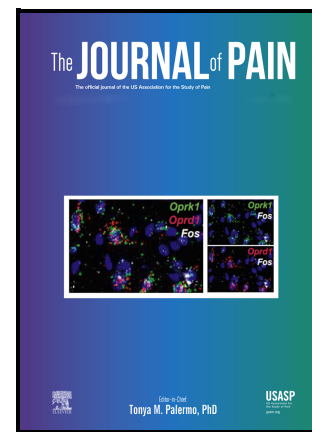


Journal Pre-proof

Trials we cannot trust: investigating their impact on systematic reviews and clinical guidelines in spinal pain
Short title: The impact of trials we cannot trust

N.E. O’Connell, R.A. Moore, G. Stewart, E. Fisher, L. Hearn, C. Eccleston, M. Wewege, A. C de C Williams



PII: S1526-5900(23)00467-4

DOI: <https://doi.org/10.1016/j.jpain.2023.07.003>

Reference: YJPAI4323

To appear in: *The Journal of Pain*

Received date: 30 May 2023

Revised date: 23 June 2023

Accepted date: 1 July 2023

Please cite this article as: N.E. O’Connell, R.A. Moore, G. Stewart, E. Fisher, L. Hearn, C. Eccleston, M. Wewege and A. C de C Williams, Trials we cannot trust: investigating their impact on systematic reviews and clinical guidelines in spinal pain
Short title: The impact of trials we cannot trust, *The Journal of Pain*, (2023) doi:<https://doi.org/10.1016/j.jpain.2023.07.003>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Published by Elsevier.

Type of article: Original Research

Title: Trials we cannot trust: investigating their impact on systematic reviews and clinical guidelines in spinal pain

Authors: N.E. O'Connell¹, R.A. Moore², G. Stewart³, E. Fisher⁴, L. Hearn⁵, C. Eccleston⁶, M. Wewege⁸, A. C de C Williams⁹,

1. Centre for Health and Wellbeing Across the Lifecourse, Department of Health Sciences, Brunel University London, Uxbridge, UK, UB8 3PH
2. Independent Researcher Newton Ferrers, Plymouth, United Kingdom Leslie Hearn
Independent Researcher London, United Kingdom
3. School of Natural and Environmental Sciences, University of Newcastle upon Tyne, Newcastle, UK
4. Centre for Pain Research, University of Bath, Claverton Down, Bath, UK
5. Cochrane Pain, Palliative and Supportive Care Review Group, Oxford University Hospitals, Oxford, UK
6. Centre for Pain Research, University of Bath, Claverton Down, Bath, UK; Department of Clinical and Health Psychology, Ghent University, Belgium; Department of Psychology, University of Helsinki, Finland
7. School of Health Sciences, Faculty of Medicine and Health, University of New South Wales, Sydney, Australia; Centre for Pain IMPACT, Neuroscience Research Australia, Sydney, Australia
8. Research Dept of Clinical, Educational & Health Psychology, University College London, London, UK

Correspondence to: Neil O'Connell

Email: neil.oconnell@brunel.ac.uk

Word Count: 4721

Twitter: @neiloconnell, @AndrewMoorepain, @metanutter, @emmafisher1, @RWISblog,
@Chris_Eccleston, @mawewege, @ACdeCWilliams

Short title: The impact of trials we cannot trust.

Abstract

We previously conducted an exploration of the trustworthiness of a group of clinical trials of cognitive behavioural therapy (CBT) and exercise in spinal pain. We identified multiple concerns in eight trials, judging them untrustworthy. In this study, we systematically explored the impact of these trials (“index trials”) on results, conclusions and recommendations of systematic reviews and clinical practice guidelines (CPGs). We conducted forward citation tracking using Google Scholar and the citationchaser tool, searched the Guidelines International Network (GIN) library and National Institute of Health and Care Excellence (NICE) archive to June 2022 to identify systematic reviews and CPGs. We explored how index trials impacted their findings. Where reviews presented meta-analyses, we extracted or conducted sensitivity analyses for the outcomes pain and disability, to explore how exclusion of index trials affected effect estimates. We developed and applied an ‘Impact Index’ to categorise the extent to which index studies impacted their results. We included 32 unique reviews and 10 CPGs. None directly raised concerns regarding the veracity of the trials. Across meta-analyses (55 comparisons), removal of index trials reduced effect sizes by a median 58% (IQR 40 to 74). 85% of comparisons were classified as highly, 3% as moderately, and 11% as minimally impacted. Nine out of 10 reviews conducting narrative synthesis drew positive conclusions regarding the intervention tested. Nine out of 10 CPGs made positive recommendations for the intervention(s) evaluated. This cohort of trials, with concerns regarding trustworthiness, has substantially impacted the results of systematic reviews and guideline recommendations.

Perspective

We found that a group of trials of CBT for spinal pain with concerns relating to their trustworthiness have had substantial impacts on the analyses and conclusions of systematic reviews and clinical practice guidelines. This highlights the need for a greater focus on the trustworthiness of studies in evidence appraisal.

Keywords

Clinical trials; spinal pain; trustworthiness; systematic reviews; clinical practice guidelines

Pre-registration

Our protocol was pre-registered on the Open Science Framework: <https://osf.io/m92ax/>

Introduction

Randomised Controlled Trials (RCTs) routinely form the basis of evidence included in systematic reviews and meta-analyses of the efficacy or effectiveness of health interventions. They are widely considered the most rigorous form of evidence to guide decisions in Clinical Practice Guidelines (CPGs) and are frequently the only evidence included in systematic reviews of clinical interventions.

RCTs are a human product and so are influenced by biases in human behaviour. Evidence-based medicine (EBM) has numerous tools and methods to assess and manage both quality and bias in research concerned with the conduct and reporting of trials, but there are few methods addressing the important question of trustworthiness of data. Trustworthiness incorporates research integrity and governance, including transparent pre-registration of protocols, appropriate ethical approval and transparent data stewardship, and potential research misconduct.¹ The latter might include fabrication or falsification of research results, or plagiarism.² If untrustworthy trials are not identified and removed during the development process of reviews and CPGs, then the conclusions and

recommendations of those reviews and guidelines are at risk of being incorrect, with potentially major impact on patient care. This issue is compounded by an academic and publishing system that is generally slow, inefficient and inconsistent in dealing with scientific error, issues of misconduct and research integrity,³ and where mistakes are often uncorrected, raising the likelihood of negative impact.⁴

We previously conducted a formal exploration of the trustworthiness of a group of 10 clinical trials of Cognitive Behavioural Therapy (CBT) and/ or physical rehabilitation for persistent spinal pain from a single author team. Three of these trials had been identified as raising concerns during the production of an earlier systematic review⁵ and had therefore been excluded. We used several tools designed to examine the plausibility of baseline characteristics and results, as well as searching for pre-registration, corrections or retractions.¹ While we found the index studies had unremarkable risk of bias profiles, we identified multiple concerns regarding the trustworthiness of eight of the ten identified trials (subsequently referred to as index trials)⁶⁻¹³ (see Supplementary Table 1 for details of those trials, Supplementary Table 2 for a summary of the key domains of trustworthiness that were explored and Supplementary Table 3 for a summary of the concerns raised for each index trial). Key concerns included issues of research governance (lack of study pre-registration, no documentary confirmation of relevant ethical approvals, a lack of sharing of data upon request, distributions of baseline variables that appeared unlikely in the context of random allocation, data anomalies (in particular, duplicate or highly similar results data across unique studies), low to no attrition of participants in some studies and implausible results (extremely large effect sizes diverging from the wider literature).. On that basis, we recommended that they should not be included in evidence syntheses and clinical guidelines in this area.¹ In this study, we explored how these index trials have affected the conclusions and recommendations of published evidence syntheses and CPGs in spinal pain.

Objectives

To explore the impact of eight identified clinical trials (index trials) of uncertain trustworthiness on systematic reviews and clinical practice guidelines.

Patient and Public Involvement

Neither patients nor the public were involved in the conception or conduct of this study.

Methods

Our protocol was pre-registered on the Open Science Framework: <https://osf.io/m92ax/>. As this study only involved evaluation of published information in the public domain, ethical approval was not required.

Inclusion criteria

We included any systematic review or CPG that cited one or more of the index trials. This included both documents that included the trial(s) in the evidence synthesis and those that identified but excluded the trial(s) from the evidence synthesis.

Search Strategy

We conducted our searches on 22nd and 23rd June 2022 without date restrictions. We used forward citation tracking on the eight index trials to find systematic reviews and CPGs that included one or more of the trials. For each trial, we searched Google Scholar and used the citationchaser¹⁴ tool (<https://estech.shinyapps.io/citationchaser>).

We further searched the Guidelines International Network (GIN) library and the National Institute of Health and Care Excellence (NICE) archive using the terms “back pain” OR “neck pain” to identify any CPGs that may have included the index trials. We planned to search the National Health and Medical Research Council's (NHMRC) guidelines portal and the National Guideline Clearinghouse of the Agency for Healthcare Research and Quality (AHRQ) but these resources are no longer active.

Two authors independently screened the titles and abstracts of search results, excluding those that were clearly irrelevant; then they independently screened the full texts of the remaining results, excluding those that were ineligible. Disagreement was resolved by consensus or recourse to a third reviewer.

Citation Summary

From Google Scholar and the citationchaser tool,¹⁴ we explored the number of unique citations received by each trial.

Data extraction

Data were extracted by one reviewer (NOC) using a standardised form. Table 1 outlines the information included in our extraction.

*****Insert Table 1 here*****

Data Synthesis

We narratively summarised the affected reviews and CPGs, with details of the affected comparisons. We synthesised narrative results for systematic reviews and CPGs separately. As only one CPG (JOA) conducted de novo meta-analyses, we included those with our synthesis of meta-analyses from systematic reviews. Where reviews or CPGs made specific practice recommendations on the basis of syntheses that included the index trials, we summarised these.

Where included reviews reported their own sensitivity analyses that excluded the index trials, we extracted those. Where reviews did not report any such analysis and where adequate data were available in reports or from authors on request, we conducted sensitivity meta-analyses for the outcomes of pain and disability at all available follow up timepoints, to explore how the exclusion of

the index trials would affect published effect estimates. We used Revman5¹⁵ to replicate the published analyses with those trials excluded. We replicated analyses using information presented in reviews regarding the meta-analysis model and compared the results of our analyses with the published analyses for any unexpected divergence that might result from different model parameters. We calculated the proportion of participants that index trials contributed to meta-analyses, the absolute and percentage change in effect sizes for each analysis and the change in effect size that resulted from the exclusion of the trial(s). To allow comparison of effect sizes across different scales we converted all to the same direction, as they all represented symptomatic improvement. We explored whether exclusion of the trials changed heterogeneity in the meta-analysis by examining the absolute change in the I^2 statistic (while not a direct measure of heterogeneity, the I^2 statistic measures the percentage of variability in effect estimates due to between-trial heterogeneity),¹⁶ it also has the benefit of being commonly reported in meta-analyses. For all analyses we used the standardised mean difference (Hedge's g) to maximise comparability. It should be noted that we conducted these analyses using the data reported in the included reviews, and did not go back to the original trials. All re-analyses were checked for errors by a second researcher.

While not planned in our original protocol, we developed an index to categorise the extent to which index studies had impacted upon the results of identified meta-analyses. This was developed *de novo* through discussion and consensus among the team and we labelled this the "Impact Index". We first considered and agreed as a team the key domains in which studies may have an impact and then agreed, through discussion and consensus what we considered to be defensible thresholds for judging the impact of index trials in each domain. These were then applied consistently across all included studies. We have not conducted a formal validation of this index. The Impact Index classifies the impact that index studies have had on a meta-analysis across four domains, as follows:

Scale of contribution: What is the proportional volume of data or weight that index trials contribute to the meta-analysis/pooled effect? We quantified this as the proportion of total participants in the meta-analysis that is contributed by index trials, though weight (%) could also be used.

