Manipulation and mobilization for treating chronic low back pain: a systematic review and meta-analysis

Ian D. Coulter, PhD^a,b,c,*  Cindy Crawford, BA^a  Eric L. Hurwitz, DC, PhD^a,d  Howard Vernon, DC, PhD^a,e  Raheleh Khorsan, PhD^f  Marika Suttorp Booth, MS^a  Patricia M. Herman, ND, PhD^a

^aRAND Corporation, 1776 Main St, Santa Monica, CA 90407-2138, USA
^bUCLA School of Dentistry, Box 951668, Los Angeles, CA 90095-1668, USA
^cSouthern California University of Health Sciences, 16200 Amber Valley Dr, Whittier, CA 90604, USA
^dOffice of Public Health Studies, University of Hawai‘i, Mānoa, 1960 East-West Rd, Biomed D104AA, Honolulu, HI 96822, USA
^eDivision of Research, Canadian Memorial Chiropractic College, 6100 Leslie St, Toronto, ON, Canada M2H 3J1
^fUCI Department of Urban Planning and Public Policy, 300 Social Ecology I, Irvine, CA 92697-7075, USA

Received 31 August 2017; revised 7 December 2017; accepted 11 January 2018

Abstract

BACKGROUND CONTEXT: Mobilization and manipulation therapies are widely used to benefit patients with chronic low back pain. However, questions remain about their efficacy, dosing, safety, and how these approaches compare with other therapies.

PURPOSE: The present study aims to determine the efficacy, effectiveness, and safety of various mobilization and manipulation therapies for treatment of chronic low back pain.

STUDY DESIGN/SETTING: This is a systematic literature review and meta-analysis.

OUTCOME MEASURES: The present study measures self-reported pain, function, health-related quality of life, and adverse events.

METHODS: We identified studies by searching multiple electronic databases from January 2000 to March 2017, examining reference lists, and communicating with experts. We selected randomized controlled trials comparing manipulation or mobilization therapies with sham, no treatment, other active therapies, and multimodal therapeutic approaches. We assessed risk of bias using Scottish Intercollegiate Guidelines Network criteria. Where possible, we pooled data using random-effects meta-analysis. Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) was applied to determine the confidence in effect estimates. This project is funded by the National Center for Complementary and Integrative Health under Award Number U19AT007912.

RESULTS: Fifty-one trials were included in the systematic review. Nine trials (1,176 patients) provided sufficient data and were judged similar enough to be pooled for meta-analysis. The standardized mean difference for a reduction of pain was SMD = −0.28, 95% confidence interval (CI) −0.47 to −0.09, p = .004; I² = 57% after treatment; within seven trials (923 patients), the reduction in disability was SMD = −0.33, 95% CI −0.63 to −0.03, p = .03; I² = 78% for manipulation or mobilization compared with other active therapies. Subgroup analyses showed that manipulation significantly reduced pain and disability, compared with other active comparators including exercise and physical therapy (SMD = −0.43, 95% CI −0.86 to 0.00; p = .05, I² = 79%; SMD = −0.86, 95% CI −1.27 to −0.45; p < .0001, I² = 46%). Mobilization interventions, compared with other active comparators including exercise regimens, significantly reduced pain (SMD = −0.20, 95% CI −0.35 to −0.04; p = .01; I² = 0%) but not disability (SMD = −0.10, 95% CI −0.28 to 0.07; p = .25; I² = 21%). Studies comparing manipulation or mobilization with sham or no treatment were too few or too heterogeneous to allow for pooling as were studies examining

* Corresponding author. RAND/Samuels Chair for Integrative Medicine, RAND Corporation, 1776 Main St, Santa Monica, CA 90407-2138, USA.
Tel: 310-393-0411.
E-mail address: coulter@rand.org (I.D. Coulter).

https://doi.org/10.1016/j.spinee.2018.01.013
1529-9430/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
relationships between dose and outcomes. Few studies assessed health-related quality of life. Twenty-six of 51 trials were multimodal studies and narratively described.

**CONCLUSION:** There is moderate-quality evidence that manipulation and mobilization are likely to reduce pain and improve function for patients with chronic low back pain; mobilization appears to produce a larger effect than mobilization. Both therapies appear safe. Multimodal programs may be a promising option. © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

**Keywords:** Chiropractic; Chronic low back pain; Manipulation; Meta-analysis; Mobilization; Systematic review

---

**Introduction**

Similar to conclusions reported most recently in BMJ [1], a recent review published in JAMA reported that among patients with acute low back pain, spinal manipulative therapy was associated with modest improvements in pain and function at up to 6 weeks, with transient minor musculoskeletal harms. However, study results showed substantial heterogeneity [2]. The review did not address the efficacy of manipulation and mobilization for chronic low back pain. Given the current interest in non-pharmacological alternatives for the treatment of chronic pain, in particular non-opioid treatments [3], a systematic review of manipulation and mobilization for chronic low back pain is timely.

