J Rehabil Med 2018; 50: Epub ahead of print

REVIEW ARTICLE

EFFECTIVENESS OF MULTIDISCIPLINARY PROGRAMMES FOR CLINICAL PAIN CONDITIONS: AN UMBRELLA REVIEW

Elena DRAGIOTI, MSc, PhD[1,2], Evangelos EVANGELOU, PhD[2,3], Britt LARSSON, MD, PhD[1] and Björn GERDLE, MD, PhD[1]
From the [1]Pain and Rehabilitation Centre, and Department of Medical and Health Sciences, Linköping University, Linköping, Sweden, [2]Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, University Campus, Ioannina, Greece and [3]Department of Epidemiology and Biostatistics, Imperial College London, London, UK

Objective: **To evaluate the strength of the evidence for multimodal/multidisciplinary rehabilitation programmes (MMRPs) for common pain outcomes.**
Data sources: **PubMed, PsychInfo, PEDro and Cochrane Library were searched from inception to August 2017.**
Study selection: **Meta-analyses of randomized controlled trials or controlled clinical trials and qualitative systematic reviews of randomized controlled trials and non-randomized controlled trials were considered eligible.**
Data extraction: **Two independent reviewers abstracted data and evaluated the methodological quality of the reviews. The strength of the evidence was graded using several criteria.**
Data synthesis: **Twelve meta-analyses, including 134 associations, and 24 qualitative systematic reviews were selected. None of the associations in meta-analyses and qualitative systematic reviews were supported by either strong or highly suggestive evidence. In meta-analyses, only 8 (6%) associations that were significant at *p*-value ≤0.05 were supported by suggestive evidence, whereas 44 (33%) associations were supported by weak evidence. Moderate evidence was found only in 4 (17%) qualitative systematic reviews, while 14 (58%) qualitative systematic reviews had limited evidence.**
Conclusion: **There is no evidence that MMRPs are effective for prevalent clinical pain conditions. The majority of the evidence remains ambiguous and susceptible to biases due to the small sample size of participants and the limited number of studies included.**

Key words: systematic review; umbrella review; meta-analysis; multimodal pain treatment; multidisciplinary treatment; pain.

Accepted Jun 7, 2018; Epub ahead of print Aug 8, 2018

J Rehabil Med 2018: 50; 00–00

Correspondence address: Elena Dragioti, Pain and Rehabilitation Centre, and Department of Medical and Health Sciences, Linköping University, SE-581 85 Linköping, Sweden. E-mail: elena.dragioti@liu.se

LAY ABSTRACT

This study evaluated the published literature regarding multimodal/multidisciplinary rehabilitation programmes (MMRPs) for pain outcomes. The study reviewed the evidence on a large scale, examining 134 associations derived from 12 meta-analyses (including 462 primary studies) and 24 qualitative systematic reviews (including 243 primary studies). The results suggest that there is a lack of robust evidence about the effectiveness of the programmes investigated; most of the published studies displayed uncertainty in effect sizes due to large heterogeneity, small sample sizes, evidence of small-study effects, excess of significant findings, or any combination of the above. Some weak evidence, especially for short-term outcomes, may be genuine, but no firm conclusions can be drawn. This study highlights the necessity for larger, better-conducted, randomized controlled trials of the effectiveness of MMRP, with a standardized formula of treatment modalities, outcome measures, pain population, pain assessments, and length of treatments.

Pain conditions, such as low back pain (LBP), neck pain (NP), spinal pain (SP), whiplash-associated disorders (WAD), widespread pain (WSP), and fibromyalgia (FMS), are highly prevalent and frequently persistent chronic conditions, which cause significant disability, distress, impaired quality of life, and work absenteeism (1–10). The prevalence of these conditions ranges from 10% to 60%, with a high variation depending on age, sex, population setting (i.e. inpatients, outpatients) and duration of pain (i.e. subacute, chronic) (11–15). A new data analysis from the 2012 National Health Interview Survey (NHIS) found that 55.7% of American adults (~126 million individuals) reported having pain (16). Moreover, the socioeconomic burden of these conditions in developed countries is enormous, due to both direct and indirect costs (10–12). Thus, effective treatments are of the utmost importance.

Over recent decades, multimodal/multidisciplinary rehabilitation programmes (MMRPs) have been studied as a promising strategy for treatment of pain (10, 17, 18). MMRPs comprise a lengthy, biopsychosocial treatment framework, which generally contains a synchronized combination of physical, educational or psychological treatments provided by a team of different professionals (5, 7, 18, 19). Several systematic reviews (SRs) and meta-analyses (MAs) support the effectiveness of MMRPs for LBP (4, 5, 8, 10, 19–23), NP (including WAD) (6, 9, 24, 25) and WSP (including FMS) (2, 26, 27). In support of this data, it has been stated that, among all pain treatments, MMRPs provide a high evidential basis for efficacy, cost-effectiveness, and lack of indu-

doi: 10.2340/16501977-2377

ced complications (28). Nonetheless, there is growing concern that these results may be influenced (29) by an array of flaws, such as the presence of between-study heterogeneity, publication bias, and selective reporting of positive results (30–35). Biases in the reported findings in SRs and MAs are not unusual in the medical literature (30–35). An up-to-date umbrella review of 247 psychotherapy MAs (including pain outcomes) found that only a small fraction (7%) were supported by strong evidence and were free from biases (35).

Although empirical studies are available, no systematic umbrella review on this topic has been performed to date. Umbrella reviews systematically evaluate the evidence on an entire topic across various SRs and MAs on multiple outcomes (36) and appraise the strength of the evidence, offering better recognition of the uncertainties, biases and knowledge gaps (37). The aim of this study was to examine if, in patients with prevalent clinical conditions, such LBP, NP, SP, WAD, and FMS (Population), do MMRPs (Intervention), compared with any other active or inactive control (Control), improve pain, disability or any other reported outcome (Outcomes). To this end, an umbrella review of SRs and MAs that evaluated the effectiveness of MMRPs for the above-mentioned pain conditions was performed to plot the evidence over time, in addition to presenting areas for further research.

METHODS

Data sources and searches

PubMed, PsycINFO, Physiotherapy Evidence Database (PEDro) and Cochrane Database of Systematic Reviews (CDSR) were searched from inception to 31 August 2017 for SRs or MAs investigating the effectiveness of an MMRP for LBP, NP, SP, WAD and WSP including FMS (see Table SI[1] for search strings). The reference lists in the relevant SRs and MAs were also hand-searched for additional articles missed by the electronic search. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations for reporting SRs and MAs were followed. The protocol for this umbrella review has been published on Prospero (Prospero record registration no: CRD42017076309).

Two independent investigators (ED, BL) screened the titles, the abstracts of the identified records, and the full-texts of the potentially eligible articles. In cases of discrepancy, a third investigator (BG) was consulted until agreement was reached.

Study selection

Qualitative SRs and MAs that tested MMRPs vs any control (e.g. treatment as usual, waiting list) or other treatment (e.g. physiotherapy, surgery) were eligible for inclusion. Reviews that used an MMRP as a control group (e.g. physiotherapy vs MMRP) were also included. If a review tested multiple treatments, this was considered eligible only in the case that separate results or analyses of MMRPs were presented. The actual definition adopted by the initial authors was used to classify whether a review examined an MMRP. In cases of absence of a clear definition, MMRP was defined as a treatment approach that includes at least 2 distinct treatment components (e.g. at least one physical and at least one educational or other psychological therapy) (7). No restrictions were set regarding the baseline characteristics (e.g. clinical setting, age or sex) and the duration of pain (e.g. acute, subacute or chronic) of the populations studied. In the case of multiple publications concerning a certain SR or MA from the same research group only the most recent or most prominent publication was used. A clear description of other exclusion criteria is provided in the Supplementary Methods and Results[1].

Data extraction and quality assessment

For all eligible reviews the following data were recorded: first author, publication year, country, type of review, examined interventions, pain condition treated, whether a definition of MMRP components was given, number of included studies, total sample size, outcomes, and main findings. For each primary study included in the MAs the following data were also recorded: first author, year of publication, study design, sample size, effect size (ES) (i.e. mean difference (MD); standardized mean difference (SMD); risk ratio (RR); odds ratio (OR)), and 95% confidence intervals (95% CI). One investigator (ED) extracted the data, which were confirmed independently by another investigator (EE). Discrepancies were resolved by discussion with a third investigator (BG).

Two independent investigators (ED, EE) assessed the methodological quality of the selected reviews using the Assessment of Multiple Systematic Reviews (AMSTAR) checklist. The AMSTAR is an 11-item instrument with values ranging from 0 to 11 related to essential features of the methodological rigor across SRs and MAs; higher scores indicate higher quality (for details see Table SII[1]). The AMSTAR scores can be also ordered as high (8–11), medium (4–7) and low quality (0–3) (38).