Impact on the pooled effect: What is the impact of index trials on the pooled estimate of treatment effect? This was quantified by calculating the absolute change in the treatment effect (SMD) observed when index trials were excluded from the analysis.

Impact on precision: What is the impact of index trials on the precision of the effect estimate? This was quantified by calculating the % change in the width of the 95% confidence intervals of the pooled effects when the index trials were excluded from the analysis or where exclusion of index trials altered the significance status of the effect at the $p < 0.05$ level. While we recognise that using p values as a marker of meaning is unsatisfactory, it is almost universally used in primary or secondary reporting of meta-analytic outcomes.

Impact on inconsistency: What is the impact of index trials on inconsistency in the meta-analysis? This was quantified by calculating the absolute change in the I^2 statistics value when index trials were excluded from the analysis or where exclusion of index trials changed the direction of effect of the point estimate.

For each domain, judgements were made on whether the impact was substantial, moderate or low. The decision thresholds for each domain are presented in Table 2. The results for each domain were then combined using a decision rule to produce an overall judgement of the level of impact. The decision rules for overall judgments were as follows:

Highly impacted:

- Substantial impact on scale of contribution +/- any effect on magnitude of effect, precision or inconsistency
- Substantial impact on magnitude of effect +/- any effect on scale of contribution, precision or inconsistency
- Moderate impact on scale of contribution +/- moderate or substantial effect on magnitude of effect, precision or inconsistency
- Moderate impact on magnitude of effect +/- moderate or substantial effect on scale of contribution, precision or inconsistency

Moderately impacted:

- Moderate impact on scale of contribution with minimal impact on magnitude of effect, precision and inconsistency
- Moderate impact magnitude of effect with minimal impact on scale of contribution, precision and inconsistency
- Minimal impact on scale of contribution or magnitude of effect with moderate or substantial impact on precision and/ or inconsistency

Minimally impacted:

Minimal effect on scale of contribution, effect magnitude, precision or inconsistency.

*****Insert Table 2 here*****

Results

Figure 1 presents the results of the searches and screening process. In summary, citation tracking identified 729 records and guideline database searches identified 147 records. After removal of

duplicates and screening of titles and abstracts, we reviewed 81 records of which we excluded 39 (for reasons see Figure 1). We finally included 42 records (32 unique systematic reviews^{5,17-47} and 10 CPGs⁴⁸⁻⁵⁷).

*****INSERT FIGURE 1 HERE *****

For the individual index trials, the number of unique citations ranged from 4 to 143. The number of identified systematic reviews that an individual trial was included in ranged from 0 to 16, and the number of CPGs that an individual trial was included in ranged from 0 to 4. The Monticone et al. (2013) study⁷ was the most cited, included in most systematic reviews, and, along with the Monticone et al. (2016) study,⁹ was included in most CPGs. Figure 2 summarises these data.

*****INSERT FIGURE 2 HERE*****

Impact on systematic reviews

Table 3 presents the characteristics of the included systematic reviews. These were published between 2015 and 2022 and were conducted by author teams from a variety of countries. Most focused on the effectiveness of psychological, multidisciplinary or multicomponent interventions for persistent pain^{5 18,- 23,25,27-,32,35,38-41,4306} or post-surgical pain.^{23,33,34,36,37} Two^{17,47} focused on the effectiveness of stabilisation exercises for back pain in people with scoliosis) and one on the effectiveness of exercise interventions for chronic low back pain.²⁶ Ten reviews reported using GRADE to evaluate the certainty of evidence^{19,20,23,24,26,29,31,32,40,42}

*****INSERT TABLE 3 HERE*****

Impact on meta-analyses

Of these reviews and one CPG,⁵³ 21 conducted pairwise meta-analyses and two undertook network meta-analyses. We were able to extract the results of sensitivity analyses where index trials were excluded from four reviews,^{5,19,20,24} of which one²⁰ provided additional data on request. One review with an included NMA²⁷ reported the results of such a sensitivity analysis conducted after a rapid response to their review by authors of this paper, raising concerns about the inclusion of identified/index trials, and we used those reported results. In one included review,³⁶ data in the reported analysis did not match the reported outcome data from the included studies which indicated that it was erroneous. As that review reported the necessary data, we re-conducted the meta-analysis with the relevant trials included and then excluded. For all other analyses we encountered no unexpected divergence between the results of our analyses with the index trials included and those reported in the included reviews.

In total, we included 55 sensitivity analyses of the impact of excluding the index trials from meta-analyses, of which 16 were conducted by the authors of the reviews,^{5,19,20,24} and 39 were conducted by us. Table 4 summarises these results.

*****INSERT TABLE 4 HERE*****

Across all meta-analyses for pain and disability at all timepoints (55 comparisons), removal of the index trials reduced effect sizes by a median 58% (IQR 40 to 74). This reflected a reduction in effect size for all comparisons. The median absolute reduction in effect size (SMD) was -0.35 (IQR -0.51 to -0.21). We saw the same pattern of reduced effect sizes when we grouped analysis by outcome (pain or disability) or clinical population (chronic pain or post-surgical pain). Figure 3 illustrates the absolute reduction in effect size observed when index trials were removed from each analysis. Even

a proportionally small contribution from index trials commonly resulted in moderate to large differences in the pooled effect size in a substantial number of cases.

*****INSERT FIGURE 3 HERE*****

Removal of index trials also led to reduction of heterogeneity (Table 4); the I^2 statistic across all meta-analyses for pain and disability at all time points (44 comparisons) reduced by a median 27% (IQR 10 to 66). A similar pattern was observed whether we grouped analysis by outcome (pain or disability) or by clinical population (chronic pain or post-surgical pain). In terms of impact on precision, removal of index trials led to a reduction in the width of the 95% confidence interval in 48 comparisons, no change in two, and an increase in the width in five. Exclusion of index trials reduced the width of the confidence intervals by 48% (27 to 65). This increase in precision after exclusion reflects the divergent nature of the effect sizes reported in these trials. These data are presented in Table 4. Twelve out of 40 statistically significant effects (at the $p < 0.05$ threshold) became non-significant after the removal of studies of interest.

Application of the Impact Index resulted in 47 of the 55 comparisons (85%) being classified as highly impacted by the inclusion of index trials, two (4%) as moderately impacted and size (11%) as minimally impacted. Sensitivity analyses using study weight in place of the proportion of participants contributing to judge the “scale of contribution” domain did not alter any overall impact judgements. Judgements by domain for each included comparison are presented in Supplementary Table 4. As noted above, four reviews conducted and reported their own analyses examining the impact of excluding the index studies from their analysis^{19,20,24} or by excluding index studies from the primary analysis.⁵ On that basis, we considered these reviews to have clearly addressed the potential impact of the index trials.

In 15 of these reviews, authors commented on the fact that studies of interest were outliers in their sample, had very large effects, and/or introduced heterogeneity to the analyses.^{5,19-}

^{21,23,24,29,33,34,39,40,42,44-46} Of these, the authors of five reviews^{19,23,24,44,45} speculated that the dose, intensity and/or aspects of the content of the intervention in those trials might explain the observed divergence, while the other reviews either did not offer an explanation or stated that the heterogeneity was unexplained.

Impact on narrative syntheses

Ten reviews^{17,18,25,28,30,31,34,41,43,44} did not conduct meta-analyses but synthesised the evidence narratively. Nine of these reported the total number of trials (range 1 to 61), and participants (range 80 to 7201) included in the review. The proportion of participants contributed by index trials ranged from 4.4 to 100% (median 9.3%). Nine of those reviews^{17,18,25,28,30,31,34,43,44} drew broadly positive conclusions regarding the effectiveness of the intervention under scrutiny, underpinned in part by evidence from index trials. Only two specifically commented on the heterogeneity in results of included trials or specifically referred to the divergent results of the index trials. Of these, one review²⁸ referred to the Monticone et al. (2013)⁷ trial as an example of the potential benefits of group-based therapy and another commented that studies with a higher number of treatment sessions (which included Monticone et al. 2013⁷ and 2016a⁹) found larger effects on pain intensity than those with fewer sessions. No review raised specific concerns regarding the veracity of the data from index trials.

Impact on overall review conclusions

Twenty (63%) of the included reviews reported broadly positive conclusions regarding the effectiveness of the intervention of interest, six (19%) reported cautiously positive conclusions which were qualified by issues of the quality of the evidence, or the size of treatments and six (19%) came

to equivocal or negative conclusions. While we could not accurately predict review authors' interpretations to analyses after the removal of the index trials, we might expect these findings to impact on either their overall conclusions and/or the certainty around those conclusions for many of those reviews.

Impact on clinical practice guidelines

We included CPGs from Belgium,⁵⁷ Canada,^{49,51} Finland,⁵⁰ Japan,⁵³ the Netherlands,⁴⁸ Russia,⁵⁶ the UK⁵⁵ and the USA.^{52,54} Seven CPGs^{48,52-57} were focused on the management of low back pain, one on the management of chronic non-malignant pain,⁵¹ one on the management of whiplash-associated disorder (WAD) and neck pain-associated disorders (NAD),⁴⁹ and one on neck pain in adults.⁵⁰ Table 5 presents the characteristics of the included CPGs. 4 CPGs^{48,49,55,57} used GRADE to evaluate the certainty of evidence.

*****INSERT TABLE 5 HERE*****

Nine CPGs presented narrative syntheses for comparisons that included index trials and one⁵³ conducted a de novo meta-analysis. The interventions of interest were described as multimodal, multidisciplinary or biopsychosocial,^{49,51,55-57} CBT combined with exercise,⁵⁴ behavioural treatment,⁴⁸ cognitive therapy⁵⁰ or general exercise.⁵² Table 6 summarises the CPG analyses that included the index trials. All but one made positive recommendations for interventions for which index trials had informed the synthesis. No guideline raised any concerns regarding the veracity of the index trials.

*****INSERT TABLE 6 HERE*****

In the Academy of Orthopaedic Physical Therapy of the American Physical Therapy Association (AAOMT) guideline,⁵² Monticone 2016a was one of four RCTs included in the evaluation of "general exercise" for low back pain. Of the four RCTs in this category, Monticone 2016 was the only study

reported to show benefit (of CBT added to general group-based exercise vs exercise alone).

Recommendations were derived from evidence from a wider range of exercise categories but the guideline did specifically recommend that physical therapists should use “general exercise”.

In the Canadian Agency for Drugs and Technologies in Health (CADTH) guideline,⁵¹ two systematic reviews and two additional RCTs, including Monticone 2017, informed the evaluation of the clinical effectiveness of multidisciplinary treatment programmes for persistent non-malignant pain. Both systematic reviews were reported to demonstrate benefit, one in the short term but not long term, and the other at long-term follow-up. The number of trials and participants in those reviews was not reported. Of the new included trials, only Monticone 2017 was reported to show a benefit of a multidisciplinary programme, including at 12-month follow-up. The guideline concluded that multidisciplinary management of chronic non-malignant pain was associated with significant improvements in pain intensity, and may be associated with significant improvements in quality of life and function.

The Canadian Chiropractic Association (CCA) and the Canadian Federation of Chiropractic (CFC) Regulatory and Educational Accrediting Boards guideline⁴⁹ included 3 RCTs in its evidence evaluation of multimodal care vs continued practitioner care for neck pain-associated disorder, of which Monticone 2012 was one. The guideline reported that Monticone 2012 did not demonstrate clinically significant effects at 1-year follow-up, but made a positive recommendation for multimodal care on the basis of the other two trials, which were reported to show benefit.

The Finnish Medical Association (FMA)⁵⁰ guideline included three RCTs and one non-randomised controlled trial in its evaluation of the effectiveness of cognitive therapies for neck pain (combined n = 813). These included the index trial Monticone 2012 (n = 80). The guideline concluded that while

cognitive therapy may be effective, there was no convincing research evidence to that effect, but reported no further detail.

