Health Psychology
2007, Vol. 26, No. 1, 1–9

DOI: 10.1037/0278-6133.26.1.1

Meta-Analysis of Psychological Interventions for Chronic Low Back Pain

Benson M. Hoffman
Duke University Medical Center

Rebecca K. Papas
Yale University School of Medicine

David K. Chatkoff
University of Michigan—Dearborn

Robert D. Kerns
Veterans Affairs Connecticut Healthcare System and Yale University

The purpose of this meta-analysis of randomized controlled trials was to evaluate the efficacy of psychological interventions for adults with noncancerous chronic low back pain (CLBP). The authors updated and expanded upon prior meta-analyses by using broad definitions of CLBP and psychological intervention, a broad data search strategy, and state-of-the-art data analysis techniques. All relevant controlled clinical trials meeting the inclusion criteria were identified primarily through a computer-aided literature search. Two independent reviewers screened abstracts and articles for inclusion criteria and extracted relevant data. Cohen's *d* effect sizes were calculated by using a random effects model. Outcomes included pain intensity, emotional functioning, physical functioning (pain interference or pain-specific disability, health-related quality of life), participant ratings of global improvement, health care utilization, health care provider visits, pain medications, and employment/disability compensation status. A total of 205 effect sizes from 22 studies were pooled in 34 analyses. Positive effects of psychological interventions, contrasted with various control groups, were noted for pain intensity, pain-related interference, health-related quality of life, and depression. Cognitive–behavioral and self-regulatory treatments were specifically found to be efficacious. Multidisciplinary approaches that included a psychological component, when compared with active control conditions, were also noted to have positive short-term effects on pain interference and positive long-term effects on return to work. The results demonstrated positive effects of psychological interventions for CLBP. The rigor of the methods used, as well as the results that reflect mild to moderate heterogeneity and minimal publication bias, suggest confidence in the conclusions of this review.

Keywords: chronic low back pain, psychological intervention, meta-analysis

Chronic pain affects approximately 30% of the U.S. population annually (Bonica & Loeser, 2000). Low back pain affects 15% to 45% of adults annually and at least 70% of adults over a lifetime (G. B. J. Andersson, 1997). Low back pain is associated with substantial disability (CDC, 2001) and negative economic impact associated with decreased work productivity, absenteeism, and increased health care utilization (Mounce, 2002).

The disabling nature of low back pain stems from somatic, psychological, and social factors (Waddell, 1987). As acute pain becomes chronic, cognitive, affective, biological, and behavioral variables interact to form a "chronic pain experience," characterized by pain, distress, and disability (Banks & Kerns, 1996). Psychosocial interventions for chronic low back pain (CLBP) specifically target these variables to decrease distress and increase functioning (Nielson & Weir, 2001; Smith & Gribbin, 2001). Psychological interventions for CLBP have become common alternatives to traditional medical and rehabilitation approaches (Campbell & Mitchell, 1996; Gatchel & Turk, 1996).

Five relatively recent published reviews report on the efficacy of psychosocial interventions for CLBP (Guzman et al., 2001; Morley, Eccleston, & Williams, 1999; Nielson & Weir, 2001; van Tulder, Koes, & Bouter, 1997; van Tulder et al., 2001). The methods for the current review were specifically informed by a critical consideration of the methods used in two of the prior reviews. Morley and colleagues (1999) conducted a meta-analysis of randomized controlled trials (RCTs) of cognitive–behavioral treatment (CBT), behavioral treatment (BT), and biofeedback and relaxation training for various chronic pain conditions. For all measurement domains (e.g., pain expression, activity level), any type of therapy was more effective than wait-list control. When compared with active treatment controls, CBT and BT were found to be effective. Using the Cochrane review system, van Tulder and his colleagues (2001) conducted a systematic review of behavioral interventions (i.e., operant, cognitive, and respondent) for CLBP. RCTs published prior to April 1999 were included. Outcome measures were pain intensity, functional status, and behavior. BTs were found to be generally indistinguishable from one another, of

Benson M. Hoffman, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center; Rebecca K. Papas, Department of Epidemiology and Public Health, Yale University School of Medicine; David K. Chatkoff, Department of Behavioral Sciences, University of Michigan at Dearborn; Robert D. Kerns, Psychology Service, Veterans Affairs Connecticut Healthcare System and Department of Psychiatry, Neurology, and Psychology, Yale University.

Corresponding concerning this article should be addressed to Robert D. Kerns, Psychology Service (116B), VA Connecticut Healthcare System, 950 Campbell Avenue, West Haven, CT 06516. E-mail: robert.kerns@med.va.gov

no incremental benefit when added to usual care, and moderately superior to no treatment, placebo, and wait-list controls. The authors noted that behavioral outcomes may have been diminished by the effect of pooling all behavioral outcomes (pain behavior, anxiety, depression, cognitive errors) into one outcome statistic, as some behavioral outcomes may respond differently to treatment than others.

None of the five recent reviews included all of the following: broad definitions of CLBP (e.g., including both "organic" and "nonspecific" low back pain); broad definitions of psychosocial interventions (e.g., not limited by theoretical orientation); assessment of methodological elements, reflecting unique aspects of behavioral research, quantitative analyses, and analyses of those who benefited from which treatments. The methodology proposed in this review attends to each of these concerns.

The primary purpose of this review was to evaluate the relative efficacy of outpatient, psychological interventions for adults with noncancerous CLBP. In contrast to prior reviews, we included studies of CLBP from both known and unknown etiologies because of poor empirical associations between the presence of CLBP and evidence of structural pathology and because of concerns about the general unreliability of medical diagnosis and causal attributions for CLBP. Outcomes including pain intensity, quality of life, and physical and emotional functioning associated with broad categories of psychosocial interventions (e.g., CBT, social support, patient education with a cognitive or behavioral component) relative to control groups (e.g., wait list, treatment as usual) were evaluated. Secondary objectives included identifying possible differences in treatment efficacy across demographic groups (i.e., age, race/ethnicity, pain duration, gender, education) as well as analyses investigating the impact of sample size and study methodology on results.

