Which specific modes of exercise training are most effective for treating low back pain? Network meta-analysis

Patrick J Owen 1,1, Clint T Miller 1,1, Niamh L Mundell 1,1, Simone J J M Verswijveren 1,1, Scott D Tagliaferri 1,1, Steven J Bowe 3, Daniel L Belavy 1

ABSTRACT

Objective Examine the effectiveness of specific modes of exercise training in non-specific chronic low back pain (NSCLBP).

Design Network meta-analysis (NMA).

Data sources MEDLINE, CINAHL, SPORTDiscus, EMBASE, CENTRAL.

Eligibility criteria Exercise training randomised controlled/trial in adults with NSCLBP.

Results Among 9543 records, 89 studies (patients=5578) were eligible for qualitative synthesis and 70 (pain), 63 (physical function), 16 (mental health) and 4 (trunk muscle strength) for NMA. The NMA consistency model revealed that the following exercise training modalities had the highest probability (surface under the cumulative ranking (SUCRA)) of being best when compared with true control: Pilates for pain (SUCRA=100%; pooled standardised mean difference (95% CI): −1.86 (−2.54 to −1.19)), resistance (SUCRA=80%; −1.14 (−1.71 to −0.56)) and stabilisation/motor control (SUCRA=80%; −1.13 (−1.53 to −0.74)) for physical function and resistance (SUCRA=80%; −1.26 (−2.10 to −0.41)) and aerobic (SUCRA=80%; −1.18 (−2.20 to −0.15)) for mental health. True control was most likely (SUCRA=10%) to be the worst treatment for all outcomes, followed by therapist hands-off control for pain (SUCRA=10%; 0.09 (−0.71 to 0.89)) and physical function (SUCRA=20%; −0.31 (−0.94 to 0.32)) and therapist hands-on control for mental health (SUCRA=20%; −0.31 (−1.31 to 0.70)). Stretching and McKenzie exercise effect sizes did not differ to true control for pain or function (p>0.095; SUCRA<40%). NMA was not possible for trunk muscle endurance or analgesic medication. The quality of the synthesised evidence was low according to Grading of Recommendations Assessment, Development and Evaluation criteria.

Summary/conclusion There is low quality evidence that Pilates, stabilisation/motor control, resistance training and aerobic exercise training are the most effective treatments, pending outcome of interest, for adults with NSCLBP. Exercise training may also be more effective than therapist hands-on treatment. Heterogeneity among studies and the fact that there are few studies with low risk of bias are both limitations.

INTRODUCTION

Low back pain is the leading cause of disability1 and the most common of all non-communicable diseases.2 Chronic low back pain (CLBP) is pain lasting 12 weeks or longer,3 localised below the costal margin and above the inferior gluteal folds, with or without leg pain.4 While CLBP makes up approximately 20%5 of all low back pain cases, it generates approximately 80% of the direct costs of low back pain.6 In up to 90% of patients with CLBP, clinicians cannot make a specific diagnosis and therefore patients are classified as having ‘non-specific’ CLBP.7 There is a need to identify and evaluate the efficacy of interventions capable of treating non-specific CLBP.

Previous pairwise meta-analyses have shown that passive treatments such as ultrasound,8 hot and cold therapy9 and massage without exercise training10 failed to reduce pain in adults with non-specific CLBP. In contrast, exercise training has collectively been shown to be effective in reducing pain when combined with non-exercise training-based treatments in adults.11–13 Similarly, pairwise meta-analyses examining specific kinds of exercise have shown that Pilates,14 stabilisation/motor control15 16 and yoga17 may reduce pain when compared with non-exercise types of exercise training10 failed to reduce pain in adults with non-specific CLBP. Exercise training has collectively been shown to be effective in reducing pain when compared with non-exercise training-based treatments in adults.11–13 Similarly, pairwise meta-analyses examining specific kinds of exercise have shown that Pilates,14 stabilisation/motor control15 16 and yoga17 may reduce pain when compared with non-exercise training comparators. However, whether specific types of exercise training are more effective in non-specific CLBP has received limited attention. One prior pairwise meta-analysis explored this question and concluded that resistance and stabilisation/motor control exercise training were effective, compared with true control (ie, no intervention), whereas aerobic and combined modalities (ie, programmes including multiple types of exercise such as aerobic, resistance and stretching) were not.12

Pairwise meta-analyses rely on the included randomised controlled/trial in adults with NSCLBP. Exercise training may also be more effective than therapist hands-on treatment. Heterogeneity among studies and the fact that there are few studies with low risk of bias are both limitations.

In a network meta-analysis (NMA), we can include...
studies that tested two or more kinds of treatment, without a control group. This enables direct comparisons of treatments (similar to pairwise meta-analyses) and also permits indirect comparisons of treatments via the network of treatments. This enables researchers to rank interventions as comparatively more or less effective.

We aimed to conduct a systematic review and NMA on the effectiveness of specific kinds of exercise training in adults with non-specific CLBP. In addition to studying the efficacy of various treatments for pain, we also examined treatment effects on subjective physical function, mental health, analgesic pharmacotherapy use as well as objective trunk muscle strength and endurance. We aimed to compare exercise training with non-exercise treatments such as treatment where the therapist uses ‘hands-on’ treatment (eg, manual therapy) and ‘hands-off’ treatment (eg, education, general practitioner management).

### METHODS

This review was conducted in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA-NMA). and was registered with PROSPERO (CRD42017068668).