Data synthesis and analysis

The main analysis in this umbrella review focused on quantitative synthesis only for SRs with quantitative synthesis or MAs of RCTs and CCTs. To this end, both fixed and random-effects models were performed to estimate the summary effect sizes (ES) and the 95% CI in each association (39). A fixed-effect model estimates a single effect that is assumed to be common in every primary study, while a random-effects model estimates the mean of a distribution of effects (40). The direction of associations presented on the original MAs was not altered, so that the results could be compared with the original results. However, to harmonize all the continuous outcomes, whenever MDs were reported transformation into SMDs were performed via standardized formula (40).

Between-study heterogeneity was appraised with the Cochran's Q statistic (41) and measured with the I^2 metric (i.e. low, moderate, large, very large for values of <25, 25–49, 50–74, >75%, respectively) (42). When heterogeneity is not present ($I^2=0$), random and fixed-effects coincide. The 95% prediction intervals (PIs) in the random effects modelling were also estimated to provide an additional account of the unexplained heterogeneity and prediction of an interval for future ES estimates (43).

The Egger's regression asymmetry test was performed to estimate small-study effects bias (44). Briefly, small-study ef-

[1]http://www.medicaljournals.se/jrm/content/?doi=10.2340/16501977-2377

fects refer to the phenomenon that smaller studies often show larger treatment effects than do large ones (44, 45). A *p*-value ≤0.10 in the Egger test, together with a summary random effects ES larger than the ES of the largest study in each association, displays evidence of small-study effects.

Excess of significant findings was assessed using the excess of significant findings test developed by Ioannidis & Trikalinos (46). This test examines whether the observed number of studies (O) with statistically significant results (*p*-value <0.05) is larger than the expected number of studies (E) (31, 35, 46). The E was taken as the sum of the statistical power estimates for each study in the MA and the power of each study was calculated with an algorithm using a non-central *t* distribution (47). Since the true ES of a meta-analysis is not known, this umbrella review assumed as the plausible true effect the ES of the largest study (48). Excess of significance bias was set at a *p*-value ≤0.10 with O>E (32, 35, 46).

Whenever the primary study data for a MA was unavailable, only the summary ESs or any other information (e.g. heterogeneity or publication bias assessment) reported by the original authors were considered. In this case, further assessments of various statistical tests (e.g. 95% PI, ES of the largest study, small-study effects or excess of significant findings) were not feasible.

The secondary analysis in this umbrella review focused on descriptive analysis for qualitative SRs and MAs excluded from the quantitative synthesis. For this analysis, studied outcomes were categorized into 5 outcome areas: (1) pain, (2) physical functioning (including disability and work status), (3) emotional functioning, (4) global measures (e.g. quality of life), and (5) other (e.g. adverse events) (49).

All analyses were performed using Stata version 12 (College Station, TX, USA) (50).

Assessment of the credibility of the evidence

The credibility of the evidence of each association provided in MAs was assessed using a number of criteria previously applied in various medical fields (31, 32, 34, 35, 51). In brief, associations that presented nominally significant random-effects summary estimates (i.e. *p*-value ≤0.05) were regarded as strong, highly suggestive, suggestive, or weak evidence (Table I). The strength of evidence of each qualitative SR or MA not included in the quantitative synthesis was also appraised in 1 of the following 4 categories: strong evidence, moderate evidence, limited evidence, and no evidence, based on modified van Tulder`s et al. criteria (Table I) (52).

RESULTS

Search results

The primary search yielded a total of 9,896 articles, which provided 89 potentially eligible articles (Fig. 1). Of these, 36 met the inclusion criteria (1–9, 17, 19–22, 24–27, 53–69), of which 13 were qualitative SRs and 23 were MAs (Table SIII[1]). The reasons for exclusion of the 53 articles (Supplementary references 1–53[1]) are summarized in Table SIV[1]. Of the 23 eligible MAs, only 12 (including 134 associations) were finally selected for quantitative synthesis (Fig. 1) (2–4, 6, 8, 17, 21–23, 54, 55, 59). Reasons for exclusion were mostly because 5 MAs were duplicate publications from the same research group, 4 MAs were updated versions of the same research group, and 2 Cochrane reviews did not provide a quantitative synthesis of data (Table SIII[1]). Primary study data were available for all MAs, with the exception of the meta-analysis by Hoffman's et al. (59).

Table SIII[1] presents the descriptive characteristics of the 36 selected SRs and MAs. All reviews were published between 1994 and 2017. Definition of the contents of MMRP was given in 21 reviews (58.3%).

Quality of selected systematic reviews and meta-analyses

The median AMSTAR quality assessment score of all 36 reviews was 7 (interquartile range (IQR)=6–9; Table SV[1]). Fifteen reached the "high-quality" level (≥8/11 of the AMSTAR checklist), while 2 reviews

Table I. Criteria of the credibility of the evidence for selected meta-analyses and qualitative systematic reviews

Category	Interpretation
	Results from meta-analyses
Convincing evidence	*p*-value $<10^{-6}$ based on random effects meta-analysis; had >350* participants; had low or moderate between-study heterogeneity ($I^2<50\%$); the largest study with nominally statistically significant ($p<0.05$); had 95% prediction interval excluding the null value; and had no evidence of small-study effects and excess significance
Highly suggestive evidence	*p*-value $<10^{-6}$ based on random effects meta-analysis; had >350* participants; and the largest study with the largest study with nominally statistically significant ($p<0.05$)
Suggestive evidence	*p*-value ≥10−6, but $p<0.001$ by random-effects; and had >350* participants
Weak evidence	All other associations with *p*-value ≤0.05
No evidence	All associations with *p*-value >0.05
	Results from qualitative systematic reviews and meta-analyses not included in quantitative synthesis
Strong evidence	At least half of a review's included high-quality randomized controlled trials (RCTs) showed generally consistent findings in at least 2 of the primary outcomes, or at least in 1 of the primary and 2 of the secondary outcomes following the intervention
Moderate evidence	A review where at least 1 high-quality RCT and in 1 or more low-quality RCTs, or at least half of a review's included low-quality RCTs showed generally consistent findings in at least 2 out of the primary outcomes, or at least in 1 of the primary and 2 of the secondary outcomes following the intervention
Limited evidence	A review where at least 1 RCT (either high or low quality) or inconsistent or contradictory evidence in multiple RCTs in at least 1 primary outcomes, or at least in 1 of the primary and 1 of the secondary outcomes following the intervention
No evidence	A review where no significant differences between intervention and control groups were reported in any of the included primary studies or evidence from 1 methodologically weak study or contradictory outcomes

*This was the necessary sample size based on a small-to-moderate effect size (standardized mean difference 0.3) with 80% power and an alpha level of 0.05 by power analysis and this was also the median number of participants in meta-analyses.

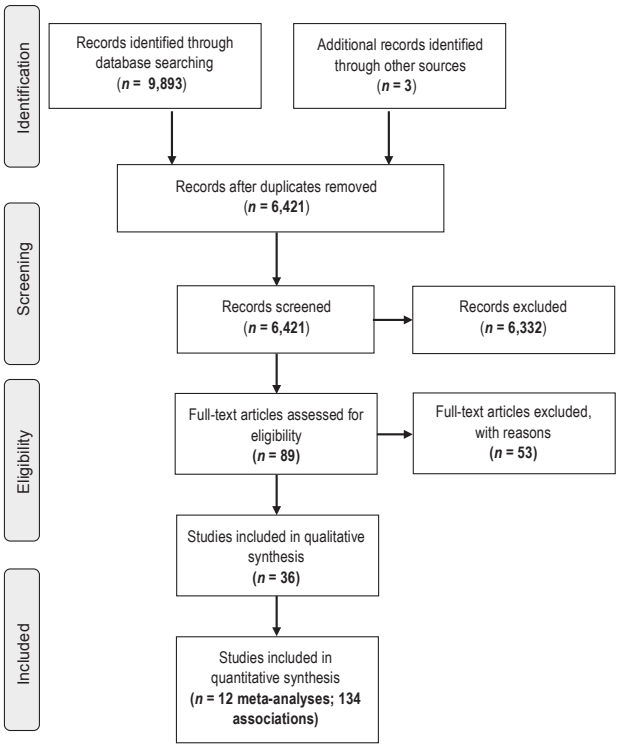

Fig. 1. Flowchart of the literature search and evaluation process of published meta-analyses and systematic reviews.

met the "low-quality" level (0–3/11). The level of agreement of AMSTAR scores was high; 90% between the 2 independent investigators.