The lifetime prevalence of low back pain in the United States may be as high as 84%. The prevalence of chronic low back pain is about 23%; it disables 11%–12% of the population [4]. A recent systematic review of the clinical course of non-specific low back pain found that in the first 3 months, 33% of patients showed recovery, but 1 year after onset, 65% still reported pain [5]. The severity, length, or duration of pain for any one individual varies, and the transition from acute to chronic low back pain is difficult to determine [6].

Pain management approaches vary greatly. Many physicians rely on non-steroidal anti-inflammatory drugs, opioid, and neurotropic medications, or steroid injections and surgery as their main tools [7]. Because of the potential or apparent risks associated with these tools [8], non-pharmacological approaches, thought to involve minimal adverse events, have become popular. In recent years, multiple studies have explored the evidence for treating chronic low back pain; options include spinal manipulation therapy, behavioral therapy, exercise therapy, transcutaneous electrical nerve stimulation, interferential currents, low-level laser therapy, and yoga [9]. Other therapies include massage, acupuncture, and superficial heat therapy (eg, thermo heat wraps, hot water bottles, heated packs filled with grain, hot towels, and electric heating pads) [10]. Manual modalities such as physiotherapy, massage, chiropractic, occupational, and osteopathic therapies, including spinal manipulation and mobilization, are often used together and alone to treat chronic non-specific low back pain [11,12].

Several systematic reviews have focused on manual therapies such as spinal manipulation and mobilization for treating back and neck pain [13–16]. Earlier work suggested that there is little or no evidence that spinal manipulative therapy was superior to other standard treatments for chronic low back pain [17,18]; however, recent systematic reviews suggest that spinal manipulation and mobilization are “viable” options for pain treatment [13,19]. However, the efficacy of manipulation and mobilization may vary depending on the duration of symptoms, how the intervention is administered (eg, whether there is additional exercise or general practitioner care, at what dosages, and follow-up periods), the comparator, and types of outcomes reported. Such variability could be considered inconsistent findings; however, the overall evidence suggests that manipulation and mobilization are effective treatment modalities compared with other therapies [13,19].

The purpose of the systematic review described here was to disentangle inconsistencies by evaluating the research according to different symptom durations across the spectrum of chronicity, the variations in treatment techniques, variations in comparators, and the impact on important patient-reported outcomes. Our goal was to better understand the effectiveness of mobilization and manipulation for chronic non-specific low back pain as compared and reported in randomized controlled trials (RCTs) since 2000. We would attempt meta-analysis when there were subsets of data similar enough to pool. This systematic review is part of a larger project investigating the appropriateness of manipulation or mobilization for the treatment of chronic low back pain and cervical pain, funded by the National Center for Complementary and Integrative Health under Award Number U19AT007912.

**Methods**

This systematic review and meta-analysis report adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines.

**Search strategy and data sources**

The systematic review builds upon previous systematic reviews of manipulation and mobilization for chronic low back pain (up through 2000), such as Bronfort et al. [13] and Shekelle et al. [18,20,21]. We designed a broad search strategy that did not define the specific population (ie, not using the word chronic or non-specific) or intervention (ie, spanning across multiple professions). In addition, we placed no limitations on control or comparators, outcomes, or study
designs, so that the breadth and variations across the research could be discovered, and the literature could inform the appropriate definitions and subgroups to consider for analysis given that inconsistencies are present. We searched PubMed or MEDLINE, Cochrane, Embase, CINAHL, PsycINFO, and ICL from January 2000 through March 2017. We drew on reference lists and consultation with subject matter experts to ensure comprehensiveness (Fig. 1). Because the National Institutes of Health–funded project focused on both chronic neck pain and chronic low back pain, we executed the search to meet both needs together (Table 1).

Scoping review

A scoping review of the literature informed the definitions of chronicity used in the review. It also clarified what is considered non-specific, and what subgroups should be considered for systematic review and meta-analysis. We reviewed articles and categorized studies according to specific populations, interventions, control or comparators, outcomes, and study designs. We excluded studies clearly not related to back pain or to an intervention involving mobilization or manipulation. We presented findings to an internal steering committee (ISC) and an external advisory committee (EAC), where evidence-informed definitions and specific research questions were devised based on the evidence base for carrying out systematic review and meta-analysis (Table 1).

Study selection

Six reviewers used study eligibility criteria to independently screen the literature in duplicate. Disagreements about inclusion were resolved through discussion and consensus, or if necessary, by the ISC. The eligibility criteria included (1) a population experiencing chronic [6,22] and non-specific [23] low back pain (as defined in Table 1); (2) an intervention, with the involvement of a therapist, consisting of either (i) manipulation (labeled as thrust), (ii) mobilization (labeled as non-thrust), or (iii) a multimodal integrative practice including manipulation or mobilization components, labeled as a multimodal “program” if the observed effect could not be attributed directly to the thrust or non-thrust intervention (eg, a study of chiropractic plus acupuncture vs. usual care would be multimodal and labeled as a “program” separate from chiropractic plus acupuncture vs. acupuncture); (3) compared with a sham treatment, no treatment, or


Note: Figure 1 addresses search strategy for low back pain as well as neck pain studies. The findings of neck pain studies are not reported here. Because the Center of Excellence for Research in CAM (CERC) project was focused on both chronic neck pain as well as chronic low back pain, the search was executed to meet both needs together to streamline the effort.