#### Search strategy

The search strategy was developed, piloted and refined based on the method guidelines for systematic reviews in the Cochrane Back and Neck Group, and previously published systematic reviews. An electronic search of MEDLINE, CINAHL, SPORTDiscus, EMBASE and CENTRAL was conducted for research published between journal inception to May 2019 using Medical Subject Headings (MeSH) for ‘pain’ and ‘exercise’ search terms in online supplementary tables 1 and 2. ‘Pain’ and ‘Exercise’ search terms were combined with ‘AND’ and search in ‘All Fields’ with the following limits: MEDLINE (All Adult: 19+ years; RCT, Human), CINAHL (Exclude MEDLINE records; Human, RCTs; Journal Article; All Adult), SPORTDiscus (Academic Journal), EMBASE (RCT; Not MEDLINE; Adult; Article) and CENTRAL (Trials). Additional searches included reviewing the reference lists of previously published systematic reviews identified via the Cochrane Database of Systematic Reviews (search terms: chronic back pain exercise; limits: none) and GoogleScholar (search terms: systematic review chronic back pain exercise; limits: previous 10 years). All results of the search were screened by PJO to exclude duplicates. The titles and abstracts of the remaining studies were independently screened by PJO and SJMV against inclusion and exclusion criteria. The full texts of those that met these criteria were further independently screened by PJO and SJMV. All disagreements were adjudicated by NLM.

#### Inclusion and exclusion criteria

For inclusion, studies were required to be published in a peer-reviewed journal (ie, grey literature excluded) in any language. All other inclusion criteria followed the Participants, Interventions, Comparators, Outcomes and Study design framework. The population group of interest were adults (≥18 years) with non-specific (no known specific pathology) chronic (≥12 weeks) low back pain (localised below the costal margin and above the inferior gluteal folds, with or without leg pain). Therefore, studies were excluded if they examined pain due to or associated with pregnancy, infection, tumour, osteoporosis, fracture, structural deformity (eg, scoliosis), inflammatory disorder, radicular syndrome or cauda equine syndrome. Studies that solely recruited patients presurgery or postsurgery were also excluded, as were those that included patients with recurrent (pain-free periods of at least 6 months) low back pain. Relevant interventions included the prescription of exercise training alone, without the addition of other treatments (eg, massage, ultrasound or hot and cold therapy) for at least 4 weeks of duration. Specific types of exercise training were determined based on group names chosen by authors and definitions presented in table 1. Non-exercise training treatment comparator groups included true control, therapist hands-on control and therapist hands-off control (table 1). Studies were required to include at least one of the outcome measures of interest: subjective pain intensity (eg, visual analogue scale), subjective physical function (eg, Oswestry Disability Index), objective trunk muscle strength (eg, lumbar extension one-repetition maximum), objective

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
<td></td>
</tr>
<tr>
<td>Resistance</td>
<td>Exercise training designed to improve the strength, power, endurance and size of skeletal muscles</td>
</tr>
<tr>
<td>Stabilisation/motor control</td>
<td>Exercise training targeting specific trunk muscles in order to improve control and coordination of the spine and pelvis</td>
</tr>
<tr>
<td>Pilates</td>
<td>Exercise training following traditional Pilates principles such as centring, concentration, control, precision, flow and breathing</td>
</tr>
<tr>
<td>Yoga</td>
<td>Exercise training following traditional yoga principles with a physical component</td>
</tr>
<tr>
<td>McKenzie</td>
<td>Exercise training following traditional McKenzie principles such as repeated passive spine movements and sustained positions performed in specific directions</td>
</tr>
<tr>
<td>Flexion</td>
<td>Exercise training consisting of controlled movements in flexion only</td>
</tr>
<tr>
<td>Aerobic</td>
<td>Exercises such as walking, cycling and jogging in any land-based mode that is designed to improve the efficiency and capacity of the cardiorespiratory system</td>
</tr>
<tr>
<td>Water-based</td>
<td>Exercise training performed in deep or shallow water</td>
</tr>
<tr>
<td>Stretching</td>
<td>Exercise training including muscle lengthening using any of the following methods: passive, static, isometric, ballistic or proprioceptive neuromuscular facilitation</td>
</tr>
<tr>
<td>Other</td>
<td>Exercise training that does not meet any of the specific types of exercise training mentioned above</td>
</tr>
<tr>
<td>Multimodal</td>
<td>Two or more of the specific types of exercise training mentioned above (not deemed multimodal if only part of warm up or cool down)</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>No intervention provided</td>
</tr>
<tr>
<td>Therapist hands-on</td>
<td>Treatment including manual therapy, chiropractic, passive physiotherapy, osteopathic, massage or acupuncture</td>
</tr>
<tr>
<td>Therapist hands-off</td>
<td>Treatment including general practitioner management, education or psychological interventions</td>
</tr>
</tbody>
</table>

---

2 of 12  
trunk muscle endurance (eg, static lumbar extension hold time), subjective analgesic pharmacotherapy use (eg, prescription medication use) or subjective mental health (eg, 36-Item Short Form Health Survey). In terms of study design, those included were parallel arm (individual-designed or cluster-designed) RCTs that compared an exercise training intervention with either a non-exercise training intervention (including true control) for pairwise comparison or another exercise training intervention for NMA with a total sample size of ≥20 patients (to reduce the risk of publication bias affecting the results).