Method

Data Sources

All relevant controlled clinical trials meeting our inclusion criteria were identified through (a) a computer-aided literature search and (b) screening for references in relevant identified publications and reviews. A call for unpublished studies in this area was also made by Robert D. Kerns to relevant health psychology Listservs. These efforts did not result in any additional data, however. The computer literature search was conducted on studies published through October 2004 through the use of the following databases: MEDLINE (1966 to September, Week 5, 2004), PsycINFO (1872 to September, Week 3, 2004), Excerpta Medica (EMBASE; 1980 to 2004, Week 40), CENTRAL (3rd quarter, 2004), and Cumulative Index to Nursing & Allied Health Literature (CINAHL; 1982 to October, Week 1, 2004). The literature search was conducted using a sensitive and specific strategy (Clark & Oxman, 2003), augmented by an exhaustive list of both broad and narrow search terms to identify psychosocial interventions. Our search terms reflected three categories: (a) randomized or controlled clinical trials, (b) psychosocial interventions (including patient education), and (c) back pain. Seminal articles, identified by Robert D. Kerns, were used as "targets" for the purpose of piloting and modifying the search strategy (Clark & Oxman, 2003). A list of search terms is available from Robert D. Kerns.

Study Selection

Searches yielded 952 unique references. Each abstract was independently reviewed by two reviewers for the inclusion criteria (listed below). Abstracts were rejected if one of the criteria was violated but were retained even if all criteria were not confirmed in the abstract. References without abstracts were kept unless exclusion criteria were evident in the title.

Following abstract review, 208 (22%) articles were sought. The primary reason for exclusion was lack of randomization. We were unable to obtain 12 articles, a majority of which were published in international journals that were not available through Veterans Affairs Connecticut Healthcare System and Yale University library systems (e.g., Kansanelakelaitoksen julkaisuja). In addition, we were unable to obtain any relevant dissertations.[1]

Articles received were reviewed and were accepted if they met all inclusion criteria. Authors of six studies were contacted to clarify whether criteria were met. Following clarification, three of those studies were excluded, one was retained, and the remaining two were excluded because there was no response from the author. Of the 196 articles retrieved, 39 articles (20%) met inclusion criteria for the meta-analysis. However, data were not extractable in five studies (i.e., mean effects for all interventions were combined, no sample size provided, no means and/or standard deviations or equivalent data provided). Personal contact with authors of two already included studies resulted in additional data. In the remaining cases, efforts were either unsuccessful in making contact or authors reported that historical data were no longer available. Hence, data were extracted from 34 articles detailing 31 separate research studies.

Inclusion criteria included English language; random assignment ("quasi-random" approaches, such as alternating sequence, were accepted); adults (i.e., age 18 and older); chronic pain of 3 months duration, or recurrent pain of 3 months duration with pain occurring more often than not during the 3-month period; nonmalignant pain; presence of a control or comparison group; presence of a pain-relevant outcome; and presence of a psychological intervention. An identifiable subsample of participants possessing the above-mentioned characteristics was acceptable, so long as subsample-specific outcome data were available.

Our definition of a pain-relevant outcome was based on recommendations from the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT; Dworkin et al., 2005; Turk et al., 2003). IMMPACT recommends the following domains: pain intensity, emotional functioning, physical functioning (pain interference or pain-specific disability, health-related quality of life), and participant ratings of global improvement or treatment satisfaction. These outcome domains were then supplemented by additional domains shown to be relevant to pain interventions: health care utilization costs (Mapel, Shainline, Paez, & Gunter, 2004), health care provider visits (H. I. Andersson, Ejlertsson, Leden, & Schersten, 1999), pain medications (Blyth, March, Nicholas, & Cousins, 2005), and employment/disability compensation status (Watson, Booker, Moores, & Main, 2004).

[1] Our attempts to obtain dissertations included the following: borrowing via interlibrary loan through the aforementioned library systems, direct negotiations with the universities at which the dissertations were housed, and attempts to directly contact each dissertation author.

We considered an intervention "psychological" if it was (a) conceived of as a psychological intervention by authors of published studies and (b) clearly based on any of the following approaches to psychological intervention: behavioral (operant or respondent approaches, in the absence of explicit mention of cognition change), cognitive–behavioral; self-regulatory (biofeedback, relaxation or hypnosis), or supportive counseling (nondirective lay or professional counseling). There was also an "other" category for additional psychologically informed interventions (e.g., psychoanalytic). Authors of published studies typically identified the psychological basis for interventions. Any disagreements about whether an intervention was "psychological" were resolved by consensus. As an example, a condition not considered psychological provided physical therapy, whereas a condition considered psychological provided aerobics, relaxation training, and a psychological pain management group.

Coding and Methodological Elements Rating

Methodological elements ratings criteria were adapted from previously published, relevant studies (Morley et al., 1999; van Tulder et al., 1997), modified for simplicity (Clark & Oxman, 2003), and tailored to the needs of biopsychosocial research (e.g., perceived credibility ratings; Nielson & Weir, 2001). Studies received a score of "met," "unmet," "unsure," or "not scored" along each dimension. A study was not scored if the methodological dimension was not applicable to the study. For example, treatment credibility was not scored when the intervention group was compared with a waiting list control group. A score of unsure was converted to unmet in subsequent analyses. Limitations of assessing methodology from the published article are the influence of writing style, primary language, and manuscript length (Clark & Oxman, 2003). However, CONSORT guidelines, now adopted by many peer-reviewed journals, suggest a minimum standard for reporting methods of clinical trials (Moher, Schulz, & Altman, 2001).

Reliability

Each abstract was independently reviewed by two researchers to determine whether the full article should be ordered. The review process relied upon a checklist, derived from the inclusion criteria, in order to protect against unintentional "drift." Each received article was fully reviewed by one researcher to determine whether the study met inclusion criteria for the meta-analysis by using the checklist. Thirteen percent of these articles were discussed among the researchers to resolve questions and to achieve consensus (e.g., due to ambiguous published details). Twenty-one percent of the independently reviewed articles were randomly selected for double review. Percentage agreement for this process was 92%. In regard to data extraction, 10 of the studies were considered pilot data and were reviewed and discussed among the researchers to achieve consensus. Each of the remaining studies was independently coded for data extraction by two researchers. Percentage agreement for coding was 94%. Any disagreement during any phase of study selection regarding coding or extraction was resolved by consensus, with involvement of a third researcher if necessary. Weekly face-to-face meetings or teleconferences were held through all phases of the study to discuss related issues.