Description of meta-analytic associations

Table SVI[1] presents the pain conditions, outcomes, characteristics and summary estimates of the 134 associations. These associations provided evidence for 4 pain conditions; namely, LBP, NP, SP and FMS, and included a total of 462 primary studies, of which only 2 were CCTs. The median number of primary studies per meta-analysis was 2 (IQR = 2–4). The median number of participants was 347 (IQR = 167–457) and the total number of participants was >1,000 in only 11 (8.2%) associations. The median length of the MMRPs was 5 weeks (IQR = 3–8). The examined outcomes are visualized in Fig. 2. A further description of the meta-analytic associations is provided in the Supplementary Methods and Results[1].

Summary effect sizes

Fig. 3 and Table SVI[1] provide summary estimates for all 134 associations. In the fixed-effect models, 71 (52.9%) associations reported ESs that were significant at p-value <0.05 (Fig. 3), of which only 4 favoured the control group. However, in 2 of those 4 MAs, the comparator was an MMRP. In the random-effect models, 52 (38.8%) associations reported ESs that were significant at p-value <0.05 (Fig. 3); all favouring the MMRPs. In 2 associations, the MMRP was also treated as a control group. Only 15 (11.2%) associations were significant at p-values < 0.001 under random-effects modelling. Of note, in 6 (4.5%) associations it was not possible to use fixed-effect models due to unavailability of the primary data. The results of the largest study in each meta-analysis are provided in the Supplementary Methods and Results[1].

In 57 (42.5%) associations the estimates of the PIs included the null value, while in 76 (56.7%) the PIs could not be estimated due to an inadequate number of included RCTs (PIs required at least 3 primary studies included in each MA to be estimated; Fig. 3). In 38 (28.4%) associations the ES of the largest study in each meta-analysis had a nominally statistically significant result. In 2 (1.5%) associations, considering the short-term outcomes of depression and disability for chronic LBP, the result was in the reverse direction (4).

Between-study heterogeneity and small-study effects

Statistically significant between-study heterogeneity (p-value ≤0.10) was found in 59 (44.0%) associations (Table SVI[1]; Fig. 3). There was large heterogeneity (I^2=50–75%) in 43 (32.1%) associations and very large heterogeneity (I^2 > 75%) in 19 (14.2%) associations of 5 outcomes for chronic and subacute LBP. A further description of the associations with high heterogeneity is provided in the Supplementary Methods and Results[1].

Small-study effects bias was found in 9 (6.7%) associations of 6 outcomes for chronic and subacute LBP (i.e. short-term episode of LBP, disability, quality of life, and coping, medium-term pain, disability and depression, and medium and long-term disability/

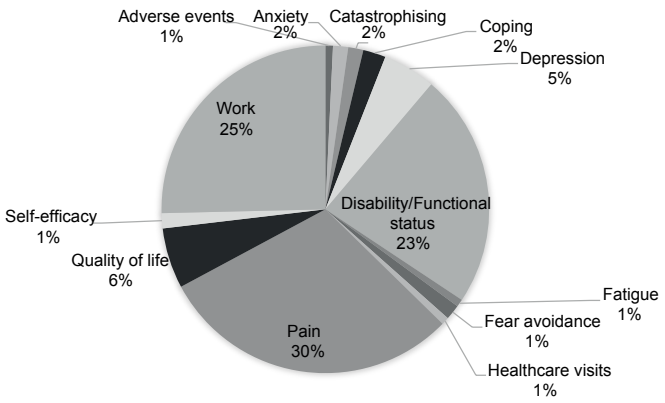

Fig. 2. Description of outcomes reported in 134 associations in meta-analyses for neck pain, low back pain and fibromyalgia.

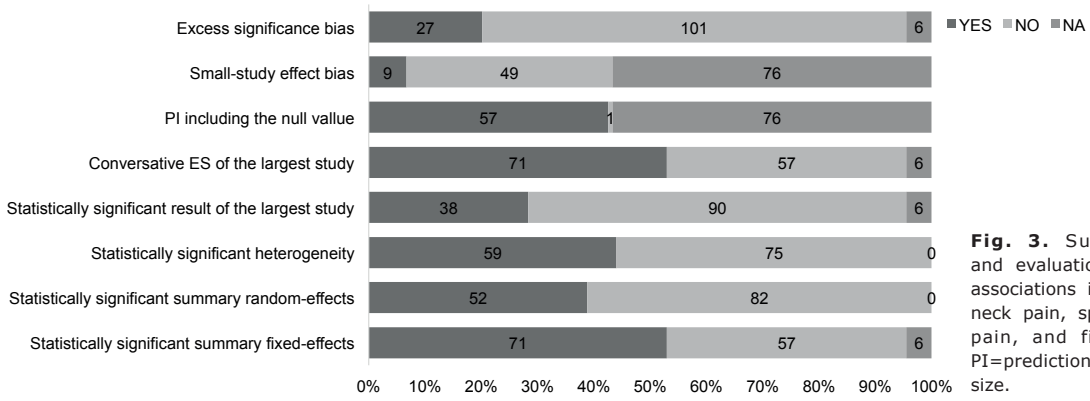

Fig. 3. Summary estimates and evaluation of biases in 134 associations in meta-analyses for neck pain, spinal pain, low back pain, and fibromyalgia Notes: PI=prediction interval, ES=effect size.

functional status and long-term return to work) (4, 6, 8, 23). Hence, an evidence of small study effects was unimportance. On the other hand, in 76 (56.7%) associations, the small-study effects could not be estimated; the Egger's test can be employed only for MAs including at least 3 primary RCTs (Fig. 3).

Excess of significant findings

An excess of significant findings ($p \leq 0.10$) was observed in 27 (20.1%) associations (Fig. 3), of 6 outcomes for chronic and subacute LBP and chronic SP. In 54 (40.3%) associations E was larger than O, indicating that an excess of significant findings was not pertinent (Table SVI[1]; Fig. 3). This test could not be estimated in only 6 associations (59). Thus, we did not detect consequential evidence of an excess of significant findings. A further description of the associations with an excess of significant findings is provided in the Supplementary Methods and Results[1].

Credibility of the evidence

The assessment of the 134 associations is presented in Table II. None (0.0%) of these associations had either convincing or highly suggestive evidence in favour of the MMRP. Only 8 (6.0%) associations had >350 participants and significant summary associations (p-value $>10^{-6}$ but <0.001) under random-effects modelling and they were classified as having suggestive evidence. Five of those associations with suggestive evidence showed beneficial effects in the short-term, 2 in the medium-term and one in the long-term. Forty-four (32.8%) were supported by weak evidence reporting nominally statistically significant random-effects associations at p-value ≤ 0.05. Thirty-eight of these displayed beneficial effects both in the short- and the long-term, whereas only 6 showed beneficial effects in the medium-term. Finally, 82 (61.2%) associations had non-significant evidence under random-effects modelling (p-value >0.05; Table SVII[1]).

Descriptive analysis and strength of the evidence of qualitative systematic reviews

Table III presents descriptive characteristics with the summary of the evidence of the 24 reviews excluded from the quantitative synthesis. These reviews included a total of 243 primary studies (median=7; IQR 3−12). A detailed descriptive analysis of qualitative SRs is provided in the Supplementary Methods and Results[1].

None of these reviews was supported by strong evidence. The criteria of moderate evidence was met by 4 (16.7%) reviews, limited evidence by 14 (58.3%) reviews, and no evidence by 6 (25.0%) reviews (Table III). Meta-analyses were not performed due to the high heterogeneity in 3 reviews and the limited number of included studies in 8 reviews. All duplicate and update MAs showed agreement on the grading of evidence observed in quantitative synthesis (Tables SII[1]).

Subgroup and sensitivity analysis

A subgroup analysis was also performed to verify whether the credibility of the evidence varies as a function based on newer (i.e. MAs published after 2010) vs older (i.e. MAs published before 2010) published MAs. This analysis showed that the newer MAs provided significantly larger associations with both suggestive and weak evidence compared with older MAs (7 vs 1 for the associations with suggestive evidence and 33 vs 11 for the associations with weak evidence; both $p<0.0001$).