The Japanese Orthopaedic Association (JOA)⁵³ guideline included Monticone 2014 in its synthesis of the effectiveness of post-operative physiotherapy after surgery for spinal stenosis. It conducted meta-analyses for back pain intensity and leg pain intensity, activities of daily living (ADL), Health-Related Quality of Life (HRQoL) and General Health at 1-year post surgery that all included Monticone 2014. We conducted sensitivity analyses where Monticone 2014 was removed from these meta-analyses. The results are presented in Supplementary Table 5. Our sensitivity analyses resulted in smaller point estimates of effect for all analyses, with greater precision for most, and reduced heterogeneity in three out of four analyses. The guideline concluded that physiotherapy was effective at alleviating pain and improving ADL and QoL, and could therefore be considered useful.

The Nederlands Huisartsen Genootschap NHG⁴⁸ guideline included a systematic review of 30 RCTS (combined n = 3438) and a further five RCTs (combined n = 889) in its evaluation of behavioural therapies for low back pain which included the index trial Monticone 2013 (n = 90). When reporting on the index trial, the guideline stated that *“the presentation of the results made it impossible to assess the clinical relevance”* of the treatment effect but did not elaborate further. Overall, it concluded that it was unclear whether there were clinically relevant benefits of cognitive behavioural treatment over waiting list or standard treatment in patients with chronic non-specific low back pain (low to very low overall quality of evidence).

The UK National Institute of Health and Care Excellence (NICE)⁵⁵ guidance in pain and the Belgian (KCE) guidance will be considered together as the KCE guidance was based on the evidence synthesis conducted by NICE. NICE included two studies of interest (Monticone 2013, 2016a) in the evaluation

of “*Multidisciplinary biopsychosocial rehabilitation (MBR) programmes for back pain*”. The guideline found the evidence for MBR programmes to be mixed, with clinical benefits seen for some comparisons, but also many instances where no benefit was observed and a few where the comparator was favoured over MBR. The guideline reports that the evidence that informed the guidelines for this intervention came “primarily” from three RCTs (combined n = 361), including Monticone 2013 and 2016a (combined n = 240). These trials all reported benefit for MBR programmes and so NICE undertook de novo threshold analyses for the cost-effectiveness of MBR programmes based on the results of these trials and concluded that the interventions in both the Monticone 2013 and 2016a studies were likely to be cost effective. NICE recommended that clinicians should “*Consider a multidisciplinary rehabilitation programme ... for people with persistent low back pain or radicular pain: or when they have psychosocial obstacles to recovery or when previous evidence-based management has not been effective.*”

The North American Spine Society NASS⁵⁴ guideline included Monticone 2013 in its evaluation of the effectiveness of interventions that address fear and avoidance. Overall, they included four RCTs for this comparison (combined n = 287), of which Monticone 2013 contributed 90 participants. Three of four studies, including Monticone 2013, reported benefits on fear and avoidance outcomes and the guideline recommended such interventions. A separate comparison of the effectiveness of adding CBT to an exercise programme included eight RCTs (combined n = 913) of which Monticone 2013 was one (90 participants). The guideline reported conflicting evidence for the addition of CBT and did not make a specific recommendation for its use.

In its evaluation of the effectiveness of multidisciplinary biopsychosocial treatment programmes, the Russian Society for the Study of Pain (RSSP) guideline⁵⁶ included 23 studies in their evidence summary. It was not clear whether these were all RCTs, and the sample size of these studies was not provided. The guideline included Monticone 2013 and 2016A, which together randomised 240

participants. Little detail was provided on how each study informed the guideline recommendation but the guideline recommended multidisciplinary interventions for chronic low back pain.

Discussion:

We previously identified concerns regarding the trustworthiness of this cohort of trials¹ relating to aspects of research governance, data anomalies and implausible results. In our current study, we found that these index trials have been included in 32 systematic reviews and have had important impacts on the results and conclusions of the majority of those. Inclusion of the index trials has exaggerated the size of estimated treatment effects, increased inconsistency in meta-analyses and altered the precision of meta-analyses. In many cases, the exclusion of index trials changed the pooled effects of meta-analyses from moderate-to-large to small or very small effect sizes. These new effect estimates are of questionable clinical significance and, in some cases, excluding index trials shifted effects from statistically significant to non-significant. Index trials have also influenced reviews undertaking narrative syntheses. While the impact on narrative syntheses is harder to quantify, it is reasonable to conclude that index trials weighted the conclusions of many of those reviews in an unduly positive direction.

We identified a number of CPGs from a range of countries and organisations that included at least one of the index trials and used them to formulate their recommendations. All CPGs made positive recommendations for the interventions for which index trials informed the syntheses. Due to the varied approaches to reporting in CPGs and the dominance of narrative approaches to syntheses, it is often not possible to ascertain the specific contribution of index trials to their conclusions and recommendations. In most included CPGs, it is reasonable to infer that the positive reported findings of the index trials contributed to recommendations that favoured psychological or multimodal therapies. In specific examples, it is clear that the index trials were crucial to such clinical

recommendations. The NICE 2016 guideline⁵⁵ clearly shows that two of the index trials^{7,9} were included in the three trials whose evidence was used for a de novo economic analysis that drove a recommendation for multidisciplinary biopsychosocial rehabilitation for low back pain. It is not unreasonable to speculate that without the index trials such a recommendation would not have been considered appropriate. That the evidence in the NICE guideline⁵⁵ was directly used in the formulation of the Belgian (KCE)⁵⁷ guideline further extends that impact.

There is a parallel in the field of hip fracture, an exploration of the impact of a cohort of trials from a different single lead author, that were affected by research misconduct, similarly found that those trials significantly distorted the findings of reviews and clinical practice guidelines⁵⁸.

As we have recently demonstrated,¹ the index trials are extreme outliers in terms of the size of reported effects of psychological therapies, specifically CBT. Just under half of the included reviews specifically referred to the outlier status of index trials and, of those, four conducted and reported sensitivity analyses that excluded the index trials. Only one (Cochrane) review excluded the index trials from the primary analyses⁵ on the basis that the observed heterogeneity was not satisfactorily explained. Other reviews did not comment further on the observed heterogeneity and a number speculated that it might be the result of specific intervention characteristics. Importantly, no review or CPG specifically raised concerns regarding the trustworthiness of the index trials.

This last observation is unsurprising. Screening for aspects of trustworthiness of studies has not yet become routine practice in evidence synthesis and indeed the development and validation of tools for this purpose is in its infancy. While systematic reviews routinely assess risk of bias, trustworthiness screening has a distinct and broader scope. It is noteworthy that risk of bias assessments for the index trials were unremarkable¹ for this field of study. As we have previously argued,¹ systematic reviewers and guideline developers need to attend to the possibility that a

range of other factors, including error, poor research governance, and/or misconduct, may affect identified studies and to develop and adopt approaches to this. Tools are beginning to emerge⁵⁹⁻⁶² that offer some structure to this task for prospective systematic reviewers, though further evaluation of their validity and performance is needed. Our results provide strong support to the argument for such screening. Using these tools, systematic reviewers might move to a process in which trustworthiness is not assumed, and where studies that do not clearly meet a threshold of trustworthiness, through pre-registration, evidence of good research governance and methodological and data transparency, are not included in the synthesis of evidence and cannot influence review conclusions.

Regardless of formal screening for trustworthiness, we propose that reviewers routinely identify and carefully scrutinise studies with divergent results in their evidence syntheses and, where possible, seek explanations from authors of those studies. Reluctance to do this can be driven by a commitment to follow a protocol and the additional resource burden to reviewers, editors and guideline developers. However, to fail to implement any approach presents a risk of the uncritical inclusion of misleading data.

Where an evidence base is dominated by small trials with generally modest effects, as is the case with psychological therapies for chronic pain⁵, the inclusion of untrustworthy data can seriously impact results.⁶³ This places conclusions regarding the effectiveness and the decisions of guideline developers in a marginal space where subtle differences in interpretation can lead to meaningful differences in recommendations.⁶⁴ The introduction of trials with unremarkable risk of bias profiles¹ but very large treatment effects can have a particularly large impact, both on pooled effect sizes and on how that body of evidence is interpreted in research and practice. Our study provides clear evidence of this.

We have used a language of trustworthiness as it accurately reflects the process of assigning a judgement of whether one has trust in the veracity of the findings. The benefits of this approach are that we are clear about where that judgement lies and the basis of that judgement. This is in line with a similar approach in GRADE in moving away from attempting an objective rating of quality to a subjective judgement of certainty of evidence⁶⁵. The risk of this approach is that it introduces terminology that could be misconstrued as a judgement of researcher behaviour or intent. As this field develops, and new tools and methods are introduced, we would also expect the terminology to be also further developed.

Our study has some specific strengths. We followed a publicly available protocol, used tools to identify citations for the index trials, and screened results of those searches with independent reviewers. We have developed a novel multidimensional index to classify the impact that index trials have had on meta-analyses and look forward to other researchers scrutinising and refining it. There are also some limitations. Data extraction was conducted by a single reviewer, though all analyses were checked by a second reviewer. The undertaking of sensitivity analyses based on the data published in the identified reviews, rather than in the original trials, risks replicating errors contained in those reviews, but the principle aim of our re-analyses was to ascertain the impact that the index trials had had on the reported analyses in those reviews, rather than to estimate the effects of interventions. The use of the I^2 statistic when considering consistency and heterogeneity in meta-analyses has been criticised, as it is not a direct measure of heterogeneity but rather of the percentage of variability in effect estimates in a meta-analysis due to between-trial heterogeneity rather than chance. However, for this study, between-trial heterogeneity driven by the inclusion of the index trials was most relevant to our purposes. As several included reviews and CPGs did not implement the GRADE approach to evaluating certainty, and performing de novo GRADE judgements from the available information reported in reviews and CPGs would have not been possible in many instances we did not evaluate the impact on GRADE judgements.

The development and application of the Impact Index was not planned in our original protocol but developed through internal consensus within the team; it has not been formally validated. We propose that it has reasonable face validity. We are aware that including aspects of statistical significance in the application of the index will be controversial. However, in practice, changes in the statistical significance of meta-analyses frequently influence the conclusions of reviews and, indeed, the interpretation of those reviews by their readers. There remains a need to formally evaluate the impact index we developed here and to more rigorously test its assumptions.

Our study represents the impact of the index studies at the point at which the searches were conducted. At the time of writing, we have raised concerns regarding these trials with the editors of their host journals: three of the trials have been formally retracted and some investigations are proceeding. Nevertheless, it is likely that some of these trials, including potentially those that have been retracted, will be included in future systematic reviews and CPGs.

Our findings have important implications for the application of evidence-based healthcare. RCTs and systematic reviews of RCTs are routinely held up as the reference standard of evidence for ascertaining the effectiveness of interventions and for underpinning clinical recommendations. We have shown here how this cohort of studies has led to substantial impacts on both the results of systematic reviews and the recommendations of CPGs, contributing to overly positive conclusions regarding the effectiveness of adding CBT to physical rehabilitation for spinal pain, with subsequent impacts on clinical decisions.

Transparency declaration

The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned (and, if relevant, registered) have been explained.

Disclosures

This study was not supported by any external funding.

Michael Wewege was supported by a Postgraduate Scholarship from the National Health and Medical Research Council of Australia and a PhD Supplementary Scholarship from Neuroscience Research Australia.

Christopher Eccleston undertook consultancy for Orion Pharma and Reckitt on the psychology of pain. He is a member of the UK Pain Parliamentary Campaign Group with a focus on raising awareness in government on the needs of pain patients

Amanda Williams undertook consultancy for Reckitt on the psychology of pain.

NOC was a member of the Guideline Development Group for the 2016 NICE guidance for low back pain and Sciatica (NG59) which is included in this study.

AW, CE, EF and LH were authors on one included review.⁵ They were not involved in any data extraction or analysis related to that review for this project.

Data Sharing Statement

The full data underpinning the analysis of the impact of studies on published meta-analyses is available via Figshare: [10.17633/rd.brunel.21427995](https://doi.org/10.17633/rd.brunel.21427995)

Contributor and guarantor information

NO'C is the guarantor and accepts overall responsibility for the work

NO'C, CE, GS, EF, AW and LH developed the original concept for the study and developed the protocol

NO'C conducted the searches; NOC and LH screened the searches

NO'C and MAW extracted and verified the extracted data

NO'C and MAW conducted the data analyses

All authors contributed to drafting of the manuscript.

