Journal of Physiotherapy 70 (2024) 51–64

Journal of PHYSIOTHERAPY

journal homepage: www.elsevier.com/locate/jphys

Research

In trials of physiotherapy for chronic low back pain, clinical relevance is rarely interpreted, with great heterogeneity in the frameworks and thresholds used: a meta-research study

Tiziano Innocenti [a,b], Tim Schleimer [c], Stefano Salvioli [b,d], Silvia Giagio [e,f], Raymond Ostelo [a,g], Alessandro Chiarotto [a,h]

[a] Department of Health Sciences, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, The Netherlands; [b] GIMBE Foundation, Italy; [c] Division of Physiotherapy, Department of Applied Health Sciences, Hochschule für Gesundheit, Bochum, Germany; [d] Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Genoa, Italy; [e] Division of Occupational Medicine, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy; [f] Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum, University of Bologna, Bologna, Italy; [g] Department of Epidemiology and Data Science, Amsterdam UMC, Location Vrije Universiteit, Amsterdam Movement Sciences Research Institute, The Netherlands; [h] Department of General Practice, Erasmus MC, University Medical Center, Rotterdam, the Netherlands

KEY WORDS

Physical therapy
Low back pain
Clinical relevance
Randomised controlled trial
Meta-research methods

ABSTRACT

Questions: How do authors of randomised controlled trials (RCTs) interpret the clinical relevance of the effects of physiotherapy interventions compared with no intervention on pain intensity, physical function and time to recovery in people with chronic low back pain (CLBP)? How can the clinical relevance be re-interpreted based on the available smallest worthwhile effect (SWE) threshold for this comparison? Are the studies in this field adequately powered? **Design**: Cross-sectional meta-research study. **Participants**: People with CLBP. **Outcome measures**: Pain intensity, physical function and time to recovery. **Results**: This review included 23 RCTs with 1,645 participants. Twenty-two and 18 studies were included in the analysis of pain intensity and physical function, respectively. No studies investigated time to recovery. Sixteen studies reported varying thresholds to interpret clinical relevance for physical function and pain intensity. Discrepancies between interpretation using the minimal important difference and SWE values were observed in five studies. Study power ranged from 9% to 98%, with only four studies having a power > 80%. **Conclusion**: Little attention is given to the interpretation of clinical relevance in RCTs comparing physiotherapy with no intervention in CLBP, with great heterogeneity in the frameworks and thresholds used. Future trials should inform patients and clinicians on whether the effect of an intervention is large enough to be worthwhile, using a reliable and comprehensive approach like available SWE estimates. **Registration**: medRxiv https://doi.org/10.1101/2022.12.14.22283454. **[Innocenti T, Schleimer T, Salvioli S, Giagio S, Ostelo R, Chiarotto A (2024) In trials of physiotherapy for chronic low back pain, clinical relevance is rarely interpreted, with great heterogeneity in the frameworks and thresholds used: a meta-research study. *Journal of Physiotherapy* 70:51–64]**

Introduction

Chronic low back pain (CLBP) is defined as pain, muscle tension or stiffness lasting > 12 weeks[1] and it is one of the leading causes of disability worldwide, accounting for high costs for healthcare systems and loss of work productivity.[2,3] Around 65% of people with low back pain (LBP) still report pain after 1 year[4] and it affects physical, emotional and social functioning.[2] Many health interventions have been proposed for patients with CLBP.[5] Randomised controlled trials (RCTs) provide the best study design, yielding accurate estimates of the effectiveness of an intervention.[6] To determine the effectiveness of an intervention for specific outcomes, various stakeholders (eg, researchers, policymakers) should judge the clinical relevance of an RCT's results.[7,8] Clinical relevance indicates whether the intervention should be used in clinical practice to improve healthcare outcomes, and it mainly involves the interpretability of between-group differences in an RCT.[9]

Many parameters have been proposed to find a threshold for interpreting the clinical relevance of between-group differences for patient-reported outcomes.[10,11] Several terms have been used to identify a threshold for clinical relevance over the years.[11,12] One of the most commonly used is the minimal important difference (MID), which is defined as 'the smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and which would lead the patient or clinician to consider a change in the management'.[13] However, the MID only partially satisfies some conditions that should be

https://doi.org/10.1016/j.jphys.2023.11.007

ensured by reasonable estimates of between-group clinical relevance.[11] First, it is calculated as a within-group change and not as a between-group difference.[11] Second, it is not intervention-specific and does not involve weighing the benefits of the intervention against its costs, risks and side effects compared with a control intervention.[8,14] Third, judgments about the effectiveness of an intervention should be based on the perspective of patients who receive the intervention, and the MID only indirectly asks patients to make that judgement.[15,16]

To overcome the issues associated with the MID, Barrett et al[11,17] described, in 2005, a new methodological approach––the benefit-harm trade-off method––to estimate the smallest worthwhile effect (SWE) of health interventions. The SWE approach captures the judgments of recipients of care; allows patients to weigh the benefits of treatment against the risks, costs and side effects; and potentially provides estimates based on an intervention-control comparison.[12] It is expressed as the smallest amount of patient-perceived benefit that an intervention would require to justify associated costs, risks and other inconveniences.[11] Two studies using the benefit-harm trade-off method[18,19] calculated SWE values for pain intensity, physical function and time to recovery for physiotherapy interventions (ie, exercise and/or manual therapy) compared with no intervention for patients with CLBP. These studies consistently found that patients need to experience $\geq$ 20% additional improvement in pain or physical function or to speed up their recovery by 10 days to consider that the effect of physiotherapy is worthwhile. In other words, the 20% value corresponds to the minimum difference people need to experience with physiotherapy compared with no intervention departing from their current level (ie, baseline score in an RCT). Exploring how the authors of published RCTs on physiotherapy for CLBP have interpreted their results and reinterpreting the clinical relevance using the SWE values for between-group differences could be relevant to future research (eg, to drive the sample size calculation in RCTs) and clinical practice (eg, to make clinicians and patients better able to judge the clinical relevance of physiotherapy interventions against no intervention in CLBP).

Therefore, the primary objectives of this study were: to evaluate whether authors have interpreted the clinical relevance of the effect of physiotherapy compared with no intervention on pain intensity, physical function and time to recovery following their a priori definition; and to re-interpret the clinical relevance of the between-group differences of the published RCTs based on the available SWE estimates for physiotherapy compared with no intervention in patients with CLBP.[18,19] As a secondary objective, this study also aimed to evaluate, for descriptive purposes, whether the studies were adequately powered or underpowered, considering the published SWE values and a power threshold of 80%.

Therefore, the research questions for this meta-research study were:

1. How do authors of RCTs interpret the clinical relevance of the effects of physiotherapy interventions compared with no intervention on pain intensity, physical function and time to recovery in people with CLBP?
2. How can the clinical relevance be re-interpreted based on the available SWE threshold for this comparison?
3. Are the studies in this field adequately powered?

Methods

A cross-sectional meta-research study was conducted. The protocol of this meta-epidemiological study was prospectively registered, and it is available as a preprint[20] and peer-reviewed publication.[21] The study followed the adaptation of the PRISMA 2009 statement for meta-epidemiological studies proposed by Murad et al[22] for the reporting of this manuscript.

Data source and search strategies

For retrieving eligible articles, a systematic literature search on the following electronic databases was conducted from inception to 15 April 2023: Medline (through the PubMed interface), PEDro, Embase (Embase.com) and Cochrane CENTRAL (www.cochranelibrary.com). In addition to the electronic database search, other potentially relevant studies were searched in the reference lists of the included articles and in relevant systematic reviews on the topic,[23,24] and in grey literature sources (ie, OpenGray and Google Scholar).

Medical subject headings (MeSH) and all relevant free text words were combined using Boolean operators (AND, OR) to retrieve papers with the relevant condition (ie, *CLBP*), intervention (eg, *exercise*, *manual therapy*) and study design (eg, *randomized controlled trial*, *controlled clinical trial*). The Cochrane MEDLINE sensitivity-maximising RCT filter was used to ensure maximum sensitivity of the search strategy, since it has very high sensitivity and a slightly better precision relative to more sensitive filters.[25] No year restrictions were applied. A detailed search strategy was created for each electronic database (see Table 1 on the eAddenda). To be consistent with the criteria used in the estimation studies for the SWE,[18,19] specific eligibility criteria were set (see Box 1).

Box 1. Eligibility criteria used for the study selection.

Inclusion criteria
- Conducted in adult patients ($>$ 18 years old) with non-specific chronic low back pain (LBP) (defined as LBP lasting for $>$ 12 weeks); RCTs including a 'mixed' population of participants with non-specific LBP and other musculoskeletal disorders (eg, specific LBP, neck pain, acute LBP) were included if $\geq$ 75% of the participants reported chronic non-specific LBP
- The intervention consisted of one or more of the following physiotherapy treatment components: any exercise therapy modality (eg, resistance exercise, strengthening exercise, aerobic exercise) or any form of manual therapy (eg, muscle energy techniques, massage therapy, spinal manipulative therapy); other treatments (eg, education, psychologically informed physiotherapy, electrotherapy or acupuncture) were included only if they represented a minor part of the intervention (ie, $\leq$ 25% of the treatment time/sessions)
- The control group was 'no intervention' (eg, waiting list or total absence of interventions); regardless of the name given to the comparison by the original authors, the study was included if no active treatment was given to the participants (eg, if authors described usual care as 'participants were instructed to follow their normal schedule of medications and physical activity', it was included)
- At least one of the outcomes was pain intensity, physical functioning or time to recovery
- The study design was a randomised controlled trial
- The follow-up time point was short-term, defined as $\leq$ 12 weeks (since the studies by Ferreira[19] and Christiansen[18] focused on 2 and 6 weeks, respectively)
- Studies written in English, French, Italian, German, Dutch, Spanish and Portuguese

Exclusion criteria
- Patients who underwent back surgery in the last year
- Patients with severe psychiatric comorbidities that impeded participation
- Other treatments (eg, pain medication, injections) combined with physiotherapy

Data selection

First, the titles and abstracts of the retrieved studies were divided into two sets. Each set was screened independently by two pairs of reviewers (TI, TS and SS, SG) to evaluate whether they met the eligibility criteria. Second, the full texts of the potentially eligible studies were downloaded and assessed for eligibility by the same two pairs. If the full text was not retrieved, the authors were contacted with a maximum of three emails. A consensus meeting was held to determine agreement on the selection; in cases of disagreement, a

third reviewer (AC) decided on inclusion. EndNote[a] was used to remove duplicates and manage the bibliography. Rayyan QCRI systematic review software[26] was used for the selection process.

