

A systematic review on the effectiveness of complementary and alternative medicine for chronic non-specific low-back pain

Sidney M. Rubinstein · Marienke van Middelkoop · Ton Kuijpers ·
Raymond Ostelo · Arianne P. Verhagen · Michiel R. de Boer ·
Bart W. Koes · Maurits W. van Tulder

Received: 7 October 2009 / Revised: 24 January 2010 / Accepted: 21 February 2010 / Published online: 14 March 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract The purpose of this systematic review was to assess the effects of spinal manipulative therapy (SMT), acupuncture and herbal medicine for chronic non-specific LBP. A comprehensive search was conducted by an experienced librarian from the Cochrane Back Review Group (CBRG) in multiple databases up to December 22, 2008. Randomised controlled trials (RCTs) of adults with chronic non-specific LBP, which evaluated at least one clinically relevant, patient-centred outcome measure were included. Two authors working independently from one

another assessed the risk of bias using the criteria recommended by the CBRG and extracted the data. The data were pooled when clinically homogeneous and statistically possible or were otherwise qualitatively described. GRADE was used to determine the quality of the evidence. In total, 35 RCTs (8 SMT, 20 acupuncture, 7 herbal medicine), which examined 8,298 patients, fulfilled the inclusion criteria. Approximately half of these (2 SMT, 8 acupuncture, 7 herbal medicine) were thought to have a low risk of bias. In general, the pooled effects for the studied interventions demonstrated short-term relief or improvement only. The lack of studies with a low-risk of bias, especially in regard to SMT precludes any strong conclusions; however, the principal findings, which are based upon low- to very-low-quality evidence, suggest that SMT does not provide a more clinically beneficial effect compared with sham, passive modalities or any other intervention for treatment of chronic low-back pain. There is evidence, however, that acupuncture provides a short-term clinically relevant effect when compared with a waiting list control or when acupuncture is added to another intervention. Although there are some good results for individual herbal medicines in short-term individual trials, the lack of homogeneity across studies did not allow for a pooled estimate of the effect. In general, these results are in agreement with other recent systematic reviews on SMT, but in contrast with others. These results are also in agreement with recent reviews on acupuncture and herbal medicine. Randomized trials with a low risk of bias and adequate sample sizes are direly needed.

Electronic supplementary material The online version of this article (doi:10.1007/s00586-010-1356-3) contains supplementary material, which is available to authorized users.

S. M. Rubinstein (✉) · R. Ostelo · M. W. van Tulder
Department of Epidemiology and Biostatistics
and EMGO Institute for Health and Care Research,
VU University Medical Center, van der Boechorststraat 7,
1081 BT Amsterdam, The Netherlands
e-mail: SM.Rubinstein@vumc.nl

M. van Middelkoop · A. P. Verhagen · B. W. Koes
Department of General Practice, Erasmus MC, University
Medical Center, Rotterdam, The Netherlands

T. Kuijpers
Dutch Institute for Health Care Improvement CBO,
Utrecht, The Netherlands

R. Ostelo · M. R. de Boer · M. W. van Tulder
Department of Health Sciences,
Faculty of Earth and Life Sciences and EMGO
Institute for Health and Care Research,
VU University, Amsterdam, The Netherlands

Keywords Systematic review ·
Spinal manipulative therapy · Acupuncture ·
Herbal medicine · Low-back pain

Introduction

Low-back pain (LBP) is a common and disabling disorder in western society, which represents a great financial burden in the form of direct costs resulting from loss of work and medical expenses, as well as indirect costs [28, 29, 76]. Effective and adequate treatment is an important issue for patients, clinicians and policy makers. In addition to traditional interventions for LBP, such as medication, exercise or behavioural therapy, therapies collectively called complementary and alternative medicine (CAM) are commonly used. This includes, for example, spinal manipulative therapy (SMT), acupuncture and herbal medicine. The effectiveness of these therapies for the treatment of chronic non-specific LBP is not without dispute; therefore, a systematic review was conducted in order to assess the effectiveness of CAM therapies for the treatment of chronic non-specific LBP.

Methods

Criteria for considering studies for this review

Types of studies

Only randomised controlled trials (RCT) on SMT, acupuncture and herbal medicine were considered.

Types of participants

A study must have fulfilled the following criteria: (1) RCT with at least 1 day of follow-up; (2) adult (≥ 18 years of age) subjects with chronic (≥ 12 weeks) non-specific LBP (including discopathy or any other non-specific degenerative pathology, such as degeneration or osteoarthritis); (3) evaluated at least one main clinically-relevant outcome measure (i.e. pain, functional status, recovery, or sick-leave) using a valid instrument; and (4) fulfilled the operational definition of the therapy evaluated. In addition, the unique contribution of the CAM therapy must have been able to be discerned, which excludes studies with a multi-modal treatment in which the comparison was another type of intervention.