The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

References

1. O'Connell NE, Moore RA, Stewart G et al. Investigating the veracity of a sample of divergent published trial data in spinal pain. *Pain* 2022; Online ahead of print. doi: 10.1097/j.pain.0000000000002659.
2. Bordewijk EM, Li W, van Eekelen R et al. Methods to assess research misconduct in health-related research: A scoping review. *J Clin Epidemiol*. 2021;136:189-202. doi: 10.1016/j.jclinepi.2021.05.012.
3. Bolland MJ, Grey A, Avenell A, Klein AA. Correcting the scientific record – A broken system? *Account Res* 2021;28:5,265-79. doi: 10.1080/08989621.2020.1852938.
4. Kataoka Y, Banno M, Tsujimoto Y et al. Retracted randomized controlled trials were cited and not corrected in systematic reviews and clinical practice guidelines. *J Clin Epidemiol* 2022;150:90-7. doi: 10.1016/j.jclinepi.2022.06.015.
5. Williams A C de C, Fisher E, Hearn L, Eccleston C. Psychological therapies for the management of chronic pain (excluding headache) in adults. *Cochrane Database Syst Rev* 2020;8:8:CD007407. doi: 10.1002/14651858.CD007407.pub4.
6. Monticone M, Baiardi P, Vanti C et al. Chronic neck pain and treatment of cognitive and behavioural factors: results of a randomised controlled clinical trial. *Eur Spine J* 2012;21(8):1558-66. doi: 10.1007/s00586-012-2287-y.
7. Monticone M, Ferrante S, Rocca B, Baiardi P, Dal Farra F, Foti C. Effect of a long-lasting multidisciplinary program on disability and fear-avoidance behaviors in patients with chronic

- low back pain: results of a randomized controlled trial. *Clin J Pain* 2013;29(11):929-38. doi: 10.1097/AJP.0b013e31827fef7e.
8. Monticone M, Ferrante S, Teli M et al. Management of catastrophising and kinesiophobia improves rehabilitation after fusion for lumbar spondylolisthesis and stenosis. A randomised controlled trial. *Eur Spine J* 2014b;23(1):87-95. doi: 10.1007/s00586-013-2889-z.
 9. Monticone M, Ambrosini E, Rocca B, Cazzaniga D, Liquori V, Foti C. Group-based task-oriented exercises aimed at managing kinesiophobia improved disability in chronic low back pain. *Eur J Pain* 2016a;20:541-51. doi: 10.1002/ejp.756.
 10. Monticone M, Ambrosini E, Cazzaniga D et al. Adults with idiopathic scoliosis improve disability after motor and cognitive rehabilitation: results of a randomised controlled trial. *Eur Spine J* 2016b;25(10):3120-9. doi: 10.1007/s00586-016-4528-y.
 11. Monticone M, Ambrosini E, Rocca B et al. Group-based multimodal exercises integrated with cognitive-behavioural therapy improve disability, pain and quality of life of subjects with chronic neck pain: a randomized controlled trial with one-year follow-up. *Clin Rehabil* 2017;31(6):742-52. doi: 10.1177/0269215516651979. [retracted]
 12. Monticone M, Ambrosini E, Rocca B et al. Multimodal exercises integrated with cognitive-behavioural therapy improve disability of patients with failed back surgery syndrome: a randomized controlled trial with one-year follow-up. *Disabil Rehabil* 2020;27:1-8. doi: 10.1080/09638288.2020.1863480. [retracted]
 13. Monticone M, Ambrosini E, Portoghese I, Rocca B. Multidisciplinary program based on early management of psychological factors reduces disability of patients with subacute low back pain. Results of a randomised controlled study with one year follow-up. *Eur J Phys Rehabil Med* 2021;57(6):959-67. doi: 10.23736/S1973-9087.21.06696-X. [retracted]
 14. Haddaway NR, Grainger MJ, Gray CT. citationchaser: An R package and Shiny app for forward and backward citations chasing in academic searching; 2021. doi:10.5281/zenodo.4543513.

15. Review Manager (RevMan) . Version 5.4. The Cochrane Collaboration; 2020.
16. Rücker G, Schwarzer G, Carpenter JR. Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Med Res Methodol* 2008; 8:79. doi: 10.1186/1471-2288-8-79.
17. Alanazi MH, Parent EC, Dennett E. Effect of stabilization exercise on back pain, disability and quality of life in adults with scoliosis: a systematic review. *Eur J Phys Rehab Med* 2017;54(5):647-53. doi: 10.23736/S1973-9087.17.05062-6.
18. Barbari V, Storari L, Ciuro A, Testa M. Effectiveness of communicative and educative strategies in chronic low back pain patients: A systematic review. *Patient Educ Couns* 2019;103(5):908-29. doi: 10.1016/j.pec.2019.11.031.
19. Bogaert L, Thys T, Depreitere B et al. Rehabilitation to improve outcomes of lumbar fusion surgery: a systematic review with meta-analysis. *Eur Spine J* 2022;31(6):1525-45. doi: 10.1007/s00586-022-07158-2.
20. Casey M, Smart KM, Segurado R, Doody C. Multidisciplinary-based Rehabilitation (MBR) Compared With Active Physical Interventions for Pain and Disability in Adults With Chronic Pain: A Systematic Review and Meta-analysis. *Clin J Pain* 2020;36(11):874-86. doi: 10.1097/AJP.0000000000000871.
21. Cheng JOS, Cheng S. Effectiveness of physical and cognitive-behavioural intervention programmes for chronic musculoskeletal pain in adults: A systematic review and meta-analysis of randomised controlled trials. *PLoS One* 2019;14(10):e0223367. doi: 10.1371/journal.pone.0223367.
22. Fadli A, Prasetya H, Kristyanto A. Effectivity Biopsychosocial intervention with Cognitive Behavioral Therapy and Exercise Therapy Program in Chronic Low Back Pain: Meta-analysis. *Indonesian J Med* 2021;6(2):177-93. doi: 10.26911/theijmed.2021.06.02.07.

23. Greenwood J, McGregor A, Jones F, Hurley M. Rehabilitation following lumbar fusion surgery: a randomised controlled feasibility study. *Eur Spine J* 2019;28(4):735-744. doi: 10.1007/s00586-019-05913-6.
24. Guerrero AVS, Maujean A, Campbell L, Sterling M. A Systematic Review and Meta-Analysis of the Effectiveness of Psychological Interventions Delivered by Physiotherapists on Pain, Disability and Psychological Outcomes in Musculoskeletal Pain Conditions. *Clin J Pain* 2018;34(9):838-57. doi: 10.1097/AJP.0000000000000601.
25. Hajjhasani A, Rouhani M, Salavati M, Hedayati R, Kahlaee AH. The Influence of Cognitive Behavioral Therapy on Pain, Quality of Life, and Depression in Patients Receiving Physical Therapy for Chronic Low Back Pain: A Systematic Review. *PM & R* 2019;11(2):167-76. doi: 10.1016/j.pmrj.2018.09.029.
26. Hayden JA, Ellis J, Ogilvie R, Malmivaara A, van Tulder M,W. Exercise therapy for chronic low back pain. *Cochrane Database Syst Rev* 2021;9(9):CD009790. doi: 10.1002/14651858.CD009790.pub2.
27. Ho EK, Chen L, Simic M et al. Psychological interventions for chronic, non-specific low back pain: systematic review with network meta-analysis. *BMJ* 2022;376:e067718. doi: 10.1136/bmj-2021-067718.
28. Ince, B. Systematic Review of the Comparative Effectiveness of Cognitive-Behavioural Therapies for Chronic Pain. *J Cogn Behav Psychother Res* 2020;9(3):248-59. doi: 10.5455/JCBPR.64324.
29. Kamper SJ, Apeldoorn AT, Chiarotto A et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. *BMJ* 2015;350:h444. doi: 10.1136/bmj.h444.
30. Knoerl R, Lavoie Smith EM, Weisberg J. Chronic pain and cognitive behavioral therapy: An integrative review. *West J Nurs Res* 2016;38(5):596-628. doi: 10.1177/0193945915615869.

31. Martinez-Calderon J, Flores-Cortes M, Morales-Asencio JM, Luque-Suarez A. Conservative Interventions Reduce Fear in Individuals With Chronic Low Back Pain: A Systematic Review. *Arch Phys Med Rehabil* 2019;101(2):329-58. doi: 10.1016/j.apmr.2019.08.470.
32. Monticone M, Ambrosini E, Cedraschi C et al. Cognitive-behavioral treatment for subacute and chronic neck pain. *Spine* 2015;40(19):1495-504. doi: 10.1097/BRS.0000000000001052.
33. Nadinda PG, van Ryckeghem DML, Peters ML. Can perioperative psychological interventions decrease the risk of postsurgical pain and disability? A systematic review and meta-analysis of randomized controlled trials. *Pain* 2022;163(7):1254-73. doi: 10.1097/j.pain.0000000000002521.
34. Nicholls JL, Azam MA, Burns LC et al. Psychological treatments for the management of postsurgical pain: a systematic review of randomized controlled trials. *Patient Relat Outcome Meas* 2018;9:49-64. doi: 10.2147/PROM.S121251.
35. O'Keeffe M, Purtill H, Kennedy N et al. Comparative effectiveness of conservative interventions for nonspecific chronic spinal pain: physical, behavioral/psychologically informed, or combined? A systematic review and meta-analysis. *J Pain* 2016;17(7):755-74. doi: 10.1016/j.jpain.2016.01.473.
36. Özden F. The Effectiveness of Physical Exercise After Lumbar Fusion Surgery: A Systematic Review and Meta-Analysis. *World Neurosurg* 2022;163:e396-e412. doi: 10.1016/j.wneu.2022.03.143.
37. Parrish JM, Jenkins NW, Parrish MS et al. The influence of cognitive behavioral therapy on lumbar spine surgery outcomes: a systematic review and meta-analysis. *Eur Spine J* 2021;30(5):1365-79. doi: 10.1007/s00586-021-06747-x.
38. Petrucci G, Papalia GF, Russo F et al. Psychological Approaches for the Integrative Care of Chronic Low Back Pain: A Systematic Review and Metanalysis. *Int J Environ Res and Pub Health* 2021;19(1):60. doi: 10.3390/ijerph19010060.