Meta-Analytic Procedure

Meta-analysis is a quantitative method for summarizing research findings. The procedures entail generating effect sizes (ES) for each relevant contrast in each study and then aggregating those ESs by outcome and response period (Lipsey & Wilson, 2001). In the present context, an ES represents the magnitude of the impact of one treatment condition (e.g., CBT), relative to that of a second treatment condition (e.g., self-regulatory treatment; SRT) or control condition (e.g., treatment as usual), on one dependent variable (e.g., pain intensity).

We calculated ES from the extracted data by using Cohen's *d* (Cohen, 1992), which is computed by subtracting the control group mean from the intervention group mean and dividing the value by the pooled sample standard deviation. If means and standard deviations were not available, and the contrast was reported as "not significant," then we coded the ES to be zero (Lipsey & Wilson, 2001). An ES value of zero was inserted in 15 of the 379 ESs calculated for this study. We weighted each ES by its inverse variance (Hedges & Olkin, 1985). Because Cohen's *d* values have been shown to be upwardly biased when based on small sample sizes (i.e., $n < 20$), we added the Hedges correction for small sample bias (Hedges, 1981). The resulting ES was computed so that a positive ES indicated superior performance for the identified "treatment" group, and a negative ES indicated superior performance for the identified "control" group. When outcomes were presented in percentages (e.g., percent disabled), odds ratios were calculated through the arcsine method (Lipsey & Wilson).

For one study (Keijsers, Groenman, Gerards, & Van Oudheusden, 1989) in which group means, but only a sample-wide (i.e., treatment and control groups combined) standard deviation was provided, we used the sample-wide standard deviation to compute an ES. This represents a conservative approach, as pooled within-group standard deviations minimize the error term relative to sample-wise standard deviations. This practice is acceptable when there is no reason to believe that the group standard deviations would have differed substantially from one another (Lipsey & Wilson, 2001). No other substitutions were made.

ESs were sorted by treatment group (i.e., CBT, SRT, BT, a combined supportive/"mixed" treatment group, and multidisciplinary treatment), control group (i.e., wait-list control, physiotherapy, attention control, and *treatment as usual* [defined as continuation of care]), time period of the contrast (i.e., posttreatment [0 to 3 months]), follow-up (3 months to 1 year), and long-term follow-up (more than 1 year), and outcome (e.g., pain intensity). ESs were also sorted to compare unimodal psychological interventions with one another and to compare unimodal psychological interventions with multidisciplinary treatments. Because of the low number of ESs in each individual control condition, we collapsed the three non-wait-list control categories into an "active control" category.

To examine the effects of interventions more broadly, we performed additional "omnibus" analyses. We investigated the impact of "any treatment" (i.e., collapsing all unimodal psychological treatments and multidisciplinatry treatments into one category) compared with active control, wait-list control, and "any control" (i.e., collapsing across active control and wait list). We also investigated the impact of "any psychological treatment" (i.e., collapsing all unimodal psychological treatments together) against the various control conditions. Finally, we collapsed all treatments

together, all controls together, all outcomes together, and all time periods together (i.e., any treatment vs. any control for any outcome at any time period).

In order to meet the assumption of independence, each study appeared only once in each pooled ES. If there was more than one ES from a study eligible for summarization, we calculated a mean study ES value (Cooper, 1998). ESs were pooled under the random effects model, in which individual ESs are thought to be estimates of a population ES, in addition to participant-level sampling error and other randomly distributed sources of error (Lipsey & Wilson, 2001). ESs were pooled for any of the aforementioned treatment–control–time period–outcome–combinations that produced at least three ESs.[2]

To examine the impact of moderator variables, we used weighted multiple meta-regression analyses for each summary statistic containing a minimum of six ESs. Study characteristics and study sample characteristics were used as predictors. All calculations were performed on SPSS (Version 11.0, 2001), with the syntax provided by Lipsey and Wilson (2001). Effects were considered statistically significant if $p < .05$, marginally significant if $.05 < p < .10$, and nonsignificant if $p > .10$.[3]

Two additional statistics were calculated for each analysis: I^2 and Fail-Safe *N*. The I^2 indicates the percentage of variance in a pooled ES that can be attributed to heterogeneity. Values of 25% are considered low, 50% moderate, and 75% high (Higgins, Thompson, Deeks, & Altman, 2003). Fail-Safe *N* is used to calculate how vulnerable meta-analysis ES findings are to the possibility of undiscovered articles. We calculated Fail-Safe *N* statistics to determine the number of undiscovered, approximately equally sized studies with an ES of zero that would be needed in order to reduce our statistically significant finding to nonsignificance, with a *p* value of .10 (Orwin, 1983).

Results

Summary of Available Data

A total of 34 articles covering 31 separate studies met our inclusion criteria, contained extractable data, and were entered into our database. Of these 31 studies, 7 studies were excluded from our analyses because they featured contrasts between conditions that were, as per our methodology for grouping studies, essentially identical: Five studies contrasted different multidisciplinary interventions against each other, and 2 studies contrasted different approaches to BT/relaxation training. Another 2 studies were excluded from our analyses because none of their contrasts pooled into groups of at least three and were therefore never summarized (i.e., CBT vs. multidisciplinary treatment and supportive counseling vs. BT). Hence, our analyses included data from 22 studies, published across 25 articles (see Table 1).

These 22 studies contained a total of 205 ESs, for an average of 9.32 ESs per study. Not all studies reported data for all 12 of our outcome domains. The percentage of studies reporting data for each of our outcomes were as follows: pain intensity (77%), pain interference (41%), depression (36%), health care utilization, visits (36%), disability—working versus not working (36%), health-related quality of life (18%), anxiety (18%), global improvement (18%), health care utilization—medications (18%), distress (14%), health care utilization—costs (3%), disability—receiving compensation (3%).