The following studies were excluded: (1) those in which specific spinal conditions were examined, specifically radiculopathies confirmed by radiodiagnostic procedures (i.e. MRI, CT), electrodiagnosis (i.e. EMG), positive Lasegue's or any other study in which subjects with clear neurological deficits were included, spondylolisthesis, vertebral spinal stenosis, ankylosing spondylitis, scoliosis, or coccydynia; (2) post-partum LBP or pelvic pain due to pregnancy; (3) post-operative studies; (4) primary or

secondary prevention studies; and (5) abstracts or non-published studies.

Types of interventions

The operational definitions as defined in the individual Cochrane reviews were maintained. SMT is defined as any hands-on treatment and includes both manipulation and mobilization [3]. Acupuncture is defined according to traditional acupuncture theory, and the needles were required to be inserted in classical meridian points, extra points or ah-shi points (painful points) [31]. Studies were excluded if the acupuncture treatment did not involve needling, such as acupressure or laser acupuncture. Herbal medicine is defined as all or part of a plant that was used for medicinal purposes, administered orally (ingestion) or applied topically. This excludes plant substances that were smoked, individual chemicals that were derived from plants or synthetic chemicals that were based on constituents of plants [32].

Types of outcome measures

Primary outcomes

The following self-reported measures were assessed: pain, back-specific functional status, perceived recovery (e.g. subjective overall improvement, proportion of patients recovered) and lost days at work (e.g. return to work status, number of days off work or with sick-leave). Physiological variables, such as spinal flexibility and number of degrees of straight leg raising were not assessed since it is believed that these outcomes correlate poorly with the clinical status of the patient.

Search methods for identification of studies

Existing Cochrane reviews for the three interventions were screened for studies fulfilling the inclusion criteria. The searches were updated from the last date that the literature was searched from these individual reviews.

Electronic searches

The primary search was conducted by an experienced librarian from the Cochrane Back Review Group (CBRG) in MEDLINE, EMBASE, CINAHL, CENTRAL (the Cochrane Central Register of Controlled Trials) and PEDro up to December 22, 2008.

Searching other resources

References from the relevant studies were also screened and experts approached in order to identify additional

primary studies not previously identified. The language was limited to English, Dutch and German. The search strategy outlined by the CBRG was followed. The full search strategy is available upon request from the primary author.

Data collection and analysis

Selection of studies

Two reviewers (SMR, MM) working independently of one another examined all citations from the electronic search, and full articles were obtained for those citations thought to fulfil the inclusion criteria. A third reviewer was consulted (MWvT), if consensus was not reached.

Assessment of risk of bias of included studies

Two reviewers working independently of one another assessed the risk of bias and performed data extraction. Risk of bias was assessed using the criteria list advised by the CBRG, which consists of 11-items evaluating internal validity. Items were scored as positive if they did fulfil the criteria, negative when bias was likely or marked as inconclusive if there was insufficient information. Differences in the scoring and data extraction were discussed during a consensus meeting. A study with a low risk of bias was defined as fulfilling six or more of the validity items. Studies with a fatal flaw were excluded from the meta-analyses: pre-defined as studies with a drop-out rate higher than 50% at the first follow-up measurement or statistically significant and clinically important baseline differences for the principal outcome measures that were not accounted for in the analyses. This definition is consistent with an earlier Cochrane review [31].

Data extraction

An a priori list of items was used for the data extraction, consisting of both descriptive data (e.g. study population, type of intervention) and quantitative data regarding the primary and secondary outcome measures. When the measure of variation was missing for the follow-up measurements, baseline measures were used. Data were not extracted when no measure of variation was presented anywhere in the study (3 older RCTs with a high risk of bias [22, 49, 53]). Missing outcome data were requested for three studies [37, 52, 77] and received from the corresponding authors for two [52, 77]. In one RCT with multiple interventions, two different forms of exercise (i.e. general exercise plus motor exercise) were combined in order to create a single pair-wise comparison [26]. In another study, data were averaged for morning and evening

pain, although there was a negligible difference between the two sets of data [15]. Data investigating the effect of different dosages of herbal medicine versus a placebo were also combined to form a single pair-wise comparison [17, 18].