39. Richmond H, Hall AM, Copsey B et al. The effectiveness of cognitive behavioural treatment for non-specific low back pain: a systematic review and meta-analysis. *PLoS One* 2015;10(8):e0134192. doi: 10.1371/journal.pone.0134192.
40. Schütze R, Rees CS, Smith A, Slater H, Campbell JM, O'Sullivan P. How Can We Best Reduce Pain Catastrophizing in Adults With Chronic Noncancer Pain? A Systematic Review and Meta-Analysis. *J Pain* 2017;19(3):233-56. doi: 10.1016/j.jpain.2017.09.010.
41. Shearer HM, Carroll LJ, Wong JJ et al. Are psychological interventions effective for the management of neck pain and whiplash-associated disorders? A systematic review by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. *Spine J* 2015;16(12):1566-81. doi: 10.1016/j.spinee.2015.08.011.
42. Szeverenyi C, Kekecs Z, Johnson A, Elkins G, Csernatony Z, Varga K. The use of adjunct psychosocial interventions can decrease postoperative pain and improve the quality of clinical care in orthopedic surgery: a systematic review and meta-analysis of randomized controlled trials. *J Pain* 2018;19(11):1231-52. doi: 10.1016/j.jpain.2018.05.006.
43. Vergeld V, Ginis KAM, Jenks AD. Psychological interventions for reducing fear avoidance beliefs among people with chronic back pain. *Rehabil Psychol* 2021;66(4):386-403. doi: 10.1037/rep0000394.
44. Vitoula K, Venneri A, Varrassi G et al. Behavioral therapy approaches for the management of low back pain: An up-to-date systematic review. *Pain Ther* 2018;7(1):1-12. doi: 10.1007/s40122-018-0099-4
45. Wilson S, Cramp F. Combining a psychological intervention with physiotherapy: A systematic review to determine the effect on physical function and quality of life for adults with chronic pain. *Phys Ther Rev* 2018;23(3):214-26. doi: 10.1080/10833196.2018.1483550.
46. Yang J, Lo WLA, Zheng F, Cheng X, Yu Q, Wang C. Evaluation of Cognitive Behavioral Therapy on Improving Pain, Fear Avoidance, and Self-Efficacy in Patients with Chronic Low Back Pain:

- A Systematic Review and Meta-Analysis. *Pain Res Man* 2022;4276175. doi: 10.1155/2022/4276175.
47. Yanyun G, Lei H, Yi Z, Jing T, Kong W, Jingsong W. The effect of Pilates exercise training for scoliosis on improving spinal deformity and quality of life: Meta-analysis of randomized controlled trials. *Medicine* 2021;100(39):e27254. doi: 10.1097/MD.00000000000027254.
48. Bons SCS, Borg MAJP, Van den Donk M et al. NHG standard M54 Non-specific low back pain. Netherlands Huisartsen Genootschap. <https://richtlijnen.nhg.org/standaarden/aspecifieke-lagerugpijn>. Accessed July 2022
49. Bussi eres AE, Stewart G, Al-Zoubi F et al. The treatment of neck pain–associated disorders and whiplash-associated disorders: a clinical practice guideline. *J Manipulative Physiol Ther* 2016;39(8):523-64. doi: 10.1016/j.jmpt.2016.08.007.
50. Finnish Medical Association Duodecim. Neck pain (adults). A working group set up by Duodecim of the Finnish Medical Society, Societas Medicinae Physicalis et Rehabilitationis Fenniae ry and the Finnish Society of General Medicine. Helsinki: Suomalainen L  k  riseura Duodecim; 2017
51. Gauthier K, Dulong C, Arg  ez C. Multidisciplinary treatment programs for patients with chronic non-malignant pain: a review of clinical effectiveness, cost-effectiveness, and guidelines – an update. Ottawa: CADTH; 2019.
52. George SZ, Fritz JM, Silfies S et al. Interventions for the management of acute and chronic low back pain: revision 2021: clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of orthopaedic physical therapy of the American Physical Therapy Association. *J Orthop Sports Phys Ther* 2021;51(11):CPG1-CPG60. doi: 10.2519/jospt.2021.0304.

53. Kawakami M, Takeshita K, Inoue G et al. Japanese Orthopaedic Association (JOA) clinical practice guidelines on the management of lumbar spinal stenosis, 2021 - Secondary publication. *J Orthop Sci* 2022; S0949-2658(22):00116-6. doi: 10.1016/j.jos.2022.03.013.
54. Kreiner DS, Matz P, Bono CM et al. Guideline summary review: an evidence-based clinical guideline for the diagnosis and treatment of low back pain. *Spine J* 2020;20(7):998-1024. doi: 10.1016/j.spinee.2020.04.006.
55. National Guideline Centre (UK) 'Low back pain and sciatica in over 16s: assessment and management'. Manchester: National Institute for Health and Care Excellence; 2016 <https://www.nice.org.uk/guidance/ng59> . Date accessed: 28/6/23
56. Parfenov VA, Y. Chronic nonspecific (musculoskeletal) low back pain. Guidelines of the Russian Society for the Study of Pain (RSSP). *Neurol Neuropsychiatr Psychosom* 2019;11(2S):7-16. doi: 10.14412/2074-2711-2019-2S-7-16.
57. Van Wambeke P, Desomer A, Ailliet L et al. Low back pain and radicular pain: assessment and management. Good Clinical Practice (GCP) Brussels: Belgian Health Care Knowledge Centre (KCE). 2017. KCE Reports 287.
58. Avenell A, Stewart F, Grey A, et al. An investigation into the impact and implications of published papers from retracted research: systematic search of affected literature. *BMJ Open* 2019;9:e031909.
59. Cochrane Pregnancy and Childbirth Review Group. Identifying and handling potentially untrustworthy trials in Pregnancy and Childbirth Cochrane Reviews. <https://pregnancy.cochrane.org/news/identifying-and-handling-potentially-untrustworthy-trials-pregnancy-and-childbirth-cochrane> . Date accessed 16/9/21
60. Grey A, Bolland MJ, Avenell A, Klein AA, Gunsalus CK. Check for publication integrity before misconduct. *Nature* 2020;577(7789):167-69. doi: 10.1038/d41586-019-03959-6.

61. Parker L, Boughton S, Lawrence R, Bero L. Experts identified warning signs of fraudulent research: a qualitative study to inform a screening tool. *J Clin Epidemiol.* 2022;151:1-17. doi: 10.1016/j.jclinepi.2022.07.006.
62. Weibel S, Popp M, Reis S, Skoetz N, Garner P, Sydenham E. Identifying and managing problematic trials: a Research Integrity Assessment (RIA) tool for randomized controlled trials in evidence synthesis. *Res Synth Methods* 2022; ahead of print. doi: 10.1002/jrsm.1599.
63. Williams ACdeC, Fisher E, Hearn L, Eccleston C. Evidence-based psychological interventions for adults with chronic pain: precision, control, quality, and equipoise. *Pain* 2021;160(8):2149-53. doi: 10.1097/j.pain.0000000000002273.
64. O'Connell NE, Cook C, Wand BM, Ward SP. Clinical guidelines for low back pain. A critical review of consensus and inconsistencies across three major guidelines. *Best Pract Res Clin Rheum* 2016; 30(6):968-80. doi: 10.1016/j.berh.2017.05.001.
65. Hultcrantz M, Rind D, Akl EA et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol.* 2017;87:4-13.
66. Faraone SV. Interpreting estimates of treatment effects: implications for managed care. *P T* 2008;33(12):700-11.

Figure Legends:

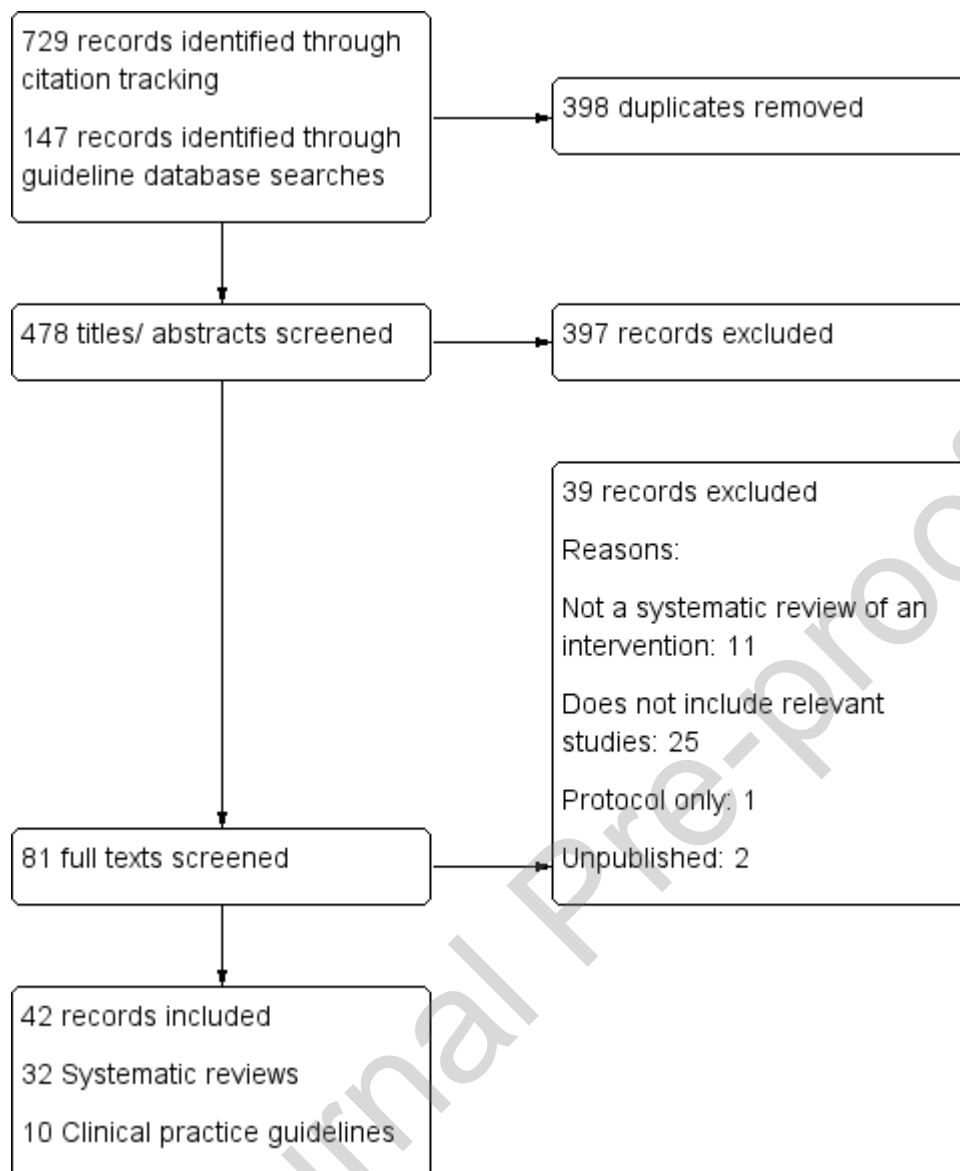


Figure 1. Flow diagram of the search screening process

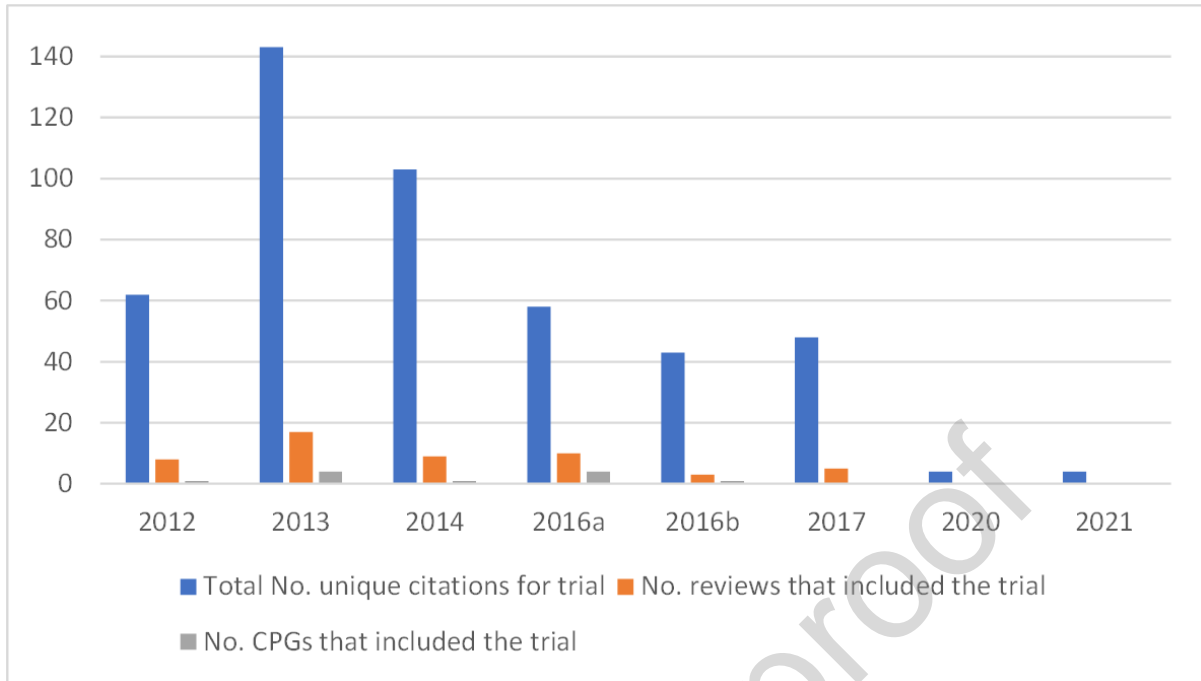


Figure 2. The number of unique citations, and appearances in systematic reviews and CPGs for each index trial.

