Critical Review

How Can We Best Reduce Pain Catastrophizing in Adults With Chronic Noncancer Pain? A Systematic Review and Meta-Analysis

Robert Schütze,* Clare Rees,* Anne Smith,† Helen Slater,† Jared M. Campbell,‡§ and Peter O’Sullivan†*

*SCHOOL OF PSYCHOLOGY AND SPEECH PATHOLOGY, †SCHOOL OF PHYSIOTHERAPY AND EXERCISE SCIENCE, CURTIN UNIVERSITY, PERTH, AUSTRALIA. ‡JOANNA BRIGGS INSTITUTE, UNIVERSITY OF ADELAIDE, ADELAIDE, AUSTRALIA. §FACULTY OF SCIENCE AND ENGINEERING, MACQUARIE UNIVERSITY, SYDNEY, AUSTRALIA.

Abstract: Pain catastrophizing (PC), defined as an exaggerated negative cognitive-affective orientation toward pain, is one of the strongest psychological predictors of pain outcomes. Although regularly included as a process variable in clinical trials, there have been no comprehensive reviews of how it can be modified. Using a registered protocol (PROSPERO 2016 CRD42016042761), we searched MEDLINE, PsychINFO, EMBASE, CINAHL, and CENTRAL up to November 2016 for all randomized controlled trials measuring PC in adults with chronic noncancer pain. Two authors independently screened studies and assessed bias risk using the Cochrane tool. Quality of evidence was rated according to Grading of Recommendations Assessment, Development and Evaluation criteria. We included 79 studies (n = 9,914), which mostly recruited participants with musculoskeletal pain and had low risk of bias. Meta-analyses (standardized mean difference) showed 9 interventions had efficacy compared with waitlist/usual care or active control, although evidence quality was often low. The best evidence (moderate-high quality) was found for cognitive-behavioral therapy, multimodal treatment, and acceptance and commitment therapy. Effects were generally of medium strength and had questionable clinical significance. When only the 8 studies targeting people with high PC were included, effects were larger and more consistent. Multimodal treatment showed the strongest effects when all studies were considered, whereas cognitive-behavioral therapy had the best evidence among targeted studies. Perspective: PC is a modifiable characteristic but most interventions produce only modest benefit unless targeted to people with high PC. More research into theory-driven interventions matched to specific patient profiles is required to improve treatment efficacy and efficiency.

Key words: Pain catastrophizing, systematic review, meta-analysis, chronic pain, pain management.

Pain catastrophizing (PC) is a negative cognitive-affective response to pain and a large body of research shows it is a significant risk marker for adverse pain and health outcomes. Elevated PC is associated with greater disability,99 pain intensity,124 depression,42 anxiety,76 work absenteeism,10 opioid misuse,12 and health-care utilization.38 A tendency to catastrophize can also predict the transition to chronicity and its maintenance,115 with the influential fear-avoidance model of pain providing an account of...
how PC facilitates pain, disability, and distress, particularly in musculoskeletal pain.\textsuperscript{56,138,141}

Experimental and clinical data show that PC is associated with a range of biological processes that could modulate nociception. These include: dysregulation of the hypothalamic-pituitary axis that is linked to central nervous system sensitization\textsuperscript{81,100}; reduced descending inhibitory control through endogenous opioid pathways\textsuperscript{55}; increased activation of brain areas associated with affective aspects of pain\textsuperscript{114}; and pain-facilitating changes in functional connectivity of the brain’s default mode network.\textsuperscript{70}

In treatment settings, PC is an important process variable that mediates improvements through interventions such as cognitive-behavioral therapy (CBT),\textsuperscript{71,113} acceptance and commitment therapy (ACT),\textsuperscript{106} exercise-based rehabilitation,\textsuperscript{116} and multidisciplinary treatment.\textsuperscript{22,102,118} Some studies using cross-lagged designs show that improvements in PC early in treatment predict later improvements in pain and disability.\textsuperscript{22,102} As a result, PC has become a key treatment target, particularly in psychological and multidisciplinary interventions for people with chronic noncancer pain. Research has tended to focus on musculoskeletal pain such as chronic low back pain (CLBP),\textsuperscript{6} neck pain,\textsuperscript{149} and osteoarthritis,\textsuperscript{13} as well as fibromyalgia,\textsuperscript{3} perioperative pain in the context of joint replacement,\textsuperscript{103} and more recently neuropathic pain.\textsuperscript{102}

However, it is still unclear how best to help people with pain to catastrophize less, because a range of different interventions produce benefit. There seems to be the most evidence for CBT, with the only meta-analytic data on PC interventions coming from the latest Cochrane review of psychological therapies for chronic pain.\textsuperscript{149} This showed that CBT reduces PC with a medium standardized mean difference (SMD) effect of \(-.53\) compared with waitlist at post-test.\textsuperscript{149} However, in a high-quality head-to-head trial comparing CBT, exercise (general aerobic and strength training), and multidisciplinary treatment combining CBT and exercise, all 3 interventions showed similar effects of moderate strength.\textsuperscript{116} This is surprising because exercise does not explicitly target unhelpful thinking processes. More recently, emerging so-called third wave psychological therapies such as ACT and mindfulness meditation have also shown efficacy for reducing PC,\textsuperscript{87,52} with some suggesting large effect sizes.\textsuperscript{77} A recent head-to-head comparison of CBT and mindfulness meditation in people with CLBP showed both were efficacious, with mindfulness slightly superior in reducing PC in the short term.\textsuperscript{131}

Although these data suggest that there are a range of different ways to reduce PC, there is no clearly superior intervention and the mechanisms that underpin this change remain unclear. To our knowledge, a meta-analytic approach that investigates all interventions measuring treatment-related changes in PC has not been conducted. The present study therefore attempts to fill this gap in the literature. Specifically, it aims to: 1) systematically review and describe randomized controlled trials (RCTs) that measure catastrophizing changes in chronic noncancer pain; 2) document and compare the pooled effects of different interventions; and 3) identify factors that moderate the efficacy of these interventions. Because of the evidence cited previously that it is not only interventions designed to target PC that show efficacy in reducing it,\textsuperscript{116} this review aims to examine all treatment-related changes in PC regardless of whether catastrophizing was specifically targeted as a primary outcome. Although it is therefore likely that many of the included studies do not primarily target PC, it allows for an examination of a wider array of possibly efficacious treatments, rather than just those intentionally designed to reduce catastrophizing in high-risk cohorts.

**Methods**

Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines\textsuperscript{82} for conducting and reporting systematic reviews were used to design this study and a review protocol was prospectively registered with PROSPERO.\textsuperscript{111}

**Data Sources and Search Strategy**

The primary search was conducted in the following databases: MEDLINE, EMBASE, PsycINFO, CINAHL, and CENTRAL (the Cochrane Central Register of Controlled Trials) up until November 2016. A search strategy was developed using free text words, questionnaire names, and MeSH headings according to published guidelines.\textsuperscript{61,105} We used validated search filters for RCTs, including the sensitivity-maximizing version of the Cochrane Highly Sensitive Search Strategy for MEDLINE,\textsuperscript{81} the CADTH filter for PsycINFO,\textsuperscript{31} and the SIGN filter for CINAHL\textsuperscript{113} (see Supplementary Table 1, which details the MEDLINE search strategy). Reference lists of retrieved studies and relevant review articles were also manually searched.