Fig. 1. Search strategy.
<table>
<thead>
<tr>
<th>Eligibility criteria</th>
<th>Reference standard definition</th>
<th>Scope driven evidence-informed definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population “chronic” low back pain</td>
<td>According to the Pain Management Task Force [6], chronic pain can be described as ongoing or recurrent pain, lasting beyond the usual course of acute illness or injury or more than 3–6 mo, and which adversely affects the individual’s well-being. In 2014, the National Institutes of Health Task Force on Research Standards for Low Back [22] recommended defining chronicity of pain as: “(1) How long has back pain been an ongoing problem for you? (2) How often has low-back pain been an ongoing problem for you over the past 6 months? A response of greater than 3 months to question 1, and a response of ‘at least half the days in the past 6 months’ to question 2 would define chronic low back pain.”</td>
<td>The majority of studies defined chronicity based on the duration of pain symptoms for 12 wk or more. Therefore, a similar definition of chronicity (≥12 wk) was adopted, and studies were categorized as those patients with &gt;12 wk, a mean duration of 6 mo, and those with &gt;12-mo pain duration.</td>
</tr>
<tr>
<td>Population “non-specific”</td>
<td>Non-specific low back pain is defined as pain not attributable to a recognizable, known specific pathology [23] (e.g., infection, tumor, osteoporosis, fracture, structural deformity, rheumatoid arthritis, radicular syndrome). Therefore, the etiology of the pain is often unknown and it is not categorized with a major pathogenic etiology.</td>
<td>The existing literature does not use standard terminology to report “non-specific” chronic pain. To guide the eligibility of studies, the following terms were specified to be outside the scope of “non-specific”: specific conditions, i.e., cancer, rheumatoid arthritis, fibromyalgia, spondylolisthesis (displacement of vertebra) and spinal stenosis (narrowing of spinal canal), pregnancy-related pain, and ankylosing spondylitis. Consensus among the internal steering committee specified the following exemptions: osteoarthritis, sciatica, radiculopathy, low back pain “of mechanical origin,” pain associated with spondylolisthesis, sacroiliac joint syndrome, trauma-induced, disc herniation, pelvic anteversion, and “occupational” low back pain.</td>
</tr>
<tr>
<td>Interventions mobilization or manipulation</td>
<td>Bronfort et al. defined mobilization as “the application of manual force to the spinal joints within the passive range of joint motion that does not involve a thrust (p. 336)” [13]. The RAND report by Coulter et al. defines mobilization as “controlled, judiciously applied force of low velocity and variable amplitude directed to spinal joint segment(s)” (p. xi) [24]. Spinal manipulation is defined as “the application of high-velocity, low amplitude manual thrusts to the spinal joints slightly beyond the passive range of joint motion,” by Bronfort et al. [13], where the RAND report by Coulter et al. defines spinal manipulation as “a controlled, judiciously applied dynamic thrust adjustment, that may include combined extension and rotation of the upper cervical spinal segments, or low-velocity and low-amplitude force with the use of a short or long lever directed to spinal joint segments within patient tolerance” (p. xi) [24].</td>
<td>The interventions in this systematic review consist of manipulation or mobilization in chiropractic settings and other non-invasive therapies including osteopathy, manual therapy and physical therapy. For simplicity, interventions were categorized into thrust and non-thrust interventions. When combined with other active interventions, they were labeled as “programs.”</td>
</tr>
<tr>
<td>Control/comparator(s)</td>
<td>This review focused on any intervention being compared with mobilization or manipulation, including any active therapy (i.e., exercise, physical therapy), manipulation (thrust), mobilization (non-thrust), sham, no treatment, usual, or standard care.</td>
<td>For the purpose of analysis, controls or comparisons were categorized as active, sham, or no treatment, or as direct comparisons between various thrust and non-thrust interventions.</td>
</tr>
<tr>
<td>Outcome(s)</td>
<td>Although pain reduction was predefined as the primary outcome of interest, the most commonly reported pain-related, patient-reported outcomes that affect health status were determined through a scoping review and thus pooled to determine which could be assessed.</td>
<td>Patient-reported outcomes that the majority of studies include to date: pain intensity or severity (as measured by a VAS or NRS) disability (as measured by the RMDQ, HRQoL) as measured by the SF-36 or safety.</td>
</tr>
<tr>
<td>Study design(s)</td>
<td>All study designs were considered for the purposes of scoping the literature.</td>
<td>Randomized controlled trials were included in the systematic review and meta-analysis. Other study designs were queried when gaps were present (i.e., safety).</td>
</tr>
</tbody>
</table>

HRQoL, health-related quality of life; NRS, numeric rating scale; RMDQ, Roland-Morris Disability Questionnaire; SF-36, Short Form 36; VAS, visual analog scale.
other active therapies, such as exercise, physiotherapy, or physical therapy; (4) an RCT, involving adult human subjects (age 18 years or older); and (5) at least one pain outcome measuring a reduction in pain intensity or severity, such as the visual analog scale (VAS) or numeric rating scale (Table 1).