**Data extraction**
Relevant publication information (ie, author, title, year, journal), study design (eg, two-arm or multi-arm parallel trial, number of assessment time points), number of patients, patient characteristics (eg, age and sex), interventions considered (table 1) and outcome measures (ie, any measure of pain, physical function, muscle strength, muscle endurance, analgesic pharmacotherapy use or mental health) were extracted by two independent assessors (PJO and ST). Extracted outcome data were preintervention and postintervention mean and SD. Data presented as medians or alternate measures of spread were converted to mean and SD.23–25 When only figures were presented (rather than numerical data within text), data were extracted using ImageJ (V1.50i, https://imagej.nih.gov/ij/) to measure the length (in pixels) of the axes to calibrate and then the length in pixels from the relevant axis to the data points of interest.26 This method was used for seven studies.27–32 When it was not possible to extract the required data, this information was requested from the authors a minimum of three times over a 4-week period. The authors of 24 studies were contacted.37 28 30 32–35 and 28 33 35 37 39 41 42 48 49 were able to supply the requested information. Similarity between extracted data from the two independent assessors (PJO and ST) was assessed via an automated code written by DB (written in the ‘R’ statistical environment V3.4.2, www.r-project.org). Any discrepancies were discussed by PJO and ST with disagreements adjudicated by NLM. Prior to commencing data extraction, this method was piloted on 10 studies chosen at random.

**Risk of bias assessment and GRADE**
Risk of bias for each individual study was assessed independently by PJO and ST using the Cochrane Collaboration Risk of Bias Tool,53 which examined potential selection bias (random sequence generation and allocation concealment), performance bias (blinding of patients and personnel), detection bias (blinding of outcome assessment), attrition bias (incomplete outcome data), reporting bias (selective outcome reporting) and other bias. For each source of bias, studies were classified as having a low, high or unclear (if reporting was not sufficient to assess a particular domain) risk. For example, studies that used a random approach to treatment allocation (eg, random number generator) were classified as low risk for this component of selection bias assessment, while those that did not use a random approach (eg, date of birth assignment) were classified as high risk. As this study involved exercise training interventions, it was not possible to blind patients to treatment allocation; thus, patient blinding was deemed as a high risk of bias for all studies and was not included in the overall risk of bias assessment of each study. All risk of bias disagreements were adjudicated by NLM. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach was used to assess the quality of the evidence behind the ranking of treatments from NMA.54

**Statistical analysis**
NMA was performed in accordance with current PRISMA NMA guidelines.58 The following steps were applied to all network meta-analyses. Step 1: A network geometry was created to explore comparative relationships among exercises and non-exercise interventions. Step 2: Consistency, whereby the treatment effects estimated from direct comparisons are consistent with those estimated from indirect comparisons, was assessed for the NMA by fitting both a consistency and inconsistency NMA and considering the results from the Wald test for inconsistency. As this test has low power, side-splitting was also conducted to further assess inconsistency. It was anticipated that there would be heterogeneity between studies; therefore, random effects meta-analysis was used. Step 3: If studies investigated multiple groups which we defined as being the same specific type of exercise training or non-exercise training intervention (eg, when exercise training load was examined), the data from the intervention groups were pooled. A minimum of three studies needed to assess an intervention type for it to be included in meta-analysis. When studies were reverse scaled (ie, higher values indicated better outcomes rather than lower values), the mean in each group was multiplied by −1 as recommended in the Cochrane Handbook.59 As all of the outcomes of interest were continuous or ordinal, but could be measured on different scales, standardised mean difference (SMD) was used as the effect estimate. Step 4: Once comparative effectiveness of the interventions was evaluated, interventions were ranked to identify superiority of the interventions. Two approaches to determine the rank order of interventions are surface under cumulative ranking (SUCRA) and the probability of being the best intervention. Superiority was considered as exercise interventions were more effective than no-intervention. Stata reports both the probabilities (ordered from best to worse) and SUCRA; however, SUCRA is considered the more precise estimation of cumulative ranking probabilities.18 55 SUCRA reports the overall probability, based on the ranking of all interventions that a given intervention is among the best treatments.55 Step 5: The GRADE approach was used to evaluate the quality of evidence from NMA.54 To further assess the transitivity assumption, preintervention pain and disability were considered as potential effect modifiers: the degree of baseline pain and disability have been identified16–59 as predictors of prognosis and treatment outcome in non-specific CLBP whereas fear of movement,60 demographic and physical variable,59 61 62 are not consistently associated with treatment outcomes or prognosis. As it is to be expected that not one individual exercise mode will be the best treatment for non-specific CLBP, for the imprecision criterion we did not require that only one exercise approach be clearly the sole best treatment. To check for the presence of bias due to small scale studies, which may lead to publication bias in NMA, a network funnel plot was generated and visually inspected using the criterion of symmetry. Network meta-analyses were conducted in Stata 15.0 (Stata, College Station, Texas, USA). Pairwise random-effects meta-analysis was also conducted to compare exercise training-based and non-exercise training-based treatments. Heterogeneity was assessed for all pairwise comparisons using the I² statistic and publication bias using the p-value of Egger’s test. Pairwise meta-analyses were implemented in the ‘R’ statistical environment (V3.4.0, www.r-project.org).

**RESULTS**
The flow of the systematic review is presented in figure 1. The electronic database search yielded 9437 records after duplicates
we were removed. Additionally, 106 records were located via the reference lists of 18 previously published systematic reviews.11 12 14–17 63–74  The examination of titles and abstracts resulted in the retrieval of 808 full-text records. Following full-text review, 89 studies were included in qualitative analysis.23–25 27–33 35 37 38 41 47 49 75–140 Of these studies, 55 (62%)24 25 27–29 31 32 35 37 41 49 73–75 140 were deemed eligible for the pairwise meta-analyses and 82 (92%)23–25 27–33 35 37 38 40 41 43 45–47 49–51 73–140 were, in total, suitable for the network meta-analyses.