**Study Inclusion**

Studies meeting the following criteria were eligible for inclusion:

1. Participants reported any kind of chronic noncancer pain, defined as pain lasting \(\geq 3\) months;
2. Participants were adults (\(\geq 18\) years old);
3. Used at least 1 experimental intervention intended to reduce clinical, rather than experimental pain or pain-related outcomes;
4. Compared experimental interventions with waitlist/usual care control or an active control using an RCT design;
5. Analyzed \(\geq 20\) participants in each treatment arm at post-test. This condition attempted to reduce the risk of bias associated with small samples and increase the likelihood that included studies would be adequately powered. This condition is consistent with the most recent Cochrane review of psychological interventions for chronic pain, which also reports on PC.\textsuperscript{149}
6. Reported on changes in PC on a validated self-report measure; and
7. Study was available in English language article published in a peer-reviewed journal.

The inclusion of only published studies was intended to maximize the quality of included data by ensuring it had
passed peer-review, thereby reducing risk of bias. As noted earlier, this review aimed to examine all treatment-related changes in PC rather than only changes associated with studies specifically designed to reduce catastrophizing. Therefore, studies treating PC as either a primary outcome or a secondary/process outcome were included, although how PC was treated within studies was recorded for the sake of moderator analysis. This is described in the section, Moderator and Subgroup Analysis.

Using these criteria, 2 independent assessors (R.S., C.R.) screened the titles and abstracts of all studies identified by the search. Clearly irrelevant studies were excluded and the full text of the remaining articles were retrieved. Any discrepancies between assessors at screening were discussed and resolved by consensus. The same process was used to independently assess the full text of potentially eligible studies. If consensus was not achieved, a third assessor (H.S.) was consulted.

### Data Extraction and Management

Data from included studies were extracted by 1 assessor (R.S.) and checked by a second (A.S.). A customized piloted data extraction form on the basis of Cochrane guidelines was used to retrieve the following information: study characteristics (design, funding, country); sample characteristics (inclusion and exclusion criteria, age, gender, pain duration, pain condition, number randomized); intervention characteristics (content, duration, format, type of therapist, total therapist contact); and outcome characteristics (PC instrument, number of participants analyzed at each time point, catastrophizing scores at each time point). For PC outcomes, means and SDs/standard errors/confidence intervals (CIs) at baseline, post-test, and follow-up were extracted, or alternatively change scores from baseline plus standard errors, were extracted. Only data from relevant PC subscales were extracted from studies using broader multidimensional measures (eg, Coping Strategies Questionnaire). Where insufficient data were reported for meta-analysis, it was requested by contacting study authors. Data screening was managed using Covidence systematic review software. Outcomes for meta-analysis were entered into Comprehensive Meta-Analysis software by 1 author (R.S.) and checked by another (A.S.).

### Risk of Bias Rating

Risk of bias was assessed using the Cochrane Collaboration’s risk of bias tool. Two authors (R.S., J.M.C.) independently assessed each study in Covidence and resolved any differences through discussion to arrive at consensus. Each study was assessed against 6 domains in the standard tool: random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), and selective reporting (reporting bias). Because many studies were expected to involve behavioral interventions in which participants cannot be blinded to intervention content, participant blinding in trials with active control groups was assessed in terms of efforts to control for expectancy effects by concealing study hypotheses. Studies only using waiting list/usual care control groups are not able to control for expectancy effects in this way and were judged at high risk of performance bias. Only participant blinding, rather than therapist blinding, was assessed because therapists delivering behavioral interventions cannot be blinded to the content they are delivering.

Because PC was always measured with self-report questionnaires, which are relatively robust to detection bias, this was rated “low” if participants completed these independently (eg, at home) and “high” if measures were administered by unblinded assessors. Where missing data due to attrition had been excluded from analysis, attrition bias was judged as “high” if the loss was ≥20% of the allocated sample. However, a “low” rating was given if intention to treat (ITT) analysis was used with robust imputation methods such as multiple imputation, or theoretically justified modeling methods that included variables as covariates that might be predictive of withdrawal. For reports for which authors did not prospectively register their trials or publish a protocol, reporting bias was judged “unclear,” because it was not possible to determine whether all planned outcomes and analyses were adequately reported.

Following Cochrane recommendations, risk of bias results were used to classify studies as either at low risk of bias, or at unclear/high risk of bias. We defined low-risk studies as those having low risk ratings on at least 3 of the 6 bias categories, and also not being judged as at high risk on any critical bias category (random sequence generation, allocation concealment, incomplete outcome data, or selective reporting).

### Data Synthesis and Analysis

Comprehensive Meta-Analysis software was used for meta-analysis. Because PC was measured with a variety of different self-report instruments as a continuous variable, pooled effect sizes were generated as SMDs. A random effects model was used because of the expected variation in interventions included and therefore likely heterogeneity in effect sizes. Heterogeneity was assessed for statistical significance using the Cochrane Q statistic and its magnitude was assessed using the I² statistic, which describes the percentage of variability due to true differences in effect sizes rather than due to chance. Values of 25%, 50%, and 75% for I² were used to classify low, moderate, and high heterogeneity, respectively. Pooled effect estimates using SMD were interpreted according to Cohen’s criteria (small ≤.2; moderate = .5; large ≥.8). Four groups of analyses were planned:

1) Waitlist/usual care controlled trials at post-test for each intervention type;
2) Waitlist/usual care controlled trials at follow-up for each intervention type;
3) Active control group trials at post-test for each intervention type; and
Moderator and Subgroup Analysis
Where possible, these meta-analyses were performed on all included studies. However, they were also performed separately on the subset of studies that targeted PC as a primary outcome and whose cohorts had clinically significant levels of mean PC at baseline (on the basis of recommendations of a score of >24 on the Pain Catastrophizing Scale [PCS]112). This was done to explore effect sizes in an emulated clinical context where interventions are commonly matched to clinical risk factors (ie, a treatment aimed at reducing PC for those with clinically significant symptoms of catastrophizing).121 This was only included as a subgroup analysis because defining clinical PC is still problematic, with published PCS cutoffs varying from 16,21 to 20,148 24,112 and 30.120 Furthermore, because PC exists on a spectrum, it is likely that even those with moderate elevations could benefit from reducing these symptoms through treatment. Therefore, although effect sizes in targeted cohorts were deemed important to document, so too were effect sizes across a spectrum of baseline catastrophizing. Pooled effects are therefore presented separately for: 1) all included studies, and 2) targeted studies (Tables 1–4).

Moderation of pooled treatment effects was also explored through meta-regression using Comprehensive Meta-Analysis. The following moderator variables were tested: risk of bias status, baseline PC, intervention duration, facilitator contact time, pain condition, year of publication, type of facilitator, whether PC was a primary outcome, PC measure, and delivery format. To use baseline PC as a moderator, scores on the various PC measures were transformed into a common scale of 0 to 100. Meta-regression significance testing was relaxed to $P < .10$ because of expected low power associated with small samples. Finally, the statistical significance of differences in effect sizes between interventions in each of the analysis domains mentioned previously was measured with analysis of variance using the weighted sum of squares Q statistic.

Publication Bias
As per Cochrane recommendations,61 funnel plots of each meta-analysis with at least 10 studies were inspected and tested for publication bias. Smaller meta-analyses ($n < 10$) were not tested because of the high probability that they would be underpowered.61 Statistical evidence of bias through asymmetry of plots was tested using the Egger-weighted regression, with a significant $P$ value suggesting possible publication bias.61 The effect of publication bias on pooled effects was estimated using Duval and Tweedie’s trim and fill method to impute likely missing studies and an adjusted effect size when these studies were included.41

Quality of Evidence
We used Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria58 to assess the quality of evidence for each intervention in the 4 analyses described previously. Starting with an assumption of high-quality evidence because all data came from RCTs,58 we downgraded evidence quality 1 category for each of the following GRADE criteria:

1) Risk of bias: >25% of participants are from studies judged at high/unclear risk of bias according to the previously mentioned criteria (Supplementary Fig 1 for risk of bias summary);

2) Inconsistency: significant heterogeneity in pooled effect ($I^2 > 50$%);

3) Indirectness of evidence: interventions not directly compared; results unlikely to generalize; surrogate outcomes used;

4) Imprecision: total participants <400 (on the basis of optimal information size for a small effect, using normative approach of $a = .05, b = .20, SMD = .21$); and

5) Publication bias: significant selective publication of evidence on the basis of the previously described criteria.