Data extraction and quality assessment

Six reviewers participated in data extraction and quality assessment of the individual studies. Population characteristics, treatment intervention(s), control or comparators, and outcomes were described for each included study. Quality assessment was performed in duplicate by reviewers; disagreements were tracked and resolved by the ISC. Risk of bias was assessed using the Scottish Intercollegiate Guidelines Network (SIGN 50) checklist for RCTs [25]. We assessed external validity using the External Validity Assessment Tool (EVAT) [26], which measures the generalizability of research to other individuals (external validity) and other settings (model validity) outside the confines of a study.

Data synthesis and analysis

The primary analysis was based on trials reporting a continuous outcome measure for pain intensity, disability, or health-related quality of life (HRQoL) [27] up to 1 month after the end of treatment. Subgroups were constructed to transparently report those studies where (1) chronicity duration is greater than 3 months or greater than 12 months, (2) intervention consisting of thrust or non-thrust, (3) and were compared with sham or no-treatment or another active intervention, for each outcome assessed. Secondary analyses were based on trials reporting a continuous measure for pain, disability, or HRQoL at follow-up points closest to 3, 6, and 12 months after treatment. Single treatment (one dose over one day) studies were excluded from any analysis. Data synthesis and analysis methods for those multimodal interventions where the effects of manipulation or mobilization could not be distinguished from the total program were not applied as these studies involved more pragmatic, program-type interventions and were heterogeneous from study to study. For simplicity and because many types and styles were included in systematic review, the authors chose to refer to the manipulation therapies as “thrust” and mobilization therapies as “non-thrust” (Table 1). We grouped studies in this way in order to attempt to create homogeneous subsets of studies, allowing us to ask questions about how interventions compare (eg, thrust vs. sham or no treatment, thrust vs. other active therapies); dose regimens and practitioner-specific techniques remained heterogeneous across studies (Supplementary Data Files 1 and 2). Some subgroups were judged too heterogeneous to pool in any meaningful way.

For all studies where data were available, we extracted sample size, mean, and standard deviation for each treatment group, at each time point reported. An unbiased estimate using the Hedges’ effect size [28] and 95% lower and upper limits was computed using intergroup differences between groups at those time points. For a reduction in pain intensity or disability, a negative effect size favors the thrust or non-thrust intervention more than the comparison arm (active comparator, sham, or no treatment group). For an increase in HRQoL, a positive effect size indicates benefit in thrust or non-thrust treatment group more than the comparison arm. This was done regardless of whether the study was considered for meta-analysis or not.

We considered a minimum of three studies judged similar enough in terms of the population, intervention, control or comparator, and outcome measure as sufficient for pooling data for meta-analysis. Standardized mean differences (SMDs) were computed using Comprehensive Meta-Analysis, Version 3.3.070 (CMA; Biostat, Englewood, NJ, USA). Meta-analyses of SMD were performed with the generic inverse model of REVMan. Because we expected heterogeneity, we used random-effects models; statistical heterogeneity was examined by I² with low, moderate, and high I² values of 25%, 50%, and 75%, respectively. Publication bias was assessed using the Begg adjusted rank correlation test [29] and the Egger regression asymmetry test [30]. Pooled effect sizes for pain and disability outcomes were translated into the VAS (0–100) using a standard deviation of 25 points, and the Roland-Morris Disability Questionnaire (0–24), using a standard deviation of 6 points, respectively, for clinical interpretation [17].

We assessed confidence in the effect estimates using the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach based on the following criteria: risk of bias, inconsistency, indirectness, imprecision, and publication bias, for each outcome [31].

Results

A total of 7,360 published citations were captured from January 2000 to March 2017, of which 6,663 were screened at title or abstract level according to the broad eligibility criteria. This search also included studies on neck pain, which are not reported here. We retrieved full text for 711 articles; of these, 334 were identified as RCTs on either neck pain or low back pain. After removing the studies on neck pain, a total of 64 RCT publications reporting on patients with chronic, non-specific low back pain were included in this systematic review. Of these, 32 reports [32–63] including 25 unique studies [32–56] were labeled as non-specific chronic low back pain unimodal (ie, thrust or non-thrust) studies; 32 reports [64–95] including 26 unique studies [64,66–90] were multimodal (ie, programs involving thrust or non-thrust) chronic low back pain studies. In addition, non-randomized studies were identified as part of a larger effort, and investigated to gain further understanding of safety issues and more pragmatic “real world” implications beyond that which might be offered in the RCTs evaluated (data not shown) (Fig. 2).
Study characteristics