A sensitivity analysis with respect to the length of the MMRP was possible only for 35 associations because the rest of the associations did not include both studies with short (≤5 weeks) and long length (>5 weeks) of MMRP (Table SVIII[1]). Sensitivity analyses that limited data to short length indicated that short length of MMRP for the outcomes of return to work short term and pain medium term, showed the largest evidence of association (highly suggestive evidence and suggestive evidence, respectively) in patients with

Table II. Assessment of the credibility of the evidence across the 134 associations in the 12 eligible meta-analyses

Author, year	Outcome	Sample size (total N)	Pain condition	Intervention/control	Significance threshold reached (under the random-effects model)g	95% prediction interval rule	Estimate of heterogeneitye	Small-study effects or excess significance bias	Random-effects summary effect size (95% CI)
Associations with convincing evidence[a]									
None of the meta-analyses was supported by strong evidence									
Associations with highly suggestive evidence[b]									
None of the meta-analyses was supported by highly suggestive evidence									
Associations with suggestive evidence[c]									
	Short-term outcomes (n = 5)								
Steffens, 2016 (8)	Episode of LBP	> 350 but < 500	LBP (prevention)	Exercise+education vs Control	> 10−6 but < 0.001	Including the null value	Not Large	Small-study effects	0.55 (0.41 to 0.74)
Kamper, 2014 (4)	Pain	> 350 but < 1,000	Chronic LBP	Multidis vs TAU	> 10^{-6} but < 0.001	Including the null value	Large	Excess significance bias	−0.55 (−0.83 to −0.27)
Kamper, 2014 (4)	Disability	> 350 but < 1,000	Chronic LBP	Multidis vs TAU	> 10^{-6} but < 0.001	Including the null value	Large	Excess significance bias	−0.41 (−0.62 to −0.19)
Van Middelkoop, 2011 (22)[f]	Pain intensity	> 350 but < 500	Chronic LBP	Multidis vs NT/WL	> 10^{-6} but < 0.001	NA	Not Large	No excess/Small-study effects NA	−0.45 (−0.67 to −0.22)
Guzman, 2002 (21)	Functional status	> 350 but < 500	Chronic LBP	Intensive (> 100h) daily Multidis with functional restoration vs Control	> 10^{-6} but < 0.001	Including the null value	Large	Neither	−0.66 (−1.02 to −0.31)
	Medium-term outcomes (n = 2)								
Kamper, 2014 (4)	Pain	> 350 but < 1,000	Chronic LBP	Multidis vs TAU	> 10^{-6} but < 0.001	Including the null value	Large	Both	−0.60 (−0.85 to −0.34)
Kamper, 2014 (4)	Disability	> 350 but < 1,000	Chronic LBP	Multidis vs TAU	> 10^{-6} but < 0.001	Including the null value	Large	Both	−0.43 (−0.66 to −0.19)
	Long-term outcomes (n = 1)								
Kamper, 2014 (4)	Work	> 1,000	Chronic LBP	Multidis vs Physical	> 10^{-6} but < 0.001	Excluding the null value	Not Large	Neither	1.87 (1.39 to 2.53)
	Associations with weak evidenced								
	Short-term outcomes (n = 19)								
Marin, 2017 (23)	Pain	< 350	Subacute LBP	Multidis vs TAU	> 0.001 but < 0.05	Including the null value	Not large	Neither	−0.40 (−0.74 to −0.06)
Marin, 2017 (23)	Disability	< 350	Subacute LBP	Multidis vs TAU	> 0.001 but < 0.05	Including the null value	Not large	Small-study effects	−0.38 (−0.63 to −0.14)
O'Keeffe, 2016 (6)	Disability	> 350 but 1,000	Chronic LBP + NP	Physical vs Physical+behavioural/ psychologically informed	> 0.001 but < 0.05	Including the null value	Large	Neither	0.27 (0.01 to 0.54)h
Kamper, 2014 (4)	Pain	> 1,000	Chronic LBP	Multidis vs Physical	> 0.001 but < 0.05	Including the null value	Very large	Neither	−0.30 (−0.54 to −0.06)
Kamper, 2014 (4)	Disability	> 1,000	Chronic LBP	Multidis vs Physical	> 0.001 but < 0.05	Including the null value	Very large	Neither	−0.39 (−0.68 to −0.10)
Kamper, 2014 (4)	Pain	< 350	Chronic LBP	Multidis vs WL	> 0.001 but < 0.05	Including the null value	Very large	Neither	−0.73 (−1.22 to −0.24)
Kamper, 2014 (4)	Disability	< 350	Chronic LBP	Multidis vs WL	> 10−6 but < 0.001	Including the null value	Not Large	Neither	−0.49 (−0.76 to −0.22)
Kamper, 2014 (4) [f]	QoL (MCS)	< 350	Chronic LBP	Multidis vs TAU	> 10−6 but < 0.001	NA	Not Large	No excess/Small-study effects NA	0.79 (0.45 to 1.14)
Kamper, 2014 (4)	Catastrophising	< 350	Chronic LBP	Multidis vs TAU	> 0.001 but < 0.05	NA	Not Large	No excess/Small-study effects NA	−0.43 (−0.83 to −0.03)
Kamper, 2014 (4)	Adverse events	> 350 but < 500	Chronic LBP	Multidis vs Surgery	> 0.001 but < 0.05	NA	Not Large	No excess/Small-study effects NA	28.25 (3.77 to 211.93)

Table II. *Cont.*

Author, year	Outcome	Sample size (total N)	Pain condition	Intervention/control	Significance threshold reached (under the random-effects model)g	95% prediction interval rule	Estimate of heterogeneitye	Small-study effects or excess significance bias	Random-effects summary effect size (95% CI)
Van Middelkoop, 2011 (22)[f]	Pain intensity	> 350 but < 500	Chronic LBP	Multidis vs Active control	> 0.001 but < 0.05	NA	Large	No excess/Small-study effects NA	−0.56 (−0.98 to −0.15)
Van Middelkoop, 2011 (22)[f]	Disability	> 350 but < 500	Chronic LBP	Multidis vs NT/WL	> 0.001 but < 0.05	NA	Not Large	No excess/Small-study effects NA	−0.34 (−0.54 to −0.15)
Norlund, 2009 (17)	Return to work	> 1,000	Subacute and chronic LBP	Multidis vs Conservative	> 0.001 but < 0.05	Including the null value	Large	Excess significance bias	1.18 (1.06 to 1.31)
Häuser, 2009 (2)	Pain	< 350	Fibromyalgia	Multidis vs Control	> 0.001 but < 0.05	Including the null value	Not Large	Neither	−0.37 (−0.62 to −0.13)
Häuser, 2009 (2)	Fatigue	< 350	Fibromyalgia	Multidis vs Control	> 0.001 but < 0.05	Including the null value	Not Large	Neither	−0.38 (−0.70 to −0.07)
Häuser, 2009 (2)	Depressive symptoms	< 350	Fibromyalgia	Multidis vs Control	$> 10^{-6}$ but < 0.001	Including the null value	Large	Neither	−0.67 (−1.08 to −0.26)
Hoffman, 2009 (59)	Pain interference	> 350 but < 500	Chronic LBP	Multidis vs Active control	> 0.001 but < 0.05	NA	Not Large	NA	0.20 (0.02 to 0.37)
Guzman, 2002 (21)[f]	Pain rating	< 350	Chronic LBP	Intensive (> 100 h) daily Multidis with functional restoration vs Control	$> 10^{-6}$ but < 0.001	NA	Not Large	No excess/Small-study effects NA	−0.57 (−0.88 to −0.26)
Guzman, 2002 (21)	Employment status	< 350	Chronic LBP	Intensive (> 100 h) daily Multidis with functional restoration vs Control	$< 10^{-6}$	NA	Not Large	No excess/Small-study effects NA	0.49 (0.31 to 0.68)
	Medium-term outcomes (n = 6)								
Kamper, 2014 (4)	Pain	> 350 but < 1,000	Chronic LBP	Multidis vs Physical	> 0.001 but < 0.05	Including the null value	Large	Excess significance bias	−0.28 (−0.54 to −0.02)
Kamper, 2014 (4)	Work	< 350	Chronic LBP	Multidis vs Physical	> 0.001 but < 0.05	Including the null value	Not Large	Neither	2.14 (1.12 to 4.10)
Kamper, 2014 (4)[f]	QoL (PCS)	< 350	Chronic LBP	Multidis vs TAU	> 0.001 but < 0.05	NA	Not Large	No excess/Small-study effects NA	0.42 (0.09 to 0.76)
Kamper, 2014 (4)[f]	QoL (MCS)	< 350	Chronic LBP	Multidis vs TAU	> 0.001 but < 0.05	NA	Not Large	No excess/mall-study effects NA	0.43 (0.09 to 0.76)
Kamper, 2014 (4)	Coping	< 350	Chronic LBP	Multidis vs Physical	> 0.001 but < 0.05	NA	Not Large	No excess/Small-study effects NA	1.09 (0.31 to 1.87)
Hoffman, 2009 (59)	Disability: working	< 350	Chronic LBP	Multidis vs Active control	> 0.001 but < 0.05	NA	Not Large	NA	0.36 (0.06 to 0.65)
	Long-term outcomes (n = 19)								
Marin, 2017 (23)	Pain	< 350	Subacute LBP	Multidis vs TAU	$> 10^{-6}$ but < 0.001	Including the null value	Not large	Excess significance bias	−0.46 (−0.70 to −0.21)
Marin, 2017 (23)	Disability	< 350	Subacute LBP	Multidis vs TAU	> 0.001 but < 0.05	Including the null value	Large	Excess significance bias	−0.44 (−0.87 to −0.01)
Marin, 2017 (23)	Return-to-work	< 350	Subacute LBP	Multidis vs TAU	> 0.001 but < 0.05	Including the null value	Not Large	Small-study effects	3.19 (1.46 to 6.98)
Marin, 2017 (23)	Sick leave periods	< 350	Subacute LBP	Multidis vs TAU	> 0.001 but < 0.05	NA	Not Large	No excess/Small-study effects NA	−0.37 (−0.73 to −0.02)
Steffens, 2016 (8)	Episode of LBP	< 350	LBP (prevention)	Exercise +education vs Control	> 0.001 but < 0.05	NA	Not Large	No excess/Small-study effects NA	0.73 (0.55 to 0.96)
O'Keeffe, 2016 (6)	Disability	> 1,000	Chronic LBP+ NP	Physical vs Physical+behavioural/ psychologically informed	> 0.001 but < 0.05	Including the null value	Large	Both	0.25 (0.07 to 0.43)h
O'Keeffe, 2016[f] (6)	Pain	> 1,000	Chronic LBP+ NP	Physical vs Physical+behavioural/ psychologically informed	> 0.001 but < 0.05	Including the null value	Not Large	Neither	0.18 (0.04 to 0.32)h