Footnotes: Each date represents a unique index trial, identified by its year of publication.

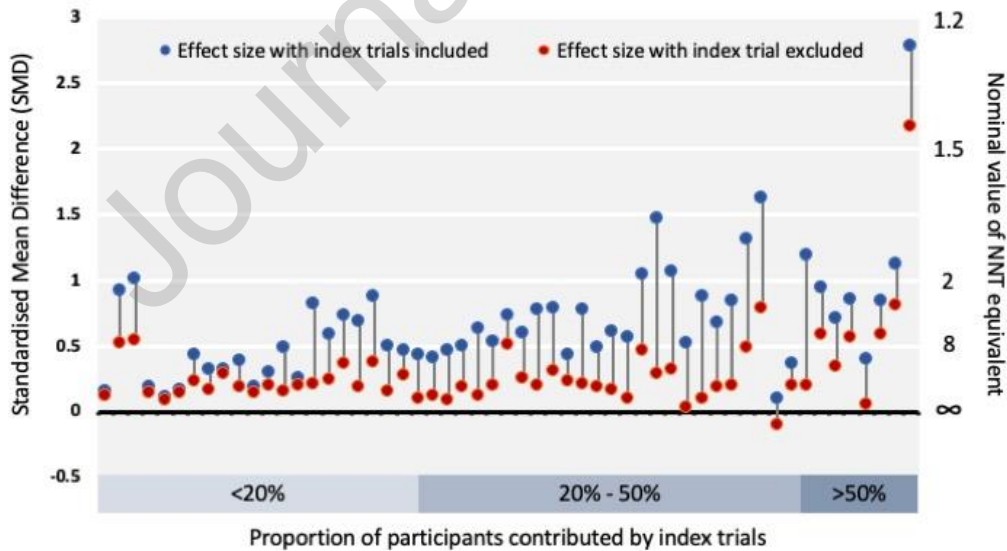


Figure 3: The effect of excluding index trials from the included meta-analyses

Footnotes: NNT estimated from the SMD using the approach described by Faraone 2008⁶⁶

Table 1. Details of data extracted from reviews and CPGs

Characteristics of the systematic review/CPG	The overall aims/ objectives of the review or guideline The country of origin of the author team/CPG The organisation developing the CPG Any declared conflicts of interest of review authors Any declared external funding. Date of publication and journal/policy body
Details of Index trial(s) inclusion in review CPG	The index trials cited in the review/CPG and how they were cited (e.g., in background, discussion, or as part of the methods and results of the review/CPG) The index trials excluded from the evidence synthesis in the review/CPG and any reported reasons for exclusion. The index trials included in the evidence synthesis in the review/CPG
Details of Index trials inclusion in analysis/ synthesis.	<p>From the index trials included in the evidence synthesis, we extracted: In which comparison(s) For which outcome(s) Details of affected comparison(s) Whether identified trial(s) were identified / recognised as divergent/ outliers Details of any reported narrative addressing that issue</p> <p>Meta-analysis (pairwise or network) Model details, number of trials, participants, weight given to trials, heterogeneity estimates, pooled effect size/precision estimates. Any exploration (details and results) of heterogeneity (subgroup analysis, sensitivity analysis, meta-regression Whether identified trial(s) were identified / recognised as divergent/ outliers. Details of any evaluation of the certainty of results in affected comparisons (e.g. using GRADE)</p> <p>Narrative synthesis. Number of trials, participants. Overall results. Details of any evaluation of the certainty of results in affected comparisons (e.g. using GRADE)</p> <p>Conclusions/ recommendations Overall conclusions of the evidence synthesis relating to affected comparisons (i.e. that include identified trial(s)) Any specific clinical recommendations made in reviews or CPGs that include the index trials</p>

Table 2: Decision thresholds for each domain of the Impact Index

	Scale of contribution	Effect magnitude SMD	Precision		Inconsistency
	Overall proportion of participants from index trials in meta-analysis	Absolute change in ES	Relationship with the null	% change in CI margin	Absolute change in I ² statistic
Low impact	<15%	<0.2		<20%	<20%
Moderate Impact	15-39%	0.2-0.49		20% to 49%	20% to 49%
High Impact	>40%	≥0.5	Changes "significance status": Changes CI to include no effect or to exclude no effect at the P<0.05 level	≥50%	≥50% or changes direction of point estimate

Table 3: Characteristics of reviews that included the index trials.

Authors, Primary	Country of origin	Review Primary Objectives	Search date	Name of trials included	Comparisons affected	meta-analysis Y/N	narrative synthesis Y/N	GRADE Y/N	Overall Review conclusions
Alanazi et al. 2018 ¹⁷	Canada, Saudi Arabia	"...to evaluate the effect of the stabilization exercise on back pain in adults with scoliosis"	To March 9 2017	2016b	active self-correction, task-oriented exercises and cognitive-behavioral therapy vs "general physiotherapy"...	N	Y	N	"Stabilization exercise, as reported in the included study, is shown to be effective in reducing back pain, disability and improving quality of life in adults with idiopathic scoliosis. However, this review highlights the paucity of literature examining the effect of exercise on back pain in adults with scoliosis and strongly suggests that further experimental research is needed aiming to ensure proper blinding as this was a common weakness."
Barbari et al. 2020 ¹⁸	Italy	"to elaborate the state of the art of scientific literature on the effectiveness of interventions that included communicative and educative strategies on three main outcomes: 1) patient's LBP awareness/knowledge, 2) maladaptive behavior modification and 3) compliance with exercise"	To between September and February 2018	2013, 2016a	CBT + physiotherapy vs physiotherapy alone	N	Y	N	"Evidence coming from high-quality RCTs sustains multimodal interventions, pain science education and graded exposure – as well as combinations of CBT, pain science education and graded exposure – on behavior modification in the mid- and long-term."
Bogaert et al. 2022 ¹⁹	Belgium	"to assess and compare the effectiveness of unimodal and multimodal rehabilitation strategies on disability, pain, and pain-related fear in patients undergoing lumbar fusion surgery for degenerative conditions and (adult) isthmic spondylolisthesis.	To April 2021	2014b	Multimodal rehabilitation intervention vs exercise therapy	Y	N	Y	"The results of this systematic review with meta-analysis encourage exercise for all patients undergoing lumbar fusion given the positive impact on disability and pain up to six months postoperative. Embedding exercise in a multimodal rehabilitation context is suggested given the additional positive effect on

		The secondary aim was to assess the effectiveness on return-to-work."							disability and pain-related fear, compared to exercise alone. It remains uncertain if these beneficial effects of exercise and multimodal rehabilitation persist in the long term."
Casey et al. 2020 ²⁰	Ireland	Systematic review examining the evidence for MBR compared with active physical interventions in chronic noncancer pain conditions	inception to November 2018	2013, 2016a, 2017	MBR vs active physical treatment	Y	N	N	"Low-quality evidence to suggest that MBR is more effective than active physical interventions for reducing pain intensity and disability in people with chronic pain, in both the short-term and long-term. As the benefits appear to be modest, other factors such as resource allocation should be considered before offering MBR programmes to people with chronic pain."
Cheng et al. 2019 ²¹	UK, Hong Kong	"to examine the effects of physical exercise cum cognitive-behavioural therapy (CBT) on alleviating pain intensity, functional disabilities, and mood/mental symptoms in those suffering with chronic musculoskeletal pain."	To December 31 2018	2012, 2013, 2016a, 2017	CBT +exercise vs exercise alone	Y	N	N	"The value of adding CBT to exercise interventions is questionable, as consistent benefits were not seen. The clinical implications and directions for future research are discussed."
Fadli et al. 2021 ²²	Indonesia	"to summarize the existing literature data regarding the effectiveness of biopsychosocial interventions with Cognitive Behavioral Therapy (CBT) and Exercise Therapy Program (ETP) in CLBP."	To 2020	2013, 2016A	CBT + exercise vs exercise alone	Y	N	N	"Biopsychosocial interventions for CBT and ETP are effective in reducing the pain and disability index of CLBP. The effectiveness of biopsychosocial interventions with CBT and ETP methods is due to the patient's efficacy in completing the rehabilitation program."
Greenwood, et al. 2016 ²³	UK	"to conduct a systematic review and meta-analysis of	To October 13, 2014	2014B	Exercise +CBT vs exercise alone	Y	N	N	"A small number of low-quality studies suggest that "complex

		current evidence evaluating the effectiveness of rehabilitation following lumbar fusion surgery (LFS)."							rehabilitation" reduces short and long-term disability and fear avoidance behavior following LFS. More, high-quality research is required to confirm the effectiveness of "complex rehabilitation" programs."
Guerrero et al. 2018 ²⁴	Australia	"we examined the effectiveness of physiotherapist delivered psychological interventions combined with physiotherapy on pain, disability and psychological outcomes for patients with musculoskeletal pain conditions"	To May 2016	2012	Combined psychological and physiotherapy vs physio/usual care alone	Y	N	Y	"The results indicate that psychological interventions delivered by physiotherapist show promise to improve health outcomes, particularly psychological outcomes, in musculoskeletal pain conditions."
Hajihassani et al. 2019 ²⁵	Iran	"To investigate the effect of adding the cognitive behavioural treatment (CBT) component to routine physical therapy (PT) on pain and depression reduction, improvement in quality of life, and enhanced function in patients with chronic low back pain (CLBP)."	To January 2018	2013	CBT + physio vs physio alone	N	Y	N	"although some patient populations benefited from receiving CBT in addition to routine PT in their CLBP rehabilitation process in terms of reducing pain and disability and enhancing functional capacity and quality of life, there were others for whom CBT did not seem advantageous. None of the investigations found that the addition of CBT assisted in reducing depression symptoms."
Hayden et al. 2021 ²⁶	Canada	"The primary objective of this systematic review is to assess the impact of exercise treatment on pain and functional limitations in adults with chronic non-specific low back pain compared to no treatment, usual care, placebo and other conservative treatments."	To April 27 2018	2016a	?	Y	N	Y	"We found moderate-certainty evidence that exercise is probably effective for treatment of chronic low back pain compared to no treatment, usual care or placebo for pain. The observed treatment effect for the exercise compared to no treatment, usual care or placebo comparisons is small for functional limitations, not meeting our threshold for

									minimal clinically important difference. We also found exercise to have improved pain (low-certainty evidence) and functional limitations outcomes (moderate-certainty evidence) compared to other conservative treatments; however, these effects were small and not clinically important when considering all comparisons together."
Ho et al. 2022 ²⁷	Australia	"To determine the comparative effectiveness and safety of psychological interventions for chronic low back pain."	To January 2021	2013, 2016a,	NMA in node "cognitive behavioural therapy delivered with physiotherapy care"	Y NMA	N	N	"For people with chronic, non-specific low back pain, psychological interventions are most effective when delivered in conjunction with physiotherapy care (mainly structured exercise). Pain education programmes (low to moderate quality evidence) and behavioural therapy (low to high quality evidence) result in the most sustainable effects of treatment; however, uncertainty remains as to their long term effectiveness."
Ince et al. 2020 ²⁸	Turkey	"to systematically review the studies investigating the effectiveness of all forms of CBT for the treatment of individuals with chronic pain."	To February 2014	2012 2013	not clearly reported	N	Y	N	"In conclusion, besides demonstrated effectiveness of cognitive behavioural therapies, there is no clear explanation for which specific treatment components are responsible for the improvement in which specific complaints."
Kamper et al. 2015 ²⁹	Australia	"To assess the long-term effects of multidisciplinary biopsychosocial rehabilitation for patients with chronic low back pain."	To February 2014	2013	Multidisciplinary biopsychosocial rehab vs physical treatment	Y	N	Y	"Patients with chronic LBP receiving MBR are likely to experience less pain and disability than those receiving usual care or a physical treatment. MBR