On the basis of these criteria, evidence from each analysis was rated as either high, moderate, low, or very low quality, defined according to GRADE.110

Results
Search
The search strategy returned 2,411 citations, with a further 54 studies identified through manual searching. As shown in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow chart (Fig 1), the final review included 79 studies.1-4,6,7,9,12-16,26-30,32,39,40,45,50,51,53,54,57,59,63,64,66,69,71,73,77-81,83-86,88-93,95-98,104,107,108,116-119,125-132,134-137,144,145,147,150,152. Eight authors were contacted.
for further information or data, and 4 responded. Sufficient data for meta-analysis was available for 77 studies. A summary table of the characteristics of included studies is provided in Supplementary Table 2.

Description of Studies
Most studies originated in Europe (64.6%) and North America (22.8%), with the following countries most strongly represented: United States (21.5%), Netherlands (16.5%), Sweden (12.7%), Spain (10.1%), and Australia (10.1%). Included studies were published between 1988 and 2016, with a median publication date of 2013 (interquartile range [IQR] = 2009–2015). Most of these RCTs used a single control group (n = 61, 77.2%), although a handful used 2 (n = 16, 20.3%) or 3 (n = 2, 2.5%) control groups. The most common measures of PC were the PCS (n = 44, 55.7%) and Coping Strategies Questionnaire (n = 28, 35.4%). Four studies (5.1%) used the Pain-Related Self-Statements scale, 49 and 3 separate studies (1.3% each) used the Cognitive Errors Questionnaire, 75 Vaginal Penetration Cognition Questionnaire, 68 and Pain Cognition List. 140 These latter 3 scales were not listed in the published review protocol because they were only discovered during systematic database searching; however investigation of their psychometric properties justified inclusion, despite this minor protocol deviation.

Only 32 studies (40.5%) reported specifically targeting PC as a primary outcome, with most treating PC as a secondary outcome or not specifying a primary outcome. There were only 8 (10.1%) targeted studies that used PC as a primary outcome and included cohorts with high baseline catastrophizing.

Participants
There were 9,914 participants (74% female) studied in total, with trial samples ranging from 40 to 341 people. Participants ranged in age from 27 to 82 years, averaging 48 years overall. Pain duration was only reported in 58 studies (73%), with means ranging from 1.2 to 23 years (overall mean = 8.7 years). Spinal pain—most often CLBP or neck pain—was the most common pain condition (n = 24, 30.4%). Mixed pain cohorts (n = 19, 24.1%) and fibromyalgia (n = 17, 21.5%) were also strongly represented, although the mixed cohorts were mainly comprised of CLBP, making spinal pain by far the dominant pain condition represented overall. Baseline PC scores were available for 75 studies and when these were converted to a 0 to 100 scale the mean score was 44.3 (SD = 13.6). This corresponds to a score of 23 on the PCS, which has a possible score of 0 to 52 (higher scores mean higher PC). When scores were dichotomized on the basis of recent evidence that a score of ≥24 on the PCS represents high PC, 112 a large proportion of studies (n = 43, 57.3%) were found to have low PC samples at baseline. Almost identical results were found when only the 44 studies using the PCS were included in this analysis.
Interventions

Seventeen different types of intervention were identified. These could be broadly grouped into: those containing mostly psychological content (n = 48, 60.8%); those involving mostly physical treatments, such as exercise, acupuncture, or manual therapy (n = 7, 8.9%); multimodal interventions involving a combination of physical and psychological content (n = 22, 27.8%); and purely pharmacological treatments (n = 2, 2.5%). Within these groupings, the most common interventions studied were CBT (n = 28, 35.4%) and multimodal interventions (n = 20, 25.3%). However, all the multimodal interventions contained a CBT component, making CBT by far the most commonly used modality. The duration of interventions varied considerably, ranging from 1 day to 28 weeks (median = 8 hours, IQR = 5.75–12.00). Similarly, the amount of contact participants had with treatment facilitators varied from no contact in the case of online treatments, to 126 hours (median = 14 hours, IQR = 3.22–24.62). Less than half of the included studies gave as much facilitator contact to the control group as they did to the experimental group (n = 33, 41.8%).

In terms of format, most interventions were delivered face-to-face, with 34 (43%) using a group format and 24 (30.4%) delivered individually. Twenty-one (26.6%) of the included studies were predominantly self-administered using some form of media (internet, smartphone, telephone, booklet), although some Web-based interventions also involved minimal therapist contact via e-mail or telephone. The most common facilitators of interventions in the experimental arms of included studies were psychologists/psychotherapists (n = 25, 31.6%) and multidisciplinary teams (n = 18, 22.8%).

Risk of Bias

A summary of the risk of bias assessment for the 79 studies reviewed is presented for each bias category in Fig 2. Most studies (n = 60, 75.9%) had “low risk” ratings for at least half of the bias categories assessed. The median number of categories that were judged low risk for each study was 4 of the 6 included. Using criteria described earlier for judging each study’s overall risk of bias, 48 studies (60.8%) were low risk whereas 31 (39.2%) were unclear/high risk (see Supplementary Fig 1, which provides a risk of bias assessment for each included study).

Risk of bias was also related to certain study characteristics. For example, the number of bias categories judged low risk for each study was positively correlated with its sample size (Spearman ρ = .233, P = .039), publication year (Spearman ρ = .239, P = .043), and number of treatment arms (Spearman ρ = .258, P = .022). This suggests that less-biased studies tended to be larger, more recent trials.

Meta-Analysis of Catastrophizing Outcomes

Intervention Versus Waitlist/Usual Care: Post-Test Outcomes

The effects of different interventions on PC compared with waitlist/usual care at post-test are summarized in Table 1, with 9 interventions showing efficacy. Meta-analysis was possible for 5 of these—ACT, CBT, exercise, mindfulness, and multimodal treatment—which is depicted with forest plots in Fig 3. One outlier28 with a very large effect (SMD = −6.78) was removed because of its outsized effect on the CBT meta-analysis, although inclusion did not alter the direction or significance of the pooled effect. As shown in Table 1, effect sizes ranged from small in the case of CBT (SMD = −.25) to very large for graded exposure (SMD = −1.74). However, the quality of evidence was very low for the interventions with large effect sizes. The best quality evidence according to GRADE criteria was found for CBT, multimodal treatment, exercise, and mindfulness (moderate quality). Multimodal treatment (medium effect) and CBT (small effect) had the largest body of evidence, but effects were inconsistent, thereby reducing confidence in their estimates. Exercise and mindfulness showed consistent medium effects but the estimates were imprecise, requiring data from more participants to justify confidence in their effect estimates. Considering only the interventions for which meta-analysis was possible, the differences in pooled effect between interventions were not quite statistically significant (Q = 9.34, df = 4, P = .053).