Statistical analysis

Comparison therapies were combined into the following main clusters of presumed effectiveness: (1) no treatment or waiting list control, (2) sham, placebo, or passive modalities, (3) CAM (SMT, acupuncture, or herbal medicine) plus any intervention versus any intervention alone and (4) any other intervention (e.g. specific exercises, standard care). Regardless of eventual heterogeneity, the following stratified analyses were planned *a priori*: (1) the intervention (SMT, acupuncture, herbal medicine); (2) outcome (pain, functional status, perceived recovery); (3) time of follow-up (defined as short-term: 1 month; intermediate: 3 and 6 months; long-term: 12 months); and (4) risk of bias (low vs. high risk of bias). Lost days at work or number of days with sick-leave were not examined as an outcome because too few studies consistently recorded this [15, 38, 52, 61]. Sensitivity analyses were conducted in order to examine heterogeneity, which was tested by the Q-test (chi-square) and I^2 . In addition, the type of contrast intervention was also examined in sensitivity analyses. Adverse effects were not assessed because this was sparsely and inconsistently reported.

Effect size calculations

A mean weighted difference (MWD) was used for pain. Scales were converted to 100-points, where necessary. In all but one study, pain was measured using a VAS or numerical rating scale. The one exception was an acupuncture study [40], which measured pain according to the von Korff Chronic Pain Grade scale [75]. A number of different instruments were used to measure functional status [i.e. Roland-Morris Disability Questionnaire (RMDQ) [5, 63], Oswestry Disability Index (ODI) [5], Hannover Functional Ability Questionnaire (HFAQ) [41, 48], the Pain Disability Index (PDI) [66] and the Aberdeen LBP scale [79]]; therefore, standardized mean differences (SMD) were calculated. A negative effect size indicated that CAM was more beneficial than the comparison therapy, i.e. it improved function or diminished pain. For dichotomous outcomes, a risk ratio (RR) was calculated and the event was defined as the number of subjects recovered. A random-effects model was used for all modelling. Funnel plots were to be examined for publication bias; however, the few number of studies identified per outcome did not make this feasible. For each treatment

comparison, an effect size and a 95% confidence interval (CI) were calculated. All analyses were conducted in Review Manager 5.0.

Quality of the evidence

GRADE (Grades of Recommendation, Assessment, Development and Evaluation) was used to evaluate the overall quality of the evidence and the strength of the recommendations [4]. Quality of the evidence for a specific outcome is based upon five principal measures: (1) limitations (due to for example, study design), (2) consistency of results, (3) directness (e.g. generalizability of the findings), (4) precision (e.g. sufficient data) and (5) other considerations, such as reporting bias. The overall quality is considered to be high when RCTs with a low risk of bias provide consistent, generalizable and precise results for a particular outcome [30]. Single studies were considered to be inconsistent and imprecise and provide “low” or “very low quality evidence” depending upon whether it was associated with a low or high risk of bias, respectively, and there were no other limitations.

The following levels of evidence were applied and defined:

High quality	Further research is very unlikely to change the level of evidence and recommendations based upon consistent findings from at least two RCTs with low risk of bias and generalizable to the population in question. There are sufficient data with narrow confidence intervals. There are no known or suspected reporting biases.
Moderate quality	Further research is likely to have an important impact on confidence in the estimate of effect and may change the estimate; one of the factors is not met.
Low quality	Further research is very likely to have an important impact on confidence in the estimate of effect and is likely to change it; two of the factors are not met.
Very low quality	Great uncertainty about the estimate; three of the factors are not met.
No evidence	No evidence from RCTs.

Clinical relevance

Findings were determined to be clinically relevant based upon the size of the effect only. A cut-off point of 25 points (0–100 scale) was considered to represent the minimally clinically important change (MCIC) for pain

[70]. An effect was considered clinically significant when the lower limit of the 95% CI was greater than this MCIC. For the SMD, effect sizes were defined as <0.2 for small, ≥ 0.2 to 0.8 for moderate and >0.8 for large effects [23].