Study-Level Descriptors

The 22 included studies were detailed in articles published between 1982 and 2003. Study sample sizes ranged from 20 to 293, with a mean of 79.42 ($SD = 59.50$). Twenty studies reported the percentage of male subjects ($M = 41.69\%$; range = 8.33–68.82%). Nineteen studies reported mean subject age ($M = 42.70$ years; range = 35.50–49.70 years), and 13 studies reported pain duration ($M = 90.86$ months; range = 4.60–154.80 months). Only 2 studies reported subject race/ethnicity, and none reported mean subject education.

Study-Level Methodological Elements Ratings

Methodological elements scores were summarized by dividing total "met" items by total possible "met" items, and scores could vary between 0 and 1.0. Among the 22 studies in our analyses, methodological elements scores varied between .17 and .64 ($M = 0.43$, $SD = 0.13$). The percentage of studies meeting each criterion is available from Robert D. Kerns.

Pooled ESs

Any psychological intervention or multidisciplinary intervention. Table 2 provides pooled ESs for combined psychological and multidisciplinary treatment groups contrasted against control conditions. In our largest omnibus-style analysis, when psychological and multidisciplinary treatment groups were collapsed together and contrasted against all control conditions, across all time periods, and across all outcome domains, treatment proved to be superior to control ($p < .05$, $d = .16$). This result was based on 135 ESs from 21 studies, averaged by study into 21 mean ESs.

When all treatment groups were collapsed together and contrasted against all control conditions at posttreatment, treatment proved superior to control on pain intensity ($p < .05$, $d = .41$), pain interference ($p < .05$, $d = .23$), and health-related quality of life ($p < .05$, $d = .41$). Treatment was not superior to control for depression ($p > .10$). When outcomes from all treatment groups combined were contrasted against only the wait-list control groups at posttreatment, treatment was superior to wait-list control for pain intensity ($p < .01$, $d = .50$) and health-related quality of life ($p < .05$, $d = .42$). A trend toward superior performance for treatment was apparent for pain interference ($p < .10$, 95% CI = −.02, .88). All treatments combined did not outperform wait-list controls on reduction in depression scores ($p > .10$). When all intervention groups were contrasted against the active control condition, treatment was not superior for pain intensity or pain interference at posttreatment or at follow-up ($ps > .10$). Similarly, treatment was not superior to active control for reduction of health care visits or medications used at follow-up ($ps < .10$). Studies pooled for contrasts of any intervention versus an active control to promote return to work are reported with multidisciplinary outcomes.

[2] There are no widely used guidelines for the minimum number of ESs required for pooling. Previous meta-analyses in this area have pooled as few as two ESs (e.g., van Tulder et al., 2001). We chose to set our minimum threshold at three in order to reduce the likelihood of a single, unusual outcome biasing our conclusions.

[3] *P* values of between .05 and .10 were noted in light of potential Type II errors as a result of the limited number of ESs in some of our pooled ES calculations.

Table 1
References of Studies (n = 22) Included in Meta-Analysis

Published studies included in meta-analysis	*N*	MES[§]	Outcome variable[*]
Alaranta et al., 1994	293	0.17	DW, PF
Basler, Jakle, & Kroner-Herwig, 1997	76	0.50	PF, PI
A. F. Bendix et al., 1996; A. F. Bendix, Bendix, Haestrup, & Busch, 1998	94	0.55	OB
A. F. Bendix, Bendix, Lund, Kirkbak, & Ostenfeld, 1997; A. F. Bendix, Bendix, Haestrup, & Busch, 1998	103	0.50	DW, PI, HM, HV
T. Bendix, Bendix, Labriola, Haestrup, & Ebbehoj, 2000	99	0.64	PI, HM, HV
Brox et al., 2003	61	0.58	DW, PF, PI
Christensen, Laurberg, & Bunger, 2003	81	0.36	DW, HV
Corey, Koepfler, Etlin, & Day, 1996	138	0.33	DW
Donaldson, Romney, Donaldson, & Skubick, 1994	36	0.58	PI
Goossens et al., 1998	148	0.42	HM, HV
Hernandez-Reif, Field, Krasnegor, & Theakston, 2001	24	0.50	MD, PI
Hodselmans, Jaegers, & Goeken, 2001	21	0.18	PF
Kankaanpaa, Taimela, Airaksinen, & Hanninen, 1999	54	0.42	PF, PI
Keijsers et al., 1989	30	0.50	MD, PF, PI
Klaber Moffett, Chase, Portek, & Ennis, 1986	78	0.25	PF, PI
Newton-John, Spence, & Schotte, 1995	44	0.36	MD, PI, PF
Nouwen, 1983	20	0.55	PI
Saarijarvi, 1991; Saarijarvi, Rytokoski, & Alanen, 1991; Saarijarvi, Alanen, Rytokoski, & Hyyppa, 1992	59	0.27	OB
Turner, 1982	36	0.33	MD, PI, HQ
Turner & Clancy, 1988	74	0.45	HQ, PI
Turner, Clancy, McQuade, & Cardenas, 1990	76	0.42	PI, MD, HQ
Turner & Jensen, 1993	102	0.58	MD, PI, HQ

Note. [§]MES = methodological element score, [*]PI = pain intensity, HQ = health-related quality of life, PF = pain interference, MD = mood depression, HV = healthcare utilization visits, HM = healthcare utilization number of medications, DW = disability working, OB = only contributed outcomes that were summarized in the omnibus test (all controls, time periods, and outcome variables).

Any psychological intervention. Table 3 provides pooled ESs for any psychological treatment contrasted against wait list control. Psychological treatment was superior to wait-list control conditions in reducing pain ($p < .01$, $d = .48$) and demonstrated a trend toward superiority for health-related quality of life ($p < .10$, 95% CI = $-.04$, .82). It was not significantly superior to wait-list control for treating depression ($p > .10$).

Multidisciplinary interventions. Table 3 provides ESs for multidisciplinary treatment contrasted against active control groups. Multidisciplinary treatment was superior to active control conditions at posttreatment at reducing pain interference ($p < .05$, $d = .36$) but not pain intensity ($p > .10$). Multidisciplinary interventions were not superior to active controls for pain intensity or pain interference at follow-up ($p > .10$). Multidisciplinary interventions were superior to active control conditions at improving the percentage who returned to work, with ES of .36 at follow-up ($p < .05$) and .53 at long-term follow-up ($p < .05$).