Data extraction

The data were extracted by two independent reviewers (TI and TS, or TI and AC) and double-checked by a third reviewer (SS) to ensure the precision and quality of the retrieved information. For clinical relevance, we checked how the authors a priori defined the MID or SWE for the between-group differences; these parameters were searched in the article text and in the sample size calculation (either in the study protocol when available or the full text), looking at the difference they used to calculate it. The following predefined decision rules were used to select data from trials to prevent selective inclusion: where trialists reported for within-group change both unadjusted and adjusted values for the same outcome, adjusted values were extracted; and where trialists reported data analysed, based on the intention-to-treat (ITT) sample and another sample (eg, per protocol, as treated), ITT-analysed data were extracted. Missing data or additional details were gathered by contacting the corresponding author of included studies with a maximum of three emails.

Data analysis

Descriptive data of the included studies were reported in tables as mean (standard deviations) or median (interquartile range) values. The following descriptive primary and secondary analyses were performed.

Primary analyses

We compared the authors' interpretation of results for clinical relevance with their results to determine if they met their a priori definitions. If the authors did not specify any threshold for clinical relevance, we used the between-group difference specified in the sample size calculation. We reported a lack of information if the authors did not report any threshold or did not report the sample size calculation (in the manuscript or protocol).

We re-interpreted the clinical relevance, comparing the between-group differences for every outcome in every RCT based on the published SWE values.[18,19] For pain intensity and physical function, the results were considered clinically relevant where there was 20% more additional improvement in the intervention group (physiotherapy) compared with the control (no intervention), starting from the baseline values of the respective outcomes when comparing physiotherapy with no intervention. For time to recovery, the result was considered clinically relevant where the between-group difference was $\geq$ 10 days. To retrieve the percentage of between-group differences, it was calculated by comparing the between-group difference with the mean baseline value of the study for each outcome. This calculation is presented with a numerical example:

Considering a trial with a mean baseline pain intensity of 60 (on a 0 to 100 scale) for both groups and a control group that reached a mean of 50 at the primary endpoint, a 20% additional improvement in the intervention group would be a mean score at follow-up of approximately 38 points. This is because the control group had an improvement of 16.7% (10 out of 60), so an additional improvement of 20% in the intervention group corresponds to an improvement from the baseline of 36.7% (16.7 reached by the control group plus 20 percentage points); that is, 22 points less in the VAS (60 multiplied by 0.367) and so 38 points as a mean value at follow-up (60 minus 22).

Secondary analysis

We calculated the power reached a posteriori for each included study considering the alpha error used in the study, sample size and a between-group difference of 20% (ie, the SWE values already published[18,19]) on the primary outcome. In this way, each study was classified as underpowered (if the power achieved was $<$ 80%) or well-powered (if the power was $\geq$ 80%). Pilot trials evaluating the feasibility of a future large RCT were excluded from this analysis.

SPSS software[b] was used for the analyses. G*Power software was used for a posteriori power calculation.[27]

Results

The search strategies retrieved 16,241 records. After removal of duplicates (n = 10,251) and title, abstract and full-text reading, we included 23 RCTs[28–50] for 30 eligible comparisons (ie, seven trials had three arms) with 1,645 participants. Figure 1 summarises the flow of the study selection process. References of excluded studies with reasons for exclusion are reported in Table 2 (see eAddenda for Table 2). Twenty-two (28 comparisons) and 18 (24 comparisons) studies were included in the analysis of pain intensity and physical function, respectively. Interventions included general and core stability exercise/stretching programs,[28,30–35,37–39,42–44,46–48] back school,[29] Pilates,[36,40,41] chiropractic interventions[45] and spinal mobilisation.[49,50] Pain intensity was measured in all but one[30] study, four studies[31,38,43,44] did not measure physical function and none of the studies measured time to recovery as an outcome. Two studies[34,48] assessing physical function and one study[48] measuring pain intensity had missing data that we were unable to retrieve. All other studies had complete datasets. The characteristics of the included studies are reported in Table 3.

Primary analysis: interpretation of clinical relevance in included studies

Nine[28,30,33,35,37,45,47,49,50] (50%) and fourteen[33–40,42,45–47,49,50] (64%) studies reported a priori (ie, for the sample size calculation) thresholds to interpret the clinical relevance for physical function and pain intensity, respectively. Six[29,31,34,41,43,46,48] studies (26%) did not report the sample size calculation. One study[49] performed a post hoc sample size calculation. The most common constructs for the clinical relevance reported a priori (ie, in the sample size calculation) were the MID (n = 7) and Cohen's threshold[51] (n = 3), while three studies[35,40,45] used the difference in mean change between groups, without explicitly referring to the SWE approach. One study[35] used a mean change between groups for the sample size calculation, but then they interpreted clinical relevance according to a different threshold (the MID values as published by Ostelo et al[52]). Lastly, two studies[47,49] used a mean within-group change, using the same threshold to discuss clinical relevance.

Nine studies[28,33,35,39,42,45,47,49] (39%) discuss the clinical relevance of the results; two studies[35,39] used two different constructs for sample size calculation and to discuss the clinical relevance of the results. Namely, they used an improvement of 20%[35] and Cohen's threshold[39] for the sample size calculation, but then they discussed clinical relevance according to the MID framework.

For pain intensity, 14 studies[33–40,42,45–47,49,50] (of 22 studies) pre-specified a cut-off value to interpret clinical relevance, and the results meet these cut-offs in all but three.[37,47,50] Re-interpreting the results according to the SWE published value (ie, 20% between-group difference), results from four studies[28,31,37,38] (17%) became not clinically relevant. Discrepancies between interpretation with MID and SWE values were observed in two studies,[38,50] where the results changed their clinical relevance if interpreted with the SWE cut-off (Table 4).

For physical function, nine studies[28,30,33,35,37,45,47,49,50] (of 18 studies) pre-specified a cut-off value to interpret clinical relevance, and the results met these cut-offs only in six comparisons from four studies.[33,35,49,50] By re-interpreting the results according to the published SWE value, results from six studies[28–30,37,41,48] (of 18 studies) were not clinically relevant. Discrepancies between trialists' interpretation (with MID) and our re-interpretation (with SWE values) were observed in three studies,[30,35,47] where the results were clinically relevant if interpreted with the SWE cut-off. The results for physical function are reported in Table 5.

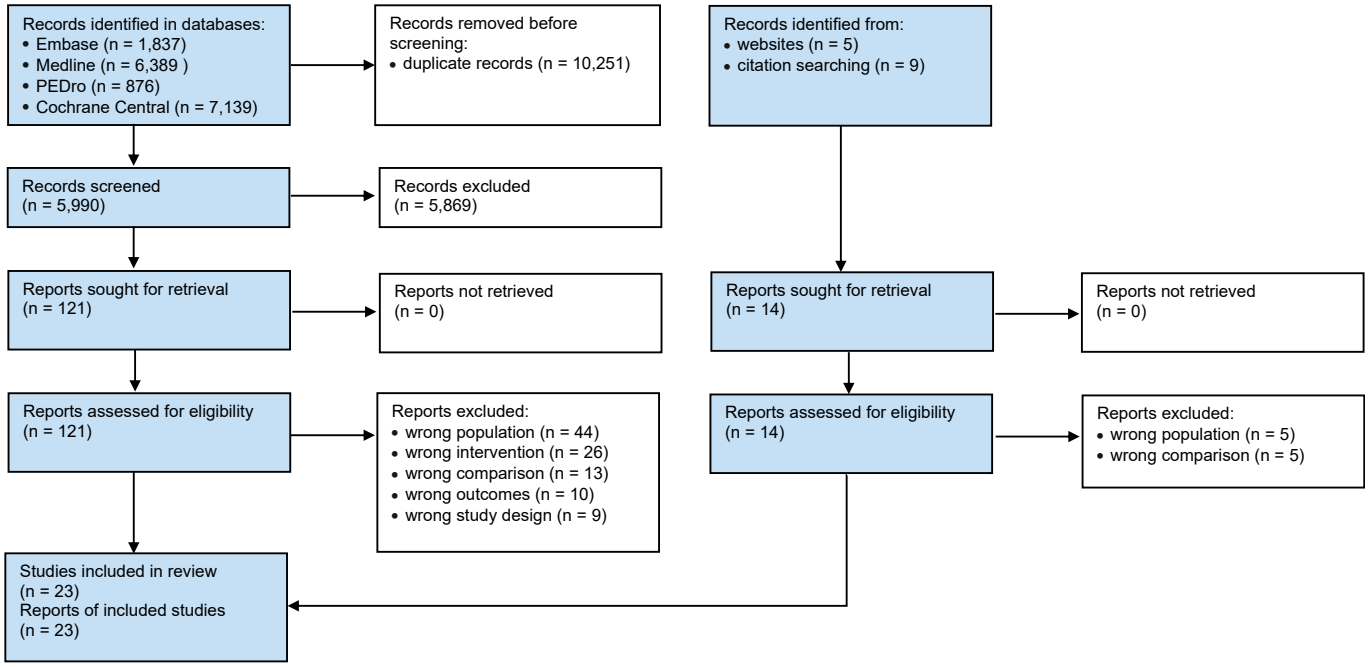

Figure 1. Flow of studies through the review.

