [-6.71, 4.84]**

Heterogeneity: Chi^2 = 1.32, df = 1 (P = 0.25); I^2 =24%

Test for overall effect: Z = 0.32 (P = 0.75)



Analysis 14.1. Comparison 14 Behavioural treatment in addition to inpatient rehabilitation versus inpatient rehabilitation, Outcome 1 Pain intensity (short term).

Review: Behavioural treatment for chronic low-back pain

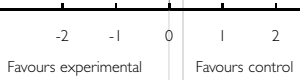
Comparison: 14 Behavioural treatment in addition to inpatient rehabilitation versus inpatient rehabilitation

Outcome: 1 Pain intensity (short term)

Study or subgroup	Behavioural N	Mean(SD)	Control N	Mean(SD)	Std. Mean Difference IV,Fixed,95% CI	Weight	Std. Mean Difference IV,Fixed,95% CI
Schweikert 2006	170	3.1 (1.2)	193	3.3 (1.2)		89.6 %	-0.17 [-0.37, 0.04]
Altmaier 1992	21	2.05 (0.74)	21	2 (0.89)		10.4 %	0.06 [-0.55, 0.66]
Total (95% CI)	**191**		**214**			**100.0 %**	**-0.14 [-0.34, 0.05]**

Heterogeneity: Chi^2 = 0.48, df = 1 (P = 0.49); I^2 =0.0%

Test for overall effect: Z = 1.43 (P = 0.15)



APPENDICES

Appendix 1. Risk of Bias assessment criteria

	Criteria for a judgment of "yes" for the sources of risk of bias
Method of randomisation	A random (unpredictable) assignment sequence. Examples of adequate methods are computer generated random sequence and pre-ordered sealed envelopes. Methods of allocation using date of birth, social insurance/security number, date in which they are invited to participate in the study, or alternation will not be regarded as appropriate.
Concealment of treatment allocation	Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.
Blinding of patients	The index and control groups are indistinguishable for the patients or the success of blinding was tested among the patients and it was successful.
Blinding of care providers	The index and control groups are indistinguishable for the care providers or the success of blinding was tested among the care providers and it was successful.
Blinding of outcome assessment	Item has a positive score if the outcome assessors are blinded regarding treatment allocation and the blinding is evaluated and adequate. If only self-reported (by the patients) outcome measures are used the blinding procedure is adequate for outcome assessors if participant blinding is scored "yes".
Incomplete outcome data addressed (drop-outs)	The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of withdrawals and drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up, and does not lead to substantial bias, a "yes" is scored.
Incomplete outcome data addressed (intention-to-treat analysis)	All randomised patients are reported/analysed in the group they were allocated to by randomisation for the most important moments of effect measurement (minus missing values) irrespective of non-compliance and co-interventions.
Free of selective reporting	The review author determines if all the results from all pre-specified outcomes have been adequately reported in the published report of the trial. This information is either obtained by comparing the protocol and the report, or in the absence of the protocol,

(*Continued*)

	assessing that the published report includes enough information to make this judgment.
Similarity of baseline characteristics	Item has a positive score if the study groups are comparable at baseline for the most important prognostic factors (for example, demographic factors, duration and severity of complaints, and value of main outcome measures).
Co-interventions avoided or similar	Item has a positive score if co-interventions are avoided in the study design or are similar among the intervention groups.
Compliance	The reviewer determines if the compliance with the interventions is acceptable, based on the reported intensity, duration, number and frequency of sessions for both the index intervention and control intervention(s).
Timing of outcome assessments	Timing of outcome assessment should be identical for all intervention groups and for all important outcome assessments.