Table II. *Cont.*

Author, year	Outcome	Sample size (total N)	Pain condition	Intervention/control	Significance threshold reached (under the random-effects model)g	95% prediction interval rule	Estimate of heterogeneitye	Small-study effects or excess significance bias	Random-effects summary effect size (95% CI)
Kamper, 2014 (4)	Pain	>350 but <1,000	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	Including the null value	Not Large	Neither	−0.21 (−0.37 to −0.04)
Kamper, 2014 (4)	Disability	>350 but <1,000	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	Including the null value	Not Large	Excess significance bias	−0.23 (−0.40 to −0.06)
Kamper, 2014 (4)	Disability	>1,000	Chronic LBP	Multidis vs Physical	>0.001 but <0.05	Including the null value	Very large	Small-study effects	−0.68 (−1.19 to −0.16)
Kamper, 2014 (4)	Catastrophizing	<350	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	−0.40 (−0.76 to −0.05)
Kamper, 2014 (4)	Fear avoidance	>350 but <500	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	Including the null value	Not Large	Neither	−0.29 (−0.49 to −0.08)
Kamper, 2014 (4)	Coping	<350	Chronic LBP	Multidis vs Physical	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	0.30 (0.06 to 0.54)
Schaafsma, 2013 (55)	Proportion off work	<350	Subacute LBP	Intense PCP vs Exercise	$>10^{-6}$ but <0.001	NA	Not Large	No excess/Small-study effects NA	0.57 (0.25 to 0.89)
Schaafsma, 2013 (55)	Time to return to work (>12 mo)	<350	Subacute LBP	Intense PCP + TAU vs TAU	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	−0.39 (−0.76 to −0.02)
Schaafsma, 2013 (55)	Time to return to work (12 mo)	>1,000	Chronic LBP	Intense PCP vs TAU	>0.001 but <0.05	Including the null value	Not Large	Neither	−0.23 (−0.42 to −0.03)
Hoffman, 2009 (59)	Disability: working	>350 but <1,000	Chronic LBP	Multidis vs Active control	>0.001 but <0.05	NA	Large	NA	0.53 (0.19 to 0.86)
Guzman, 2002 (21)	Functional status (60 mo)	<350	Chronic LBP	Intensive (>100 h) daily Multidis with functional restoration vs Control	>0.001 but <0.05	NA	Large	No excess/Small-study effects NA	−0.79 (−1.29 to −0.29)
Guzman, 2002 (21)	Employment status (12 mo)	<350	Chronic LBP	Intensive (>100 h) daily Multidis with functional restoration vs Control	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	0.34 (0.16 to 0.74)

[a]Convincing evidence criteria: >350 participants, significant summary associations ($p<10^{-6}$) per random-effects calculation, prediction intervals not including the null, heterogeneity not large ($I^2<50\%$), no evidence of small-study effects and no evidence of excess of significance bias.
[b]Highly suggestive evidence criteria: >350 participants, significant summary associations ($p<10^{-6}$) per random-effects calculation, and 95% prediction interval not including the null value.
[c]Suggestive evidence criteria: >350 participants and significant summary associations ($p>10^{-6}$ but <0.001) per random-effects calculation.
[d]Weak evidence criteria: all other treatment effects with $p\leq0.05$.
[e]Heterogeneity was categorized as not large ($I^2<50\%$), large ($I^2\geq50\%$ but $I^2<75\%$), and very large ($I^2\geq75\%$).
[f]On these comparisons MD is reported, instead of SMD.
[g]Random effects refer to summary effect (95% CI) using the random-effects model. The direction is arbitrary.
[h]Favour control, but in these 3 meta-analyses the control group was a multidisciplinary programme.
CBT: cognitive behavioural treatment; QoL: quality of life; PCS: physical component summary; MCS: mental component summary; LBP: low back pain; mo: months; NP: neck pain; Multidis: multidisciplinary programme; MBPSR: multidisciplinary bio-psychosocial rehabilitation programmes; PCP: physical conditioning programme; NT: no treatment; WL: waiting list; TAU: treatment as usual; CI: confidence interval; Control: not specified control group; NA: not applicable, because only 2 studies were available or information on included studies was not provided.

Table III. Descriptive characteristics with the summary of the evidence of the 24 qualitative systematic reviews and meta-analyses not included in quantitative synthesis

Author, year	Condition treated	Included studies, *n*	Total sample size, *n*	Outcomes, *n*	Outcomes (Symbol)					Combination of all 3 core health areas (i.e. physical, mental and social health)	Strength of the evidence
					Pain	Physical health /Disability/Work	Emotional health	Global/ Social health	Other		
Sutton, 2016 (9)	WAD	18	2,502	6	+	+	+	+	+	+	Limited
Brady, 2016 (53)	CLBP/CNP /CSP/WSP /FMS	4	349	7	+	+	+	–	–	+	Limited
Kamper, 2015 (5)	CLBP	41	6,858	4	+	+	–	–	+	–	Moderate
Teasell, 2010 (24)	WAD	3	2,248	8	+	+	+	+	+	+	Limited
Teasell, 2010 (24)	WAD	9	367	11	+	+	+	+	+	+	Limited
Schaafsma, 2010 (56)	CLBP	19	3,371	3	–	+	–	–	–	–	Limited
Ravenek, 2010 (57)	CLBP	12	1,913	3	+	+	–	–	–	–	Limited
Sarzi-Puttini, 2008 (58)	FMS	12	919	8	+	+	+	+	+	+	Limited
Scascighini, 2008 (7)	CLBP/FMS	35	2,407	10	+	+	+	+	+	+	Moderate
van Koulil, 2007 (27)	FMS	6	681	3	+	+	+	–	–	+	Limited
van Geen, 2007 (19)	CLBP	10	1,958	4	+	+	–	+	–	+	Limited
Burckhardt, 2006 (26)	FMS	10	1,340	4	+	+	+	–	–	–	Moderate
Tveito, 2004 (60)	LBP	2	271	8	+	+	+	+	–	+	No evidence
Karjalainen, 2003 (61)	LBP	2	233	7	+	+	–	+	+	+	No evidence
Karjalainen, 2003 (62)	CNP	3	177	1	+	+	–	+	+	–	No evidence
Schonstein, 2003 (63)	LBP	18	3,280	5	–	+	+	–	–	–	Limited
Schonstein, 2003 (64)	LBP	7	552	1	–	+	–	–	–	–	Limited
Guzmán, 2001 (20)	CLBP	10	1,964	5	+	+	+	+	+	+	Moderate
Karjalainen, 2001 (69)	CNP	3	177	1	+	+	–	+	+	–	No evidence
Peeters, 2001 (66)	WAD	1	60	4	+	+	–	–	+	–	Limited
Karjalainen, 2001 (65)	LBP	2	233	6	+	+	+	+	+	+	Limited
Karjalainen, 2000 (67)	LBP	2	233	6	+	+	+	+	+	+	Limited
Karjalainen, 2000 (68)	FMS	7	1,050	6	+	+	+	+	+	+	No evidence
Feuerstein, 1994 (1)	CLBP	7	1,025	1	–	+	–	–	–	–	No evidence

WAD: whiplash-associated disorders; CLBP: chronic low back pain; CNP: chronic neck pain; CSP: chronic spinal pain; WSP: widespread pain; FMS: fibromyalgia syndrome; LBP: low back pain; +: a positive symbol indicates that a certain outcome was assessed; –: a negative symbol indicates that a certain outcome was not assessed.