Moderator and subgroup analysis. Multimodal treatment and CBT were the only interventions with sufficient studies for meta-regression using the moderator variables described earlier. For CBT, baseline PC was a significant moderator of pooled effect (Q = 3.56, df = 1, P = .06), favoring high baseline PC. The pooled effect of

![Figure 2](image-url)
Table 1. Effects of Different Interventions on PC Compared With Waitlist/Usual Care at Post-Test

<table>
<thead>
<tr>
<th>INTERVENTION</th>
<th>PARTICIPANTS (STUDIES)</th>
<th>SMD EFFECT (95% CI)</th>
<th>PCS CHANGE (95% CI)</th>
<th>HETEROGENEITY Q, I²</th>
<th>MODERATORS</th>
<th>EVIDENCE QUALITY (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All included studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT&lt;sup&gt;16,71&lt;/sup&gt;</td>
<td>263 (2)</td>
<td>−.71 * (−1.38 to −0.04)</td>
<td>−6.6 (−12.9 to −4.4)</td>
<td>6.71*, 85.10</td>
<td>Baseline PC</td>
<td>Low§,¶</td>
</tr>
<tr>
<td>CBT&lt;sup&gt;3,5,9-11,14,26,38,46,50,51,57-59,62&lt;/sup&gt;</td>
<td>1,933 (15)</td>
<td>−.25 ** (−.41 to −.10)</td>
<td>−2.3 (−3.8 to −.9)</td>
<td>35.80**, 60.90</td>
<td>Very low§,¶,‖</td>
<td></td>
</tr>
<tr>
<td>EFT&lt;sup&gt;8&lt;/sup&gt;</td>
<td>66 (1)</td>
<td>−.87 ** (−1.38 to −.36)</td>
<td>−8.1 (−12.9 to −3.4)</td>
<td>.22, .00</td>
<td>Very low §,¶,‖,§</td>
<td></td>
</tr>
<tr>
<td>Exercise&lt;sup&gt;44,51&lt;/sup&gt;</td>
<td>277 (2)</td>
<td>−.38*** (−.59 to −.17)</td>
<td>−3.5 (−5.5 to −1.6)</td>
<td>.08, .00</td>
<td>Very low §,¶,‖,§</td>
<td></td>
</tr>
<tr>
<td>Graded exposure&lt;sup&gt;30&lt;/sup&gt;</td>
<td>70 (1)</td>
<td>−1.74*** (−2.92 to −1.19)</td>
<td>−16.3 (−21.4 to 11.1)</td>
<td>18.91**, 68.28</td>
<td>Baseline PC; PC primary outcome</td>
<td>Moderate¶</td>
</tr>
<tr>
<td>Hypnosis&lt;sup&gt;48&lt;/sup&gt;</td>
<td>59 (1)</td>
<td>−.32 (−.84 to −.19)</td>
<td>−3.0 (−7.8 to 1.8)</td>
<td>.04, .00</td>
<td>Very low §,¶,‖,§</td>
<td></td>
</tr>
<tr>
<td>Manual therapy&lt;sup&gt;45&lt;/sup&gt;</td>
<td>48 (1)</td>
<td>−1.56*** (−2.21 to −0.92)</td>
<td>−14.6 (−20.6 to −8.6)</td>
<td>.04, .00</td>
<td>Very low §,¶,‖,§</td>
<td></td>
</tr>
<tr>
<td>Mindfulness&lt;sup&gt;19,58&lt;/sup&gt;</td>
<td>338 (2)</td>
<td>−.46*** (−.67 to −.24)</td>
<td>−4.3 (−6.3 to −2.2)</td>
<td>.08, .00</td>
<td>Moderate¶</td>
<td></td>
</tr>
<tr>
<td>Multimodal&lt;sup&gt;16,17,51-53,63,66&lt;/sup&gt;</td>
<td>737 (7)</td>
<td>−.63*** (−.89 to −.38)</td>
<td>−5.9 (−8.3 to −3.5)</td>
<td>.04, .00</td>
<td>Moderate¶</td>
<td></td>
</tr>
<tr>
<td>Yoga&lt;sup&gt;15&lt;/sup&gt;</td>
<td>53 (1)</td>
<td>−.71 * (−1.27 to −.15)</td>
<td>−6.6 (−11.9 to −1.4)</td>
<td>.04, .00</td>
<td>Moderate¶</td>
<td></td>
</tr>
<tr>
<td>Only studies targeting elevated PC † †</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT&lt;sup&gt;15,46,59&lt;/sup&gt;</td>
<td>288 (4)</td>
<td>−.45 * (−.85 to −.06)</td>
<td>−4.2 (−7.9 to −.6)</td>
<td>8.36*, 64.11</td>
<td>Very low §,¶,‖,§</td>
<td></td>
</tr>
<tr>
<td>Multimodal&lt;sup&gt;16,17,53&lt;/sup&gt;</td>
<td>375 (3)</td>
<td>−.88*** (−1.09 to −.66)</td>
<td>−8.2 (−10.2 to −6.2)</td>
<td>.42, .00</td>
<td>Moderate¶</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: EFT, emotional freedom technique.
* P < .05.
** P < .01.
*** P < .001.
†† Studies included in each pooled effect.
§ Change in PCS score calculated by multiplying SMD by average SD of included studies that used PCS (SD = 9.34).
¶ Downgraded because of inconsistency.
‖ Downgraded because of imprecision.
§§ Downgraded because of risk of bias.
† † Targeted interventions are those that treat PC as a primary outcome and have cohorts with clinically significant levels of catastrophizing (>24 equivalent on the PCS).
Figure 3. Pooled effects on PC of different interventions versus waitlist/usual care at post-test.
CBT among studies with high baseline PC was larger (SMD = −.36, 95% CI = −.55 to −.16, \( P < .001 \)) and less heterogeneous (\( Q = 18.75, \text{ df } = 8, \ P = .02, \ I^2 = 57.34 \)). Multimodal interventions were also moderated by baseline PC (\( Q = 4.78, \text{ df } = 1, \ P = .03 \)) as well as whether PC was targeted as a primary outcome. As shown in Table 1, these moderation effects are reflected in the subgroup analyses, where including only high baseline PC studies that targeted catastrophizing produced larger effect sizes for CBT (SMD = −.45) and multimodal treatment (SMD = −.88). A reduction in heterogeneity was also observed for multimodal treatments, but not CBT.

**Publication bias.** Only CBT had sufficient studies for an intervention-specific funnel plot. As shown in Fig 4, there was no evidence of asymmetry (Egger test = −1.52, df = 13, \( P = .13 \)) and no missing studies according to Duval and Tweedie’s trim and fill method.

**Intervention Versus Waitlist/Usual Care: Follow-Up Outcomes**

There were fewer studies providing follow-up data than post-test data; however, the efficacy of 5 different interventions compared with waitlist/usual care at 6 to 12 months is shown in Table 2. All interventions—ACT, CBT, hypnosis, mindfulness, and multimodal treatment—had significant medium effects. Meta-analysis was possible for ACT, CBT, and multimodal treatment, as depicted with forest plots in Fig 5. The outlier\(^2\)\(^8\) exerting an outsized effect on the post-test CBT meta-analysis was removed for the same reason. The quality of evidence ranged from very low (hypnosis) to moderate (CBT, multimodal). Again, CBT had the most data (6 studies), but the considerable heterogeneity of effect sizes reduced confidence in the pooled effect estimate. As shown in Fig 5, this heterogeneity was mainly because of the influence of 1 study with large effects.\(^3\) Multimodal treatment had more consistent effects, however, the estimate was imprecise in the absence of a larger sample. The pooled effects of the meta-analyzed interventions were not significantly different from each other (\( Q = 1.34, \text{ df } = 2, \ P = .51 \)).

**Moderator and subgroup analysis.** The only intervention with sufficient studies for meta-regression was CBT. Baseline PC was a significant moderator of pooled effect (\( Q = 3.42, \text{ df } = 1, \ P = .06 \)), favoring high baseline PC. Recalculating the pooled effect of CBT for studies with high baseline PC produced a larger effect (SMD = −.69, 95% CI = −1.31 to −.06, \( P < .05 \)) but increased heterogeneity (\( Q = 4.88, \text{ df } = 2, \ P = .03, \ I^2 = 79.50 \)). Similarly, the subgroup analysis of only targeted studies (Table 2) resulted in a larger CBT effect (SMD = −1.01), although this was only on the basis of 1 study.