Appendix 2. Electronic database searches

MEDLINE

1. randomised controlled trial.pt.
2. controlled clinical trial.pt.
3. randomized.ab.
4. placebo.ab,ti.
5. drug therapy.fs.
6. randomly.ab,ti.
7. trial.ab,ti.
8. groups.ab,ti.
9. or/1-8
10. (animals not (humans and animals)).sh.
11. 9 not 10
12. dorsalgia.ti,ab.
13. exp Back Pain/
14. backache.ti,ab.
15. exp Low Back Pain/
16. (lumbar adj pain).ti,ab.
17. coccyx.ti,ab.
18. coccydynia.ti,ab.
19. sciatica.ti,ab.
20. sciatica/
21. spondylosis.ti,ab.
22. lumbago.ti,ab.
23. or/12-22
24. exp Behavior Therapy/
25. Conditioning, Operant/
26. exp "Reinforcement (Psychology)"/
27. behavior therapy.mp.

28. operant conditioning.mp.
29. respondent treatment.mp.
30. behavioral therapy.mp.
31. behavioural therapy.mp.
32. Cognitive Therapy/
33. cognitive therapy.mp.
34. cognitive treatment.mp.
35. behavior treatment.mp.
36. relaxation.mp. or exp Relaxation/
37. graded activity.mp.
38. or/24-37
39. 38 and 11 and 23
40. 39
41. limit 40 to yr="2007 - 2009"

EMBASE

1. Clinical Article/
2. exp Clinical Study/
3. Clinical Trial/
4. Controlled Study/
5. Randomized Controlled Trial/
6. Major Clinical Study/
7. Double Blind Procedure/
8. Multicenter Study/
9. Single Blind Procedure/
10. Phase 3 Clinical Trial/
11. Phase 4 Clinical Trial/
12. crossover procedure/
13. placebo/
14. or/1-13
15. allocat$.mp.
16. assign$.mp.
17. blind$.mp.
18. (clinic$ adj25 (study or trial)).mp.
19. compar$.mp.
20. control$.mp.
21. cross?over.mp.
22. factorial$.mp.
23. follow?up.mp.
24. placebo$.mp.
25. prospectiv$.mp.
26. random$.mp.
27. ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).mp.
28. trial.mp.
29. (versus or vs).mp.
30. or/15-29
31. 14 and 30
32. human/
33. Nonhuman/
34. exp ANIMAL/
35. Animal Experiment/
36. 33 or 34 or 35
37. 32 not 36
38. 31 not 36

39. 37 and 38
40. 38 or 39
41. dorsalgia.mp.
42. back pain.mp.
43. exp BACKACHE/
44. (lumbar adj pain).mp.
45. coccyx.mp.
46. coccydynia.mp.
47. sciatica.mp.
48. exp ISCHIALGIA/
49. spondylosis.mp.
50. lumbago.mp.
51. exp Low Back Pain/
52. or/41-51
53. exp Behavior Therapy/
54. exp Cognitive Therapy/
55. exp CONDITIONING/
56. behavior therapy.mp.
57. behavior treatment.mp.
58. behavioural therapy.mp.
59. behavioural treatment.mp.
60. cognitive therapy.mp.
61. cognitive treatment.mp.
62. exp instrumental conditioning/
63. operant conditioning.mp.
64. operant treatment.mp.
65. relaxation.mp.
66. or/53-65
67. 40 and 52 and 66
68. 67
69. limit 68 to yr="2007 - 2009"

CINAHL

S58 S56 and S41 and S23 Limiters - Published Date from: 200701-200912
S57 S56 and S41 and S23
S56 S55 or S54 or S53 or S52 or S51 or S50 or S49 or S48 or S47 or S46 or S45 or S44 or S43 or S42
S55
S54 ("relaxation") or (MH "Relaxation")
S53 "behavioural treatment"
S52 "behavioral treatment"
S51 "behavioural therapy"
S50 "behavioral therapy"
S49 "respondent treatment"
S48 "operant conditioning"
S47 "behaviour therapy"
S46 "behavior therapy"
S45 (MH "Cognitive Therapy")
S44 (MH "Reinforcement (Psychology)+")
S43 (MH "Conditioning (Psychology)")
S42 (MH "Behavior Therapy+")
S41 S40 or S39 or S38 or S37 or S36 or S35 or S34 or S33 or S32 or S31 or S30 or S29 or S28 or S27 or S26 or S25 or S24
S40 "lumbago"
S39 (MH "Spondylolysis")
S38 (MH "Spondylolisthesis")