CLBP. Sensitivity analysis that limited data to long length indicated that long length of MMRP for the outcomes of disability medium- and long-term, and pain long-term showed the largest evidence of association (both weak evidence) in patients with CLBP.

DISCUSSION

This study appraised the strength of the evidence across published SRs and MAs of MMRPs for prevalent clinical pain conditions. Primary analysis found that, among 134 associations, less than half produced significant results at p-value ≤ 0.05 under random-effects modelling. The proportion of significant results reduced to almost 11% when a stricter threshold was applied (p-value < 0.001). In addition, none of the statistically significant results presented either convincing or highly suggestive evidence. Only a trivial quantity was supported by suggestive evidence. These pertained to MMRPs associations merely for LBP and mainly for short-term outcomes. However, only one of those associations regarding the long-term effects on work absenteeism inferred by both statistically significant results and absence of biases (4, 5). The remaining associations with statistically significant results were supported by weak evidence, of which the vast majority showed both short-term and long-term beneficial effects. These results were further confirmed by secondary analysis of the 24 qualitative SRs or duplicate MAs not included in the quantitative synthesis. Likewise, none of these reviews was supported by strong evidence. Moderate evidence was found in only 4 reviews, while two-thirds of those had limited evidence. However, the MAs published after 2010 showed larger associations in terms of both suggestive and weak evidence, compared with older MAs published before 2010. Sensitivity analysis that limited data to short length specified that short length of MMRP provided larger evidence of association (highly suggestive evidence and suggestive evidence) compared with long length of MMRP (weak evidence) in patients with CLBP.

This study pinpoints concerns about the robustness of the empirical evidence regarding the effectiveness of MMRPs. Some of the evidence, although limited, may reveal probable associations between MRRPs and the outcomes of pain and disability. The possibility that MMRPs increases the odds of return to work sounds promising and should be tested in future large RCTs. Furthermore, these results highlight that MMRPs may have more favourable effects on short-term outcomes compared with medium- and long-term outcomes; assumptions that require further assessment, e.g. with respect to methods for maintaining gains after MMRPs. Consequently, stakeholders, such as clinicians, researchers, and health policymakers, should be aware that

findings stemming from few MAs with restricted numbers of RCTs must be used with caution. Indeed, there is ongoing discussion regarding meaningful clinical interpretation of the results of the published MAs and their reported outcomes (70). Health policymakers and expert panels should be aware that the evidence is limited, and adjust for the cost-effectiveness of these treatments. Concerns regarding the economic burden of MMRPs have been described repeatedly in the literature (4, 5, 71). However, adjusting for costs may not be as simple as that; the implementation of larger RCTs may be not be practical due to cost barriers. On the other hand, the consideration of such costs should be balanced against healthcare costs and societal costs, e.g. within the social insurance system and in the workplace.

The method used to grade the evidence presents some difficulties in comparing the current results directly with previous research. However, the method used here generally complies with a current SR on behalf of the American College of Physicians Clinical Practice Guideline (72). In that review, adopting the criteria of the Agency for Healthcare Research and Quality, the authors found low-to-moderate evidence for MMRPs on LBP (72). Similarly, the majority of reviewed SRs and MAs used in this study (some also based on the GRADE approach) conclude that it is possible that MMRP may have benefits; however, there is no convincing evidence (4–7, 9, 17, 18, 21, 26, 57, 61, 62, 66–68). Only a meta-analysis of Hauser et al. (2) reported strong evidence on short-term effects on key symptoms of FMS; a finding not supported by our evaluation. In particular, this finding failed to achieve strong evidence, principally because the small sample size of the participants (<350) and the PIs under the random-effect modelling included the null value. Additional SRs from other medical fields using GRADE have also produced similar results, e.g. a review of stroke rehabilitation resulted in a weak recommendation regarding acupuncture (73). One may argue that we used a low threshold of the sample size to evaluate the evidence compared with other studies (32, 34, 35, 74). The threshold of above 1,000 cases is used mainly in genetic association studies (51, 74), but there are other fields that, by definition, cannot recruit such sample sizes. In the literature, lower sample sizes (e.g. ≥ 200) for the assessment of the quality of evidence have been also proposed (75).

At first glance, the failure of both SRs and MAs to reach the criteria of strong evidence might be discouraging; however, cautious examination of the results may reveal some optimistic inferences. More than 60% of the published associations displayed non-significant effects. This may indicate that data dredging, also known as "*p*-value hacking" (76) is less common in the MMRP literature. In a previously published umbrella review of psychotherapy treatments, the significant effects were in favour of the psychotherapy by 80%, while the *p*-value threshold below 0.001 was found in 65% of associations (35). By the same logic, the finding that the majority of associations encompassed a low risk of biased results may indicate that the publication bias favouring positive results, selection bias or outcome reporting bias are less likely to occur in the MMRP field. However, a large body of work advises that there are a number of diverse possible reasons for heterogeneity, small-study effects or excess of significant biases, and the presence of such biases cannot be determined based only on negative assessments (31, 32, 34, 43, 44, 46, 77). It is also possible that, due to the small number of included studies per MA, the application of such statistical tests is scanty.

It is important to note that the amount of substantial heterogeneity was high, a not unexpected finding, considering the great variability of MMRP components and reported outcomes (7, 18). Similar figures have been reported previously in the psychotherapy field (35, 78) or other medical areas (32, 79). A SR of Cochrane reviews of physiotherapy and occupational therapy, for instance, found that in 52% of these reviews no meta-analysis was performed, mainly due to heterogeneity obstacles (30). In addition, calculation of the 95% prediction intervals, which indicates the possible future treatment effect in an individual study setting (43, 80), revealed that the null value was excluded in only 1 meta-analysis. This may indicate that unexplained sources of heterogeneity remain.

To the best of our knowledge, this umbrella review is the first and the largest comprehensive summary of the published literature regarding MMRPs for common clinically important pain conditions. In addition, this is the first study to assess the existing evidence by applying standardized methodology and state-of-the-art approaches based on rigorous criteria to appraise the results from both MAs and SRs (51). The only published overview of SRs in this field only critically summarized the available evidence (18). Furthermore, the methodological quality of the selected MAs and SRs was assessed in the current study with the AMSTAR tool, which has good reliability, construct validity, and feasibility (38).

Limitations

This study has a number of limitations. As with any umbrella review, no firm conclusions can be reached about the sources of heterogeneity and the other possible biases, i.e. small-study effects or excess of significant findings. Our statistical tests only can offer an indication of their existence and cannot explain their aetiology effectively (44, 46, 77). However, such an

examination was outside of the aims of the current study. One may argue that different lengths of MMRPs may be one of the explanations of the heterogeneity of studies. A previous SR concludes that, in the literature, the relationship between dose of MMRP and outcome effect is limited (29). In addition, the sensitivity analysis did not reveal a common pattern in terms of the credibility of the evidence. The current study also did not evaluate the homogeneity of MAs and SRs in terms of PICO and the limitations in the PICO description. Therefore, this study was limited to providing evidence at a "micro level" perspective in terms of variation within the pain conditions (e.g. definitions), characteristics of patient populations (e.g. co-morbidities), behavioural factors (e.g. smoking), environmental factors (e.g. working status), equity-related factors (e.g. income), treatment characteristics (e.g. education and competence of staff), country-specific factors (e.g. health and social care system), and in the outcome measures. Thus, we cannot exclude the possibility that absence of statistical heterogeneity also means absence of clinical heterogeneity in published MAs. Thus, only when thorough data on PICO of the original studies is available, can a clear decision be made as to whether a MA is justified. Another limitation lies in the fact that some overlap (27 out of 462; 6%), in terms of primary RCTs, mostly in the case of quantitative synthesis, could not be avoided; however, the final set of primary RCTs in each MA was considerably different, thus providing dissimilar summary estimates. A further weakness, which is a common problem in umbrella reviews, is that the results of this study are derived only from published SRs and MAs and, therefore, could have missed some information derived from single RCTs not included in these reviews or from unpublished data. The quality of primary studies included in the SRs and MAs was also not examined, although this is one of the central aims of the original SRs and MAs. Finally, albeit that the methodological quality of the included qualitative SRs and MAs was satisfactory, we did not contact the original authors to elucidate whether particular methodological issues were actually examined; hence, errors may have been introduced.