Results

Description of studies

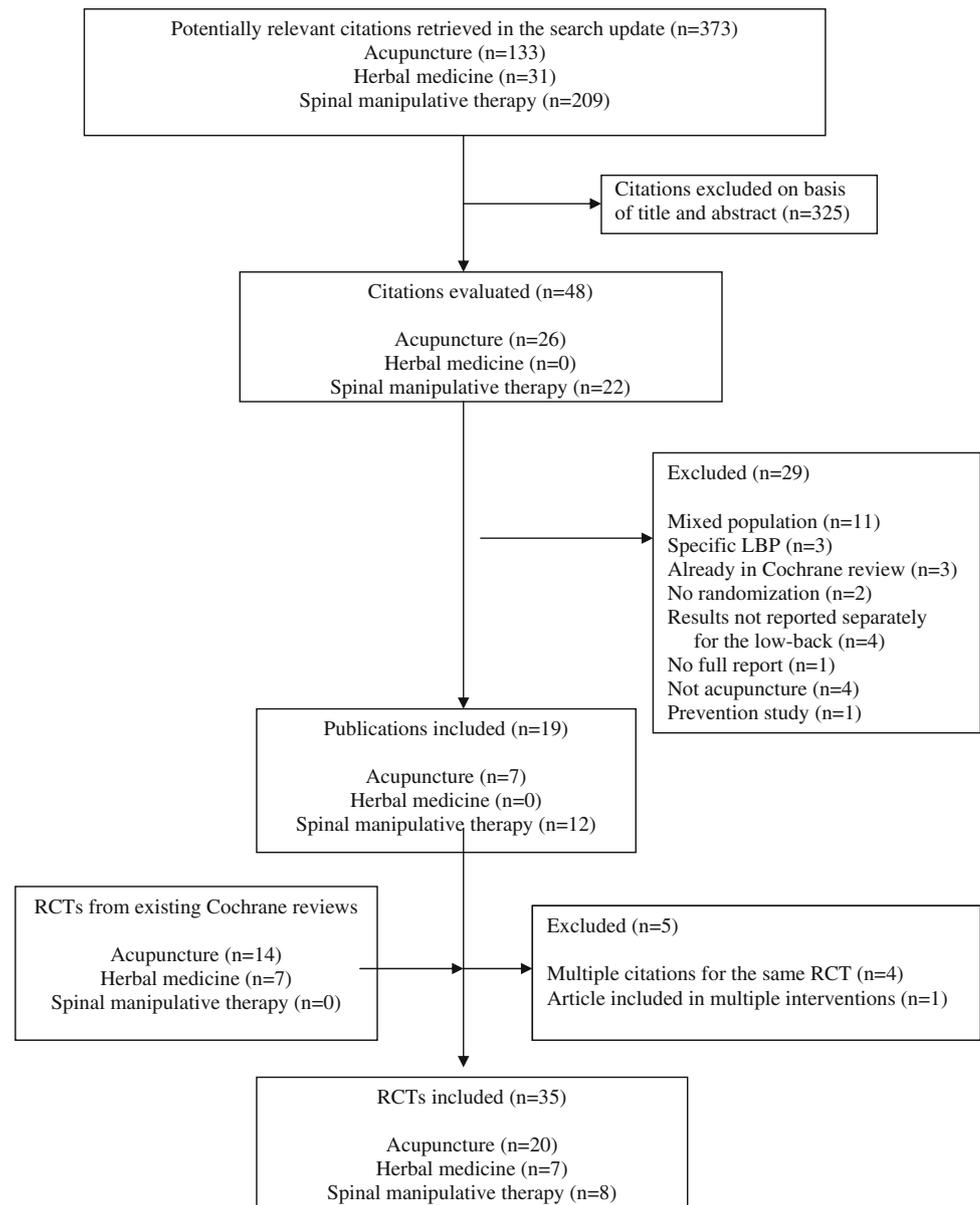
Results of the search

In total, 373 potentially relevant titles and abstracts were identified in the search update for the three interventions and screened for potential inclusion from the CBRG register (Fig. 1). This represents new studies not previously identified in the existing Cochrane reviews. After reviewing the full text from 48 citations, 19 new publications were identified, representing 14 new studies on CAM published since the respective Cochrane reviews.

Included studies

Spinal manipulative therapy

Studies were searched in the CBRG register from January 2000, which resulted in the identification of 209 new studies potentially fulfilling the inclusion criteria. After reviewing the full text of 22 articles, 12 articles were included [13, 14, 16, 26, 33–35, 37, 52, 59, 60, 77] representing eight RCTs fulfilling the inclusion criteria [16, 26, 35, 37, 52, 59, 60, 77]. Multiple publications were identified for Gudavalli et al. [13, 14, 37] and Muller et al. [33, 34, 60]. The long-term study for Muller and Giles [60] was used in subsequent referencing, while the principal study from Gudavalli et al. [37] was referenced. None of the 39 RCTs from the Cochrane review fulfilled the inclusion criteria; thus, all included publications on SMT were published in the last decade [3]. All of the earlier publications, which included chronic LBP, examined a mixed population, except for two studies. These were also excluded because one study examined a post-surgical population [68], while the other study [62] examined just one SMT treatment in the “manipulation” arm of the trial which also included weekly injections. In total, 1,393 patients were examined in the eight RCTs identified. Sample sizes ranged from 30 to 323 [median 178 (IQR 97–240)]. Treatment consisted of manipulative, mobilization, or muscle energy techniques and was conducted by physical or manual therapists in most studies [26, 35, 59], chiropractors [33, 34, 37, 60, 77] or osteopaths [16, 52].