Secondary analysis: power calculation

All studies pre-specified the threshold used for statistical significance (0.05). Seven studies[29,36,37,39,41,46,48] did not report which was the primary outcome. In these seven studies, we decided to perform a secondary analysis considering physical function as the outcome (ie, arguably the most relevant outcomes in LBP RCTs[53]) for power calculation. Study power ranged from 9% to 98% (median = 44, IQR = 53). Only four studies[40,41,43,44] were well powered (ie, power > 80%). Results from this secondary analysis are reported in Table 6.

Discussion

Twenty-three studies investigated the effectiveness of physiotherapy interventions compared with no intervention in people with CLBP. The current study highlights that, despite the effectiveness of physiotherapy compared with no intervention in patients with CLBP seeming to be clinically relevant, little attention was given to that by the studies' authors, with great heterogeneity in frameworks and thresholds used (Tables 3 to 5). When framework and thresholds differ between trials while focusing on the same comparison, they may complicate the interpretation of a treatment's clinical effect and, more generally, the ability to compare results across different treatments and diverse populations. It is believed that this is the first study to compare the clinical relevance of physiotherapy for pain intensity and physical function of previous RCTs with the published SWE of 20%.[18,19] No other systematic reviews with the same criteria were found; therefore, comparison with other literature is limited.

Among studies that calculated sample size (n = 16), about half (60%) used a threshold for clinical relevance in their calculation. In most (90%) of those that did so, thresholds for meaningful within-group change (eg, a reduction of 2 points on a 0 to 10 scale of pain intensity[52]) were confused with the evaluation of what constitutes a meaningful between-group difference. The determination of the clinical importance of group differences in RCTs depends on many factors, including: the magnitude of the group difference observed in the trial and its associated confidence intervals; the availability of other treatments; adverse events associated with the intervention; and an overall evaluation of the benefit–risk profile, assessed directly by patients including cost and inconveniences.[12,16,54,55] All but two studies[40,45] in the current analysis used frameworks (eg, MID) that do not consider the abovementioned characteristics.

In this sample, the clinical relevance of the intervention was confirmed after our re-interpretation with SWE values in most (79%) of the studies. It can still be argued that it could potentially not be the case if other research fields, interventions and comparisons were considered. For example, a recent publication by Henderson and Riddle[56] found that the SWE was superior to the MID for estimating acceptable benefits of knee arthroplasty, highlighting that patients expect a substantially greater change in outcome scores than expected using the published MID estimates. It is vital to use a reliable, comprehensive, patient-driven and context-specific approach to consider and interpret clinical relevance in clinical trials; managers, health funders, clinicians and other providers of physiotherapy services should be informed about whether the effect of an intervention is large enough to be worthwhile. This can help them decide which interventions to bring to the shared decision-making conversation.

Another interesting finding of this study is that while all studies commented on the statistical significance of the results, few discussed the clinical relevance of the results. It is well acknowledged that the interpretation of changes in trial outcomes needs to go beyond a simple discussion of statistical significance to include clinical relevance.[16] Statistically significant evidence of a treatment's effectiveness in a clinical trial is insufficient to indicate that the magnitude of the treatment effect is clinically important; instead, the evaluation should involve determining whether the benefits of treatment are meaningful to society (eg, reducing healthcare costs).[9]

Lastly, these results underline the need for a carefully planned sample size calculation. Most of the studies (83%) in this review were underpowered if a SWE threshold of 20% was used. Issues about the power of LBP trials are well acknowledged in the literature.[57,58] In an analysis of LBP trials published between 1980 and 2012, Froud and colleagues[57] found that sample size calculations were reported in only 41% of trials, and only 5% had sufficient power to detect clinically meaningful effects. Researchers should plan their trials to detect the SWE of the intervention, trying to avoid type II errors.[12]

This study's results are generalisable only to CLBP patients when physiotherapy interventions (exercise and/or manual therapy) are compared with no intervention. It should be considered that the SWE may vary between countries due to differences in healthcare

Table 3
Characteristics of the included studies.

Study	Eligibility criteria		Intervention group [a]			Comparison group [a]			Clinical relevance framework
	Inclusion	Exclusion	Characteristics	Pain intensity	Physical function	Characteristics	Pain intensity	Physical function	
Smeets et al 2006[28]	Age 18 to 65 y; LBP ± radiation to the leg for > 3 mth resulting in functional limitations (RMDQ > 3); ability to walk ≥ 100 m without interruption	Vertebral fracture; spinal inflammatory disease, infections or malignancy; current nerve root pathology; spondylolysis or spondylolisthesis; lumbar spondylodesis; medical contraindication to exercise; ongoing diagnostic procedures or treatment for LBP; or a clear treatment preference	n = 53 Age *(y)* = 43 Aerobic training and dynamic strengthening exercises	0 to 100 VAS Baseline: 51.2 (26.6) 10-wk change score between-group: −8.7 (−16.9 to −0.5)	RMDQ Baseline: 14.2 (3.7) 10-wk change score between-group: −2.4 (−4.2 to −0.7)	n = 21 Age *(y)* = 41 Waiting list	0 to 100 VAS Baseline: 51.0 (25.4) 10 wk: 53.4 (22.6)	RMDQ Baseline: 14.0 (3.9) 10 wk: 13.9 (4.8)	Not specified
Andrade et al 2008[29]	LBP for > 3 months; pain at the time of the study; cognitive capacity to give informed consent	Pregnancy; herniated disc; infectious or inflammatory spondylitis; tumours; fractures; thoracic, shoulder or neck pain; or fibromyalgia	n = 34 Age *(y)* = 45 Back school	0 to 10 VAS Baseline: 5.0 (2.7) 4 wk: 3.2 (2.8)	RMDQ Baseline: 10.5 (5.9) 4 wk: 8.1 (6.1)	n = 36 Age *(y)* = 45 Waiting list	0 to 10 VAS Baseline: 5.6 (2.6) 4 wk: 4.9 (1.9)	Baseline: 11.0 (4.7) 4 wk: 9.8 (4.9)	Not reported
Harts et al 2008[30]	Male army recruits aged 18 to 54 y; LBP > 12 wk; available to visit department one to two times/wk over 8 wk; willing to abandon other treatment interventions for the lower back during the intervention period	Surgery in the last 2 y; severe back pain hindering maximal isometric strength efforts; or pain radiating below the knee with signs of nerve root compression	n = 20 Age *(y)* = 44 High-intensity training n = 21 Age *(y)* = 42 Low-intensity training	Not measured Not measured	RMDQ Baseline: 6.2 (4.4) 8 wk: 3.4 (4.0) Baseline: 7.6 (4.6) 8 wk: 6.1 (6.0)	n = 21 Age *(y)* = 41 Waiting list	Not measured	RMDQ Baseline: 6.5 (3.9) 8 wk: 5.2 (3.9)	MID
Hall et al 2011[47]	Age 18 to 70 y; persistent nonspecific LBP with ≥ 'moderate' pain or activity limitation as determined by their response to questions 7 or 8 on the SF-36	Known or suspected serious spinal pathology; any contraindication to exercise; or scheduled for spinal surgery	n = 80 Age *(y)* = 43 Tai Chi with balance, strengthening, stretching and body awareness	0 to 10 NRS Baseline: 4.4 (4.0 to 4.9) 10 wk: 3.4 (2.9 to 3.8)	RMDQ Baseline: 10.2 (9.1 to 11.3) 10 wk: 7.0 (5.9 to 8.1)	n = 80 Age *(y)* = 44 Waiting list	0 to 10 NRS Baseline: 4.4 (4.0 to 4.9) 10 wk: 4.7 (4.2 to 5.1)	RMDQ Baseline: 9.1 (8.0 to 10.2) 10 wk: 8.1 (7.0 to 9.3)	MID
Lee et al 2011[31]	Age ≥ 21 y; LBP > 2 months without pain referral into the lower extremities	Psychological illness that might interfere with the study protocol; overt neurological signs (sensory deficits or motor paralysis); or pregnancy	n = 13 Age *(y)* = 50 Core stability	0 to 100 VAS Baseline: 41.1 (12.8) 4 wk: 31.9 (14.2)	Not measured	n = 19 Age *(y)* = 47 Continuation of daily activities	0 to 100 VAS Baseline: 32.8 (11.9) 4 wk: 20.9 (11.6)	Not measured	Not reported
Masharawi et al 2013[32]	Age 45 to 65 y; LBP ≥ 12 wk; able to give informed consent; understood instructions; willing to cooperate with treatment	LBP from systemic or structural pathology; diagnosed with inflammatory joint disease; or overt neurological signs	n = 20 Age *(y)* = 52 Active non-weight-bearing group exercise	0 to 10 VAS Baseline: 4.0 (1.4) 4 wk: 1.7 (0.8)	RMDQ Baseline: 14.2 (5.2) 4 wk: 9.3 (5.8)	n = 20 Age *(y)* = 54 Waiting list	0 to 10 VAS Baseline: 3.9 (1.6) 4 wk: 3.9 (1.5)	RMDQ Baseline: 14.9 (6.0) 4 wk: 14.4 (5.8)	Not reported

Table 3 (*Continued*)