Characteristics of included studies are detailed in Supplementary Data Files 1 and 2. The included RCTs examining either a uni- or multimodal intervention of thrust or non-thrust for patients with chronic low back pain were published between 2000 and 2014. No studies meeting eligibility criteria were found between 2015 and 2017. The total number of participants across the 51 unique studies was 8,748, ranging from 19 to 1,334 participants in each study. The average age of participants was 42 years, ranging from 29 to 59 years. On average, there were more females than males. For the unimodal (Supplementary Data File 1) studies, participants reported average pain duration of 3 months or more in 60% of included studies, 6 months or more in 12% of studies; in 28% of studies, participants described chronic pain as more than 1 year. Multimodal studies (Supplementary Data File 2) reported participant average pain duration of 3 months or more in 30% of the included studies; fewer studies described participants’ chronic pain lasting 6 months (19%). Lastly, 51% of included studies described participants’ chronic pain as lasting more than 1 year.

Of the 25 unimodal studies, 60% were identified as thrust interventions, 28% were non-thrust interventions, and 12% used a combination of both. Some of these studies were multimodal by design, but were included in this subgroup for the purposes of analysis because the effect of each intervention could be distinguished. A variety of interventions were...
combined to serve more as “programs” for the included multimodal studies. The most prevalent interventions used in combination with a thrust or non-thrust intervention were prescribed exercises; others included stretches, massage, ultrasound, education, or advice therapy. Combined treatment dosages varied from one session for 3 minutes to 16 sessions of 45 minutes each over the course of 8 weeks. The majority of studies (84%) had active control or comparators (eg, acupuncture, physical therapy, exercise, usual care); the remaining studies compared the intervention with sham or no treatment. Active control or comparator dosages varied from one session for 3 minutes to 16 sessions of 45 minutes each over the course of 8 weeks.

The studies reported outcomes related to pain intensity or severity, disability, and HRQoL, all of which were considered critical outcomes for evaluation. The most common outcome measures were evaluated using the VAS (26 of 51) and the numeric pain rating scale (12 of 51) for pain reduction, the Roland-Morris Disability Questionnaire (21 of 41) or Oswestry Disability Index for disability (16 of 41), and the Short Form-36 for HRQoL (15 of 17) (Supplemental Data Files 1 and 2).

**Methodological quality**

Overall, risk of bias was not considered serious across either the unimodal or the multimodal studies. Of the 25 unimodal RCT studies [32–56], 3 (12%) were given a SIGN 50 score of high quality (++) [32–34], 18 (72%) as acceptable quality (+) [35–52], and 4 (16%) as low quality (0) [53–56]. Among the 26 multimodal studies, 3 (12%) were rated high quality (+++) [88–90], 20 (77%) acceptable quality (+) [68–87,94], and 3 (11%) low quality (0) [64,65,67,96]. The most prevalent poorly addressed quality criteria related to pitfalls in reporting group differences, intention-to-treat analyses, and multisite similarities, respectively. Overall, all EVAT categories were addressed adequately. The source population (44 of 51) and recruitment of participants (39 of 51) were transparently described and reflective of the population from which they were drawn. Twenty-three of 51 studies described the staff, places, and facilities where treatment occurred, but other studies lacked the details required to fully understand the clinical applicability to real-world settings. The majority of studies involved physical therapists, chiropractors, physicians, naturopathic, or osteopathic clinicians. In some, multiple therapists delivered the interventions. Practitioner characteristics were well described in many of studies. Treatment locations varied, including private clinics, university settings, hospitals, or other medical facilities (Table 2).

**Adverse events**

Of the 25 unimodal RCTs, 5 reported that no adverse events occurred during the study period; 2 reported minor adverse events—typically worsening symptoms. Another study reported that 2% of patients experienced serious adverse events. However, none of these symptoms was determined to be treatment-related, and the frequency of adverse events in the treatment and control groups was not significantly different. The remaining 17 studies did not provide any information on adverse events in the publications.

None of the multimodal studies reported a serious adverse event. Ten studies failed to report on adverse events and 10 reported that none had occurred. Of the remaining six, the adverse events were noted as mild, including temporary treatment-related soreness, tiredness, or worsening of existing complaints. Of studies that did report adverse events, the authors failed to describe how an event was determined to be adverse, how data were collected, and the intervals for data collection. It appears that this is perhaps spontaneous self-reporting by the subjects in the studies (Supplemental Data Files 1 and 2).