Fig. 1 Selection process for articles on complementary and alternative medicine (CAM)

Acupuncture

Studies were searched in the CBRG register from January 2003, which resulted in the identification of 133 new studies potentially fulfilling the inclusion criteria. After reviewing the full text of 26 articles, seven RCTs were found to fulfil the inclusion criteria [11, 40, 42, 43, 60, 64, 78], one of which is included with another intervention (SMT) and also represents the long-term follow-up of an earlier study included in the Cochrane review [60]. Of the 35 RCTs included in the Cochrane review [31], 15 publications representing 14 RCTs fulfilled the inclusion criteria [15, 22, 24, 33, 34, 36, 38, 46, 49, 50, 53, 57, 58, 67, 79]. Multiple publications were identified for Brinkhaus et al.

[9–11], Haake et al. [39, 40], Mendelson et al. [56, 57] and Muller and Giles [33, 34, 60] which represent protocols and preliminary or earlier findings. The most prominent of these publications are cited. In total, 20 RCTs were included and 5,590 patients were examined. Sample sizes ranged from 17 to 3,093 [median 56 (IQR 50–110)].

Herbal medicine

Studies were searched in the CBRG register from January 2005, which resulted in the identification of 31 new studies potentially fulfilling the inclusion criteria. After reviewing the titles and abstracts, no new studies were included. Of the ten RCTs included in the Cochrane Review, [32] seven

RCTs fulfilled the inclusion criteria [17–21, 27, 45]. In total, 1,315 patients were examined. Sample sizes ranged from 88 to 320 [median 197 (IQR 118–228)]. Five RCTs compared herbal medicine, consisting of willow bark [17], a capsicum plaster [27, 45], or Harpagophytum [18, 21] to a placebo. Two RCTs compared herbal medicine, consisting of willow bark [19], or Harpagophytum [20] to a COX-2 inhibitor (Rofecoxib).

In total, 35 RCTs (8 SMT [16, 26, 35, 37, 52, 59, 60, 77], 20 acupuncture [11, 15, 22, 24, 36, 38, 40, 42, 43, 46, 49, 50, 53, 57, 58, 60, 64, 67, 78, 79], 7 herbal medicine [17–21, 27, 45]) fulfilled the inclusion criteria.

A detailed description of the included and excluded CAM studies are available in Appendix 1 and 2 of Electronic Supplementary Material.

Risk of bias

An overview of the risk of bias can be found in Table 1.

Spinal manipulative therapy

Two studies were considered to have a low risk of bias [26, 52]. Two studies were excluded from the meta-analyses because of a fatal flaw [16, 60]. In the study by Chown et al. [16], overall, only 41 and 27% of the participants responded at the 6-week- and 12-month follow-up measurements, respectively. In the study by Muller and Giles [60], drop-out at the first follow-up period was 49% for medication and 52% for acupuncture [60].

Acupuncture

Eight studies were considered to have a low risk of bias [11, 15, 40, 42, 50, 58, 64, 79]. In total, two studies were considered to have fatal flaws, one of which is described above [36, 60]. Another study was considered to have a fatal flaw because the two groups examined were fundamentally different at baseline for all the major outcome variables [36].

Herbal medicine

All seven studies included in this review were considered to have a low risk of bias [17–21, 27, 45]. Thus, in total, 17 RCTs (2 SMT [26, 52], 8 acupuncture [11, 15, 40, 42, 50, 58, 64, 79], 7 herbal medicine [17–21, 27, 45]) were thought to have a low risk of bias. The most prevalent methodological shortcomings, which were identified in well over half of the studies, were the lack of description on the allocation procedure and use of co-interventions or whether compliance with the protocol was considered acceptable. Most studies conducted intention-to-treat

analyses, which could be defined here as a complete case analysis. For virtually all studies, missing data were not imputed. In most of the SMT studies, neither the patient or care provider was blinded, as opposed to acupuncture and herbal medicine.

Effects of interventions

Spinal manipulative therapy

In total, six RCTs without a fatal flaw (2 with a low risk of bias) were included which examined the effects of SMT versus sham SMT [52], ineffectual or passive modalities [35, 59], SMT plus an intervention versus the intervention alone [52] and versus any other intervention [26, 35, 37, 77]. Data were available for pain [26, 35, 37, 52, 59, 77], and functional status, consisting of the ODI [35, 59] and RMDQ [26, 37, 52, 77]. One RCT examined perceived recovery, but analysed it as a continuous outcome [26]. A summary of the effect estimates can be found in Table 2. In addition, figures pertaining to the meta-analyses are available for all the studied interventions in Appendix 3 of Electronic Supplementary Material.