Study characteristics
A detailed summary of each included study (n=89)23–25 27–33 35 37 38 40 41 43 45–47 49–51 73–140 is presented in online supplementary table 3. Sample size ranged from 20 to 240 patients (mean age range, 20–70 years; one study did not report age) and study duration ranged from 4 to 24 weeks. Eleven studies (12%)25 28 30 50 95 105 106 110 121 123 and 16 studies (18%)20 31 38 43 47 49 80 93 94 116 118 119 122 124 135 139 only included females and males, respectively, whereas the remainder included both sexes. Twelve studies (13%) did not report sex.40 47 73 76 82 89 97 103 104 111 115 136

The average duration of pain reported at baseline ranged from 12 to 725 weeks (0.2–14.0 years) and average pain intensity was 21–79 points when normalised on a 100-point scale. Notably, 13 studies (15%)29 34 45 50 93 95 98 120 122 124 126 127 139 did not report a measure of baseline pain intensity.

Among the 89 studies included (patients: n=5578), there were 131 exercise training interventions (patients: n=3924) and 59 non-exercise training comparators (patients: n=1654; online supplementary table 3). Exercise training interventions included: aerobic (studies: n=5, patients: n=127),33 83 100 113 127 multi-modal (studies: n=23, patients: n=756),23–25 32 38 40 45 50 75 76 82 88 91 99 102 109 114 118 124 129 131 134 137 Pilates (studies: n=13, patients: n=350),3 27 29 30 33 34 51 86 94 105 111 139 resistance (studies: n=12, patients: n=472),43 78 79 81 90 92 93 100 101 127 130 stabilisation/motor control (studies: n=39, patients: n=1062),25 28 38 43 49 50 73 76 82 84 85 89–91 93–95 103–106 108–110 115 118 122–125 128 129 134 136 water-based (studies: n=6, patients: n=144)37 41 45 88 120 137 and yoga (studies: n=6, patients: n=387).35 46 114 126 132 133 Non-exercise training interventions included: true control (studies: n=33, patients: n=733),27 28 31 34 35 37 41 49 77 78 80–83 92–95 98–108 110–113 115–117 119–121 123 125 126 128 130 132 133 135 138–140 therapist hands-off control (studies: n=14, patients: n=506)25 30 47 81 83 92 98 99 101 106 113 123 128 130 and therapist hands-on control (studies: n=12, patients: n=413).24 29 32 46 51 84 102 107 119 126 132 133 The frequency of exercise training per week ranged from 1 to 7 days, whereas therapist hands-on controls were performed on 0.3–5 days per week. The risk of bias assessment for each individual study is presented in online supplementary table 4 and summary data in figure 2. Overall, studies tended to exhibit a low risk of selective outcome reporting (60%), random sequence generation (60%), other bias (57%), but not allocation concealment (29%), blinding of patients and personnel (0%), blinding of outcome assessment (42%) and incomplete outcome data (34%).
Figure 3  Network meta-analysis maps of studies examining the efficacy of exercise training in patients with non-specific chronic low back pain on pain, physical function, mental health and muscle strength. CON: non-exercise control, INT: exercise training intervention. The size of the nodes relates to the number of participants in that intervention type and the thickness of lines between interventions relates to the number of studies for that comparison.

Pain
Seventy studies assessed pain and were eligible for NMA of pain (figure 3). The results from the consistency NMA provided evidence that when compared with true control, Pilates (p<0.001), aerobic (p=0.006), stabilisation/motor control (p<0.001), multimodal (p<0.001), resistance (p=0.002) and ‘other’ (p<0.001) exercise training all result in lower pain following intervention (table 2). The results of the consistency model indicated that Pilates (SUCRA: 100%), aerobic (SUCRA: 80%) and stabilisation/motor control (80%) exercise training were among the best interventions for pain. True control (SUCRA: 10%) and therapist hands-off control (SUCRA: 10%) were most likely to be the least effective. The Wald test for inconsistency in the network was not significant ($\chi^2=40.7$, p=0.057; see also online supplementary table 5). The comparison-adjusted funnel plot did not provide evidence for apparent publication bias (online supplementary figure 1). Data from pairwise meta-analysis gave evidence for considerable heterogeneity (min $I^2=29\%$, median $I^2=90\%$, max $I^2=95\%$; online supplementary table 6). The quality of the evidence for the ranks of the treatment was low (table 3). Pairwise meta-analysis provided evidence that exercise training (all) was more effective than therapist hands-off control (SMD (95% CI): −1.06 (−1.62 to −0.51), p<0.001, $I^2=87\%$, studies: n=8) and therapist hands-on control (SMD (95% CI): −0.57 (−1.05 to −0.08), p=0.023, $I^2=93\%$, studies: n=11; online supplementary table 6).