The authors relied on the Food and Drug Administration’s definition of an adverse event as any adverse experience

---

**Table 2**

<table>
<thead>
<tr>
<th>Percentage (n)</th>
<th>Single modal studies</th>
<th>Multimodal studies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poor</td>
<td>Adequate</td>
</tr>
<tr>
<td>Appropriate and clearly focused question</td>
<td>—</td>
<td>64% (16)</td>
</tr>
<tr>
<td>Randomization</td>
<td>24% (6)</td>
<td>52% (13)</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td>28% (7)</td>
<td>56% (14)</td>
</tr>
<tr>
<td>Blinding</td>
<td>28% (7)</td>
<td>60% (15)</td>
</tr>
<tr>
<td>Percentage of dropouts</td>
<td>20% (5)</td>
<td>12% (3)</td>
</tr>
<tr>
<td>Baseline similarities</td>
<td>4% (1)</td>
<td>28% (7)</td>
</tr>
<tr>
<td>Group differences</td>
<td>36% (9)</td>
<td>56% (14)</td>
</tr>
<tr>
<td>Outcome reliability/validity</td>
<td>—</td>
<td>20% (5)</td>
</tr>
<tr>
<td>Intention-to-treat analyses</td>
<td>36% (9)</td>
<td>12% (3)</td>
</tr>
<tr>
<td>Multisite similarities</td>
<td>28% (7)</td>
<td>4% (1)</td>
</tr>
</tbody>
</table>

EVAT, External Validity Assessment Tool; NA, not applicable; SIGN, Scottish Intercollegiate Guidelines Network.
during treatment resulting in death, life-threatening adverse experience, hospitalization or prolongation of existing hospitalization, or persistent or significant disability or incapacity (Supplementary Data Files 1 and 2). Of the non-RCT studies examined for the larger project effort, the majority neglected to report on adverse events. Those that did reported either new complaints or worsening of existing complaints after initial treatment; however, these complaints were not associated with worse long-term outcomes and were unlikely to be serious [97–99].

Data synthesis

Studies comparing thrust or non-thrust with a sham or no treatment control were heterogeneous and could not be pooled for analysis in any meaningful way. In addition, dose studies that are included in the systematic review were heterogeneous and omitted from any analysis. As noted previously, we excluded multimodal studies from analysis because of their inherent heterogeneity. The remainder of studies for consideration consisted of those thrust and non-thrust interventions compared with another active therapy, consisting of exercise or physical therapy where treatment was over multiple sessions and post-treatment assessment was closest to 1 month (4–8 weeks from baseline) across outcomes of pain reduction, disability, and enhanced HRQoL.

Reduction in pain

Nine trials (1,176 patients) with a reduction in pain at post-treatment closest to 1 month from baseline and with continuous data were included in the meta-analysis. The pooled SMD across all studies showed a statistically significant larger reduction in pain from thrust or non-thrust interventions compared with an active comparator being that of exercise or physical therapy (SMD = −0.28, 95% CI −0.47 to −0.09, p = .004; I² = 57%). Statistical subgroup analyses showed that when compared with active therapies (SMD = −0.43, 95% CI −0.86 to 0.00; p = .05, I² = 79%), the data favor thrust interventions for patients with chronic low back pain (5 trials; 481 patients). Among five trials (695 patients), non-thrust interventions, compared with other active comparators, showed a statistically significantly larger reduction in pain (SMD = −0.20, 95% CI −0.35 to −0.04; p = .01; I² = 0%). Translated into the VAS, the reduction in pain intensity achieved by thrust interventions equates to 10.75 points larger on a 0–100-mm scale than its comparators; non-thrust interventions equate to a 5.0-point reduction higher on a 0–100-mm scale compared with other active comparators. Fig. 3 details the overall analysis and the subgroups analyzed.

Secondary analyses comparing varying durations of chronicity or dose were not feasible given the available literature but are specified if documented in the study in Supplementary Data File 1. Data were available for 3 and 6 months.
follow-up in both thrust and non-thrust compared with other active comparators for pooling into meta-analysis. Thrust interventions compared with other active comparators (3 trials, 370 patients) trended toward an increasing effect over time at 3 months follow-up (SMD = −0.68, 95% CI −1.14 to −0.23; I² = 73.7%) and 6 months (3 trials, 223 patients) follow-up (SMD = −0.72, 95% CI −0.99 to −0.45; I² = 0%) (data not shown). Non-thrust interventions at 3 and 6 months follow-up did not significantly change from post-treatment.

Reduction in disability

Seven trials (923 patients) with continuous data were included in the meta-analysis for a reduction in disability at post-treatment closest to 1 month (Fig. 4). The pooled SMD across all studies showed a statistically significant larger reduction in disability from thrust or non-thrust interventions compared with an active comparator (SMD = −0.33, 95% CI −0.63 to −0.03; p = 0.03; I² = 78%). Subgroup analysis showed a statistically significant larger effect in favor of thrust (3 trials; 225 patients) compared with other active comparators (SMD = −0.86, 95% CI −1.27 to −0.45; p < 0.0001, I² = 46%); however, analysis of non-thrust interventions (5 trials; 698 patients), when compared with other active comparators, did not show a statistically significantly larger reduction in disability after treatment (SMD = −0.10, 95% CI −0.28 to 0.07; p = 0.25; I² = 21%). Translated into the Roland-Morris Disability Questionnaire, the reduction in disability by thrust interventions equates to a 5.16-point larger reduction on a 0–24-mm scale than its comparators; non-thrust interventions equates to a 0.6-point reduction larger on a 0–24-mm scale compared with other active comparators. Data were available for 6 months follow-up comparing thrust with other active comparators; the pooled estimate across three trials (223 patients) was an SMD = −0.71 (95% CI −0.98 to −0.44; I² = 0%) (data not shown). Non-thrust interventions at 3 and 6 months follow-up did not significantly change from post-treatment (data not shown).