Versus no treatment or waiting list control No studies examined this contrast.

Versus sham, placebo or passive modalities In total, 3 RCTs (1 with a low risk of bias [52]) were identified, which compared SMT to care consisting of an educational booklet [35], sham manipulation [52] and ultrasound [59]. For pain, data could be pooled for two studies at 3 months only providing very low-quality evidence (serious limitations, indirect, imprecise), which demonstrated no significant effect (MWD 1.81, 95% CI –7.13 to 10.75) [35, 52]. Only one study measured the long-term effects, which demonstrated no significant effect (very low quality evidence) [35]. The only study with a low risk of bias demonstrated no significant effect at the short-term or intermediate follow-up (low quality evidence) [52]. For functional status, data could be pooled at all follow-up measurements. A moderate, significant effect was observed at 1 month from two RCTs [52, 59] providing very low-quality evidence (serious limitations, indirect, imprecise) in favour of SMT (SMD –0.36, 95% CI –0.66 to –0.06); however, the only study with a low risk of bias demonstrated no significant effect (low quality evidence) [52]. At all other follow-up measurements, no significant effect was observed, also when examined for risk of bias. No studies reported recovery.

SMT plus an intervention versus intervention alone One RCTs (with a low risk of bias [52]) was identified, which

Table 1 Risk of bias for studies on complementary and alternative medicine

Author	A	B	C	D	E	F	G	H	I	J	K	Total	Comments
Spinal manipulative therapy													
Chown et al. [16] ^a	+	?	+	-	-	-	?	-	-	+	?	3	Fatal flaw
Ferreira et al. [26]	+	+	+	-	-	-	+	+	+	+	+	8	
Goldby et al. [35]	+	?	+	-	-	-	?	-	+	+	?	4	
Gudavalli et al. [37] ^b	+	+	+	-	-	-	?	?	-	+	±	4+	
Licciardone et al. [52]	+	+	+	+	-	+	-	?	-	+	?	6	
Mohseni-Bandpei et al. [59]	?	?	+	-	-	-	?	?	-	+	?	2	
Muller and Giles [60] ^c	?	+	+	-	-	-	?	-	-	+	+	4	Fatal flaw
Wilkey et al. [77]	?	+	-	-	-	-	?	-	+	+	?	3	
Acupuncture													
Brinkhaus et al. [10, 11]	+	?	+	-	-	-	+	?	+	+	+	6	
Carlsson and Sjolund [15] ^d	+	+	?	+	-	+	?	?	±	+	+	6+	
Coan et al [22]	+	+	?	-	-	-	?	-	-	-	-	2	
Ding [24]	?	-	?	+	-	+	?	?	+	+	-	4	
Giles and Muller [33] ^c	+	+	+	-	-	-	+	?	-	+	+	6	Fatal flaw
Giles and Muller [34] ^c	?	+	?	-	-	-	?	-	-	+	-	2	Fatal flaw
Grant et al. [36] ^c	+	+	-	-	-	-	+	?	+	+	?	5	Fatal flaw
Gunn et al. [38]	-	?	?	-	-	-	?	?	+	-	-	1	
Haake et al. [40]	+	+	+	+	-	+	?	?	+	+	+	8	
Itoh et al. [42]	+	?	+	+	-	+	+	?	-	+	?	6	
Itoh et al. [43]	+	?	+	+	-	+	?	?	-	+	?	5	
Kerr et al. [46] ^f	+	?	?	+	-	+	?	?	-	+	-	4	
Lehmann et al. [49]	?	?	?	-	-	-	+	?	-	+	-	2	
Leibing et al. [50]	+	+	+	+	-	+	+	?	-	+	?	7	
MacDonald et al. [53]	?	?	+	+	-	+	?	?	+	?	+	4	
Mendelson et al. [57]	?	?	+	+	-	+	?	?	+	+	-	5	
Meng et al. [58] ^g	+	+	±	-	-	-	+	?	+	+	+	6+	
Muller and Giles [60] ^c	+	+	+	-	-	-	?	-	-	+	+	5	Fatal flaw
Sator-Katzenschlager et al. [64]	+	?	+	+	+	+	?	?	+	+	+	8	
Thomas and Lundberg [67]	?	?	+	-	-	-	-	+	?	+	+	4	
Witt et al. [78]	+	?	+	-	-	-	?	?	+	+	?	4	
Yeung et al. [79]	?	+	+	-	-	-	+	+	+	+	+	7	
Herbal medicine													
Chrubasik et al. [20]	+	?	+	+	+	+	?	?	+	+	+	8	
Chrubasik et al. [19]	+	+	+	-	-	-	+	?	+	+	-	6	
Chrubasik et al. [17]	+	?	-	+	+	+	?	?	+	+	+	7	
Chrubasik et al. [18]	+	?	+	+	+	+	?	?	+	+	+	8	
Chrubasik et al. [21]	+	?	+	+	+	+	?	?	+	+	-	7	
Frerick et al. [27]	+	+	?	?	?	?	+	+	+	?	+	6	
Keitel et al. [45]	?	?	?	+	?	+	+	+	+	+	+	7	