**Publication bias.** There were not enough studies to reliably test for publication bias in any of the meta-analyses at follow-up.

**Intervention Versus Active Control: Post-Test Outcomes**

Ten different interventions were tested against active control interventions, as represented by the 40 studies and 4,191 participants in Table 3. Only half of these showed efficacy—ACT, CBT, exercise, hypnosis, and multimodal treatment—and effects were moderate except for multimodal treatment, which had a large effect. Meta-analysis was possible for the 7 interventions depicted with forest plots in Fig 6. Omitted from these analyses were studies that compared different variants of the same type of intervention.\(^28,39,50,89,93,95,118,132,136,144,150\) As shown in Table 3, there was high-quality evidence for ACT (SMD = −.44) on the basis of 4 studies. Again, CBT provided the most data (12 studies); however, several studies had an unclear or high risk of bias, so the quality of this evidence (SMD = −.47) was downgraded to
Table 2. Effects of Different Interventions on PC Compared With Waitlist/Usual Care at Follow-Up (6–12 Months)

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Participants (Studies)</th>
<th>SMD Effect (95% CI)</th>
<th>PCS Change (95% CI)</th>
<th>Heterogeneity Q, I²</th>
<th>Moderators</th>
<th>Evidence Quality (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All included studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT⁵⁶,⁷¹</td>
<td>263 (2)</td>
<td>−.60* (−1.06 to −.14)</td>
<td>−5.6 (−9.9 to −1.3)</td>
<td>3.27, 69.45</td>
<td></td>
<td>Low§,¶</td>
</tr>
<tr>
<td>CBT³,⁴,⁹,¹⁰,⁵⁷,⁵⁸</td>
<td>1,116 (6)</td>
<td>−.39***(−.59 to −.19)</td>
<td>−3.6 (−5.5 to −1.8)</td>
<td>13.04*, 61.64</td>
<td>Baseline PC</td>
<td>Moderate¶</td>
</tr>
<tr>
<td>Hypnosis⁶⁰</td>
<td>59 (1)</td>
<td>−.69* (−1.22 to −.17)</td>
<td>−6.4 (−11.4 to −1.6)</td>
<td></td>
<td></td>
<td>Very low§,¶</td>
</tr>
<tr>
<td>Mindfulness⁶⁸</td>
<td>229 (1)</td>
<td>−.46***(−.67 to −.24)</td>
<td>−4.3 (−6.2 to −2.2)</td>
<td></td>
<td></td>
<td>Low§,¶</td>
</tr>
<tr>
<td>Multimodal¹⁶,¹⁷</td>
<td>285 (2)</td>
<td>−.56***(−.80 to −.32)</td>
<td>−5.2 (−7.5 to −3.0)</td>
<td>.44, .00</td>
<td></td>
<td>Moderate¶</td>
</tr>
<tr>
<td>Only studies targeting elevated PC† †</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT³</td>
<td>95 (1)</td>
<td>−1.01***(−1.44 to −.59)</td>
<td>−9.4 (−13.4 to −5.5)</td>
<td></td>
<td></td>
<td>Moderate¶</td>
</tr>
<tr>
<td>Multimodal¹⁶,¹⁷</td>
<td>285 (2)</td>
<td>−.56***(−.80 to −.32)</td>
<td>−5.2 (−7.5 to −3.0)</td>
<td>.44, .00</td>
<td></td>
<td>Moderate¶</td>
</tr>
</tbody>
</table>

Abbreviation: EFT, emotional freedom technique.
*P < .05.
**P < .01.
***P < .001.
†Studies included in each pooled effect.
‡Change in PCS score calculated by multiplying SMD by average standard deviation of included studies that used PCS (SD = 9.34).
§Downgraded due to inconsistency.
¶Downgraded due to imprecision.
‖Downgraded due to risk of bias.
††Targeted interventions are those that treat PC as a primary outcome and have cohorts with clinically significant levels of catastrophizing (>24 equivalent on the PCS).
moderate. Multimodal treatment had moderate-quality evidence of a large effect (SMD = −1.00), which was downgraded because of high heterogeneity. All other interventions had low- or very low-quality evidence. Differences in pooled effects across meta-analyzed interventions were not statistically significant (Q = 9.74, df = 6, P = .136).

**Moderator and subgroup analysis.** There were sufficient studies for meta-regression of ACT, CBT, education, and multimodal treatments. For ACT, facilitator contact time significantly moderated treatment effect (Q = 4.02, df = 1, P = .04), favoring more contact. Delivery format also moderated the effect (Q = 3.76, df = 1, P = .05), with face-to-face interventions superior to internet interventions. For CBT, facilitator type moderated effect (Q = 8.53, df = 4, P = .07), favoring psychologists. Whether PC was targeted as a primary outcome also influenced pooled effects (Q = 5.66, df = 1, P = .02), and limiting the CBT analysis to studies targeting PC produced larger effects (SMD = −.65, 95% CI = −.85 to −.45, P < .001) with less heterogeneity (Q = 2.64, df = 3, P < .001, I² = .00). Baseline PC moderated the effect of education interventions (Q = 5.57, df = 1, P = .02), although unexpectedly favoring low baseline PC. Contact time influenced education effects as well (Q = 3.65, df = 1, P = .06), favoring more contact. Finally, for multimodal interventions baseline PC influenced outcome, favoring high PC as expected (Q = 4.33, df = 1, P = .04). Facilitator type also moderated multimodal outcomes, with multidisciplinary teams superior to physiotherapists (Q = 5.66, df = 1, P = .02). Limiting this meta-analysis to studies of multimodal interventions led only by multidisciplinary teams for people with high baseline PC increased the pooled effect (SMD = −1.52, 95% CI = −2.45 to −.59, P < .001) but heterogeneity remained high (Q = 102.81, df = 1, P < .001, I² = 96.11).

The subgroup analyses of targeted high PC interventions showed that only CBT was effective. It had stronger and more consistent effects in this cohort (SMD = −.84)
### Table 3. Effects of Different Interventions on PC Compared With Active Controls at Post-Test