Improved health-related quality of life

Too few studies reported on health-related quality of life to allow for pooling an overall effect estimate at any time point. Where assessed across studies, however, data are detailed in Supplementary Data File 1.

Confidence in the effect estimates

Overall, risk of bias was not of serious concern across analyses or subgroups pooled for each outcome assessed. As expected, we detected statistically significant heterogeneity across the overall analysis for a reduction in pain (p = 0.009) and a reduction in disability (p = 0.0001). Heterogeneity was likely due to pooling various types of intervention techniques, dose regimens, and their comparators. Within the subgroups when compared with active comparators,
heterogeneity remained significant for thrust versus other active therapies. The comparators in these studies consisted of various exercise regimens and physical therapy. There were too few studies to pool according to specific chronicity duration or according to specific predefined dosing cutoffs. The results appeared precise; however, sample size pooled remains small. We detected no publication bias according to either the Begg or Egger test for either the overall analysis or according to subgroups (data not shown). Overall, confidence in the reported effects was graded as moderate for a reduction in pain and disability after treatment.

Discussion

During the period January 2000 to March 2017, the methodological quality of the RCT studies for mobilization and manipulation for chronic low back pain is adequate overall; however, studies remain heterogeneous in terms of dose, duration, techniques involved with varying interventions and different practitioners with perhaps different training and backgrounds, controls or comparators being used across studies, and duration of chronicity of patients included. The source populations and recruitment of the participants seem to reflect the population of interest for the study (external validity); however, the staff, places, and facilities in which patients are receiving therapy are not well described, making extrapolation for real-life settings challenging (model validity).

Mobilization and manipulation appear to be safe, based on what was reported in the literature. A small-moderate effect was found in favor of manipulation for patients with chronic low back pain, with pain duration of at least 3 months or more. This effect seems to increase over time at 3 and 6 months follow-up for reducing pain compared with other active comparators, namely exercise and physical therapy comparators. Manipulation was also shown to reduce disability. The quality of the body of evidence is moderate for both of these outcomes. In the case of studies that defined chronic pain as 12 months or more, there is insufficient evidence to draw any conclusions solely on this chronicity duration from the current literature. Evidence to support mobilization interventions does not seem to be as strong as evidence to support manipulation intervention for chronic low back pain. Although there are several large studies on mobilization compared with active comparators for chronic low back pain, the majority did not show statistically significant differences in favor of mobilization interventions compared with other active comparators. We are moderately confident in the effect estimates being reported.

There is currently a gap in the evidence concerning the efficacy of manipulation and mobilization compared with sham or no treatment on pain or disability in the population studied. Pooling across other subgroups was limited because there were too few sufficiently similar studies. In addition, we were not able to draw definitive conclusions about patients’ HRQoL due to data limitations. Unlike the unimodal studies, which evaluated the results from the thrust or non-thrust interventions, the body of evidence from multimodal studies included a variety of interventions and integrated programs. For example, with exercise, individuals were allowed to choose their at-home routine or practitioners prescribed specific treatments. These types of programs may be attractive to patients because the programs may be similar to what would occur in real practice.

Overall completeness and applicability

The European Workgroup guidelines recommend a referral for spinal manipulation therapy, including mobilization, for patients who are suffering from chronic back pain [100]. In the United States, similar recommendations exist in favor of manual therapies including manipulation and mobilization for chronic low back pain [101–103]. However, the recommendations regarding manual therapies for chronic low back pain continue to show some variation depending on country or region of origin. In most guidelines, manipulation is recommended or presented as a therapeutic option. Other guidelines do not recommend it. It is not known why there are such inconsistencies across guidelines [103–105]. Guidelines may have depended largely on panelists’ interpretations, which have been based on insufficient or inconclusive evidence or reflected methodological flaws in the reported studies. Other factors that may influence guideline recommendations include local and national political variance or bias [103].

Similar to the practice guidelines, recent systematic reviews have reported favorable evidence for treating chronic non-specific low back pain using manipulation and mobilization, including chiropractic [13,106], osteopathic manipulation therapy [107], and physical therapy [108]. However, as with practice guidelines, these systematic reviews conclude that the scientific evidence is challenged by heterogeneity in the types of populations and interventions being studied, lacks long-term outcomes, includes insufficient data to explore subgroup effects, and has methodological bias that can limit and complicate the interpretation of the results [11]. Indeed, most systematic reviews conclude that it is difficult to draw definitive conclusions regarding the risk-benefit of manual therapies in patients with chronic non-specific pain.