Study	Eligibility criteria		Intervention group [a]			Comparison group [a]			Clinical relevance framework
	Inclusion	Exclusion	Characteristics	Pain intensity	Physical function	Characteristics	Pain intensity	Physical function	
Steele et al 2013[33]	LBP >12 wk; no contraindication to resistance training	Any medical condition for which movement therapy might be contraindicated	n = 12 Age (*y*) = 46 Exercise program full ROM n = 10 Age (*y*) = 42 Exercise program limited ROM	0 to 100 VAS Baseline: 46.7 (25.5) 12-wk change score within-group: −30.3 (25.8) Baseline: 41.3 (22.9) 12-wk change score within-group: −16.3 (11.0)	ODI Baseline: 36.2 (11.1) 12-wk change score within-group: −18.2 (6.6) Baseline: 26.9 (13.6) 12-wk change score within-group: −12.0 (5.2)	n = 9 Age (*y*) = 41 No training	0 to 100 VAS Baseline: 19.2 (15.5) 12-wk change score within-group: 6.7 (14.9)	ODI Baseline: 26.2 (7.3) 12-wk change score within-group: −3.0 (6.9)	MID
Moussouli et al 2014[34]	LBP ≥ 6 mth; generally good health	Spinal stenosis; spinal inflammatory disease; fracture, spondylolysis or spondylolisthesis; genetic structural abnormality in the spine; daily intensive LBP; pregnancy; or medication affecting heart rate and/or blood pressure	n = 13 Age (*y*) = 54 Isometric exercise n = 13 Age (*y*) = 53 Isotonic exercise	SF-36 subscale Baseline: 49.3 (23.0) 4 wk: 64.9 (14.2) Baseline: 50.7 (19.7) (SF-36 subscale) 4 wk: 60.1 (19.0) (SF-36 subscale)	Not reported Not reported	n = 13 Age (*y*) = 63 Continuation of daily activities	SF-36 subscale Baseline: 68.4 (17.3) 4 wk: 48.6 (31.0)	Not reported	Cohen's d threshold
Lawand et al 2015[35]	Age 18 to 65 y; LBP > 3 mth that worsens with movement and improves with rest; LBP severity 3 to 8/10.	Nerve root pain; motor impairment; inflammatory spondyloarthropathy; spondylolisthesis; fibromyalgia; previous back surgery; fracture; pregnancy, recent or current physiotherapy; or recent change in analgesia	n = 31 Age (*y*) = 49 Global postural re-education	0 to 10 VAS Baseline: 6.4 (1.6) 12 wk: 3.1 (2.3)	RMDQ Baseline: 12.6 (4.8) 12 wk: 7.2 (5.2)	n = 30 Age (*y*) = 47 Waiting list	0 to 10 VAS Baseline: 6.3 (1.6) 12 wk: 6.1 (2.1)	RMDQ Baseline: 11.9 (5.0) 12 wk: 10.9 (5.5)	MID
Kofotolis et al 2016[36]	Females aged 25 to 65 y; new nonspecific LBP > 12 wk; inability to resume daily activities in prior 3 wk	Acute LBP; spinal stenosis or surgery; inflammatory disease affecting the spine; fracture; spondylolysis or spondylolisthesis; genetic spinal structure abnormality; pregnancy, use of medication that affects heart rate and/or blood pressure; or pelvic girdle pain	n = 40 Age (*y*) = 41 Pilates n = 40 Age (*y*) = 41 General strengthening exercise	SF-36 subscale Baseline: 38.5 (12.6) 8 wk: 79.1 (7.9) Baseline: 39.4 (14.5) 8 wk: 71.3 (11.3)	RMDQ Baseline: 11.3 (4.1) 8 wk: 3.3 (1.8) Baseline: 12.4 (3.7) 8 wk: 4.9 (1.6)	n = 40 Age (*y*) = 43 Continuation of daily activities	SF-36 subscale Baseline: 36.9 (15.5) 8 wk: 41.6 (16.0)	RMDQ Baseline: 11.3 (5.4) 8 wk: 10.1 (4.5)	Cohen's d threshold
Segal-Snir et al 2016[37]	Females aged 40 to70 y; LBP > 12 wk; no physical treatment of LBP for 12 wk	LBP as a result of systemic or structural pathology; inflammatory joint disease; or overt neurological signs	n = 20 Age (*y*) = 57 Group exercise	0 to 10 VAS Baseline: 7.0 (2.3) 4 wk: 7.0 (3.0)	RMDQ Baseline: 13.0 (6.0) 4 wk: 11.0 (6.4)	n = 15 Age (*y*) = 55 Waiting list	0 to 10 VAS Baseline: 8.0 (1.4) 4 wk: 7.0 (1.9)	RMDQ Baseline: 14.0 (6.3) 4 wk: 14.0 (6.3)	Based on authors' previous study

Table 3 (*Continued*)

Study	Eligibility criteria		Intervention group [a]			Comparison group [a]			Clinical relevance framework
	Inclusion	Exclusion	Characteristics	Pain intensity	Physical function	Characteristics	Pain intensity	Physical function	
Xia et al 2016[50]	Aged 21 to 54 y; LBP ≥ 4 wk; RMDQ ≥ 6; had musculoskeletal LBP without radiation, or with radiation to either the proximal or distal extremity, consistent with the Quebec Task Force Classification for Spinal Disorders categories 1 to 3 [d]	Nerve root compression; neurological signs; lumbar spine stenosis; past back surgery; chronic pain syndrome, LBP from fracture, infection or visceral disease; comorbid conditions that could complicate the prognosis of LBP, including pregnancy or narcotic or other drug abuse; major clinical depression; inflammatory arthropathies involving the spine; bleeding disorders; significant osteoporosis; involved in litigation related to this LBP episode; or receiving disability for any health-related condition	n = 72 Age *(y)* = 40 High-velocity low amplitude spinal manipulation n = 72 Age *(y)* = 39 Low-velocity low amplitude spinal manipulation	0 to 100 VAS Baseline: 56.5 (22.5) 3 wk: 31.0 (25.2 to 36.7) Baseline: 54.5 (20.3) 3 wk: 35.2 (29.5 to 40.9)	RMDQ Baseline: 9.8 (3.6) 3 wk: 5.8 (4.8 to 6.8) Baseline: 9.5 (3.0) 3 wk: 5.8 (4.8 to 6.7)	n = 48 Age *(y)* = 40 Waiting list	0 to 100 VAS Baseline: 56.1 (19.6) 3 wk: 48.0 (41.2 to 54.8)	RMDQ Baseline: 9.7 (3.0) 3 wk: 8.9 (7.7 to 10.0)	MID
Arampatzis et al 2017[38]	Age 18 to 50 y; LBP ≥ 12 wk on ≥ half of the days; no specific pathology from infection, structural deformity, tumour, trauma or inflammatory disorder; moderate LBP (2 to 6/10); restriction during daily activities	Past spinal operation; prolapse or herniated disc; arthritis; mental, neurological or cardiovascular diseases; sensory motor deficits; or continuous dependency on pain relief medication or physiotherapy treatment	n = 20 Age *(y)* = 32 General exercises plus specific low-back training	0 to 10 VAS Baseline: 4.0 (1.4) 13 wk: 3.0 (1.8)	Not measured	n = 20 Age *(y)* = 31 Regular routines	0 to 10 VAS Baseline: 4.2 (1.7) 13 wk: 3.9 (1.9)	Not measured	Cohen's d threshold
Cortell-Tormo et al 2018[39]	Women aged 20 to 55 y; LBP > 3 months without associated leg pain; non-specific pain (soft tissue in origin) at L1 to L5; no history of formal exercise training	Concurrent treatment; spinal tumour, infection or inflammatory disease; spinal or lower-limb surgery; spinal fractures or structural deformities such as spondylolisthesis and spondylolysis; contraindications to exercise; nerve root compromise; or receiving medications other than analgesics and nonsteroidal anti-inflammatory drugs	n = 12 Age *(y)* = 36 Functional resistance training	0 to 10 VAS Baseline: 4.0 (1.8) 12 wk: 1.5 (1.5)	ODI Baseline: 15.5 (8.4) 12 wk: 6.0 (6.1)	n = 12 Age *(y)* = 36 Continuation of daily activities	0 to 10 VAS Baseline: 4.5 (1.6) 12 wk: 4.4 (1.4)	ODI Baseline: 14.0 (12.0) 12 wk: 14.5 (11.1)	MID

Table 3 (*Continued*)