Physical function
Sixty-three studies were eligible for the NMA of subjective physical function (figure 3). The results from the consistency NMA provide evidence that when compared with true control, stabilisation/motor control (p<0.001), resistance (p<0.001), water-based (p=0.004), Pilates (p=0.001), yoga (p=0.015), multimodal (p=0.002), aerobic (p=0.029) and ‘other’ (p=0.017) exercise training resulted in improved physical function following intervention (table 2). The results showed that stabilisation/motor control (SUCRA: 80%) and resistance (SUCRA: 80%) exercise training had the highest probability of being the best treatments, followed by water-based (SUCRA: 79%), Pilates (SUCRA: 79%) and yoga (SUCRA: 70%) exercise training. True control (SUCRA: 0%) and therapist hands-off control (SUCRA: 20%) were most likely to be the least effective. There was evidence of inconsistency in the network ($\chi^2=46.3,$
Table 2  Network meta-analysis consistency models for pain, physical function, mental health and muscle strength in studies examining the efficacy of exercise training in patients with non-specific chronic low back pain

<table>
<thead>
<tr>
<th>Treatment type</th>
<th>Comparison to ‘no treatment’</th>
<th>Likelihood (%) of being…</th>
<th>Best</th>
<th>Worst</th>
<th>SUCRA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON: True control</td>
<td>–</td>
<td>–</td>
<td>0.0</td>
<td>37.0</td>
<td>10</td>
</tr>
<tr>
<td>INT: Aerobic</td>
<td>-1.41 (–2.43 to –0.40)</td>
<td>0.006</td>
<td>16.6</td>
<td>0.2</td>
<td>80</td>
</tr>
<tr>
<td>INT: Other</td>
<td>-1.07 (–1.67 to –0.47)</td>
<td>&lt;0.001</td>
<td>1.3</td>
<td>0.0</td>
<td>60</td>
</tr>
<tr>
<td>CON: Hands-off</td>
<td>0.09 (–0.71 to 0.89)</td>
<td>0.827</td>
<td>0.0</td>
<td>56.6</td>
<td>10</td>
</tr>
<tr>
<td>CON: Hands-on</td>
<td>-0.69 (–1.39 to 0.01)</td>
<td>0.054</td>
<td>0.0</td>
<td>1.1</td>
<td>30</td>
</tr>
<tr>
<td>INT: Water-based</td>
<td>-0.91 (–1.85 to 0.03)</td>
<td>0.057</td>
<td>2.5</td>
<td>1.1</td>
<td>50</td>
</tr>
<tr>
<td>INT: McKenzie</td>
<td>-0.81 (–1.74 to 0.11)</td>
<td>0.083</td>
<td>1.1</td>
<td>1.8</td>
<td>40</td>
</tr>
<tr>
<td>INT: Multimodal</td>
<td>-0.99 (–1.55 to –0.44)</td>
<td>&lt;0.001</td>
<td>0.1</td>
<td>0.0</td>
<td>50</td>
</tr>
<tr>
<td>INT: Pilates</td>
<td>-1.86 (–2.54 to –1.19)</td>
<td>&lt;0.001</td>
<td>68.9</td>
<td>0.0</td>
<td>100</td>
</tr>
<tr>
<td>INT: Resistance</td>
<td>-1.14 (–1.86 to –0.42)</td>
<td>0.002</td>
<td>2.4</td>
<td>0.0</td>
<td>60</td>
</tr>
<tr>
<td>INT: Stabilisation</td>
<td>-1.31 (–1.75 to –0.87)</td>
<td>&lt;0.001</td>
<td>2.0</td>
<td>0.0</td>
<td>80</td>
</tr>
<tr>
<td>INT: Stretching</td>
<td>-0.62 (–1.38 to 0.13)</td>
<td>0.107</td>
<td>0.3</td>
<td>1.7</td>
<td>30</td>
</tr>
<tr>
<td>INT: Yoga</td>
<td>-0.99 (–2.00 to 0.03)</td>
<td>0.056</td>
<td>4.7</td>
<td>0.5</td>
<td>50</td>
</tr>
<tr>
<td><strong>Physical function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON: True control</td>
<td>–</td>
<td>–</td>
<td>0.0</td>
<td>59.4</td>
<td>0</td>
</tr>
<tr>
<td>INT: Aerobic</td>
<td>-0.85 (–1.62 to –0.09)</td>
<td>0.029</td>
<td>9.3</td>
<td>1.0</td>
<td>60</td>
</tr>
<tr>
<td>INT: Other</td>
<td>-0.72 (–1.32 to –0.13)</td>
<td>0.017</td>
<td>2.4</td>
<td>0.5</td>
<td>50</td>
</tr>
<tr>
<td>CON: Hands-off</td>
<td>-0.31 (–0.94 to 0.32)</td>
<td>0.340</td>
<td>0.0</td>
<td>12.7</td>
<td>20</td>
</tr>
<tr>
<td>CON: Hands-on</td>
<td>-0.49 (–1.08 to 0.10)</td>
<td>0.102</td>
<td>0.0</td>
<td>3.2</td>
<td>30</td>
</tr>
<tr>
<td>INT: Water-based</td>
<td>-1.00 (–1.67 to –0.32)</td>
<td>0.004</td>
<td>18.1</td>
<td>0.1</td>
<td>70</td>
</tr>
<tr>
<td>INT: McKenzie</td>
<td>-0.27 (–1.02 to 0.47)</td>
<td>0.471</td>
<td>0.4</td>
<td>20.0</td>
<td>20</td>
</tr>
<tr>
<td>INT: Multimodal</td>
<td>-0.75 (–1.23 to –0.27)</td>
<td>0.002</td>
<td>0.5</td>
<td>0.0</td>
<td>50</td>
</tr>
<tr>
<td>INT: Pilates</td>
<td>-0.92 (–1.48 to –0.36)</td>
<td>0.001</td>
<td>8.0</td>
<td>0.0</td>
<td>70</td>
</tr>
<tr>
<td>INT: Resistance</td>
<td>-1.14 (–1.71 to –0.56)</td>
<td>&lt;0.001</td>
<td>27.5</td>
<td>0.0</td>
<td>80</td>
</tr>
<tr>
<td>INT: Stabilisation</td>
<td>-1.13 (–1.53 to –0.74)</td>
<td>&lt;0.001</td>
<td>18.7</td>
<td>0.0</td>
<td>80</td>
</tr>
<tr>
<td>INT: Stretching</td>
<td>-0.55 (–1.20 to 0.10)</td>
<td>0.095</td>
<td>1.3</td>
<td>2.9</td>
<td>40</td>
</tr>
<tr>
<td>INT: Yoga</td>
<td>-0.92 (–1.66 to –0.18)</td>
<td>0.015</td>
<td>13.9</td>
<td>0.1</td>
<td>70</td>
</tr>
<tr>
<td><strong>Mental health</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON: True control</td>
<td>–</td>
<td>–</td>
<td>0.0</td>
<td>68.1</td>
<td>10</td>
</tr>
<tr>
<td>INT: Aerobic</td>
<td>-1.18 (–2.20 to –0.15)</td>
<td>0.024</td>
<td>32.9</td>
<td>0.2</td>
<td>80</td>
</tr>
<tr>
<td>CON: Hands-on</td>
<td>-0.31 (–1.31 to 0.70)</td>
<td>0.552</td>
<td>0.2</td>
<td>26.6</td>
<td>20</td>
</tr>
<tr>
<td>INT: Multimodal</td>
<td>-0.75 (–1.61 to 0.11)</td>
<td>0.086</td>
<td>8.3</td>
<td>2.8</td>
<td>50</td>
</tr>
<tr>
<td>INT: Pilates</td>
<td>-0.84 (–1.71 to 0.02)</td>
<td>0.056</td>
<td>13.4</td>
<td>1.7</td>
<td>60</td>
</tr>
<tr>
<td>INT: Resistance</td>
<td>-1.26 (–2.10 to –0.41)</td>
<td>0.003</td>
<td>40.0</td>
<td>0.0</td>
<td>80</td>
</tr>
<tr>
<td>INT: Stabilisation</td>
<td>-0.78 (–1.49 to –0.07)</td>
<td>0.031</td>
<td>5.2</td>
<td>0.5</td>
<td>50</td>
</tr>
<tr>
<td><strong>Muscle strength</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CON: True control</td>
<td>–</td>
<td>–</td>
<td>8.0</td>
<td>65.6</td>
<td>20</td>
</tr>
<tr>
<td>INT: Resistance</td>
<td>0.18 (–0.27 to 0.64)</td>
<td>0.427</td>
<td>42.9</td>
<td>17.0</td>
<td>60</td>
</tr>
<tr>
<td>INT: Stabilisation</td>
<td>0.21 (–0.28 to 0.69)</td>
<td>0.399</td>
<td>49.1</td>
<td>17.4</td>
<td>70</td>
</tr>
</tbody>
</table>