Individual psychological interventions. Table 4 provides ESs for individual psychological interventions contrasted against various comparison groups. CBT proved superior to wait-list controls at reducing posttreatment pain intensity ($p < .01$, $d = .62$), but not health-related quality of life or depression ($p > .10$). SRT proved superior to wait-list controls at reducing posttreatment pain intensity ($p < .001$, $d = .75$,) and depression ($p < .05$, $d = .81$). CBT was marginally less effective than SRT for reducing depression at posttreatment ($p < .10$, 95% CI = $-.83$, .01) but not for pain intensity at posttreatment or follow-up periods ($ps > .10$).

Moderator Analyses

Six pooled ESs provided the minimum number of contrasts (six) for regressing moderators (percentage male, sample size, study methodological elements, and age) on ES. Of the 24 regression analyses run, only 2 were statistically significant. For the contrast pitting any treatment versus any control on depression at posttreatment, percentage male and study methodology predicted smaller ESs ($p < .01$, $B = -.93$; $p < .05$, $B = -.69$, respectively).

I^2 and Fail-Safe N

Study homogeneity was generally in the mild-to-moderate range. Of the 34 I^2 statistics calculated, none exceeded 75%. Ten exceed 50%, suggesting moderate-to-high levels of heterogeneity. Seven fell between 25% and 50%, suggesting low-to-moderate levels of heterogeneity. Seventeen fell below 25%, suggesting low levels of heterogeneity and therefore a high degree of concordance between studies.

Eighteen Fail-Safe *N* statistics were calculated (one for each statistically significant finding). In 12 of the 18 analyses, the number of unpublished or unlocated studies with ES of zero

Table 2
Effect Sizes Between Psychological and Multidisciplinary Treatments Combined Versus Comparison Groups for Each Outcome Variable and Time Period

Outcome variable	Comparison group	*k*	*N*	*d*	95% CI	*z*	*p*	I^2	Fail-safe *N*
All time periods									
All outcomes	Any control	21	1,700	.16	.03, .30	2.33	.02	.32	8.74
At posttreatment									
Pain intensity	Any control	11	614	.27	.06, .48	2.47	.01	.20	5.52
HRQOL	Any control	4	288	.41	.00, .83	1.98	.05	.32	.81
Pain interference	Any control	7	596	.23	.06, .39	2.71	.01	0	4.53
Depression	Any control	6	312	.31	−.10, .71	1.50	.13	.44	
Pain intensity	Wait list	7	382	.50	.23, .77	3.64	.00	0	8.49
HRQOL	Wait list	4	288	.42	.00, .83	1.98	.05	.32	.81
Pain interference	Wait list	3	95	.43	−.02, .88	1.86	.06	0	.39
Depression	Wait list	5	288	.35	−.15, .84	1.38	.17	.55	
Pain intensity	Active control	5	308	.06	−.18, .30	.51	.61	0	
Pain interference	Active control	4	501	−.10	−.33, .13	−.88	.38	.27	
At follow-up									
Pain intensity	Active control	5	393	.16	−.27, .59	.72	.47	.71	
Pain interference	Active control	3	408	.08	−.28, .44	.44	.66	.53	
Disability: working	Active control	3	245	.36	.06, .65	2.37	.02	0	1.33
Health care visits	Active control	4	431	−.03	−31, .26	−.15	.88	.34	
Medications	Active control	3	350	.12	−.14, .38	.93	.35	0	
At long-term follow-up									
Disability: working	Active control	4	609	.53	.19, .86	3.10	.03	.66	3.55

Note. CI = confidence interval; HRQOL = health-related quality of life.

needed to reduce the significant findings beyond the point of marginal significance (i.e., $p > .10$) was more than 50% of the number found. For example, 11 studies were found that pitted any treatment against any control for posttreatment pain intensity. An additional six studies with ES of zero would be needed to render this finding nonsignificant.

Discussion

This study is the first to use state-of-the-art meta-analytic techniques to derive indices of the efficacy of psychological interventions, broadly defined and delivered either alone or as a component of a multidisciplinary intervention for CLBP of known and un-

Table 3
Effect Sizes Between any Psychological Treatment or Multidisciplinary Treatment Versus Comparison Groups

Outcome variable	Comparison group	*k*	*N*	*d*	95% CI	*z*	*p*	I^2	Fail-safe *N*
Any psychological intervention at posttreatment									
Pain intensity	Wait list	6	352	.52	.23, .82	3.53	.00	0	6.88
HRQOL	Wait list	4	288	.39	−.04, .82	1.78	.08	.38	.33
Depression	Wait list	4	258	.34	−.32, .99	1.02	.31	.69	
Any multidisciplinary intervention at posttreatment									
Pain intensity	Active control	4	284	.12	−.13, .38	.98	.33	0	
Pain interference	Active control	4	501	.20	.02, .37	2.18	.03	0	1.33
Any multidisciplinary intervention at follow-up									
Pain intensity	Active control	5	393	.15	−.29, .59	.66	.51	.72	
Pain interference	Active control	3	408	.09	−.26, .44	.52	.60	.52	
Disability: working	Active control	3	245	.36	.06, .65	2.37	.02	0	1.33
Any multidisciplinary intervention at long-term follow-up									
Disability: working	Active control	4	609	.53	.19, .86	3.10	.03	.66	3.55

Note. CI = confidence interval; HRQOL = health-related quality of life.