S37 (MH “Thoracic Vertebrae”)
S36 (MH “Lumbar Vertebrae”)
S35 coccydynia
S34 “sciatica”
S33 “coccyx”
S32 (MH “Sciatica”)
S31 (MH “Coccyx”)
S30 “lumbar N5 pain”
S29 “”lumbarW1pain“”
S28 “lumbar W1 pain”
S27 “backache”
S26 (MH “Low Back Pain”)
S25 (MH “Back Pain+”)
S24 dorsalgia
S23 S21 not S22
S22 (MH “Animals+”)
S21 S20 or S19 or S18 or S17 or S16 or S15 or S14 or S13 or S12 or S11 or S10 or S9 or S8 or S7 or S6 or S5 or S4 or S3 or S2 or S1
S20 “volunteer*”
S19 prospectiv*
S18 “control*”
S17 “follow-up stud*”
S16 (MH “Prospective Studies+”)
S15 (MH “Evaluation Research+”)
S14 (MH “Comparative Studies”)
S13 “latin square”
S12 (MH “Study Design+”)
S11 (MH “Random Sample+”)
S10 “random*”
S9 “placebo*”
S8 (MH “Placebos”)
S7 (MH “Placebo Effect”)
S6 “triple-blind”
S5 “single-blind”
S4 “double-blind”
S3 “”clinical W8 trial“”
S2 “randomi?ed controlled trial*”
S1 (MH “Clinical Trials+”)

PsycINFO

((DE=(“relaxation therapy” or “behavior therapy” or “relaxation”)) or (DE=(“conditioning” or “operant conditioning” or “behavior therapy”)) or (DE=(“behavior therapy” or “behavior modification” or “psychotherapy” or “cognitive behavior therapy”))) AND ((KW=(Randomi?ed controlled trial*) OR KW=(clinical trial*) OR KW=(clin* near trail*) OR KW= (sing* near blind*) OR KW=(sing* near mask*) OR (doub* near blind*) OR KW=(doubl* NEAR mask*) OR KW=(trebl* near mask*) OR KW=(trebl* near mask*) OR KW=(tripl* near blind*) OR KW=(tripl* near mask*) OR KW=(placebo*) OR KW=(random*) OR DE=(research design) OR KW=(Latin square) OR KW=(comparative stud*) OR KW=(evaluation stud*) OR KW=(follow up stud*) OR DE=(prospective stud*)OR KW=(control*) OR KW=(prospective*) OR KW=(volunteer*)) AND (DE=(back) OR DE=(back pain) OR DE=(neck)))

Date Range: 2007-2009

Cochrane Back Review Group Trials Register

1. behav*
2. cognit*
3. relax*
4. graded activity
5. reinforcement

6. respondent
February 2009

FEEDBACK

Comment on the Back Group behavioural review (*The Cochrane Library* 2000, issue 2)

Summary

The commentator thinks there is a discrepancy between the author's presentation and discussion on the studies, and their conclusions. The commentator states that the results indicate that behavioral therapy gives a moderate effect only with back patients that receive no other treatment, but that effect is not proved in comparison with other treatment modalities. The commentator goes on to suggest that the whole treatment effect of behavioral treatment in this patient group is a placebo effect and no real treatment effect. The commentator states, "Behavioral treatment should not be recommended outside the setting of clinical trials."

Reply

We indeed clearly state in the review that behavioral therapy only gives moderate effect in the comparison with no treatment or waiting list control and that there were no differences identified when behavioral therapy was compared to other treatments. Moreover, comments like this and ongoing discussion regarding the potential beneficial effect of behavioral treatment (among other in guideline committees) prompted us to disaggregate the analyses regarding the comparison of behavioural treatment versus waiting list control, no treatment or placebo for this up-date. The main reason for this is that we wanted to assess what components or types of behavioural treatment were effective. This update shows that especially a combined cognitive-respondent therapy or progressive relaxation are effective for short-term pain reduction.