Study	Eligibility criteria		Intervention group [a]			Comparison group [a]			Clinical relevance framework
	Inclusion	Exclusion	Characteristics	Pain intensity	Physical function	Characteristics	Pain intensity	Physical function	
Cruz-Diaz et al 2017[40]	Aged 18 to 50 years; LBP ≥ 12 wk; pain 3 to 7/10	Radiculopathy or other damages to the spine such as fractures, stenosis, or tumours; habitual Pilates practitioners; receiving other physiotherapy treatment now or in past 6 mth; pregnancy or not enough physical autonomy	n = 34 Age *(y)* = 37 Pilates mat n = 34 Age *(y)* = 35 Equipment-based Pilates	0 to 10 VAS Baseline: 4.6 (1.2) 6 wk: 3.3 (1.6) Baseline: 4.9 (1.1) 6 wk: 2.1 (1.3)	RMDQ Baseline: 11.4 (5.0) 6 wk: 7.9 (5.1) Baseline: 11.2 (5.1) 6 wk: 6.7 (5.1)	n = 34 Age *(y)* = 36 No intervention	0 to 10 VAS Baseline: 4.8 (1.0) 6 wk: 5.1 (1.0)	RMDQ Baseline: 10.5 (4.9) 6 wk: 10.3 (5.2)	Not specified
Keane et al 2017[48]	Aged 18 to 70 y, LBP ≥3 mth; no surgical intervention to the lower back; no specific injury to the lower back; not pregnant	Acute LBP lasting 1 to 6 wk; osteoporosis; stenosis; fracture; prior back surgery; or spondylosis or spondylitis	n = 10 Age *(y)* = 51 Land-based stretching exercise	0 to 10 VAS Baseline: 5.7 [b] 12 wk: 2.6 (1.9) [c]	Modified ODI Baseline: 34.0 [b] 12 wk: 25.8 (20.1) [c]	n = 9 Age *(y)* = 41 No intervention	0 to 10 VAS Baseline: 4.4 [b] 12 wk: 4.9 (3.3) [c]	Modified ODI Baseline: 40.0 [b] 12 wk: 31.6 (18.2) [c]	Not reported
Mazloum et al 2018[41]	Aged 18 to 55 y; non-specific LBP ≥ 3 mth; clinical evaluation indicates suitable to exercise; willingness to participate	History of trauma or spinal surgery; any misalignment or specific condition in the lumbar spine; spondylosis or spondylolisthesis; neurological or psychological conditions; or receiving physical therapy or other treatment interventions in prior 6 mth	n = 20 Age *(y)* = 37 Selective Pilates n = 20 Age *(y)* = 37 Extension-based exercises	0 to 10 VAS Baseline: 6.8 (1.4) 6 wk: 3.4 (1.0) Baseline: 7.2 (1.3) 6 wk: 5.3 (1.3)	ODI Baseline: 30.8 (1.2) 6 wk: 22.7 (3.1) Baseline: 27.2 (7.6) 6 wk: 23.2 (7.6)	n = 20 Age *(y)* = 39 Continuation of daily activities	0 to 10 VAS Baseline: 6.5 (1.2) 6 wk: 6.6 (1.3)	ODI Baseline: 26.2 (5.5) 6 wk: 26.6 (4.9)	Not reported
Tavares et al 2017[49]	Aged 18 to 55 y; nonspecific, continuous and recurrent LBP ≥ 3 mth	Pregnancy or red flag signs (neoplasia, spinal fracture, spinal osteomyelitis, infection, cauda equina syndrome, rheumatic diseases, diseases impairing cognition)	n = 20 Age *(y)* = 39 Spinal mobilisation	0 to 10 NRS Baseline: 4.8 (3.5 to 6.1) 5 wk: 0.2 (–0.1 to 0.6)	ODI Baseline: 11.3 (8.4 to 14.3) 5 wk: 3.1 (1.2 to 5.0)	n = 20 Age *(y)* = 31 No intervention	0 to 10 NRS Baseline: 4.1 (3.1 to 5.0) 5 wk: 3.8 (3.1 to 4.6)	ODI Baseline: 7.1 (5.2 to 9.0) 5 wk: 4.5 (3.1 to 6.0)	MID
Noormoham-madpour et al 2018[42]	Aged 18 to 55 y; ≥ 1 y in the nursing profession; non-specific LBP ≥ 3 months in the past 6 mth	Serious spinal cord involvement; pregnancy; spinal trauma; spinal or abdominal surgery; systemic disease (eg, systemic scleroderma or muscular dystrophy); spinal deformity (eg, scoliosis or kyphosis); abdominal wall hernia; or core stability exercises in the past 6 mth	n = 20 Age *(y)* = 43 Retraining of trunk muscles	0 to 100 VAS Baseline: 38.4 (21.7) 8 wk: 4.0 (5.4)	RMDQ Baseline: 7.8 (3.4) 8 wk: 1.7 (2.4)	n = 18 Age *(y)* = 41 Waiting list	0 to 100 VAS Baseline: 36.2 (27.2) 8 wk: 25.2 (17.7)	RMDQ Baseline: 9.5 (4.9) 8 wk: 7.9 (3.3)	MID

Table 3 (*Continued*)

Study	Eligibility criteria		Intervention group [a]			Comparison group [a]			Clinical relevance framework
	Inclusion	Exclusion	Characteristics	Pain intensity	Physical function	Characteristics	Pain intensity	Physical function	
Liu et al 2019[43]	Aged ≥ 50 y; LBP ≥ 3 mth; independent ambulation; able to participate in the training	LBP due to a tumour, rheumatoid arthritis or infection; neurological, musculoskeletal or psychiatric disorder; or cerebrovascular disease	n = 15 Age *(y)* = 63 Core stability (Tai Chi)	0 to 10 VAS Baseline: 5.7 (0.7) 4 wk: 4.3 (0.8)	Not measured	n = 13 Age *(y)* = 63 Continuation of daily activities	0 to 10 VAS Baseline: 5.9 (0.9) 4 wk: 5.9 (0.8)	Not measured	Not reported
Shariat et al 2019[44]	Age 20 to 50 y; LBP ≥ 3 mth with ≥ two chronic symptoms (eg, pain lifting heavy loads or postural changes); absence of pain in lower limb during physical testing; absence of disc tearing, presence of low intensity pain signals between vertebral area, and absence of joint burst following MRI	Use of corticosteroids; or had symptoms (dizziness, unconscious feeling, paralysis) prior to the start of the study	n = 19 Age *(y)* = not reported 13 simple exercise movements	FRI Baseline: 13.0 (2.0) 6 wk: 4.0 (1.0)	Not measured	n = 19 Age *(y)* = not reported No intervention	FRI Baseline: 12.0 (2.0) 6 wk: 11.0 (2.0)	Not measured	Not reported
Vining et al 2020[45]	Active-duty military personnel aged 18 to 40 y; LBP with severity of 2/10 over the prior 24 h	Chiropractic care within 30 days of consent; pregnancy; undergoing disability evaluation; patellar height was < 38 cm or > 65 cm; LBP from a visceral source; conditions contraindicating spinal manipulation; spinal fracture or spinal surgery in prior 6 mth; confirmed stenosis or neurogenic claudication; chronic pain syndrome; inflammatory spinal arthropathy; urgent or emergent medical condition; or unsafe for physical performance tests	n = 55 Age *(y)* = 31 Chiropractic care	0 to 10 VAS Baseline: 3.1 (1.5) 4-wk change score within-group: −0.9 (−1.3 to −0.6)	RMDQ Baseline: 6.5 (4.0) 4-wk change score within-group: −3.2 (−4.1 to −2.2)	n = 55 Age *(y)* = 30 Waiting list	0 to 10 VAS Baseline: 3.1 (1.4) 4-wk change score within-group: −0 (−0.4 to 0.3)	RMDQ Baseline: 7.6 (5.1) 4-wk change score within-group: −1.0 (−2.0 to −0.1)	Not specified
Hatefi et al 2021[46]	Females aged 25 to 40 y; LBP ≥ 3 mth not radiating past buttock; bilateral tightness of hip flexor muscles	Systemic neuromuscular, metabolic or neurological disorders; pregnancy; current other LBP treatment; or past surgery for disc herniation, spinal bifida or spinal stenosis	n = 15 Age *(y)* = 26 Stretching exercises	0 to 10 VAS Baseline: 4.7 (1.1) 8 wk: 3.1 (1.3)	ODI Baseline: 29.5 (8.7) 8 wk: 16.7 (5.9)	n = 15 Age *(y)* = 26 No intervention	0 to 10 VAS Baseline: 5.1 (1.4) 8 wk: 4.8 (1.1)	ODI Baseline: 28.0 (4.7) 8 wk: 24.8 (4.9)	MID

FRI = Functional Rating Index, LBP = low back pain, MID = minimal important difference, MRI = magnetic resonance imaging, NRS = Numerical Rating Scale, ODI = Oswestry Disability Index, RMDQ = Roland-Morris Disability Questionnaire, SF-36 = Short Form 36, SWE = smallest worthwhile effect, VAS = visual analogue scale

[a] Mean and either SD or 95% CI are reported, depending on what was available in the original study.

[b] Neither SD nor confidence intervals were reported for the baseline score.

[c] SD for follow-up was extracted from figures.

[d] Most had symptoms for > 3 mth.

Table 4
Results of the primary analysis: clinical relevance on the pain intensity outcome.