<table>
<thead>
<tr>
<th>INTRODUCTION</th>
<th>PARTICIPANTS (STUDIES)</th>
<th>SMD EFFECT (95% CI)</th>
<th>PCS CHANGE (95% CI)</th>
<th>HETEROGENEITY Q, I²</th>
<th>MODERATORS</th>
<th>EVIDENCE QUALITY (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All included studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT[12,29,56,71]</td>
<td>474 (4)</td>
<td>−.44** (−.69 to −.19)</td>
<td>−4.1 (−6.4 to −1.8)</td>
<td>5.49, 44.32</td>
<td>Contact; format</td>
<td>High</td>
</tr>
<tr>
<td>Acupuncture[65]</td>
<td>126 (1)</td>
<td>−.01 (−.36 to .34)</td>
<td>−.1 (−3.4 to −3.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT[7,14,18,21,31,35,39,44,50,51]</td>
<td>1,251 (12)</td>
<td>−.47*** (−.62 to −.33)</td>
<td>−4.4 (−5.8 to −3.1)</td>
<td>16.18, 31.99</td>
<td>Facilitator; PC primary outcome</td>
<td>Moderate §, ¶</td>
</tr>
<tr>
<td>Education[22,28,36,43]</td>
<td>284 (4)</td>
<td>−.52 (−1.14 to .09)</td>
<td>−4.9 (−10.6 to −8)</td>
<td>18.67***, 83.94</td>
<td>Baseline PC; contact</td>
<td>Very low §, ¶, †</td>
</tr>
<tr>
<td>Exercise[64]</td>
<td>139 (1)</td>
<td>−.36* (−.69 to −.02)</td>
<td>−3.4 (−6.4 to −2)</td>
<td></td>
<td></td>
<td>Low §, ¶</td>
</tr>
<tr>
<td>Graded exposure[23]</td>
<td>77 (1)</td>
<td>−.34 (−.79 to .11)</td>
<td>−3.2 (−7.4 to −1.0)</td>
<td></td>
<td></td>
<td>Low §, ¶</td>
</tr>
<tr>
<td>Hypnosis[22]</td>
<td>169 (2)</td>
<td>−.47** (−.78 to −.16)</td>
<td>−4.4 (−7.3 to −1.5)</td>
<td>.04, .00</td>
<td></td>
<td>Low §, ¶</td>
</tr>
<tr>
<td>Mindfulness[20,23,67]</td>
<td>281 (3)</td>
<td>−.13 (−.37 to .10)</td>
<td>−1.2 (−3.5 to −9)</td>
<td>.26, .00</td>
<td></td>
<td>Low §, ¶</td>
</tr>
<tr>
<td>Multimodal[16,34,37,40-42,45,47,52]</td>
<td>1,258 (10)</td>
<td>−1.00*** (−1.54 to −.46)</td>
<td>−9.3 (−14.4 to −4.3)</td>
<td>176.93***, 94.91</td>
<td>Baseline PC; facilitator</td>
<td>Moderate §, ¶</td>
</tr>
<tr>
<td>Pharmacotherapy[27,54]</td>
<td>132 (2)</td>
<td>−.02 (−.37 to .32)</td>
<td>−.2 (−3.5 to 3.0)</td>
<td>.52, .00</td>
<td></td>
<td>Low §, ¶</td>
</tr>
</tbody>
</table>

Only studies targeting elevated PC††

<table>
<thead>
<tr>
<th>INTRODUCTION</th>
<th>PARTICIPANTS (STUDIES)</th>
<th>SMD EFFECT (95% CI)</th>
<th>PCS CHANGE (95% CI)</th>
<th>HETEROGENEITY Q, I²</th>
<th>MODERATORS</th>
<th>EVIDENCE QUALITY (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBT[3,44]</td>
<td>146 (2)</td>
<td>−.84*** (−1.18 to −.50)</td>
<td>−7.8 (−11.0 to −4.7)</td>
<td>.23, .00</td>
<td></td>
<td>Low §, ¶</td>
</tr>
<tr>
<td>Education[28]</td>
<td>105 (1)</td>
<td>.20 (−.18 to .58)</td>
<td>−1.9 (−1.7 to 5.4)</td>
<td></td>
<td></td>
<td>Low §, ¶</td>
</tr>
<tr>
<td>Pharmacotherapy[24]</td>
<td>70 (1)</td>
<td>−.14 (−.61 to .33)</td>
<td>−8.2 (−10.2 to −6.2)</td>
<td></td>
<td></td>
<td>Low §, ¶</td>
</tr>
</tbody>
</table>

Abbreviation: EFT, emotional freedom technique.

*P < .05.

**P < .01.

***P < .001.

†Studies included in each pooled effect.

‡Change in PCS score calculated by multiplying SMD by average standard deviation of included studies that used PCS (SD = 9.34).

§Downgraded due to inconsistency.

¶Downgraded due to imprecision.

‖Downgraded due to risk of bias.

††Targeted interventions are those that treat PC as a primary outcome and have cohorts with clinically significant levels of catastrophizing (>24 equivalent on the PCS).
Figure 6. Pooled effects on PC of different interventions versus active control at post-test.

<table>
<thead>
<tr>
<th>Study name</th>
<th>Std diff in means</th>
<th>p-Value</th>
<th>Sample size</th>
<th>Std diff in means and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptance and Commitment Therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buhmann 2013a</td>
<td>-0.51</td>
<td>0.03</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>Kristiansson 2013</td>
<td>-0.40</td>
<td>0.02</td>
<td>59</td>
<td>65</td>
</tr>
<tr>
<td>Lattievo 2014</td>
<td>-0.77</td>
<td>0.00</td>
<td>51</td>
<td>53</td>
</tr>
<tr>
<td>Trompetter 2015</td>
<td>-0.18</td>
<td>0.23</td>
<td>82</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>-0.44</td>
<td>0.00</td>
<td>240</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td>-1.20</td>
<td>-0.65</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Cognitive Behavior Therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aida 2011</td>
<td>-0.46</td>
<td>0.00</td>
<td>39</td>
<td>30</td>
</tr>
<tr>
<td>Bergmo 2016</td>
<td>-0.17</td>
<td>0.46</td>
<td>39</td>
<td>30</td>
</tr>
<tr>
<td>Buhmann 2013b</td>
<td>-0.76</td>
<td>0.00</td>
<td>36</td>
<td>38</td>
</tr>
<tr>
<td>Buhmann 2015</td>
<td>-0.86</td>
<td>0.01</td>
<td>28</td>
<td>24</td>
</tr>
<tr>
<td>Chiacchi 2016</td>
<td>-0.55</td>
<td>0.00</td>
<td>95</td>
<td>104</td>
</tr>
<tr>
<td>Ernst 2008</td>
<td>-0.17</td>
<td>0.20</td>
<td>114</td>
<td>103</td>
</tr>
<tr>
<td>Gustavsson 2010</td>
<td>-0.33</td>
<td>0.04</td>
<td>77</td>
<td>79</td>
</tr>
<tr>
<td>Martinez 2014</td>
<td>-0.82</td>
<td>0.02</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>Naylor 2006</td>
<td>-0.73</td>
<td>0.01</td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td>Thors 2011</td>
<td>-0.17</td>
<td>0.52</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>Turner 2006</td>
<td>-0.43</td>
<td>0.03</td>
<td>61</td>
<td>65</td>
</tr>
<tr>
<td>Turner 2011</td>
<td>-0.83</td>
<td>0.01</td>
<td>47</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>-0.47</td>
<td>0.00</td>
<td>654</td>
<td>617</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gallagher 2013</td>
<td>-0.56</td>
<td>0.01</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>Itten 2014</td>
<td>-0.56</td>
<td>0.01</td>
<td>53</td>
<td>52</td>
</tr>
<tr>
<td>Meese 2010</td>
<td>-0.56</td>
<td>0.06</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>Moseley 2004</td>
<td>-0.52</td>
<td>0.00</td>
<td>143</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td>-2.00</td>
<td>-1.00</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td>Hypnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albrecht 2009</td>
<td>-0.42</td>
<td>0.19</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Ten Kule 1996</td>
<td>-0.69</td>
<td>0.01</td>
<td>69</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>-0.47</td>
<td>0.00</td>
<td>69</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>-1.00</td>
<td>-0.50</td>
<td>0.80</td>
<td>0.50</td>
</tr>
<tr>
<td>Mindfulness meditation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Donat 2015</td>
<td>-0.03</td>
<td>0.60</td>
<td>62</td>
<td>62</td>
</tr>
<tr>
<td>Garnt 2012</td>
<td>-0.08</td>
<td>0.73</td>
<td>34</td>
<td>32</td>
</tr>
<tr>
<td>Zwinka 2006</td>
<td>-0.13</td>
<td>0.28</td>
<td>47</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>-1.00</td>
<td>-0.50</td>
<td>0.80</td>
<td>0.50</td>
</tr>
<tr>
<td>Multimodal treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alld 2010</td>
<td>-1.23</td>
<td>0.24</td>
<td>53</td>
<td>54</td>
</tr>
<tr>
<td>Bene 2016</td>
<td>-0.16</td>
<td>0.35</td>
<td>68</td>
<td>64</td>
</tr>
<tr>
<td>Mott 2014</td>
<td>-1.19</td>
<td>0.90</td>
<td>65</td>
<td>66</td>
</tr>
<tr>
<td>Mott 2015a</td>
<td>-1.92</td>
<td>0.90</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Mott 2015b</td>
<td>-1.41</td>
<td>0.90</td>
<td>75</td>
<td>76</td>
</tr>
<tr>
<td>Mott 2015c</td>
<td>-1.95</td>
<td>0.90</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>Nicholas 2013</td>
<td>-0.06</td>
<td>0.93</td>
<td>49</td>
<td>53</td>
</tr>
<tr>
<td>Overman 2015</td>
<td>-0.11</td>
<td>0.53</td>
<td>85</td>
<td>87</td>
</tr>
<tr>
<td>Soners 2012</td>
<td>-0.55</td>
<td>0.00</td>
<td>62</td>
<td>59</td>
</tr>
<tr>
<td>Van der Ve 2015</td>
<td>-0.88</td>
<td>0.00</td>
<td>49</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>-1.50</td>
<td>0.00</td>
<td>634</td>
<td>632</td>
</tr>
<tr>
<td>Pharmacotherapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haug 2016</td>
<td>-0.11</td>
<td>0.67</td>
<td>33</td>
<td>29</td>
</tr>
<tr>
<td>Tse 2015</td>
<td>-0.02</td>
<td>0.80</td>
<td>35</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>-1.00</td>
<td>-0.50</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Favours Experimental Favours Control
but evidence quality was downgraded to low because of the small sample.