The significance of the difference of the effect estimate (SMD) compared with ‘no treatment’ control is indicated by the p value. The likelihood of being the best or worst treatment is indicated as a percentage and this is best summarised by the SUCRA which ranges from 0% to 100%. The SUCRA measure assesses the likely ranks of all treatments and the higher the SUCRA percentage, the more likely that therapy is one of the most effective.

CON, control; INT, exercise training intervention; SMD, standardised mean difference; SUCRA, surface under the cumulative ranking curves.

p=0.049; online supplementary table 5). The comparison-adjusted funnel plot did not provide evidence for apparent publication bias (online supplementary figure 1). Data from pairwise meta-analysis gave evidence for considerable heterogeneity (min I²=0%, median I²=84%, max I²=97%; online supplementary table 7). The quality of the evidence for the ranks of the treatment was low (table 3). Pairwise meta-analysis showed that exercise training (all) was more effective than therapist hands-off control (SMD (95% CI): −0.46 (–0.78 to –0.14), p=0.010, I²=73%, studies: n=9) and therapist hands-on control (SMD (95% CI): −0.55 (–0.94 to –0.15), p=0.010, I²=88%, studies: n=10; online supplementary table 7).

**Mental health**

Twenty-four studies72 77 78 81 86 88 93 100 101 103 105 107 109 112–115 125 127–129 131 132 138 assessed mental health outcomes and 16 studies78 83 86 93 100 101 105 109 113 115 125 127–129 131 138 were eligible for the NMA of mental health (figure 3). The results from the consistency NMA showed that resistance (p=0.003), aerobic (p=0.024) and stabilisation/motor control (p=0.031) exercise training resulted in improved mental health following intervention compared with true control (table 2). The results indicated that resistance (SUCRA: 80%) and aerobic (SUCRA: 80%) exercise training had the highest probability of being the best interventions for mental health. True control (SUCRA: 10%)
and therapist hands-on control (SUCRA: 20%) were most likely to be the least effective. There was evidence of inconsistency in the network ($\chi^2 = 54.8, p<0.001$; online supplementary table 5). The comparison-adjusted funnel plot did not provide evidence for apparent publication bias (online supplementary figure 1). The quality of the evidence for the ranks of the treatment was low (table 3). Data from pairwise meta-analysis presented evidence that exercise training (all) improved mental health in comparison to therapist hands-off control (SMD (95% CI): $-0.53$ ($-0.88$ to $-0.18$), $p=0.003$, $I^2=26\%$, studies: $n=3$) and therapist hands-on control (SMD (95% CI): $-0.79$ ($-1.56$ to $-0.03$), $p=0.042$, $I^2=92\%$, studies: $n=4$; online supplementary table 8).