Study	Intervention	Pain intensity			Cut-off for clinical relevance[b], mean (scale)	Meets a priori cut-off	Meets the SWE values (20%)
		Within-group change, mean (%)		Difference in between-group change[a], mean (%)			
		Intervention	Comparison				
Smeets et al 2006[28]	Aerobic training and dynamic strengthening exercises	Only between-group change is reported	Only between-group change is reported	−8.7 (−17%)	Not reported	NA	No
Andrade et al 2008[29]	Back school	−1.8 (−36.0%)	−0.7 (−12.5%)	−1.1 (−23.5%)	Not reported	NA	Yes
Hall et al 2011[47]	Tai Chi	−1.0 (−22.7%)	0.3 (6.8%)	−1.3 (−29.5%)	1 (0 to 10 NRS)	Yes	Yes
Lee et al 2011[31]	Core stability	−9.2 (−22.4%)	−11.9 (−36.3%)	−2.7 (−13.9%)	Not reported	NA	No
Masharawi et al 2013[32]	Active non-weight bearing group exercise	−2.3 (−57.5%)	0 (0%)	−2.3 (−57.5%)	Not reported	NA	Yes
Steele et al 2013[33]	Exercise program full ROM	−30.3 (−65.0%)	6.7 (35.0%)	−38.0 (−100.0%)	15 (0 to 100 VAS)	Yes	Yes
	Exercise program limited ROM	−16.3 (39.5%)	6.7 (35.0%)	−23.0 (−74.5%)	15 (0 to 100 VAS)	Yes	Yes
Moussouli et al 2014[34]	Isometric exercise	−15.6 (−31.6%)	10.0 (14.6%)	−25.6 (−46.2%)	Cohen's d values [c]	Yes (large)	Yes
	Isotonic exercise	−9.4 (−18.5%)	10.0 (14.6%)	−19.4 (−33.1%)	Cohen's d values [c]	Yes (large)	Yes
Lawand et al 2015[35]	Global postural re-education	−3.3 (−51.6%)	−0.2 (−3.2%)	−3.1 (−48.4%)	20% (0 to 10 VAS)	Yes	Yes
Kofotolis et al 2016[36]	Pilates	−40.6 (−105%)	−4.7 (−12.7%)	−35.9 (−92.3%)	Cohen's d values [c]	Yes (large)	Yes
	General strengthening exercise	−31.9 (−81.0%)	−4.7 (−12.7%)	−27.2 (−68.3%)	Not reported	NA	Yes
Segal-Snir et al 2016[37]	Group exercise	0 (0%)	−1.0 (−12.5%)	1.0 (12.5%)	4 (0 to 10 VAS)	No	No
Xia et al 2016[50]	High-velocity, low-amplitude spinal manipulation	−25.5 (−45.1%)	−8.1 (−14.4%)	−17.4 (−30.7%)	15 (0 to 100 VAS)	Yes	Yes
	Low-velocity, low-amplitude spinal manipulation	−19.3 (−35.4%)	−8.1 (−14.4%)	−11.2 (−21%)	15 (0 to 100 VAS)	No	Yes
Arampatzis et al 2017[38]	General exercises and specific low-back training	−1.0 (−25.0%)	−0.3 (−7.1%)	−0.7 (−17.9%)	0.25 (Cohen's d)	Yes	No
Cortell-Tormo et al 2018[39]	Functional resistance training	−2.5 (−62.5%)	−0.1 (−2.2%)	−2.4 (−60.3%)	2 (0 to 10 VAS)	Yes	Yes
Cruz-Diaz et al 2017[40]	Mat-based Pilates	−1.3 (−28.3%)	0.3 (6.2%)	−1.6 (−34.5%)	20% (0 to 10 VAS)	Yes	Yes
	Equipment-based Pilates	−2.8 (−57.1%)	0.3 (6.2%)	−3.1 (−63.3%)	20% (0 to 10 VAS)	Yes	Yes
Keane et al 2017[48]	Land-based stretching exercise	−3.1 (−54.4%)	0.5 (11.4%)	−3.6 (−65.8%)	Not reported	NA	Yes
Mazloum et al 2018[41]	Selective Pilates	−3.4 (−50.0%)	0.1 (6.5%)	−3.5 (−56.5%)	Not reported	NA	Yes
	Extension-based exercises	−1.9 (−26.4%)	0.1 (6.5%)	−2.4 (−35.9%)	Not reported	NA	Yes
Tavares et al 2017[49]	Spinal mobilisation	−4.6 (−95.8%)	−0.3 (−6.2%)	−4.3 (−89.6%)	30% difference before and after treatment (0 to 10 NRS)	Yes	Yes
Noormohammadpour et al 2018[42]	Retraining of trunk muscles	−34.4 (−89.6%)	−11.0 (−30.4%)	−23.4 (−59.2%)	20 (0 to 100 VAS)	Yes	Yes
Liu et al 2019[43]	Core stability	−1.4 (−24.6%)	0 (0%)	−1.4 (−24.6%)	Not reported	NA	Yes
Shariat et al 2019[44]	13 simple exercise movements	−9.0 (−69.0%)	−1 (−8.3%)	−8.0 (.60.7%)	Not reported	NA	Yes
Vining et al 2020[45]	Chiropractic care	−0.9 (−29%)	0 (0%)	−0.9 (−29%)	≥ 20% difference in mean change between groups	Yes	Yes
Hatefi et al 2021[46]	Stretching exercises	−1.6 (−34.0%)	−0.3 (−6.0%)	−1.3 (−28%)	1.7 (0 to 10 VAS)	Yes	Yes

NA = not applicable, NRS = Numerical Rating Scale, VAS = visual analogue scale
[a] Negative values represent an improvement in favour of the intervention.
[b] Reported a priori by study authors.
[c] Cohen's d thresholds are 0.20 = small, 0.50 = medium and 0.80 = large effect.

Table 5
Results of the primary analysis: clinical relevance on the physical function outcome.

Study	Intervention	Physical function			Cut-off for clinical relevance[b], mean (scale)	Meets a priori cut-off	Meets the SWE values (20%)
		Within-group change, mean (%)		Difference in between-group change[a], mean (%)			
		Intervention	Comparison				
Smeets et al 2006[28]	Aerobic training and dynamic strengthening exercises	Only between-group change is reported	Only between-group change is reported	−2.4 (−17%)	2.5 (RMDQ)	No	No
Andrade et al 2008[29]	Back school	−2.4 (−23.0%)	−1.2 (−11.0%)	−1.2 (−12.0%)	Not reported	NA	No
Harts et al 2008[30]	High–intensity training	−2.8 (−45.0%)	−1.3 (−20.0%)	−1.5 (−25.0%)	2 (RMDQ)	No	Yes
	Low–intensity training	−1.5 (−19.7%)	−1.3 (−20.0%)	0.2 (0.3%)	2 (RMDQ)	No	No
Hall et al 2011[47]	Tai Chi	−3.2 (−31.4%)	−1.0 (−11.0%)	−2.2 (−20.4%)	1 (RMDQ)	Yes	Yes
Masharawi et al 2013[32]	Active non-weight-bearing group exercise	−4.9 (−34.5%)	−0.5 (−3.3%)	−4.4 (−31.4%)	Not reported	NA	Yes
Steele et al 2013[33]	Exercise program full ROM	−18.2 (−50.0%)	6.7 (35.0%)	−24.9 (−85.0%)	10 (ODI)	Yes	Yes
	Exercise program limited ROM	−12.0 (44.6%)	6.7 (35.0%)	−18.7 (−79.6%)	10 (ODI)	Yes	Yes
Lawand et al 2015[35]	Global postural re-education	−5.4 (−42.9%)	−1.0 (−8.4%)	−4.4 (−34.5%)	5 (RMDQ)	No	Yes
Kofotolis et al 2016[36]	Pilates	−8.0 (−70.1%)	−1.2 (−10.6%)	−6.8 (−59.5%)	Not reported	NA	Yes
	General strengthening exercise	−7.5 (−60.4%)	−1.2 (−10.6%)	−6.3 (−49.8%)	Not reported	NA	Yes
Segal-Snir et al 2016[37]	Group exercise	−2.0 (−15.4%)	0 (0%)	−2.0 (−15.4%)	11 (RMDQ)	No	No
Xia et al 2016[50]	High-velocity, low-amplitude spinal manipulation	−4.0 (−40.8%)	−0.8 (−8.3%)	−3.2 (−32.5%)	2 to 3 points (RMDQ)	Yes	Yes
	Low-velocity, low-amplitude spinal manipulation	−3.7 (−38.9%)	−0.8 (−8.3%)	−2.9 (−30.6%)	2 to 3 points (RMDQ)	Yes	Yes
Cortell-Tormo et al 2018[39]	Functional resistance training	−9.5 (−61.3%)	0.5 (3.5%)	−10.0 (−64.5%)	Not reported	NA	Yes
Cruz-Diaz et al 2017[40]	Mat-based Pilates	−3.5 (−30.1%)	−0.2 (−1.9%)	−3.3 (−28.2%)	Not reported	NA	Yes
	Equipment-based Pilates	−4.5 (−40.2%)	−0.2 (−1.9%)	−4.3 (−38.3%)	Not reported	NA	Yes
Keane et al 2017[48]	Land-based stretching exercise	−8.2 (−24.1%)	−8.4 (−21%)	0.2 (−3.1%)	Not reported	NA	No
Mazloum et al 2018[41]	Selective Pilates	−8.1 (−26.2%)	0.4 (1.5%)	−8.5 (−27.7%)	Not reported	NA	Yes
	Extension-based exercises	−4.0 (−14.7%)	0.4 (1.5%)	−4.4 (−16.2%)	Not reported	NA	No
Tavares et al 2017[49]	Spinal mobilisation	−8.2 (−72.6%)	−2.6 (−36.6%)	−5.6 (−36%)	30% difference before and after treatment (ODI)	Yes	Yes
Noormohammadpour et al 2018[42]	Retraining of trunk muscles	−6.1 (−78.0%)	−1.6 (−16.8%)	−4.5 (−61.2%)	Not reported	NA	Yes
Vining et al 2020[45]	Chiropractic care	−3.2 (−49%)	−1.0 (−13,2%)	−2.2 (−36%)	≥ 20% difference in mean change between groups	Yes	Yes
Hatefi et al 2021[46]	Stretching exercises	−12.8 (−43.4%)	−3.2 (−11.4%)	−9.6 (−32%)	Not reported	NA	Yes

NA = not applicable, ODI = Oswestry Disability Index, RMDQ = Roland-Morris Disability Questionnaire
[a] Negative values represent an improvement in favour of the intervention.
[b] Reported a priori by study authors.

Table 6
Results of secondary analysis. Power calculation of the included studies using the 20% smallest worthwhile effect threshold as between-group difference.

Study	p-value cut-off	Primary outcome	Study power calculated a posteriori with SWE = 20% [a]	Well-powered or underpowered [b]
Smeets et al 2006[28]	0.05	Physical function	37%	Underpowered
Andrade et al 2008[29]	0.05	Not reported [c]	55%	Underpowered
Harts et al 2008[30]	0.05	Physical function	20%	Underpowered
Hall et al 2011[47]	0.05	Other: bothersomeness [c]	60%	Underpowered
Lee et al 2011[31]	0.05	Pain intensity	15%	Underpowered
Steele et al 2013[33]	0.05	Physical function	18%	Underpowered
Moussouli et al 2014[34]	0.05	Pain intensity	12%	Underpowered
Lawand et al 2015[35]	0.05	Pain intensity	64%	Underpowered
Kofotolis et al 2016[36]	0.05	Not reported [c]	51%	Underpowered
Segal-Snir et al 2016[37]	0.05	Not reported [c]	35%	Underpowered
Xia et al 2016[50]	0.05	Physical function	70%	Underpowered
Arampatzis et al 2017[38]	0.05	Pain intensity	33%	Underpowered
Cortell-Tormo et al 2018[39]	0.05	Not reported [c]	9%	Underpowered
Cruz-Diaz et al 2017[40]	0.05	Pain intensity	98%	Well powered
Keane et al 2017[48]	0.05	Not reported [c]	15%	Underpowered
Mazloum et al 2018[41]	0.05	Not reported [c]	90%	Well powered
Tavares et al 2017[49]	0.05	Pain intensity	13%	Underpowered
Noormohammadpour et al 2018[42]	0.05	Pain intensity	13%	Underpowered
Liu et al 2019[43]	0.05	Pain intensity	96%	Well powered
Shariat et al 2019[44]	0.05	Pain intensity	90%	Well powered
Vining et al 2020[45]	0.05	Other: muscle strength [c]	63%	Underpowered
Hatefi et al 2021[46]	0.05	Not reported [c]	74%	Underpowered