**Publication bias.** There were enough studies to test for publication bias in the CBT and multimodal intervention meta-analyses. As shown in Fig 7, the funnel plot for CBT displayed some asymmetry (Egger test = −2.22, df = 10, \( P = .045 \)), but trim and fill did not impute any missing studies that would alter the pooled effect size, suggesting the effect of any possible publication bias is trivial. As shown in Fig 8, the funnel plot for multimodal interventions also showed slight asymmetry (Egger test = −22.13, df = 8, \( P = .026 \)), and trim and fill suggested 1 study in the same direction of the pooled effect was missing, which again suggests trivial effects of any publication bias.

**Intervention Versus Active Control: Follow-Up Outcomes**

Six different interventions were tested against active control groups at follow-up (22 studies, \( n = 2,653 \)). However, as shown in Table 4, only 3 of these were efficacious: ACT, CBT, and multimodal treatment. These 3 interventions were suitable for meta-analysis, with forest

---

**Figure 7.** Funnel plot of observed (white) and imputed (black) studies in comparison: CBT versus active control at post-test (\( n = 12 \) studies).

**Figure 8.** Funnel plot of observed (white) and imputed (black) studies in comparison: multimodal interventions versus active control at post-test (\( n = 10 \) studies).
## Table 4. Effects of Different Interventions on PC Compared With Active Controls at Follow-Up (6–12 Months)

<table>
<thead>
<tr>
<th>Intervention†</th>
<th>Participants (Studies)</th>
<th>SMD Effect (95% CI)</th>
<th>PCS Change (95% CI)‡</th>
<th>Heterogeneity Q, I²</th>
<th>Moderators</th>
<th>Evidence Quality (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All included studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT12,29,56,71</td>
<td>399 (3)</td>
<td>−.35** (−.59 to −.11)</td>
<td>−3.3 (−5.5 to −1.0)</td>
<td>2.86, 30.16</td>
<td>Pain condition; PC primary outcome</td>
<td>High</td>
</tr>
<tr>
<td>CBT3,7,18,21,25,35,55,61</td>
<td>928 (8)</td>
<td>−.30** (−.51 to −.09)</td>
<td>−2.8 (−4.8 to −.8)</td>
<td>16.08*, 56.46</td>
<td></td>
<td>Moderate§</td>
</tr>
<tr>
<td>Education28</td>
<td>105 (1)</td>
<td>.15 (−.23 to .54)</td>
<td>1.4 (−2.1 to 5.0)</td>
<td></td>
<td>Very low§, ¶, ‖</td>
<td></td>
</tr>
<tr>
<td>Graded exposure33</td>
<td>73 (1)</td>
<td>−.07 (−.53 to .39)</td>
<td>−.7 (−5.0 to 3.6)</td>
<td></td>
<td>Very low§, ¶, ‖</td>
<td></td>
</tr>
<tr>
<td>Mindfulness20</td>
<td>124 (1)</td>
<td>.00 (−.35 to .36)</td>
<td>.0 (−3.3 to 3.4)</td>
<td></td>
<td>Low§, ¶, ‖</td>
<td></td>
</tr>
<tr>
<td>Multimodal1,6,34,37,40,42,47</td>
<td>1,024 (8)</td>
<td>−1.39** (−2.27 to −.51)</td>
<td>−13.0 (−21.2 to −4.8)</td>
<td>268.24***, 97.39</td>
<td>Baseline PC; facilitator</td>
<td>Moderate§</td>
</tr>
<tr>
<td>Only studies targeting elevated PC ††</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBT3</td>
<td>95 (1)</td>
<td>−.73** (−1.15 to −.32)</td>
<td>−6.8 (−10.7 to −3.0)</td>
<td></td>
<td>Moderate¶</td>
<td></td>
</tr>
<tr>
<td>Education28</td>
<td>105 (1)</td>
<td>.15 (−.23 to .54)</td>
<td>−1.4 (−2.1 to 5.0)</td>
<td></td>
<td>Low¶, ‖</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: EFT, emotional freedom technique.

*P < .05.
**P < .01.
***P < .001.
†Studies included in each pooled effect.
‡Change in PCS score calculated by multiplying SMD by average standard deviation of included studies that used PCS (SD = 9.34).
§Downgraded due to inconsistency.
¶Downgraded due to risk of bias.
‖Downgraded due to imprecision.
††Targeted interventions are those that treat PC as a primary outcome and have cohorts with clinically significant levels of catastrophizing (>24 equivalent on the PCS).
Like results at post-test, there was high-quality evidence of a medium effect for ACT (SMD = -0.35). Moderate-quality evidence was found for CBT having a medium effect (SMD = -0.30), which was downgraded because of heterogeneity. Multimodal treatment was downgraded for the same reason, yielding moderate-quality evidence for a large effect (SMD = -1.39). Comparison of pooled effect sizes showed a significant difference in favor of multimodal treatment (Q = 7.51, df = 2, P = .023).

### Moderator and subgroup analysis.

There were sufficient CBT and multimodal studies for regression-based moderator analysis. Pain condition moderated the effect of CBT (Q = 14.74, df = 5, P = .01), favoring spinal pain and fibromyalgia over vulvodynia and temporomandibular disorders. Targeting PC as a primary outcome also moderated CBT effects (Q = 14.74, df = 5, P = .01). Limiting this meta-analysis to studies targeting PC increased the CBT effect (SMD = -0.54, 95% CI = -0.82 to -0.26, P < .001) and decreased its heterogeneity (Q = 1.31, df = 1, P = .25, I² = 25.42). For multimodal interventions, baseline PC moderated effects (Q = 19.50, df = 1, P < .001), favoring high PC. Facilitator type also moderated effects (Q = 7.85, df = 1, P = .005), favoring multidisciplinary teams over physiotherapists. Limiting this meta-analysis to studies with high baseline PC and multidisciplinary facilitators increased effects (SMD = -2.95, 95% CI = -4.20 to -1.71, P < .001), but heterogeneity remained high (Q = 39.12, df = 2, P < .001, I² = 94.89). Subgroup analyses of only targeted interventions echoed post-test findings, with only

---

**Figure 9.** Pooled effects on PC of different interventions versus active control at follow-up (6–12 months).
CBT showing efficacy, albeit with greater strength (SMD = −.73).