Table 4
Effect Sizes Between Specific Psychological Treatments Versus Comparison Groups

Outcome variable	Comparison group	*k*	*N*	*d*	95% CI	*z*	*p*	I^2	Fail-safe *N*
Cognitive–behavioral treatment at posttreatment									
Pain intensity	Wait list	4	256	.62	.25, .98	3.32	.00	.06	4.07
HRQOL	Wait list	3	212	.40	−.27, 1.07	1.18	.24	.62	
Depression	Wait list	3	182	.34	−.51, 1.19	.79	.43	.73	
Pain intensity	Self-regulatory	4	218	−.13	−.50, .23	−.72	.47	0	
Depression	Self-regulatory	3	182	−.41	−.83, .01	−1.89	.06	0	.45
Cognitive–behavioral treatment at follow-up									
Pain intensity	Self-regulatory	3	182	−.10	−.55, .35	−.44	.66	0	
Self-regulatory treatment at posttreatment									
Pain intensity	Wait list	4	202	.75	.35, 1.15	3.68	.00	0	4.95
Depression	Wait list	3	182	.81	.11, 1.52	2.27	.02	0	1.14

Note. CI = confidence interval; HRQOL = health-related quality of life.

known etiology. Our review focused solely on analyses of results from RCTs, a design considered to provide the strongest test of the efficacy of a specific intervention. The methods used in the current analyses were particularly rigorous and were designed to provide the most reliable, albeit conservative, conclusions about intervention efficacy. These methods encourage confidence in any positive effects of psychological interventions for CLBP that were demonstrated by this review, particularly findings entailing mild to moderate heterogeneity and for which minimal publication bias was deemed to influence results.

In total, there were 34 analyses run and 18 significant findings. As per the largest omnibus test, psychological interventions for CLBP, whether alone or as part of multidisciplinary treatment, are effective. Although the ES was small, it was robust with respect to associated heterogeneity and Fail-Safe *N* statistics. When psychological interventions were compared with wait-list controls, moderate ESs were noted for pain intensity and health-related quality of life. Small effects were noted for reducing pain interference and moderate effects for lowering work-related disability when compared against active controls. Moderate to large positive effects on pain intensity were demonstrated for specific psychological interventions described as CBT, whereas large effects were demonstrated for treatments described as SRT. Only SRT was shown to significantly reduce depression at posttreatment when compared with a wait-list control, and there was a trend toward significance for reducing depression when compared against efficacy of CBT. Pain intensity demonstrated the most robust effect across various control groups, an effect consistent with prior reviews.

In general, results of the current meta-analyses found little support for the comparative efficacy of psychological interventions relative to other active treatment conditions. Two exceptions to these findings are noteworthy. As was reported by Flor and her colleagues (1992), behavioral outcomes, including return to work, were significantly better for multidisciplinary, compared with unimodal, treatments. Our analyses extend these findings by demonstrating that multidisciplinary programs that included psychological interventions were superior to other active treatment conditions at improving work-related outcomes at both short-term and long-term follow-up. An examination of effects of multidisciplinary treatment versus wait-list control, which likely would have yielded larger effects, was precluded by the dearth of data. It is also noteworthy that our analyses provided evidence of relatively strong effects of SRT such as biofeedback and relaxation training and the demonstration that these interventions may outperform CBT in relieving depressive symptom severity. These findings encourage continued examination of these effects.

Our lack of substantive findings for moderator effects suggests that treatment effects do not vary by patient and study characteristics. Our results were generally consistent with previous research (Flor, Fydrich, & Turk, 1992). However, we were unable to examine several moderator effects because of insufficient data (i.e., education, pain duration, and race/ethnicity). The examination of race/ethnicity has not been attempted in previous reviews. It is unfortunate that so few studies reported these data so as to preclude such an analysis in our study. Growing evidence of race/ethnic differences in pain perception and pain care (Green, Baker, Sato, Washington, & Smith, 2003) encourages explicit consideration of these differences in future research.

Our study extends the findings of van Tulder and his colleagues (2001). Examination of the two review methods is particularly relevant to the consideration of apparent differences in their findings. Our strategy began with the careful selection of search terms that was designed to identify studies of broad psychological interventions, whereas the review by van Tulder and his colleagues was limited to studies examining the efficacy of interventions that were explicitly informed by learning theory and characterized as operant, cognitive, or respondent in nature. Ultimately, our search strategy and the broader time span led to the identification of a substantially larger number of studies that were considered in our analyses (922 vs. 244), and a larger number of studies included in the statistical analyses of effects (22 vs. 12); hence, our study involved an 83% increase in number of studies. Our study included 58% of those quantitatively summarized by van Tulder

and colleagues; the remaining five articles included in the van Tulder et al. study were excluded by us or did not pool. Twenty-seven percent of our studies were published after April 1999, the cutoff for the previous review. We quantitatively summarized 34 comparisons, nearly three times as many as in the previous study. This was despite our limiting the calculation of ES to comparisons in which there were at least three independent effects to be pooled. Of particular importance, we also used a conservative approach to the consideration of reports of nonsignificant effects in individual studies as well as examined measures of study heterogeneity and of publication bias associated with findings. Finally, our approach to calculate mean effects within a study provided a more reliable approach (using, e.g., 135 contrasts in our largest omnibus test).

As emphasized by the authors of the previously published systematic reviews, study methodology remains a concern. Although there are significant limitations to the use of methodological assessments, as mentioned above, it remains an important topic as indicated by CONSORT reporting guidelines. A particular shortcoming is the failure of all but a few published trials to report on data confirming treatment fidelity (Lichstein, Reidel, & Grieve, 1994). Investigators are encouraged to provide evidence that treatments are delivered as intended (e.g., treatment manuals are followed by therapists), received by the participants (e.g., understood the material), and enacted (e.g., adherence to therapist recommendations for practice of pain coping skills). Investigators should also provide evidence on concurrent treatments that may confound attributions of effects to incorporate an evaluation of perceived credibility or treatment satisfaction and to conform to other CONSORT reporting guidelines.

Investigators are also encouraged to adopt the recommendations of the IMMPACT group for the assessment of "core" outcome domains. When comparing two treatments, researchers are encouraged to consider adding a wait-list or placebo control condition in order to enhance the quantitative evaluation of treatment efficacy.

Limitations of our study were the lack of inclusion of dissertations and unpublished studies. Although we made efforts to include both of these data, this lack of data may have biased our findings. Nevertheless, results of the Fail-Safe *N* analyses suggest that publication bias had minimal influence on many of our findings.