[a] The power of each included study was calculated considering the p-value cut-off used in the study, the number of participants for each group, the between-group difference of 20% on the primary outcome, assuming the standard deviation of the control group at the follow-up.
[b] Each study was classified as underpowered if the power achieved was $< 80\%$ and well-powered if the difference was $\geq 80\%$.
[c] When the primary outcome was not reported or different from pain intensity, physical function and time to recovery, the power was calculated based on physical functioning.

systems.[59] Also, it should be considered that the SWE of the studies by Ferreira et al[19] and Christiansen et al[18] had an interquartile range of 10 to 30%. Using the range instead of the median (ie, 20%) could change the interpretation. Therefore, we decided to focus our analyses (Tables 4 and 5) on the point estimates of within-group changes and between-group differences without taking the 95% confidence intervals of these changes/differences into account. The reason for this is that the SWE 20% difference is not provided alongside a confidence interval (although with an interquartile range), which would make the comparison with confidence intervals difficult. Lastly, physiotherapy was defined according to the script used in the articles by Ferreira[19] and Christiansen,[18] in which it was defined as a mix of exercise and manual therapy. We included studies mainly focused on exercise therapy since we did not retrieve any study in which exercise was delivered together with manual therapy (and compared against no intervention). However, exercise therapy is one of the most frequently recommended interventions for people with CLBP[60] and a recent study from Hansford et al[61] found the same SWE median value (ie, 20%) for exercise compared with no intervention.

The results of this review could have important clinical implications and enable recommendations to be made to clinicians and researchers. First, clinicians are encouraged to prudently evaluate and consider clinical relevance when they read and assess an RCT involving patients with CLBP. Researchers should carefully plan RCTs to detect the SWE of intervention, informing readers on whether the effect of an intervention is large enough to be worthwhile. This should be done using approaches based on benefit-harm trade-off methods (eg, SWE) rather than anchor-based methods (eg, MID) that cannot tell us whether patients feel that the effect of an intervention is large enough to make the costs, risks and inconveniences associated with intervention worthwhile.[12] Moreover, this should be done a priori and clearly specified in the trial protocol.

In conclusion, physiotherapy interventions seem to provide clinically relevant results when compared with no intervention in people with CLBP. However, in RCTs comparing physiotherapy with no intervention, little attention is given to the interpretation of clinical relevance, with great heterogeneity in frameworks and thresholds. It is vital to use a reliable, comprehensive, patient-driven and context-specific approach (eg, the SWE approach) to consider and interpret clinical relevance in clinical trials because patients and clinicians should be informed on whether the effect of an intervention is large enough to be worthwhile.

What was already known on this topic: To determine the effectiveness of an intervention on specific outcomes, clinicians, researchers and policymakers should judge the clinical relevance of a trial's results. Clinical relevance indicates whether the intervention should be used in clinical practice to improve healthcare outcomes.
What this study adds: Physiotherapy interventions provide clinically relevant results when compared with no intervention in people with chronic low back pain. However, in trials comparing physiotherapy with no intervention, little attention is given to the interpretation of clinical relevance, with great heterogeneity in frameworks and thresholds used.

Footnotes: [a] EndNote software, Clarivate, Philadelphia, USA.
[b] IBM SPSS Statistics for Macintosh, Version 28.0, IBM Corp, Armonk, USA.
eAddenda: 2 (Tables 1 and 2).
Ethics approval: Not applicable.
Competing interests: Nil.
Source(s) of support: This research received no specific grant from any funding agency in public, commercial or not-for-profit sectors.
Acknowledgements: Nil.
Provenance: Not invited. Peer reviewed.
Correspondence: Tiziano Innocenti, Department of Health Sciences, Vrije Universiteit, Amsterdam, The Netherlands. Email: t.innocenti@vu.nl

References

1. Chiarotto A, Koes BW. Nonspecific low back pain. *New Engl J Med*. 2022;386:1732–1740. https://doi.org/10.1056/NEJMcp2032396.
2. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, et al. What low back pain is and why we need to pay attention. *Lancet*. 2018;391:2356–2367. https://doi.org/10.1016/S0140-6736(18)30480-X.
3. Collaborators GBDLBP. Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. *Lancet Rheumatol*. 2023;5:e316–e329. https://doi.org/10.1016/S2665-9913(23)00098-X.
4. Itz CJ, Geurts JW, van Kleef M, Nelemans P. Clinical course of non-specific low back pain: a systematic review of prospective cohort studies set in primary care. *Eur J Pain*. 2013;17:5–15. https://doi.org/10.1002/j.1532-2149.2012.00170.x.