Future MMRPs should focus on some major methodological issues that appear to challenge the reported evidence. Many RCTs report on several outcomes, which are seldom divided into primary and secondary outcomes, e.g. one Swedish SR (not included here) included an average of 9 outcomes (81). MMRP is a complex treatment with broad goals and as a result, it is highly unlikely that changes in 9 outcomes are independent of each other. The question arises as to how to determine whether positive results are obtained in an RCT of MMRP; evaluating a single outcome at a time, as done here and in most RCTs, SRs and MAs, may not be the most accurate process, since the treatment was not designed to target only a single outcome. Moreover, small changes in 9 outcomes may be more important for the patient than one prominent change in 1 out of 9 outcomes.

This study suggests that, although the exact components of MMRPs are difficult to grasp even in RCTs, a standardized protocol of MMRPs components and outcomes, which could be applied to any MMRP study, might be more usable for making concrete comparisons in future effectiveness studies. Two topical SRs found that the components of the MMPR were described only in general terms, and the outcome domains were measured inconsistently across studies (7, 49); characteristics of MMRPs studies also noted in our evaluation. A further concern applies to the question of whether the patient groups included in different RCTs are indeed comparable; they may have chronic LBP, but the presence of comorbidities and long-term sick leave may be unequal among these patients. Hence, there is a lack of taxonomy of chronic pain patients applicable in clinical settings and in research. The present study also recommends that, notwithstanding the costs, there is a need for more, larger, and better-conducted, RCTs on the effectiveness of MMRPs. An in-depth examination of possible reasons for heterogeneity, including the length of the MMRPs and the homogeneity of PICOs, in future MA may lead to a better understanding of the variations between studies. Finally, data regarding adverse events, and more studies in other pain groups, are also necessary.

Conclusion

The results of this study indicate an absence of strong empirical evidence for MMRPs for common pain conditions. In contrast, the available evidence, although limited, did not manifest a high risk of biased results. Nonetheless, it cannot be ruled out that those biases may be hidden by the small number of studies and small sample sizes. The use of an identical formula for treatment modalities, outcome measures, and length of MMRPs may facilitate comparisons of MMRP effectiveness across future studies. Larger and more rigorous RCTs are, therefore, required.

ACKNOWLEDGEMENTS

Conflicts of interest statement: The authors have no conflicts of interest to declare. BG received a research grant from AFA Insurance; AFA Insurance is a commercial founder, which is owned by Sweden's labour markets parties: the Confederation of the Swedish Enterprise, the Swedish Trade Union Confederation (LO) and The Council for Negotiation and Co-operation (PTK). They insure employees within the private sector, municipalities

and county councils. AFA Insurance do not seek to generate a profit, which implies that no dividends are paid to the shareholders. AFA Insurance had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript

REFERENCES

1. Feuerstein M, Menz L, Zastowny T, Barron BA. Chronic back pain and work disability: Vocational outcomes following multidisciplinary rehabilitation. J Occup Rehabil 1994; 4: 229–251.
2. Hauser W, Bernardy K, Arnold B, Offenbacher M, Schiltenwolf M. Efficacy of multicomponent treatment in fibromyalgia syndrome: a meta-analysis of randomized controlled clinical trials. Arthritis Rheum 2009; 61: 216–224.
3. Henschke N, Ostelo RW, van Tulder MW, Vlaeyen JW, Morley S, Assendelft WJ, et al. Behavioural treatment for chronic low-back pain. Cochrane Database Syst Rev 2010; 7: CD002014.
4. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev 2014; 9: CD000963.
5. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. BMJ 2015; 350: h444.
6. O'Keeffe M, Purtill H, Kennedy N, Conneely M, Hurley J, O'Sullivan P, 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: 755–774.
7. Scascighini L, Toma V, Dober-Spielmann S, Sprott H. Multidisciplinary treatment for chronic pain: a systematic review of interventions and outcomes. Rheumatology (Oxford) 2008; 47: 670–678.
8. Steffens D, Maher CG, Pereira LS, Stevens ML, Oliveira VC, Chapple M, et al. Prevention of low back pain: a systematic review and meta-analysis. JAMA Intern Med 2016; 176: 199–208.
9. Sutton DA, Cote P, Wong JJ, Varatharajan S, Randhawa KA, Yu H, et al. Is multimodal care effective for the management of patients with whiplash-associated disorders or neck pain and associated disorders? A systematic review by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. Spine J 2016; 16: 1541–1565.
10. Turk DC, Wilson HD, Cahana A. Treatment of chronic non-cancer pain. Lancet 2011; 377: 2226–2235.
11. Livshits G, Popham M, Malkin I, Sambrook PN, Macgregor AJ, Spector T, et al. Lumbar disc degeneration and genetic factors are the main risk factors for low back pain in women: the UK Twin Spine Study. Ann Rheum Dis 2011; 70: 1740–1745.
12. Fejer R, Kyvik KO, Hartvigsen J. The prevalence of neck pain in the world population: a systematic critical review of the literature. Eur Spine J 2006; 15: 834–848.
13. Kyhlback M, Thierfelder T, Soderlund A. Prognostic factors in whiplash-associated disorders. Int J Rehabil Res 2002; 25: 181–187.
14. Heidari F, Afshari M, Moosazadeh M. Prevalence of fibromyalgia in general population and patients, a systematic review and meta-analysis. Rheumatol Int 2017; 37: 1527–1539.
15. Mansfield KE, Sim J, Jordan JL, Jordan KP. A systematic review and meta-analysis of the prevalence of chronic widespread pain in the general population. Pain 2016; 157: 55–64.
16. Nahin RL. Estimates of pain prevalence and severity in adults: United States, 2012. J Pain 2015; 16: 769–80.
17. Norlund A, Ropponen A, Alexanderson K. Multidisciplinary interventions: review of studies of return to work after rehabilitation for low back pain. J Rehabil Med 2009; 41: 115–121.
18. Momsen AM, Rasmussen JO, Nielsen CV, Iversen MD, Lund H. Multidisciplinary team care in rehabilitation: an overview of reviews. J Rehabil Med 2012; 44: 901–912.
19. van Geen JW, Edelaar MJ, Janssen M, van Eijk JT. The long-term effect of multidisciplinary back training: a systematic review. Spine (Phila Pa 1976) 2007; 32: 249–255.
20. Guzman J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary rehabilitation for chronic low back pain: systematic review. BMJ 2001; 322: 1511–1516.
21. Guzman J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary bio-psycho-social rehabilitation for chronic low back pain. Cochrane Database Syst Rev 2002; 1: CD000963.
22. van Middelkoop M, Rubinstein SM, Kuijpers T, Verhagen AP, Ostelo R, Koes BW, et al. A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. Eur Spine J 2011; 20: 19–39.
23. Marin TJ, Van Eerd D, Irvin E, Couban R, Koes BW, Malmivaara A, et al. Multidisciplinary biopsychosocial rehabilitation for subacute low back pain. Cochrane Database Syst Rev 2017; 6: CD002193.
24. Teasell RW, McClure JA, Walton D, Pretty J, Salter K, Meyer M, et al. A research synthesis of therapeutic interventions for whiplash-associated disorder (WAD): part 4 – noninvasive interventions for chronic WAD. Pain Res Manag 2010; 15: 313–322.
25. Teasell RW, McClure JA, Walton D, Pretty J, Salter K, Meyer M, et al. A research synthesis of therapeutic interventions for whiplash-associated disorder (WAD): part 3 - interventions for subacute WAD. Pain Res Manag 2010; 15: 305–312.
26. Burckhardt CS. Multidisciplinary approaches for management of fibromyalgia. Curr Pharm Des 2006; 12: 59–66.
27. van Koulil S, Effting M, Kraaimaat FW, van Lankveld W, van Helmond T, Cats H, et al. Cognitive-behavioural therapies and exercise programmes for patients with fibromyalgia: state of the art and future directions. Ann Rheum Dis 2007; 66: 571–581.
28. Schatman M. Interdisciplinary Chronic Pain Management: International Perspectives. Pain: Clinical Updates. 2012; 20: 1–6. Available from: https: //www.iasp-pain.org/PublicationsNews/NewsletterIssue.aspx?ItemNumber=2065
29. Waterschoot FP, Dijkstra PU, Hollak N, de Vries HJ, Geertzen JH, Reneman MF. Dose or content? Effectiveness of pain rehabilitation programs for patients with chronic low back pain: a systematic review. Pain 2014; 155: 179–189.
30. van den Ende CH, Steultjens EM, Bouter LM, Dekker J. Clinical heterogeneity was a common problem in Cochrane reviews of physiotherapy and occupational therapy. J Clin Epidemiol 2006; 59: 914–919.
31. Bortolato B, Kohler CA, Evangelou E, Leon-Caballero J, Solmi M, Stubbs B, et al. Systematic assessment of environmental risk factors for bipolar disorder: an umbrella review of systematic reviews and meta-analyses. Bipolar Disord 2017; 19: 84–96.
32. Bellou V, Belbasis L, Tzoulaki I, Evangelou E, Ioannidis JP. Environmental risk factors and Parkinson's disease: an umbrella review of meta-analyses. Parkinsonism Relat Disord 2016; 23: 1–9.
33. Dragioti E, Dimoliatis I, Evangelou E. Disclosure of researcher allegiance in meta-analyses and randomised controlled trials of psychotherapy: a systematic appraisal. BMJ Open 2015; 5: e007206.
34. Tzoulaki I, Siontis KC, Evangelou E, Ioannidis JP. Bias in associations of emerging biomarkers with cardiovascular disease. JAMA Intern Med 2013; 173: 664–671.
35. Dragioti E, Karathanos V, Gerdle B, Evangelou E. Does psychotherapy work? An umbrella review of meta-analyses of randomized controlled trials. Acta Psychiatr Scand 2017; 136: 236–246.
36. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc 2015; 13: 132–140.
37. Ioannidis J. Next-generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. Br J Sports Med 2017; 51: 1456–1458.
38. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009; 62: 1013–1020.

39. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7: 177–188.
40. Egger M, Smith GD, Altman DG. Systematic reviews in health care: meta-analysis in context. 2nd edn. London: BMJ Books; 2001.
41. Cochran WG. The combination of estimates from different experiments. Biometrics 1954; 10: 101.
42. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002; 21: 1539–1558.
43. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011; 342: d549.
44. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997; 315: 629–634.
45. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011; 343: d4002.
46. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials 2007; 4: 245–253.
47. Lubin JH, Gail MH. On power and sample size for studying features of the relative odds of disease. Am J Epidemiol 1990; 131: 552–566.
48. Ioannidis JPA. Clarifications on the application and interpretation of the test for excess significance and its extensions. J Mathemat Psychol 2013; 57: 184–187.
49. Deckert S, Kaiser U, Kopkow C, Trautmann F, Sabatowski R, Schmitt J. A systematic review of the outcomes reported in multimodal pain therapy for chronic pain. Eur J Pain 2016; 20: 51–63.
50. StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP. 2013.
51. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ 2009; 181: 488–493.
52. van Tulder MW, Koes BW, Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain. A systematic review of randomized controlled trials of the most common interventions. Spine (Phila Pa 1976) 1997; 22: 2128–2156.
53. Brady B, Veljanova I, Chipchase L. Are multidisciplinary interventions multicultural? A topical review of the pain literature as it relates to culturally diverse patient groups. Pain 2016; 157: 321–328.
54. Monticone M, Cedraschi C, Ambrosini E, Rocca B, Fiorentini R, Restelli M, et al. Cognitive-behavioural treatment for subacute and chronic neck pain. Cochrane Database Syst Rev 2015; 5: CD010664.
55. Schaafsma FG, Whelan K, van der Beek AJ, van der Es-Lambeek LC, Ojajarvi A, Verbeek JH. Physical conditioning as part of a return to work strategy to reduce sickness absence for workers with back pain. Cochrane Database Syst Rev 2013; 8: CD001822.
56. Schaafsma F, Schonstein E, Whelan KM, Ulvestad E, Kenny DT, Verbeek JH. Physical conditioning programs for improving work outcomes in workers with back pain. Cochrane Database Syst Rev 2010; 1: CD001822.
57. Ravenek MJ, Hughes ID, Ivanovich N, Tyrer K, Desrochers C, Klinger L, et al. A systematic review of multidisciplinary outcomes in the management of chronic low back pain. Work 2010; 35: 349–367.
58. Sarzi-Puttini P, Buskila D, Carrabba M, Doria A, Atzeni F. Treatment strategy in fibromyalgia syndrome: where are we now? Semin Arthritis Rheum 2008; 37: 353–365.
59. Hoffman BM, Papas RK, Chatkoff DK, Kerns RD. Meta-analysis of psychological interventions for chronic low back pain. Health Psychol 2007; 26: 1–9.
60. Tveito TH, Hysing M, Eriksen HR. Low back pain interventions at the workplace: a systematic literature review. Occup Med (Lond) 2004; 54: 3–13.
61. Karjalainen K, Malmivaara A, van Tulder M, Roine R, Jauhiainen M, Hurri H, et al. Multidisciplinary biopsychosocial rehabilitation for subacute low back pain among working age adults. Cochrane Database Syst Rev2003; 2: CD002193.
62. Karjalainen K, Malmivaara A, van Tulder M, Roine R, Jauhiainen M, Hurri H, et al. Multidisciplinary biopsychosocial rehabilitation for neck and shoulder pain among working age adults. Cochrane Database Syst Rev 2003; 2: CD002194.
63. Schonstein E, Kenny DT, Keating J, Koes BW. Work conditioning, work hardening and functional restoration for workers with back and neck pain. Cochrane Database Syst Rev 2003; 1: CD001822.
64. Schonstein E, Kenny D, Keating J, Koes B, Herbert RD. Physical conditioning programs for workers with back and neck pain: a cochrane systematic review. Spine (Phila Pa 1976) 2003; 28: 391–395.
65. Karjalainen K, Malmivaara A, van Tulder M, Roine R, Jauhiainen M, Hurri H, et al. Multidisciplinary biopsychosocial rehabilitation for subacute low back pain in working-age adults: a systematic review within the framework of the Cochrane Collaboration Back Review Group. Spine (Phila Pa 1976) 2001; 26: 262–269.
66. Peeters GG, Verhagen AP, de Bie RA, Oostendorp RA. The efficacy of conservative treatment in patients with whiplash injury: a systematic review of clinical trials. Spine (Phila Pa 1976) 2001; 26: 64–73.
67. Karjalainen K, Malmivaara A, van Tulder M, Roine R, Jauhiainen M, Hurri H, et al. Multidisciplinary biopsychosocial rehabilitation for subacute low back pain among working age adults. Cochrane Database Syst Rev 2000; 3: CD002193.
68. Karjalainen K, Malmivaara A, van Tulder M, Roine R, Jauhiainen M, Hurri H, et al. Multidisciplinary rehabilitation for fibromyalgia and musculoskeletal pain in working age adults. Cochrane Database Syst Rev 2000; 2: CD001984.
69. Karjalainen K, Malmivaara A, van Tulder M, Roine R, Jauhiainen M, Hurri H, et al. Multidisciplinary biopsychosocial rehabilitation for neck and shoulder pain among working age adults: a systematic review within the framework of the Cochrane Collaboration Back Review Group. Spine (Phila Pa 1976) 2001; 26: 174–181.
70. Jensen SA, Corralejo SM. Measurement issues: large effect sizes do not mean most people get better – clinical significance and the importance of individual results. Child Adolesc Mental Health 2017; 22: 163–166.
71. Thomsen AB, Sorensen J, Sjogren P, Eriksen J. Economic evaluation of multidisciplinary pain management in chronic pain patients: a qualitative systematic review. J Pain Symptom Manage 2001; 22: 688–698.
72. Chou R, Deyo R, Friedly J, Skelly A, Hashimoto R, Weimer M, et al. Nonpharmacologic therapies for low back pain: a systematic review for an American College of Physicians Clinical Practice Guideline. Ann Intern Med 2017; 166: 493–505.
73. Xin Z, Xue-Ting L, De-Ying K. GRADE in systematic reviews of acupuncture for stroke rehabilitation: recommendations based on high-quality evidence. Sci Rep 2015; 5: 16582.
74. Tsilidis KK, Papatheodorou SI, Evangelou E, Ioannidis JP. Evaluation of excess statistical significance in meta-analyses of 98 biomarker associations with cancer risk. J Natl Cancer Inst 2012; 104: 1867–1878.
75. Pollock A, Farmer SE, Brady MC, Langhorne P, Mead GE, Mehrholz J, et al. An algorithm was developed to assign GRADE levels of evidence to comparisons within systematic reviews. J Clin Epidemiol 2016; 70: 106–110.
76. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PLoS Biol 2015; 13: e1002106.
77. Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. BMJ 2007; 335: 914–916.
78. Markozannes G, Aretouli E, Rintou E, Dragioti E, Damigos D, Ntzani E, et al. An umbrella review of the literature on the effectiveness of psychological interventions for pain reduction. BMC Psychol 2017; 5: 31.
79. Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, Howells DW, et al. Evaluation of excess significance bias in animal studies of neurological diseases. PLoS Biol 2013; 11: e1001609.
80. Ades AE, Lu G, Higgins JP. The interpretation of random-effects meta-analysis in decision models. Med Decis Making 2005; 25: 646–654.
81. Swedish Council on Health Technology Assessment (SBU). Metoder för behandling av långvarig smärta- en systematisk litteraturöversikt. [Methods for treatment of chronic pain – a systematic review of the literature]. Stockholm: SBU; 2006 (in Swedish).