In summary, results of the current meta-analysis provide support for the efficacy of psychological interventions in reducing self-reported pain, pain-related interference, depression, disability, and increasing health-related quality of life among persons with CLBP. The robust nature of these findings should encourage confidence among clinicians and researchers alike. Inconsistent effects of these interventions on emotional functioning underscore a challenge to the field. It can be hypothesized that the efficacy of psychological interventions can be enhanced through refinement designed to address ethnic/racial diversity and other putative mechanisms involved in the development and perpetuation of CLBP such as pain catastrophizing (Sullivan, Thorn, Rodgers, & Ward, 2004), fear avoidance (Vlaeyen & Linton, 2000), psychiatric comorbidities (Banks & Kerns, 1996), and pain-relevant communication (Kerns, Haythornthwaite, Southwick, & Giller, 1990). Continued examination of stepped care models that emphasize early intervention (Von Korff, 1999) and promotion of a self-management approach (Kerns & Habib, 2004) are also encouraged.

References

Alaranta, H., Rytokoski, U., Rissanen, A., Talo, S., Ronnemaa, T., Puukka, P., et al. (1994). Intensive physical and psychosocial training program for patients with chronic low back pain. A controlled clinical trial. *Spine, 19,* 1339–1349.

Andersson, G. B. J. (1997). The epidemiology of spinal disorders. In J. W. Frymoyer (Ed.), *The adult spine: Principles and practices* (2nd ed., pp. 93–141). New York: Raven Press.

Andersson, H. I., Ejlertsson, G., Leden, I., & Schersten, B. (1999). Impact of chronic pain on health care seeking, self care, and medication. Results from a population-based Swedish study. *Journal of Epidemiology and Community Health, 53,* 503–509.

Banks, S. M., & Kerns, R. D. (1996). Explaining high rates of depression in chronic pain: A diathesis-stress framework. *Psychological Bulletin, 119,* 95–110.

Basler, H. D., Jakle, C., & Kroner-Herwig, B. (1997). 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, 31,* 113–124.

Bendix, A. F., Bendix, T., Haestrup, C., & Busch, E. (1998). A prospective, randomized 5-year follow-up study of functional restoration in chronic low back pain patients. *European Spine Journal, 7,* 111–119.

Bendix, A. F., Bendix, T., Lund, C., Kirkbak, S., & Ostenfeld, S. (1997). Comparison of three intensive programs for chronic low back pain patients: A prospective, randomized, observer-blinded study with one-year follow-up. *Scandinavian Journal of Rehabilitation Medicine, 29,* 81–89.

Bendix, A. F., Bendix, T., Vaegter, K., Lund, C., Frolund, L., & Holm, L. (1996). Multidisciplinary intensive treatment for chronic low back pain: A randomized, prospective study. *Cleveland Clinic Journal of Medicine, 63,* 62–69.

Bendix, T., Bendix, A., Labriola, M., Haestrup, C., & Ebbehoj, N. (2000). Functional restoration versus outpatient physical training in chronic low back pain: A randomized comparative study. *Spine, 25,* 2494–2500.

Blyth, F. M., March, L. M., Nicholas, M. K., & Cousins, M. J. (2005). Self-management of chronic pain: A population-based study. *Pain, 113,* 285–292.

Bonica, J. J., & Loeser, J. D. (2000). History of pain concepts and therapies. In J. D. Loeser, S. H. Butler, C. R. Chapman, & D. C. Turk (Eds.), *Bonica's management of pain* (3rd ed., pp. 3–16). Philadelphia: Lippincott Williams & Wilkins.

Brox, J. I., Sorensen, R., Friis, A., Nygaard, O., Indahl, A., Keller, A., et al. (2003). Randomized clinical trial of lumbar instrumented fusion and cognitive intervention and exercises in patients with chronic low back pain and disc degeneration. *Spine, 28,* 1913–1921.

Campbell, J. N., & Mitchell, M. J. (1996). *Pain treatment centers at a crossroads: A practical and conceptual reappraisal.* Seattle, WA: IASP Press.

Centers for Disease Control (CDC). (2001). *Morbidity and Mortality Weekly Report, 50,* 120–125.

Christensen, F. B., Laurberg, I., & Bunger, C. E. (2003). Importance of the back-cafe concept to rehabilitation after lumbar spinal fusion: A randomized clinical study with a 2-year follow-up. *Spine, 28,* 2561–2569.

Clark, M., & Oxman, A. D. (2003). *Cochrane reviewers' handbook 4.2.0.* Retrieved January 5, 2004, from www.cochrane.dk/cochrane/handbook/handbook.htm

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand Oaks, CA: Sage.

Corey, D. T., Koepfler, L. E., Etlin, D., & Day, H. I. (1996). A limited functional restoration program for injured workers: A randomized trial. *Journal of Occupational Rehabilitation, 6,* 239–249.

Donaldson, S., Romney, D., Donaldson, M., & Skubick, D. (1994). Randomized study of the application of single motor unit biofeedback

training to chronic low back pain. *Journal of Occupational Rehabilitation, 4,* 23–37.

Dworkin, R. H., Turk, D. C., Farrar, J. T., Haythornthwaite, J. A., Jensen, M. P., Katz, N. P., et al. (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain, 113,* 9–19.

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

Gatchel, R. J., & Turk, D. C. (1996). *Psychological treatments for pain: A practitioner's handbook.* New York: Guilford Press.

Goossens, M. E. J., Rutten-Van Molken, M. P. M., Kole-Snijders, A. M. J., Vlaeyen, J. W. S., Van Breukelen, G., & Leidl, R. (1998). Health economic assessment of behavioural rehabilitation in chronic low back pain: A randomised clinical trial. *Health Economics, 7,* 39–51.

Green, C. R., Baker, T. A., Sato, Y., Washington, T. L., & Smith, E. M. (2003). Race and chronic pain: A comparative study of young black and white Americans presenting for management. *Journal of Pain, 4,* 176–183.

Guzman, J., Esmail, R., Karjalainen, K., Malmivaara, A., Irvin, E., & Bombardier, C. (2001). Multidisciplinary rehabilitation for chronic low back pain: Systematic review. *British Medical Journal, 322,* 1511–1516.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hernandez-Reif, M., Field, T., Krasnegor, J., & Theakston, H. (2001). Lower back pain is reduced and range of motion increased after massage therapy. *International Journal of Neuroscience, 106,* 131–145.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistence in meta-analysis. *British Medical Journal, 327,* 557–560.