We disagree with the commentator when he states that, "Behavioural treatment should not be recommended outside the setting of clinical trials." In our opinion many treatment modalities are used for chronic low back pain and there is not one treatment option that has been proven beyond doubt to be more effective than other treatment modalities. It is unlikely that the effectiveness of behavioral treatment is merely a placebo effect. As we have stated in our recommendations, we do agree that more research is needed in order to improve the understanding what specific patient populations benefit most from what kind of behavioral treatment.

Contributors

Commentator - Ottar Grimstad
Criticism Editor - Dr. Alf Nachemson

WHAT'S NEW

Last assessed as up-to-date: 31 July 2009.

Date	Event	Description
19 January 2011	Amended	Contact details updated.

HISTORY

Protocol first published: Issue 3, 1998

Review first published: Issue 2, 2000

Date	Event	Description
12 May 2010	New citation required and conclusions have changed	Reviewer comments incorporated into the manuscript. As a result, two RCTs from the original review were excluded (Bru 1994; Lindström 1992) as they did not meet the inclusion criteria and were not included in the analyses. Removing these RCTs and responding to these amendments did not change the conclusions of the updated review.
1 March 2009	New search has been performed	We included 11 new randomised controlled trials in this update. The risk of bias and clinical relevance of all included RCTs was assessed using the updated criterion of the Cochrane Back Review Group. The GRADE approach was incorporated and used to judge the quality of the evidence for a particular comparison.
30 October 2004	New citation required but conclusions have not changed	Neither of the changes listed below changed the conclusion of the original review dramatically. However, by disaggregating the different treatment effects, and analysing treatment effects of the various types of treatment separately, we tried to assess which type of behavioural treatment was most effective, which components of behavioural treatment are necessary, and which are superfluous.
18 October 2004	Feedback has been incorporated	See feedback section Feedback received: 15/06/01 Response to feedback: 18/01/04
1 October 2003	New search has been performed	We included one additional randomised controlled study in this update. Another new element is that the original comparison of "all types of Cognitive Behavioural Treatment versus Waiting List Control, No treatment or Placebo" is now divided into various types of treatments (operant, cognitive and behavioural).

CONTRIBUTIONS OF AUTHORS

Nicholas Henschke and Raymond Ostelo identified and selected studies, assessed the methodological quality of studies, performed the data extraction, and conducted the data analyses.

Maurits van Tulder, Johan Vlaeyen, Stephen Morley, Willem Assendelft, and Chris Main were involved in final decisions regarding in- and exclusion of studies, and in defining the domains.

All authors read and approved the final version of the manuscript.

DECLARATIONS OF INTEREST

Johan Vlaeyen is co-author of four of the trials included in this review, and Maurits van Tulder is the co-author of one of the included trials. Risk of bias assessment and data extraction of these trials was done by two other review authors (Nicholas Henschke and Raymond Ostelo).

SOURCES OF SUPPORT

Internal sources

- Vrije Universiteit, EMGO Institute, Netherlands.
- Univeristy of Maastricht, Netherlands.
- University of Leeds, UK.
- LUMC, Netherlands.

External sources

- Dutch Health Insurance Board, Netherlands.
- National Health & Medical Research Council, Australia.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Thirty RCTs were included in this updated systematic review. In addition to eleven new trials identified since the previous review, an updated risk of bias assessment was performed, and the GRADE approach was used to judge the quality of the evidence. Note that the text supporting the risk of bias assessment was only included for the 11 newly added trials, while the assessment was completed for all 30 trials.

INDEX TERMS

Medical Subject Headings (MeSH)

Behavior Therapy [*methods]; Chronic Disease; Conditioning, Operant; Depression [therapy]; Low Back Pain [psychology; *therapy]; Muscle Relaxation; Randomized Controlled Trials as Topic; Waiting Lists

MeSH check words

Humans