**Muscle strength**

Eight studies were eligible for the NMA of objectively measured trunk muscle strength (figure 3). The results from the consistency NMA did not give evidence for a significant impact of resistance or stabilisation/motor control exercise training on muscle strength (table 2). Stabilisation/motor control (SUCRA: 70%) and resistance (SUCRA: 80%) exercise training were most likely to be the best interventions for increasing muscle strength. There was no evidence of inconsistency in the network ($\chi^2 = 0.1, p=0.750$; online supplementary table 5). The comparison-adjusted funnel plot did not provide evidence for apparent publication bias (online supplementary figure 1). The quality of the evidence for the ranks of the treatment was low (table 3). Pairwise meta-analysis presented evidence that exercise training (all) improved muscle strength compared with control (all) only (SMD (95% CI): $0.29$ ($0.00$ to $0.58$), $p=0.050$, $I^2=14\%$, studies: $n=6$; online supplementary table 9).

**Muscle endurance**

Eight studies assessed trunk muscle endurance as an outcome measure, but NMA was not possible for objectively measured trunk muscle endurance as there were only two intervention types (true control and Pilates exercise training) with more than two studies for this outcome. Pairwise meta-analysis presented evidence that exercise training (seven exercise training types (aerobic (n=1), other (n=1), McKenzie (n=1), Pilates (n=2), resistance (n=1), stretching (n=1), water-based (n=1)) in five studies) improved muscle endurance compared with true control (SMD (95% CI): $1.37$ ($0.69$ to $2.45$), $p<0.001$, $I^2=87\%$, studies: $n=5$; online supplementary table 10).

**Analgesic pharmacotherapy use**

Only four studies reported analgesic pharmacotherapy use as an outcome and these were not reported in a format that permitted further analysis (ie, measured at follow-up only, variance not reported and examined as a categorical variable).

**DISCUSSION**

In this first NMA of exercise training in CLBP, we found that, depending on outcome of interest, Pilates (for pain), resistance and aerobic (for mental health) and resistance and stabilisation/motor control (for physical function) exercise training were the most effective interventions. The effect of stretching and McKenzie exercise on pain and self-report physical function did not significantly differ to no-intervention control. Limitations on these findings was that the quality of the evidence was low according to GRADE criteria. True control was the least effective treatment for all outcomes, followed by therapist hands-off for pain and physical function and therapist hands-on treatment for mental health.

**Reducing pain**

It is unlikely that one kind of exercise training is the single best approach to treating non-specific CLBP. Our study provides evidence that ‘active therapies’, such as Pilates, resistance, stabilisation/motor control and aerobic exercise training, where the patient is guided, actively encouraged to move and exercise in a progressive fashion are the most effective. This notion is further supported by our findings that true control, as well as therapist hands-on and -off, were most likely to be the least effective treatments. The evidence for the therapeutic use of exercise training for the management of chronic musculoskeletal pain (eg, non-specific CLBP) continues to grow as the field of pain science moves towards the biopsychosocial approach. Using this approach considers both patient needs and clinician competencies and contends that exercise training interventions should be individualised based on patient presentation, goals and modality preferences. The results of this NMA can be used by clinicians to help guide their selection of exercise interventions to patients with non-specific CLBP.

As mentioned above, our NMA for reducing pain identified Pilates, stabilisation/motor control and aerobic exercise training as the three treatments most likely to be the best. If the pooled SMD of these comparisons are considered as effect sizes, all three of these findings are large (ie, $>0.8$). These interpretations limit the comparisons between these interventions in terms predictive efficacy in the therapeutic setting, as well as determination of whether or not these findings reflect clinically meaningful changes. Considering transformation to more commonly used clinical outcomes in an alternative suggestion in the Cochrane Handbook and was previously considered for the visual analogue scale for pain intensity using a SD of 20. Using this approach suggests that Pilates ($-37$ points), stabilisation/motor control ($-26$ points) and aerobic exercise training ($-28$ points) can all reduce pain intensity by a clinically meaningful amount (ie, $-20$ points). These three treatments differed to therapist hands-on treatment ($+2$ points) and therapist hands-off treatment ($-14$ points) by a clinically significant extent. While the differences among these three likely best exercise treatments fell below this threshold, Pilates and stabilisation/motor control differed by 20 points or more to stretching ($-12$ points) and McKenzie ($-16$ points) exercise. Overall, there is low-quality evidence that there may be a differential effect.

### Table 3 Summary of confidence in ranking of treatments for outcomes (GRADE approach) in studies examining the efficacy of exercise training in patients with non-specific chronic low back pain

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Confidence</th>
<th>Reason for downgrade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>Low</td>
<td>Study limitations,* Inconsistency†</td>
</tr>
<tr>
<td>Physical function</td>
<td>Low</td>
<td>Study limitations, I Inconsistency†</td>
</tr>
<tr>
<td>Mental health</td>
<td>Low</td>
<td>Study limitations, § Inconsistency‡</td>
</tr>
<tr>
<td>Muscle strength</td>
<td>Low</td>
<td>Study limitations, ¶ Imprecision**</td>
</tr>
</tbody>
</table>

*Evidence came from 91% unclear to high risk of bias studies. See online supplementary figure 2 for additional data on the assessment of transitivity.
†High heterogeneity or inconsistency present.
‡Evidence came from 89% unclear to high risk of bias studies.
§Evidence came from 88% unclear to high risk of bias studies.
¶All evidence from unclear to high risk of bias studies.
**Imprecision of clear most or least favoured treatments.
between specific modes of exercise for impacting pain intensity in non-specific CLBP.