Hodselmans, A. P., Jaegers, S. M., & Goeken, L. N. (2001). Short-term outcomes of a back school program for chronic low back pain. *Archives of Physical Medicine & Rehabilitation, 82,* 1099–1105.

Kankaanpaa, M., Taimela, S., Airaksinen, O., & Hanninen, O. (1999). The efficacy of active rehabilitation in chronic low back pain: Effect on pain intensity, self-experienced disability, and lumbar fatigability. *Spine, 24,* 1034–1042.

Keijsers, J. F. E., Groenman, N. H., Gerards, F. M., & Van Oudheusden, E. (1989). A back school in the Netherlands: Evaluating the results. *Patient Education & Counseling, 14,* 31–44.

Kerns, R. D., & Habib, S. (2004). A critical review of the pain readiness to change model. *Journal of Pain, 5,* 357–367.

Kerns, R. D., Haythornthwaite, J., Southwick, S., & Giller, E. L. J. (1990). The role of marital interaction in chronic pain and depressive symptom severity. *Journal of Psychosomatic Research, 34,* 401–408.

Klaber Moffett, J. A., Chase, S. M., Portek, I., & Ennis, J. R. (1986). A controlled, prospective study to evaluate the effectiveness of a back school in the relief of chronic low back pain. *Spine, 11,* 120–122.

Lichstein, K. L., Reidel, B. W., & Grieve, R. (1994). Fair tests of clinical trials: A treatment implementation model. *Advances in Behavioral Research and Therapy, 16,* 1–29.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Mapel, D. W., Shainline, M., Paez, K., & Gunter, M. (2004). Hospital, pharmacy, and outpatient costs for osteoarthritis and chronic back pain. *Journal of Rheumatology, 31,* 573–583.

Moher, D., Schulz, K. F., & Altman, D. G. (2001). *Revised recommendations for improving the quality of reports of parallel group randomized trials.* Retrieved September 1, 2004, from http://www.consort-statement.org/statement/revisedstatement.htm

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

Mounce, K. (2002). Back pain. *Rheumatology, 41,* 1–5.

Newton-John, T. R. O., Spence, S. H., & Schotte, D. (1995). Cognitive–behavioral therapy versus EMG biofeedback in the treatment of chronic low back pain. *Behaviour Research & Therapy, 33,* 691–697.

Nielson, W. R., & Weir, R. (2001). Biopsychosocial approaches to the treatment of chronic pain. *Clinical Journal of Pain, 17*(Suppl. 4), S114–S127.

Nouwen, A. (1983). EMG biofeedback used to reduce standing levels of paraspinal muscle tension in chronic low back pain. *Pain, 17,* 353–360.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal for Educational Statistics, 8,* 157–159.

Saarijarvi, S. (1991). A controlled study of couple therapy in chronic low back pain patients. Effects on marital satisfaction, psychological distress and health attitudes. *Journal of Psychosomatic Research, 35,* 265–272.

Saarijarvi, S., Alanen, E., Rytokoski, U., & Hyyppa, M. T. (1992). Couple therapy improves mental well-being in chronic low back pain patients. A controlled, five year follow-up study. *Journal of Psychosomatic Research, 36,* 651–656.

Saarijarvi, S., Rytokoski, U., & Alanen, E. (1991). A controlled study of couple therapy in chronic low back pain patients. No improvement of disability. *Journal of Psychosomatic Research, 35,* 671–677.

Smith, B., & Gribbin, M. (2001). Introduction. *Clinical Journal of Pain, 14*(Suppl. 4), S1–S4.

SPSS. (2001). SPSS for Windows. Chicago, IL: Author.

Sullivan, M. J., Thorn, B., Rodgers, W., & Ward, L. C. (2004). Path model of psychological antecedents to pain experience: Experimental and clinical findings. *Clinical Journal of Pain, 20,* 164–173.

Turk, D. C., Dworkin, R. H., Allen, R. R., Bellamy, N., Brandenburg, N., Carr, D. B., et al. (2003). Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain, 106,* 337–345.

Turner, J. A. (1982). Comparison of group progressive-relaxation training and cognitive-behavioral group therapy for chronic low back pain. *Journal of Consulting and Clinical Psychology, 50,* 757–765.

Turner, J. A., & Clancy, S. (1988). Comparison of operant behavioral and cognitive-behavioral group treatment for chronic low back pain. *Journal of Consulting and Clinical Psychology, 56,* 261–266.

Turner, J. A., Clancy, S., McQuade, K. J., & Cardenas, D. D. (1990). Effectiveness of behavioral therapy for chronic low back pain: A component analysis. *Journal of Consulting and Clinical Psychology, 58,* 573–579.

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

van Tulder, M. W., Koes, B. W., & Bouter, L. M. (1997). Conservative treatment of acute and chronic nonspecific low back pain: A systematic review of randomized controlled trials of the most common interventions. *Spine, 22,* 2128–2156.

van Tulder, M. W., Ostelo, R., Vlaeyen, J. W. S., Linton, S. J., Morley, S. J., & Assendelft, W. J. J. (2001). Behavioral treatment for chronic low back pain: A systematic review within the framework of the Cochrane Back Review Group. *Spine, 26,* 270–281.

Vlaeyen, J. W., & Linton, S. J. (2000). Fear-avoidance and its consequences in chronic musculoskeletal pain: A state of the art. *Pain, 85,* 317–332.

Von Korff, M. (1999). Pain management in primary care: An individualized stepped-care approach. In R. J. Gatchel & D. C. Turk (Eds.), *Psychosocial factors in pain: Critical perspectives* (pp. 360–373). New York: Guilford Press.

Waddell, G. (1987). A new clinical model for the treatment of low back pain. *Spine, 12,* 632–644.

Watson, P. J., Booker, C. K., Moores, L., & Main, C. J. (2004). Returning the chronically unemployed with low back pain to employment. *European Journal of Pain, 8,* 359–369.