Knoerl et al. 2015 ³⁰	USA	"to determine (a) which CBT doses, delivery methods, strategies, and follow-up periods have been explored in recent intervention studies of individuals with chronic pain"	Between 2005 and 2015	2012, 2013	CBT	N	Y	N	also has a positive influence on work status compared to physical treatment. Effects are of a modest magnitude and should be balanced against the time and resource requirements of MBR programs. people with indicators of significant psychosocial impact are referred to MBR."
Martinez-Calderon et al. 2020 ³¹	Spain	"To systematically review and critically appraise the effectiveness of conservative and surgical interventions to reduce fear in studies of people with chronic low back pain, based on the analysis of randomized controlled trials for which fear was a primary or secondary outcome."	To May 2019	2013, 2014B, 2016A	Multidisciplinary interventions	N	Y	Y	"The results of this review demonstrated that CBT was effective for pain intensity in 43% of the trials and was an effective treatment for many pain-related variables recommended by IMMPACT such as physical functioning, anxiety, depression, and quality of life."
Monticone et al. 2015 ³²	Italy	"To assess the effects of CBT among individuals with subacute and chronic NP. Specifically, the following comparisons were investigated: (1) cognitive-behavioural therapy versus placebo, no treatment, or waiting list controls; (2)	To November 2014	2012	CBT in addition to another intervention versus the other intervention alone. CBT in addition to another intervention versus the other intervention alone	Y	N	Y	"Multidisciplinary and psychological interventions as well as exercise reduced kinesiphobia. Fear-avoidance beliefs were reduced by the aforementioned interventions, manual therapy, and electrotherapy. A multidisciplinary intervention reduced the fear of falling. There was moderate evidence of multidisciplinary interventions and exercise to reduce kinesiphobia."
									"When comparing both CBT to other types of interventions and CBT in addition to another intervention to the other intervention alone, no differences were found."

		cognitive-behavioural therapy versus other types of interventions; (3) cognitive-behavioural therapy in addition to another intervention (e.g. physiotherapy) versus the other intervention alone."								
Nadinda et al. 2022 ³³	Belgium	"to investigate the efficacy of perioperative psychological interventions in reducing (sub)acute postsurgical pain and CPSP and disability in adults."	To March 2020	2014B	Psychological interventions vs control (+moderator analyses)	Y	N	N	"Psychological interventions can be effective in reducing (sub) acute postsurgical pain and CPSP and disability. These results underscore the possible benefits of integrating psychological services into multidisciplinary acute and transitional pain teams...implementing psychological interventions in the perioperative period may have the potential to reduce the humanitarian and economic burden of CPSP."	
Nicholls et al. 2018 ³⁴	Canada	"To identify randomized controlled trials evaluating the efficacy of these psychotherapy approaches on pain-related surgical outcomes."	To 2017	2014B	Effects of CBT on pain-related outcomes	N	Y	N	"This systematic review provides preliminary evidence that CBT-based psychological interventions reduce PSP intensity and disability. Future research should further clarify the efficacy and optimal delivery of CBT and newer psychological approaches to PSP."	
O'Keeffe et al. 2016 ³⁵	Ireland	"to assess the comparative effectiveness of physical, behavioural and/or psychologically informed, and combined interventions on pain and disability in patients with NSCSP."	To January 2013	2012	Effect of physical versus combined interventions on pain and disability	Y	N	N	"No clinically significant differences were found for pain and disability between physical, behavioral/psychologically informed, and combined interventions for NSCSP."	
Özden et al. 2022 ³⁶	Turkey	"to systematically review the effect	To October	2014b	cognitive behavioral	Y	N	N	"The results of the present systematic	

		of exercise interventions and conduct its meta-analysis in patients after LFS."	- December 2021		training in addition to exercise vs exercise				review and meta-analysis reported that cognitive therapy applied in addition to exercise could provide more effective results in physical and psychological parameters."
Parrish et al. 2021 ³⁷	USA	"to investigate the influence of CBT on patient reported outcomes among lumbar spine surgery patients."	To December 2019	2014B	The effect of CBT on back pain, leg pain, disability vs usual care	Y	N	N	"Compared to usual care or alternative therapy control arms, CBT delivered the most improvement with overall quality of life and psychological outcomes. Among appropriately selected patients, CBT could improve perioperative disability, pain, quality of life, and psychological health following lumbar spine surgery."
Petrucci et al. 2022 ³⁸	Italy	"to identify and to describe the most common psychological approaches used to treat patients who suffer from CLBP"	Not reported	2013	CBT vs control	Y	N	N	"CBT and MBSR have proven their significant effectiveness to improve pain intensity and quality of life compared to controls. These approaches also demonstrated their efficacy in reducing disability and fear-avoidance, but without significant results. Our findings suggest that CBT and MBSR modify pain-related outcomes and that they could be implemented in clinical practice."
Richmond et al. 2015 ³⁹	UK	"To assess whether cognitive behavioural (CB) approaches improve disability, pain, quality of life and/or work disability for patients with low back pain (LBP) of any duration and of any age."	To November 2014	2013	CBT vs guideline based active treatment	Y	Y	N	"CB interventions yield long-term improvements in pain, disability and quality of life in comparison to no treatment and other guideline-based active treatments for patients with LBP of any duration and of any age."
Schütze et al. 2018 ⁴⁰	Australia	"to examine all treatment-related changes in PC regardless of whether catastrophizing was specifically	To November 2016	2914b, 2016A, 2016B, 2017	Multimodal treatment vs active control	Y	N	Y	"The best evidence (moderate-high quality) was found for Cognitive Behavior Therapy (CBT), multimodal treatment, and

		targeted as a primary outcome.”							Acceptance and Commitment Therapy (ACT). Effects were generally of medium strength and had questionable clinical significance.”
Shearer et al. 2016 ⁴¹	Canada	“to update findings of the NPTF and evaluate the effectiveness of psychological interventions for the management of neck pain and associated disorders (NAD) or whiplash associated disorders (WAD).”	To January 2015	2012	Physical therapist-provided CBT for Persistent grades I–II NAD, 12 month follow up	N	Y	N	“We found no clear evidence to support the use of relaxation training, biofeedback, or CBT for reducing pain and improving disability scores in patients with persistent NAD or WAD.”
Szeverenyi et al. 2018 ⁴²	Hungary, Sweden, USA	“to assess the effectiveness of psychosocial techniques to decrease postoperative pain and improve perioperative clinical care in orthopedic surgery. A systematic review and meta-analysis was performed to evaluate the effects of psychosocial methods among adults undergoing orthopedic surgeries.”	To September 2016	2014B	psychosocial interventions for post operative pain	Y	N	Y	“The results indicate that psychosocial interventions, especially patient education and relaxation training, may reduce perioperative side effects and improve recovery in patients undergoing orthopedic procedures, but the quality of evidence is generally low.”
Vergeld et al. 2021 ⁴³	Germany	“to systematically locate and synthesize the current evidence regarding the effectiveness of psychological interventions on fear avoidance beliefs and fear avoidance behavior in patients with CBP.”	To October 2019	2013	CBT interventions vs active control group or other control group	N	Y	N	“At this point, the evidence is inconclusive as to which psychological interventions are most effective to treat FAB among people with CBP. Although there is some promising evidence to support CBT, additional research is needed to determine which components of CBT are most effective.”
Vitoula et al. 2018 ⁴⁴	Greece	“to overview the behavioral approaches that can be used in the management of patients with low back pain. Approaches such as	To February 2018	2013, 2016A	CBT vs ?	N	Y	N	“This systematic review indicates the following key points: 1. Behavioral therapy approaches are effective in patients with LBP particularly in altering pain

									<p>electromyography (EMG) biofeedback, cognitive behavioral therapy, and mindfulness-based stress reduction are discussed as non-pharmacological options in the management of low back pain.”</p> <p>perception and helping patients to regain their functionality. 2. Treatment outcomes can be improved if the treatments are personalized to individual patients’ needs [73, 74]. 3. A multidisciplinary approach is the future. Multidisciplinary rehabilitation includes more than just physical treatment. A team approach accounting for several aspects within the biopsychosocial model is more likely to help individuals with chronic LBP compared to standard care alone. 4. CBT is the type of psychotherapy that has been most studied in patients with LBP. Although most of the other behavioral therapy interventions have been tried in randomized trials in other conditions, more trials of such approaches are needed in patients with LBP. 5. Future research, however, must focus on the improvement of specific outcomes, using not only measures of pain intensity but also using measures of pain acceptance, reduction of medication used, disability, and quality of life to assess efficacy.”</p>
Williams et al. 2020 ⁵	UK	“To determine the clinical efficacy and safety of psychological interventions for chronic pain in adults (age > 18 years) compared with active controls, or waiting	To April 16 2020	2013, 2016A, 2017	CBT vs active care	Y	N	Y	<p>“We found sufficient evidence across a large evidence base (59 studies, over 5000 participants) that CBT has small or very small beneficial effects for reducing pain, disability, and distress in chronic pain, but we found</p>

Wilson et al. 2018 ⁴⁵	UK	list/treatment as usual (TAU). "to determine whether the addition of psychological approaches to physiotherapy is more effective in improving physical functioning and quality of life than physiotherapy alone."	To Septem ber 2018	2012, 2013, 2017	Physiotherap y + psychologica l vs physiotherap y alone	Y	Y	N	insufficient evidence to assess AEs" "There is evidence that combining physiotherapy and psychological approaches improves physical function in chronic pain in comparison with physiotherapy alone. Further examination of this field is required to inform changes in practice and to develop treatment methods."
Yang et al. 2022 ⁴⁶	China	"investigated the effectiveness of CBT on pain, disability, fear avoidance, and self-efficacy in patients with CLBP."	To Novemb er 20 2021	2013	"CBT vs other therapies"	Y	N	N	"CBT is beneficial in patients with CLBP for improving pain, disability, fear avoidance, and self-efficacy in CLBP patients. Further study is recommended to investigate the long-term benefits of CBT."
Yanyun et al. 2021 ⁴⁷	China	"to systematically review the published evidence to determine whether Pilates exercise training is an efficacious therapy for scoliosis."	To Decemb er 2020	2016B	Pilates vs other	Y	N	N	"Pilates exercise training may reduce the Cobb angle and trunk rotation, relieve pain, increase trunk ROM, and improve QOL for patients with scoliosis. Due to the poor quality of the evidence, however, these results should be interpreted with caution."

Table 4: Summary of results of sensitivity analyses exploring the effect of removing studies of interest from published meta-analyses.