Criteria items: A Was the method of randomization adequate? B Was the treatment allocation concealed? C Were the groups similar at baseline regarding the most important prognostic indicators? D Was the patient blinded to the intervention? E Was the care provider blinded to the intervention? F Was the outcome assessor blinded to the intervention? G Were co-interventions avoided or similar? H Was the compliance acceptable in all groups? I Was the drop-out rate described and acceptable? J Was the timing of the outcome assessment similar in all groups? K Were all randomized participants analysed in the group to which they were allocated?

^a Only 41 and 27% of the participants responded at the 6-week and 12-month follow-up measurements, respectively
^b ITT analysis was conducted only at the first follow-up measurement (at 4 weeks); subsequent analyses were “per-protocol”
^c Muller et al. is the long-term follow-up to Giles et al.; 52% drop-out during treatment in the acupuncture group, 51 and 44% of the data available for the medication group at 9 weeks and 12 months, respectively; Overall, just 60 and 54% of the data were available for the follow-up measurements
^d ITT analysis at 1 month; subsequent measurements did not include all subjects
^e The two groups were fundamentally different at baseline with respect to all principal outcome measures, which was not corrected for in the analyses
^f Co-interventions might have influenced the results: 76 and 66.7% of the patients were followed for the short-term and intermediate follow-ups
^g Small difference in pain baseline measure; important difference in RDQ for the acupuncture and control group

Table 2 Summary effect estimates for spinal manipulative therapy

Outcome or subgroup	Studies	Participants	Statistical method	Effect estimate
1. Spinal manipulative therapy versus no treatment or waiting list control				
No studies were identified				
2. Spinal manipulative therapy versus sham/placebo/passive modalities				
2.1 Pain	3		Mean difference (IV, Random, 95% CI)	Subtotals only
2.1.1 Pain at 1 month	2	177	Mean difference (IV, Random, 95% CI)	−4.32 [−25.36, 16.73]
2.1.2 Pain at 3 months	2	177	Mean difference (IV, Random, 95% CI)	1.81 [−7.13, 10.75]
2.1.3 Pain at 6 months	3	225	Mean difference (IV, Random, 95% CI)	−0.41 [−15.05, 14.23]
2.1.4 Pain at 12 months	1	102	Mean difference (IV, Random, 95% CI)	5.20 [−9.56, 19.96]
2.2 Disability	3		Std. Mean difference (IV, Random, 95% CI)	Subtotals only
2.2.1 Disability at 1 month	2	177	Std. Mean Difference (IV, Random, 95% CI)	−0.36 [−0.66, −0.06]
2.2.2 Disability at 3 months	2	177	Standard mean difference (IV, Random, 95% CI)	0.13 [−0.19, 0.45]
2.2.3 Disability at 6 months	3	225	Standard mean difference (IV, Random, 95% CI)	0.06 [−0.26, 0.39]
2.2.4 Disability at 12 months	1	102	Standard mean difference (IV, Random, 95% CI)	0.13 [−0.30, 0.57]
2.3 Recovery	0	0	Risk ratio (M-H, Random, 95% CI)	Not estimable
3. Spinal manipulative therapy + intervention versus intervention alone				
3.1 Pain	1		Mean Difference (IV, Random, 95% CI)	Subtotals only
3.1.1 Pain at 1 month	1	59	Mean difference (IV, Random, 95% CI)	−8.80 [−21.43, 3.83]
3.1.2 Pain at 3 months	1	52	Mean difference (IV, Random, 95% CI)	−14.20 [−26.89, −1.51]
3.1.3 Pain at 6 months	1	47	Mean difference (IV, Random, 95% CI)	−4.90 [−18.68, 8.88]
3.1.4 Pain at 12 months	0	0	Mean difference (IV, Random, 95% CI)	Not estimable
3.2 Disability	1		Standard mean difference (IV, Random, 95% CI)	Subtotals only
3.2.1 Disability at 1 month	1	59	Std. Mean Difference (IV, Random, 95% CI)	−0.27 [−0.84, 0.29]
3.2.2 Disability at 3 months	1	52	Standard mean difference (IV, Random, 95% CI)	0.04 [−0.55, 0.63]
3.2.3 Disability at 6 months	1	47	Standard mean difference (IV, Random, 95% CI)	−0.19 [−0.80, 0.43]
3.2.4 Disability at 12 months	0	0	Standard mean difference (IV, Random, 95% CI)	Not estimable
3.3 Recovery	0	0	Risk ratio (M-H, Random, 95% CI)	Not estimable
4. Spinal manipulative therapy versus any other intervention				
4.1 Pain	4		Mean difference (IV, Random, 95% CI)	Subtotals only
4.1.1 Pain at 1 month	2	265	Mean difference (IV, Random, 95% CI)	−3.28 [−5.73, −0.82]
4.1.2 Pain at 3 months	4	580	Mean difference (IV, Random, 95% CI)	−1.91 [−5.37, 1.55]
4.1.3 Pain at 6 months	3	524	Mean difference (IV, Random, 95% CI)	1.04 [−10.50, 12.58]
4.1.4 Pain at 12 months	3	530	Mean difference (IV, Random, 95% CI)	−0.32 [−4.92, 4.29]
4.2 Disability	4		Standard mean difference (IV, Random, 95% CI)	Subtotals only
4.2.1 Disability at 1 month	2	226	Standard mean difference (IV, Random, 95% CI)	−0.49 [−1.41, 0.43]
4.2.2 Disability at 3 months	4	579	Standard mean difference (IV, Random, 95% CI)	−0.20 [−0.54, 0.14]
4.2.3 Disability at 6 months	3	528	Standard mean difference (IV, Random, 95% CI)	−0.05 [−0.34, 0.24]
4.2.4 Disability at 12 months	3	529	Standard mean difference (IV, Random, 95% CI)	0.04 [−0.15, 0.23]
4.3 Recovery	0	0	Risk ratio (M-H, Random, 95% CI)	Not estimable