### Improving physical function

If we focus our attention beyond the outcome of pain, our study suggested that resistance and stabilisation/motor control exercise training forms had the highest probably of improving physical function (ie, disability as measured by questionnaire). In addition to these modalities, large effect sizes (ie, >0.8) in favour of reductions in disability were also observed for yoga, Pilates, water-based and aerobic exercise training. These suggest that a range of distinctively different exercise training modalities may reduce disability in patients with non-specific CLBP; clinicians who prescribe exercise training should work with patients to identify a modality suitable for their capabilities and interests to increase the likelihood of efficacy. An 11-point reduction on the Oswestry Disability Index is considered clinically significant. On the basis of the preintervention data from the included studies the typical SD of the Oswestry Disability Index is 12 (online supplementary table 11). Converting the estimated effect sizes in table 2 to Oswestry Disability Index units, yoga (−11 percentage points), Pilates (−11 percentage points), water-based (−12 percentage points), resistance (−14 percentage points) and stabilisation (−14 percentage points) exercise training showed a clinically significant change in self-report physical function compared with true control. This was not observed for McKenzie exercise (−3 percentage points) and stretching exercise (−7 percentage points).

### Improving mental health

Aerobic and resistance exercise training emerged from our NMA as the most likely to improve mental health based on large effect sizes (ie, >0.8). These analyses were limited by the lower number of studies when compared with other measures, such as pain and physical function. Only 16 studies, including 5 exercise treatments and 2 control interventions, were eligible for NMA. Mental health issues, such as depression and anxiety, are seen in 36% and 29% of people with CLBP, respectively. Given these comorbidities are associated with pain intensity, we suggest that future exercise training studies report these measures. We also suggest that standardised assessment tools are considered, as the majority of studies included in the current study assessed mental health via different tools.

### What is the role for trunk muscles in treatment?

Trunk muscle strength and endurance are known risk factors for future back pain. Despite the clinical relevance of this outcome, only four studies, all of which examined trunk muscle strength, were eligible for the NMA. While the analysis provided low-quality evidence that resistance and stabilisation exercise training could improve trunk muscle strength, the effect sizes of these interventions versus no intervention control were not significant. NMA was not possible for trunk muscle endurance. Trunk muscle strength and endurance have the potential to be objective outcome measures to differentiate treatment approaches and future trials should include these measures to elucidate the potential efficacy of exercise training.

### Use of analgesics as an outcome of clinical trials

Analgesic pharmacotherapy use is of clinical relevance and directly relevant to the affected individual, yet only four studies reported analgesic pharmacotherapy use and none were eligible for meta-analysis. We recommend that the use of analgesic pharmacotherapy be considered as an outcome in future RCTs.

### Strengths and limitations

The strengths of this study included that searches were not limited by publication date or language, and studies included were not restricted to a specific type of intervention or comparator. Considering numerous outcome measures of clinical relevance in those with non-specific CLBP also strengthened this study; in particular, those measures that pertain to pharmacotherapy use and mental health are often omitted despite the high prevalence of mental health issues and analgesic use in this susceptible population group.

We report several limitations. We did not consider variations in safety between exercise modalities. Underreporting of adverse events is a known issue associated with the reporting of exercise training studies. Nonetheless, there is low-quality evidence that exercise does not cause serious harms and that when adverse events are reported, they are limited to muscle soreness and increased pain.

A high proportion (78%) of studies had a high risk of bias for at least one type of bias, even after excluding performance bias, which was high for all studies due to the inability to blind patients to exercise training. Most included studies (76%) provided insufficient information to appropriately determine risk of bias, leaving only seven studies (8%) at low risk of bias. In an attempt to overcome this limitation, we applied the GRADE approach, which suggests that with higher quality evidence, one can be more confident that the effect estimates are the true effects and additional evidence is less likely to result in a change in the estimate. Given that the levels of evidence were low, it is possible that the effect sizes and rankings of the treatments will change as more evidence is obtained.

### CONCLUSION

In conclusion, our study provided evidence that various exercise training approaches are effective and should be incorporated into usual care for adults with non-specific CLBP due to its potential for improving pain, physical function, muscle strength and mental health. Importantly, exercise training was more effective than hands-on therapist treatment for reducing pain and improving physical function and mental health. However, despite our identifying numerous studies that examined exercise training, we were unable to determine whether exercise training improved trunk muscle strength, trunk muscle endurance and reducing analgesic pharmacotherapy use; these outcomes were not often reported.

Examination of specific kinds of exercise training was limited by the number of studies available and variability in reporting. There was low-quality evidence that true control (ie, no intervention), therapist hands-off treatment (eg, general practitioner management, education or psychological interventions) and therapist hands-on (eg, manual therapy, chiropractic or passive physiotherapy) are most likely to be ineffective interventions for non-specific CLBP. There was low-quality evidence that stretching and McKenzie exercise training were not effective for pain or physical function in people with non-specific CLBP. Finally, there was low-quality evidence that Pilates, stabilisation/motor control, resistance and aerobic exercise training were able to improve pain, physical function and mental health in people with non-specific CLBP. Collectively, our findings provide evidence that active exercise therapies may be an effective treatment of non-specific CLBP in adults.
Worldwide, low back pain is the leading cause of disability and most common non-communicable disease. Exercise training is an effect treatment for non-specific chronic low back pain, but the best mode of exercise training is unknown. Exercise training may also be more effective than hand-on therapist treatments.

What are the new findings

Pilates, stabilisation/motor control, aerobic and resistance exercise training are possibly the most effective treatments, pending outcome of interest, for adults with non-specific chronic low back pain.


REFERENCES

Review


11 of 12