Group	Outcome (s)	Effect size: Number of comparisons	Effect size (SMD) (median IQR) with index trials included	Effect size (SMD) (median IQR) with index trials excluded	Absolute change in effect size (SMD), median (IQR)	% reduction in effect size, median (IQR)	% change in width of 95% confidence intervals, median (IQR)	I2 number of comparisons	Absolute reduction in I2 statistic, median (IQR)
All studies	Pain and Disability	55	-0.6 (-0.85 to -0.42)	-0.19 (-0.33 to -0.15)	-0.35 (-0.51 to -0.21)	58.3 (39.8 to 73.7)	-47.6 (-64.6 to -26.5)	44	27.3 (10.3 to 66.3)
	Pain only	29	-0.52 (-0.73 to -0.4)	-0.19 (-0.29 to -0.15)	-0.35 (-0.49 to -0.19)	62.3 (41.3 to 73.8)	-47.8 (-65.2 to -25)	24	28.5 (11 to 74.1)
	Disability only	26	-0.73 (-0.87 to -0.45)	-0.22 (-0.43 to -0.18)	-0.35 (-0.51 to -0.27)	58.2 (40.3 to 72.8)	-43.5 (-63.4 to -28.5)	20	26.3 (8 to 43)

Interventions for chronic pain.	Pain and Disability	38	-0.52 (-0.91 to 0.37)	-0.19 (-0.31 to -0.15)	-0.35 (-0.51 to -0.19)	56.9 (41.9 to 72.8)	-44.5 (-60 to -20)	36	24 (10.3 to 46.3)
	Pain only	19	-0.48 (-0.83 to -0.34)	-0.19 (-0.30 to -0.15)	-0.34 (-0.44 to -0.17)	50.7 (39 to 69.9)	-48.9 (-59 to -15)	18	26.5 (11.5 to 61)
	Disability only	19	-0.63 (-0.95 to -0.4)	-0.19 (-0.33 to -0.16)	-0.35 (-0.55 to -0.21)	58 (46.5 to 74)	-40 (-60.7 to -23.1)	18	22.5 (8 to 35.5)
Interventions for post-surgical pain	Pain and disability	17	-0.71 (-0.84 to -0.52)	-0.23 (-0.34 to -0.18)	-0.36 (-0.49 to -0.29)	60.3 (38.3 to 73.8)	-47.8 (-73.4 to -40.2)	8	88.5 (20.8 to 95.3)
	Pain only	10	-0.64 (-0.73 to -0.45)	-0.20 (-0.28 to -0.11)	-0.41 (-0.49 to -0.25)	72.9 (47.9 to 84.1)	-47.7 (-74.9 to -41.5)	6	61 (8.3 to 95.8)
	Disability only	7	-0.84 (-0.85 to -0.69)	-0.31 (-0.57 to -0.22)	-0.35 (-0.42 to -0.3)	58.3 (36.2 to 61.8)	-53 (-72.1 to -37.1)	2	88.5 (87.3 to 89.8)

Table 5. Characteristics of Clinical Practice Guidelines that included index trials.

Organisation	Title Primary	Year published	Country of origin	Review Objectives	Search dates	Index trials included	Interventions and/or Comparisons affected	meta-analysis? Y/N	narrative synthesis Y/N	GRADE Y/N
Academy of Orthopaedic Physical Therapy of the American Physical Therapy Association (AAOPT) ⁵²	Interventions for the Management of Acute and Chronic Low Back Pain: Revision 2021	2021	USA	CPG for acute and chronic LBP	To June 25 2020	Monticone 2016a	General exercise to other exercise training interventions	N	Y	N
Belgian Health Care Knowledge Centre (KCE) ⁵⁷	Low back Pain and radicular Pain: Assessment and Management	2017	Belgium	The guidelines provides recommendations based on current scientific evidence for the evaluation and management of low back Pain and radicular Pain in adult population. Clinicians.	To Dec 2015	Monticone 2013, 2016a	Multidisciplinary biopsychosocial treatment programs	N	Y	Y
Canadian Agency for Drugs and Technologies in Health (CADTH) ⁵¹	Multidisciplinary treatment programs for patients with chronic non-malignant Pain: a review of	2019	Canada	to review the comparative clinical effectiveness, cost-effectiveness, and evidence-based	To April 2019	Monticone 2017	Multidisciplinary treatment programs	N	Y	N

	clinical effectiveness, cost-effectiveness, and guidelines: an update			guidelines regarding the use of multidisciplinary treatment programs for patients with chronic, non-malignant Pain in outpatient settings						
Canadian Chiropractic Association and the Canadian Federation of Chiropractic Regulatory and Educational Accrediting Boards (CCA/CFC) ⁴⁹	The Treatment of Neck Pain-Associated Disorders and Whiplash-Associated Disorders: A Clinical Practice Guideline	2016	Canada	to determine which treatments or combinations of treatments are more effective for managing NAD and WAD.	To December 24 2015	Montic one 2012	Multimodal care vs continued practitioner care for persistent grades I to III NAD?	N	Y	Y
Dutch General Practitioners Association. Netherlands Huisartsen Genootschap (NHG) ⁴⁸	Non-specific low back Pain (M54)	2017	Netherlands	CPG for the diagnosis and management of non-specific low back Pain.	To June 2014	Montic one 2013	Behavioral treatment and exercise therapy vs. exercise therapy	N	Y	Y
Finnish Medical Association, Association of Physical Medicine and Rehabilitation Fennia (FMA) ⁵⁰	Neck Pain	2017	Finland	CPG for the management of neck Pain	NR	Montic one 2012	Cognitive therapy	N	Y	N
Japanese Orthopaedic Association (JOA) ⁵³	Japanese Orthopaedic Association (JOA) clinical practice guidelines on the management of lumbar spinal	2022	Japan	biopsychosocial rehabilitation for patients with chronic low back Pain.	2008 to 2019	Montic one 2014B	Effects of postoperative physiotherapy 1 year after surgery	Y	Y	N

											stenosis, 2021 - Secondary publication
National Institute of Health and Care Excellence (NICE) ⁵⁵	NG59 Low back Pain and sciatica in over 16s: assessment and manage- ment	2016	UK	Clinical guideline for the assessment and manage- ment of low back Pain, with or without sciatica	to Dec 2015	Montic one 2013, 2016a	Multidiscip linary biopsych social treatment programs	N	Y	Y	
North America n Spine Society (NASS) ⁵⁴	Guideline summary review: an evidence- based clinical guideline for the diagnosis and treatment of low back Pain	2020	USA	To provide an evidence- based educational tool to assist spine specialists when making clinical decisions for adult patients with nonspecific low back Pain.	NR	Montic one 2013	Exercise therapy alone versus exercise with cognitive behavioral therapy (CBT) Interventio ns that address fear- avoidance behaviors	N	Y	N	
Russian Society for the Study of Pain (RSSP) ⁵⁶	Chronic nonspecific (musculosk eletal) low back Pain. Guidelines of the Russian Society for the Study of Pain (RSSP)	2019	Russia	Chronic nonspecific (musculosk eletal) low back Pain. Guidelines of the Russian Society for the Study of Pain	NR	Montic one 2013. 2016A	Multidiscip linary biopsych social treatment programs	N	Y	N	

Footnotes: Recommendations based on evidence that included a study of interest. NR= not reported

Table 6: Summary of guideline analyses that included Index trials.

CPG	Affected research question or comparison	Monticone trials included	Synthesis Summary	Recommendation or conclusions
Academy of Orthopaedic Physical Therapy of the American Physical Therapy Association (AAOPT) ⁵²	How effective is general exercise compared to other exercise training interventions?	2016A	4 trials of exercise training included. Combined N not reported. N contributed by trial of interest= 150	"Physical therapists should use exercise training interventions, including trunk muscle strengthening and endurance, multimodal exercise interventions, specific trunk muscle activation exercise, aerobic exercise, aquatic exercise, and general exercise, for patients with chronic LBP."
Belgian Health Care Knowledge Centre (KCE) ⁵⁷	What is the effectiveness biopsychosocial treatment programs for low back Pain	2013, 2016a	3 trials of multidisciplinary biopsychosocial treatment programs included.	"Consider a multidisciplinary rehabilitation programme, which combines a physical and a psychological component, incorporating a cognitive behavioural approach, and which takes into account a person's specific needs and capabilities, for people with persistent low back Pain or radicular Pain: or when they have psychosocial obstacles to recovery or when

	with or without sciatica?		Combined N= 361 N contributed by index trials = 240	previous evidence-based management has not been effective."
Canadian Agency for Drugs and Technologies in Health (CADTH) ⁵¹	To review the comparative clinical effectiveness, cost-effectiveness, and evidence-based guidelines regarding the use of multidisciplinary treatment programs for patients with chronic, non-malignant Pain in outpatient settings	2017	2 Systematic review and 2 RCTS included. RCTS combined N=329. N contributed by trial of interest= 170	"Overall, findings from the included studies suggested that the multidisciplinary management of chronic non-malignant Pain was associated with significant improvements in Pain intensity, and may be associated with significant improvements in quality of life and function."
Canadian Chiropractic Association and the Canadian Federation of Chiropractic Regulatory and Educational Accrediting Boards (CCA/CFC) ⁴⁹	Should multimodal care vs continued practitioner care be used for persistent grades I to III NAD?	2012	2 RCTs, combined N=357. N contributed by trial of interest = 90. Trial of interest not considered to show evidence of clinically important effects. "The addition of a cognitive behavioural treatment did not provide greater outcomes than multimodal care alone."	"For patients presenting with persistent neck Pain grades I to III, we suggest clinicians offer multimodal care* and/ or practitioner advice† based on patient preference. (Weak recommendation, low-quality evidence)"
Dutch General Practitioners Association ⁴⁸	Behavioural treatment and exercise therapy vs. exercise therapy	2013	1 systematic review with 30 RCTs (combined N= 3438) + 5 additional RCTs (combined N=889). N contributed by index trials = 90	"It is unclear whether there are clinically relevant benefits of cognitive behavioural treatment over waiting list or standard treatment in patients with chronic non-specific low back Pain."
Finnish Medical Association, Association of Physical Medicine and Rehabilitation Fennia ⁵⁰	The effectiveness of cognitive therapies for neck Pain	2012	3 RCTS AND 1 controlled clinical trial (combined N=813) N contributed by index trials =80	"Cognitive therapy may be effective in the treatment of neck Pain, but there is no convincing research evidence of the effectiveness of such therapy".
Japanese Orthopaedic Association (JOA) ⁵³	What are the effects of postoperative physiotherapy 1 year after surgery for spinal stenosis?	2014B	Meta-analyses: Physiotherapy for leg Pain at one year. 5 RCTS, combined N= 595	"Physiotherapy for patients undergoing surgical treatment for LSS is effective for alleviating Pain and improving ADL and QOL 3 months after surgery and adverse events are rare. Thus, postoperative physiotherapy can be considered useful."

			N contributed by index trials = 117	
			Physiotherapy for back Pain at one year. 4 RCTs, combined N=473	
			N contributed by index trials = 117	
National Institute of Health and Care Excellence (NICE) ⁵⁵	What is the effectiveness biopsychosocial treatment programs for low back Pain with or without sciatica?	2013, 2016a	3 RCTs of multidisciplinary biopsychosocial treatment programs included. Combined N=361 N contributed by index trials = 240	"Consider a multidisciplinary rehabilitation programme, which combines a physical and a psychological component, incorporating a cognitive behavioural approach, and which takes into account a person's specific needs and capabilities, for people with persistent low back Pain or radicular Pain: or when they have psychosocial obstacles to recovery or when previous evidence-based management has not been effective."
North American Spine Society (NASS) ⁵⁴	What are outcomes, including duration of Pain, intensity of Pain, functional outcomes and return-to-work status, for exercise therapy alone versus exercise with cognitive behavioral therapy (CBT)?	2013	8 RCTs of the addition of CBT to exercise. Combined N=913 N contributed by index trials = 90	"There is conflicting evidence that addition of CBT to an exercise program results in significant improvement in Pain and function compared with exercise alone in patients with chronic low back Pain." "Treatments targeting fear avoidance combined with physical therapy are recommended compared to physical therapy alone to improve low back Pain in the first 6 months."
			4 RCTs of interventions to reduce fear avoidance. Combined N=287 N contributed by index trials = 90	
Russian Society for the Study of Pain (RSSP) ⁵⁶	What is the effectiveness of Multidisciplinary biopsychosocial treatment programs?	2013, 2016A	23 studies referenced in the evidence summary for multidisciplinary biopsychosocial treatment programs. Combined N or further details not reported. N contributed by index trials = 240	"Multidisciplinary programs, which include physical therapy (therapeutic exercises), psychological methods (primarily cognitive-behavioral therapy), and educational conversations / lectures (schools) for patients are recommended for the treatment of chronic LBP. The use of multidisciplinary programs allows to improve the main indicators of the patient's condition: intensity of Pain, functional activity, professional activity, psychological and physical quality of life."

Highlights

- A group of trials with trust concerns had major impacts on the results of systematic reviews and clinical guidelines.
- They substantially impacted effect sizes and influenced the conclusions and recommendations drawn.
- There is a need for a greater focus on the trustworthiness of studies in evidence appraisal.