Publication bias. There were not enough studies to reliably test for publication bias in any of the meta-analyses at follow-up.

Discussion

To our knowledge this is the first systematic review to focus on reductions in PC using any type of intervention. We had 3 related aims: 1) to systematically review and describe RCTs that measure catastrophizing changes in chronic noncancer pain, 2) to document and compare the pooled effects of different interventions, and 3) to identify factors that moderate the efficacy of these interventions. Considering the first of these, there is clearly a large body of literature examining treatment-related changes in PC, providing strong evidence that PC is a modifiable characteristic. Our review, using strict methodological inclusion criteria, yielded a considerable 79 studies representing 9,914 people with mostly musculoskeletal pain, although it is possible that this is an incomplete retrieval of available data. However, only a minority (40.5%) of these studies targeted PC as a primary outcome. This probably reflects the fact that PC is usually seen as a process variable and many of the included studies were secondary mediation analyses of broader trials. Only 8 studies targeted PC and also included cohorts with high PC, suggesting there are relatively few high-quality trials matching treatments to this particular risk profile.

Although the methodological quality of the included research was generally good, with most studies (61%) receiving a low risk of bias rating, there was scope for improving performance bias, attrition bias, and reporting bias. For example, more of the studies using active controls could have made explicit attempts to control for expectancy effects by concealing study hypotheses. There were also still too many studies either not accounting for attrition with ITT analysis or using ITT with unreliable imputation methods such as last observation carried forward. Finally, it was disappointing that so few recent studies had a low risk of reporting bias because publication guidelines have required prospective registration of clinical trials since at least 2005.

The second aim of this review has perhaps the most clinical utility—documenting and comparing effect sizes across interventions. Our general finding when considering all included studies was that several interventions work modestly well in reducing PC and multimodal treatments combining CBT and exercise may work best, although further high-quality research is needed to confirm this. Nine different interventions showed efficacy at post-test when compared with waitlist/usual care, although considering only treatments with at least moderate quality, there was evidence for: ACT, CBT, exercise, mindfulness, and multimodal treatment. When compared with active controls, 3 treatments stood out: CBT, multimodal treatment, and ACT. An encouraging finding was that treatment benefits were largely maintained at follow-up, which may suggest that these interventions involve skill acquisition that translates to at least medium-term behavior change.

When all 79 studies were considered, most of the pooled effect sizes observed were of medium strength (SMD = −.3−.8). Converting this to scores on the most common scale of PC suggests reductions of approximately 3 to 7 points on the 52-point PCS. The obvious question is whether this is clinically meaningful, which is difficult to answer with only group-level data reported in the included studies. Although several methods for assessing clinical significance exist, one common approach looks for reliable change unlikely to be caused by measurement error, along with movement from a clinical range to a nonclinical range. There is evidence that minimum reliable change on the PCS is approximately 20%, whereas a conservative clinical cutoff for the PCS is a score of 20. On this basis, minimum clinically significant change would be a reduction of 5 points on the PCS (baseline 24, post-test 19), which equates to an effect size of SMD = −.54, because the average SD of the PCS in this review was 9.34. Considering only interventions with at least moderate-quality evidence, this condition was only satisfied by multimodal treatment when all interventions were considered, and by CBT as well as multimodal treatment when only targeted studies were considered.

There are several plausible reasons that multimodal treatments might have shown larger effect sizes. First, because multimodal treatment usually combined CBT and exercise, it is possible that these components had additive effects. Considering the fear-avoidance model, if a person’s catastrophizing involved magnified threat cognitions relating to injury or functional limitations, exercising without catastrophic outcomes could function as a form of behavioral experiment aiding cognitive restructuring through the provision of disconfirmatory evidence. Similarly, exercise may help to shift attention away from rumination because of its attentional demands and mood effects, whereas the use of exercise as a self-management tool could increase self-efficacy and thereby reduce helplessness. Finally, because of evidence of a bidirectional relationship between pain intensity and PC, the modulation of descending inhibitory control mechanisms associated with paced exercise may indirectly reduce PC via pain reduction. These effects could occur in addition to the positive effects of traditional CBT components. However, another possibility on the basis of patient-treatment matching models, is that these broader spectrum multimodal treatments have a greater chance of matching at least 1 treatment component to a patient strength or deficit.

Unfortunately, the evidence for multimodal treatments in this review is marred by a high level of heterogeneity in effect sizes. This was largely because of the influence of much larger effect sizes for studies coming from a single research group. It is difficult to account for these differences in terms of treatment content or other factors and, although there is no methodological reason to exclude these studies, this unexplained heterogeneity lowers confidence in the
pooled effect estimate. Indeed, heterogeneity was a significant problem for other interventions and was the most common reason for evidence quality to be downgraded. This likely relates to the lack of consistency among interventions, although the absence of detailed manualization of many treatments included in this review makes it difficult to compare intervention content.

It was possible to reduce heterogeneity by including moderator variables. This highlights our third aim, which was to document moderators of PC treatment effect. Where meta-regression was possible, the most consistent moderators were baseline PC and whether PC was a primary outcome. Indeed, in the subgroup analyses of only studies targeting high PC, effect sizes were significantly higher and heterogeneity lower. For example, CBT versus active control at post-test increased from \( SMD = -0.47 \) to \( SMD = -0.84 \) and CBT was the only effective treatment in the active control group analyses. However, this probably reflects the lack of targeted studies using other interventions, pointing to the need for more research using other interventions with targeted samples.

In general, it is likely that effect estimates from meta-analyses that included all 79 eligible studies were diluted by the lower baseline PC scores in many samples, again suggesting more targeted studies using high-risk cohorts are needed. Further research is also needed to explore whether people with moderate levels of PC still benefit from a reduction in PC, or whether others benefit from resilience-oriented early interventions that prevent future clinical catastrophizing.

Improving the efficacy and efficiency of our PC treatments may require re-examining the construct and, to echo several commentators, clarifying psychological treatment mechanisms as well as developing algorithms for how to match treatment components to patient profiles. The limit, activate, enhance model of psychosocial pain management moderation provides one useful framework for operationalizing this. It suggests treatments need to be tailored to: 1) limit a person’s maladaptive coping responses, 2) activate or increase their healthy behaviors, and 3) enhance outcomes by optimizing existing strengths.

Catastrophizing is usually seen as something to limit within this model; however, it is possible that subgroups of people with elevated PC may also exist, requiring differentially targeted interventions. For example, cognitive restructuring in CBT may act as a “limit-oriented” therapy to help someone high on the magnification subscale of the PCS who also lacks a clear understanding of their pain. In addition, pain neuro-

Conclusions

A large body of evidence shows PC is a modifiable characteristic. Several interventions show efficacy; however, ignoring the poorer-quality evidence, 3 treatments stand out: CBT, multimodal treatment, and ACT. Effect sizes were generally modest and in many cases may not be clinically meaningful. Treatments are most likely to produce clinically significant benefits when they are targeted to people with high levels of catastrophizing and CBT has the best evidence in these cohorts. Future research should focus on testing theory-driven interventions for PC in targeted samples of people with elevated catastrophizing while matching treatment components to specific patient characteristics.

Acknowledgments

The authors are grateful to Diana Blackwood for help with the search strategy and Neil Wurmel for help with the figures.

Supplementary Data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jpain.2017.09.010.

References


4. Amris K, Waehrens EE, Christensen R, Bliddal H, Danneskiold-Samsoe B: Group IMproS: Interdisciplinary rehabilitation of patients with chronic widespread pain: Primary


90. Müller R, Gertz KJ, Molton IR, Terrill AL, Bombardier CH, Ehde DM, Jensen MP: Effects of a tailored positive psychology intervention on well-being and pain in individuals with