As stated previously, we relied on the evidence provided by Bronfort et al. [13] and Shekelle et al. [18,20,21] as a starting point for our analysis. We used the definitions of manipulation and mobilization based on Bronfort et al. [13] and Coulter et al. [24]. However, our methodology generated a list of terms a priori to consider for inclusion criteria to meet our definitions. Therefore, our comprehensiveness may have increased the number of RCTs in this report but may have also increased the heterogeneity for the pooled estimates across those studies. Bronfort et al. [13] identified 31 total low back pain trials. Of these, 11 trials (n=1,472) assessed chronic low back pain and 14 trials (n=3,068) investigated a mix of patients with acute and chronic low back pain.
Because of heterogeneity across studies (ie, too dissimilar in terms of patient characteristics, outcome measures, time points, and type of treatment comparisons), we did not statistically pool these studies. However, the results from Bronfort et al. [13] were generally in favor of spinal manipulation or mobilization for treating chronic low back pain. Bronfort et al. [13] and Shekelle and Coulter [21] suggested that recommendations for spinal manipulation may be made with some degree of confidence. They identified gaps in the current literature base that need to be filled in future work, such as the need for future trials to examine well-defined subgroups of patients and further address the value of manipulation and mobilization to establish optimal number of treatment visits. Our review attempted to explore this, but research evidence remains lacking. We found that methodological flaws in the RCTs we analyzed—lack of power (low precision due to sample size) and some inconsistency—affected our statistical analysis and the overall quality of the body of evidence. Indeed, better designed studies, more homogeneous groupings, clinically relevant patient-based outcomes, and larger clinical trials are needed to inform practice decisions regarding spinal manipulation and mobilization for patients with chronic non-specific low back pain. However, this review suggests that we can have moderate confidence in the estimate of the effect across the studies for each outcome evaluated and subgroup assessed, and the effect seems to increase over time, especially for manipulation therapy.

Strengths and limitations

This review had several strengths, including (1) the involvement of an ISC to contribute to the question development and the population, intervention, control/comparator and outcome (PIFCO) framework, as well as to provide guidance throughout the review with an EAC; (2) the use of a systematic, explicit, and transparent methodology, incorporating the evaluation of internal validity (risk of bias), external and model validity, meta-analysis according to patient-reported outcomes, and GRADE framework applied to determine the overall quality of evidence for each critical outcome evaluated; and (3) an independent methodological review team to carry out each of the technical steps involved in the review phases. None of the study authors reported any conflict of interests.

The current search, scope review, and analysis leverage previous efforts and add to the literature by including both manipulation and mobilization interventions not only in chiropractic settings but also in other non-invasive therapy settings such as osteopathy, manual therapy, and physical therapy. Our approach also addresses the complexity of chronicity and non-specificity for populations being studied. We attempted to sort the literature in the most homogeneous fashion, predefining eligibility criteria and specifying very precise definitions with subject matter expertise to drive the systematic review. However, clinical heterogeneity between study groups remains a confounder. Indeed, this is not surprising since chronic pain is a multifactorial condition associated with specific and non-specific medical disorders. Non-specific chronic low back pain is difficult to evaluate, and the nature of the pain and its underlying pathophysiology are poorly understood. Therefore, chronic non-specific pain, by definition, is heterogeneous. Adequate assessment of pain and use of validated tools are essential prerequisites of a successful pain treatment plan and research study [109]. In addition, the styles, techniques, and dosing or duration of treatment involving manipulation and mobilization vary considerably, and what might work well for one individual may not for another. Stakeholders, including patients, need to be involved to help focus the research on that which could be most impactful to them.

Conclusion

There is moderate-quality evidence that manipulation (ie, thrust) interventions may produce small-moderate reduction in pain intensity compared with other active comparators such as exercise. Thrust interventions are also likely to reduce disability for patients compared with these active comparator interventions. The effect seems to increase over time at 3 and 6 months follow-up. There is moderate-quality evidence that mobilization (ie, non-thrust) interventions are likely to have minimal effect compared with other active comparators in terms of reducing pain intensity or disability. Multimodal programs may be promising options. More research is needed to assess other important patient reported outcomes in order to strengthen the evidence base regarding mobilization and manipulation for reducing disability and increasing HRQoL for patients with chronic low back pain. The research to date is still heterogeneous, and questions remain about optimal treatment duration, dose requirements, practitioners to be involved, and the kinds of patients who may benefit the most.

Acknowledgments

The authors would like to acknowledge the following individuals for their technical and administrative support throughout the project: Cindy Lentino, Viviane Enslein, Chris Baur, John Bingham, Holly Chittum, Laurie Davidson, and Judy Bearer.

Supplementary material

Supplementary material related to this article can be found at https://doi.org/10.1016/j.spinee.2018.01.013.

References

[44] Parker J, Heinking KP, Kappler RE. Efficacy of osteopathic manipulative treatment for low back pain in euhydrated and...