5. Artus M, van der Windt DA, Jordan KP, Hay EM. Low back pain symptoms show a similar pattern of improvement following a wide range of primary care treatments: a systematic review of randomized clinical trials. *Rheumatology*. 2010;49:2346–2356. https://doi.org/10.1093/rheumatology/keq245.
6. Kendall JM. Designing a research project: randomised controlled trials and their principles. *Emerg Med J*. 2003;20:164–168. https://doi.org/10.1136/emj.20.2.164.
7. Leung W-C. Balancing statistical and clinical significance in evaluating treatment effects. *Postgrad Med J*. 2001;77:201–204. https://doi.org/10.1136/pmj.77.905.201.
8. Kamper SJ. Interpreting Outcomes 2-Statistical Significance and Clinical Meaningfulness: Linking Evidence to Practice. *J Orthop Sports Phys Ther*. 2019;49:559–560. https://doi.org/10.2519/jospt.2019.0704.
9. Smith SM, Dworkin RH, Turk DC, McDermott M, Eccleston C, Farrar JT, et al. Interpretation of chronic pain clinical trial outcomes: IMMPACT recommended considerations. *Pain*. 2020;161:2446–2461. https://doi.org/10.1097/j.pain.0000000000001952.
10. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407–415. https://doi.org/10.1016/0197-2456(89)90005-6.
11. Barrett B, Brown D, Mundt M, Brown R. Sufficiently important difference: expanding the framework of clinical significance. *Med Decis Making*. 2005;25:250–261. https://doi.org/10.1177/0272989X05276863.
12. Ferreira ML, Herbert RD, Ferreira PH, Latimer J, Ostelo RW, Nascimento DP, et al. A critical review of methods used to determine the smallest worthwhile effect of interventions for low back pain. *J Clin Epidemiol*. 2012;65:253–261. https://doi.org/10.1016/j.jclinepi.2011.06.018.
13. Schunemann HJ, Guyatt GH. Commentary–goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res*. 2005;40:593–597. https://doi.org/10.1111/j.1475-6773.2005.00374.x.
14. Kamper SJ. Interpreting Outcomes 3-Clinical Meaningfulness: Linking Evidence to Practice. *J Orthop Sports Phys Ther*. 2019;49:677–678. https://doi.org/10.2519/jospt.2019.0705.
15. Ferreira ML, Herbert RD. What does 'clinically important' really mean? *Aust J Physiother*. 2008;54:229–230. https://doi.org/10.1016/s0004-9514(08)70001-1.
16. Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2009;146:238–244. https://doi.org/10.1016/j.pain.2009.08.019.
17. Barrett B, Brown R, Mundt M, Dye L, Alt J, Safdar N, et al. Using benefit harm tradeoffs to estimate sufficiently important difference: the case of the common cold. *Med Decis Making*. 2005;25:47–55. https://doi.org/10.1177/0272989X04273147.
18. Christiansen DH, de Vos Andersen NB, Poulsen PH, Ostelo RW. The smallest worthwhile effect of primary care physiotherapy did not differ across musculoskeletal pain sites. *J Clin Epidemiol*. 2018;101:44–52. https://doi.org/10.1016/j.jclinepi.2018.05.019.
19. Ferreira ML, Herbert RD, Ferreira PH, Latimer J, Ostelo RW, Grotle M, et al. The smallest worthwhile effect of nonsteroidal anti-inflammatory drugs and physiotherapy for chronic low back pain: a benefit-harm trade-off study. *J Clin Epidemiol*. 2013;66:1397–1404. https://doi.org/10.1016/j.jclinepi.2013.02.018.
20. Innocenti T, Schleimer T, Salvioli S, Giagio S, Ostelo R, Chiarotto A. The interpretation of clinical relevance in randomised clinical trials in patients with chronic low back pain: protocol for a meta-research study. *medRxiv*. 2023:2022.12.14.22283454. https://doi.org/10.1101/2022.12.14.22283454.
21. Innocenti T, Schleimer T, Salvioli S, Giagio S, Ostelo R, Chiarotto A. The interpretation of clinical relevance in randomised clinical trials in patients with chronic low back pain: protocol for a meta-research study. *MethodsX*. 2023/01/01/ 2023;10:102239. https://doi.org/10.1016/j.mex.2023.102239.
22. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. *Evid Based Med*. 2017;22:139–142. https://doi.org/10.1136/ebmed-2017-110713.
23. Hayden JA, Ellis J, Ogilvie R, Malmivaara A, van Tulder MW. Exercise therapy for chronic low back pain. *Cochrane Database Syst Rev*. 2021;9:CD009790. https://doi.org/10.1002/14651858.CD009790.pub2.
24. Rubinstein SM, de Zoete A, van Middelkoop M, Assendelft WJJ, de Boer MR, van Tulder MW. Benefits and harms of spinal manipulative therapy for the treatment of chronic low back pain: systematic review and meta-analysis of randomised controlled trials. *BMJ*. 2019;364:l689. https://doi.org/10.1136/bmj.l689.
25. Glanville J, Kotas E, Featherstone R, Dooley G. Which are the most sensitive search filters to identify randomized controlled trials in MEDLINE? *J Med Libr Assoc*. 2020;108:556–563. https://doi.org/10.5195/jmla.2020.912.
26. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev*. 2016;5:210. https://doi.org/10.1186/s13643-016-0384-4.
27. Faul F, Erdfelder E, Lang AG, Buchner AG. *Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39:175–191. https://doi.org/10.3758/bf03193146.
28. Smeets RJEM, Vlaeyen JWS, Hidding A, Kester AD, van der Heijden GJ, van Geel AC, et al. Active rehabilitation for chronic low back pain: Cognitive-behavioral, physical, or both? First direct post-treatment results from a randomized controlled trial [ISRCTN22714229]. *BMC Musculoskel Disord*. 2006;7:5. https://doi.org/10.1186/1471-2474-7-5.
29. Andrade SC, Araújo AG, Vilar MJ. Escola de coluna para pacientes com lombalgia crônica inespecífica: benefícios da associação de exercícios e educação ao paciente. *Acta Reumatol Port*. 2008;33:443–450.
30. Harts CC, Helmhout PH, de Bie RA, Staal JB. A high-intensity lumbar extensor strengthening program is little better than a low-intensity program or a waiting list control group for chronic low back pain: a randomised clinical trial. *Aust J Physiother*. 2008;54:23–31. https://doi.org/10.1016/s0004-9514(08)70062-x.
31. Lee T, Kim YH, Sung PS. A comparison of pain level and entropy changes following core stability exercise intervention. *Med Sci Monitor*. 2011;17:Cr362. https://doi.org/10.12659/msm.881846.
32. Masharawi Y, Nadaf N. The effect of non-weight bearing group-exercising on females with non-specific chronic low back pain: a randomized single blind controlled pilot study. *J Back Musculoskel Rehabil*. 2013;26:353–359. https://doi.org/10.3233/bmr-130391.
33. Steele J, Bruce-Low S, Smith D, Jessop D, Osborne N. A randomized controlled trial of limited range of motion lumbar extension exercise in chronic low back pain. *Spine*. 2013;38:1245–1252. https://doi.org/10.1097/BRS.0b013e318291b526.
34. Moussouli M, Vlachopoulos SP, Kofotolis ND, Theodorakis Y, Malliou P, Kellis E. Effects of stabilization exercises on health-related quality of life in women with chronic low back pain. *J Phys Act Health*. 2014;11:1295–1303. https://doi.org/10.1123/jpah.2012-0426.
35. Lawand P, Lombardi Júnior I, Jones A, Sardim C, Ribeiro LH, Natour J. Effect of a muscle stretching program using the global postural reeducation method for patients with chronic low back pain: A randomized controlled trial. *Joint Bone Spine*. 2015;82:272–277. https://doi.org/10.1016/j.jbspin.2015.01.015.
36. Kofotolis N, Kellis E, Vlachopoulos SP, Gouitas I, Theodorakis Y. Effects of Pilates and trunk strengthening exercises on health-related quality of life in women with chronic low back pain. *J Back Musculoskel Rehabil*. 2016;29:649–659. https://doi.org/10.3233/bmr-160665.
37. Segal-Snir Y, Lubetzky VA, Masharawi Y. Rotation exercise classes did not improve function in women with non-specific chronic low back pain: A randomized single blind controlled study. *J Back Musculoskel Rehabil*. 2016;29:467–475. https://doi.org/10.3233/bmr-150642.
38. Arampatzis A, Schroll A, Catalá MM, Laube G, Schüler S, Dreinhofer K. A random-perturbation therapy in chronic non-specific low-back pain patients: a randomised controlled trial. *Eur J Appl Physiol*. 2017;117:2547–2560. https://doi.org/10.1007/s00421-017-3742-6.
39. Cortell-Tormo JM, Sánchez PT, Chulvi-Medrano I, Tortosa-Martínez J, Manchado-López C, et al. Effects of functional resistance training on fitness and quality of life in females with chronic nonspecific low-back pain. *J Back Musculoskel Rehabil*. 2018;31:95–105. https://doi.org/10.3233/bmr-169684.
40. Cruz-Díaz D, Bergamin M, Gobbo S, Martínez-Amat A, Hita-Contreras F. Comparative effects of 12 weeks of equipment based and mat Pilates in patients with chronic low back pain on pain, function and transversus abdominis activation. A randomized controlled trial. *Complement Ther Med*. 2017;33:72–77. https://doi.org/10.1016/j.ctim.2017.06.004.
41. Mazloum V, Sahebozamani M, Barati A, Nakhaee N, Rabiei P. The effects of selective Pilates versus extension-based exercises on rehabilitation of low back pain. *J Bodyw Mov Ther*. 2018;22:999–1003. https://doi.org/10.1016/j.jbmt.2017.09.012.
42. Noormohammadpour P, Kordi M, Mansournia MA, Akbari-Fakhrabadi M, Kordi R. The role of a multi-step core stability exercise program in the treatment of nurses with chronic low back pain: a single-blinded randomized controlled trial. *Asian Spine J*. 2018;12:490–502. https://doi.org/10.4184/asj.2018.12.3.490.
43. Liu J, Yeung A, Xiao T, Tian X, Kong Z, Zou L, et al. Chen-style tai chi for individuals (aged 50 years old or above) with chronic non-specific low back pain: a randomized controlled trial. *Int J Environ Res Public Health*. 2019;16:517. https://doi.org/10.3390/ijerph16030517.
44. Shariat A, Alizadeh R, Moradi V, Afsharnia E, Hakakzadeh A, Ansari NN, et al. The impact of modified exercise and relaxation therapy on chronic lower back pain in office workers: a randomized clinical trial. *J Exerc Rehabil*. 2019;15:703–708. https://doi.org/10.12965/jer.1938490.245.
45. Vining R, Long CR, Minkalis A, Gudavalli MR, Xia T, Walter J, et al. Effects of chiropractic care on strength, balance, and endurance in active-duty US military personnel with low back pain: a randomized controlled trial. *J Altern Complement Med*. 2020;26:592–601. https://doi.org/10.1089/acm.2020.0107.
46. Hatefi M, Babakhani F, Ashrafizadeh M. The effect of static stretching exercises on hip range of motion, pain, and disability in patients with non-specific low back pain. *J Exp Orthop*. 2021;8:55. https://doi.org/10.1186/s40634-021-00371-w.
47. Hall AM, Maher CG, Lam P, Ferreira M, Latimer J. Tai chi exercise for treatment of pain and disability in people with persistent low back pain: a randomized controlled trial. *Arthritis Care Res*. 2011;63:1576–1583. https://doi.org/10.1002/acr.20594.
48. Keane LG. Comparing AquaStretch with supervised land based stretching for chronic lower back pain. *J Bodyw Mov Ther*. 2017;21:297–305. https://doi.org/10.1016/j.jbmt.2016.07.004.
49. Tavares FAG, Chaves TC, Silva ED, Guerreiro GD, Gonçalves JF, Alves de Albuquerque AA. Immediate effects of joint mobilization compared to sham and control intervention for pain intensity and disability in chronic low back pain patients: randomized controlled clinical trial. *Rev Dor São Paulo*. 2017;18:2–7.
50. Xia T, Long CR, Gudavalli MR, Wilder DG, Vining RD, Rowell RM, et al. Similar effects of thrust and nonthrust spinal manipulation found in adults with subacute and chronic low back pain: a controlled trial with adaptive allocation. *Spine*. 2016;41:E702–E709. https://doi.org/10.1097/BRS.0000000000001373.
51. Cohen J. *Statistical Power Analysis for the Behavioural Science*. 2nd ed. NJ, USA: Lawrence Erlbaum Associates, Inc; 1988.
52. Ostelo RW, Deyo RA, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine*. 2008;33:90–94. https://doi.org/10.1097/BRS.0b013e31815e3a10.
53. Chiarotto A, Deyo RA, Terwee CB, Boers M, Buchbinder R, Corbin TP, et al. Core outcome domains for clinical trials in non-specific low back pain. *Eur Spine J*. 2015;24:1127–1142. https://doi.org/10.1007/s00586-015-3892-3.
54. Dworkin JD, McKeown A, Farrar JT, Gilron I, Hunsinger M, Kerns RD, et al. Deficiencies in reporting of statistical methodology in recent randomized trials of nonpharmacologic pain treatments: ACTTION systematic review. *J Clin Epidemiol*. 2016;72:56–65. https://doi.org/10.1016/j.jclinepi.2015.10.019.
55. Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS. Interpreting treatment effects in randomised trials. *BMJ*. 1998;316:690–693. https://doi.org/10.1136/bmj.316.7132.690.
56. Henderson N, Riddle DL. The smallest worthwhile effect is superior to the MCID for estimating acceptable benefits of knee arthroplasty. *J Clin Epidemiol*. 2022;152:201–208. https://doi.org/10.1016/j.jclinepi.2022.10.019.
57. Froud R, Rajendran D, Patel S, Bright P, Bjørkli T, Eldridge S, et al. The power of low back pain trials: a systematic review of power, sample size, and reporting of

sample size calculations over time, in trials published between 1980 and 2012. *Spine*. 2017;42:E680–E686. https://doi.org/10.1097/brs.0000000000001953.
58. Castellini G, Gianola S, Banfi G, Bonovas S, Moja L. Mechanical low back pain: secular trend and intervention topics of randomized controlled trials. *Physiother Can*. 2016;68:61–63. https://doi.org/10.3138/ptc.2014-58.
59. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol*. 2010;63:524–534. https://doi.org/10.1016/j.jclinepi.2009.08.010.
60. Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin CW, Chenot JF, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. *Eur Spine J*. 2018;27:2791–2803. https://doi.org/10.1007/s00586-018-5673-2.
61. Hansford HJ, Jones MD, Cashin AG, Ostelo RW, Chiarotto A, Williams SA, et al. The smallest worthwhile effect on pain intensity of nonsteroidal anti-inflammatory drugs and exercise therapy for acute and chronic low back pain: a benefit-harm trade-off study. *J Physiother*. 2023;69:240–248. https://doi.org/10.1016/j.jphys.2023.08.006.