examined the effects of SMT when added to usual care. This study demonstrated significant pain relief at 3 months in favour of SMT (MWD −14.20, 95% CI −26.89 to −1.51) [low quality evidence (indirect, imprecise)]. No other significant association was found. For functional status, no significant effect was found at any interval (low quality evidence).

Versus any other intervention In total, four RCTs (1 with a low risk of bias [26]) were identified, which included

interventions, such as exercise [26, 35, 37] and treatment in a hospital outpatient pain clinic [77]. Data could be pooled for pain at every follow-up measurement, except 6 months. A small, significant, but not clinically relevant effect was observed at 1 month from two RCTs [37, 77] with a high risk of bias in favour of SMT (MWD −3.28, 95% CI −5.73 to −0.82) [low quality evidence (serious limitations, indirect)]. At 3 and 12 months, no significant effect was observed [very low quality evidence (serious limitations, indirect, inconsistent)]. For functional status, there was

substantial heterogeneity at the short-term and intermediate follow-up. Three RCTs reported data for the long-term follow up, but the effect was non-significant (very low quality evidence) [26, 35, 37]. In a sensitivity analysis, the effects of SMT on pain and functional status were examined in those studies which compared SMT to exercise only [26, 35, 37]. A small significant, but not clinically relevant effect was seen for pain relief at 1 month in favour of SMT (1 RCT; MWD -6.00 ; 95% CI -11.50 to -0.50) (very low quality evidence). No significant effect was seen for functional status at any of the follow-up measurements (very low quality evidence). The only study with a low risk of bias demonstrated no significant effect on pain relief or functional improvement at any of the intermediate or long-term follow-up measurements (data were not collected for the short-term follow-up) (low quality evidence) [26].

Versus other types of SMT No studies examined this contrast.

Acupuncture

In total, 18 RCTs without a fatal flaw (8 with a low risk of bias) were identified which examined the effects of acupuncture versus waiting list control [22, 67, 78], sham acupuncture [11, 40, 42, 50, 57], placebo including TENS [15, 46, 49, 53], acupuncture plus another intervention versus the intervention alone [38, 50, 58, 78, 79], versus any other intervention [40] and other types of acupuncture [24, 43, 64]. Data were available which examined pain [11, 15, 22, 40, 42, 43, 46, 49, 50, 53, 57, 58, 64, 67, 78, 79]. In one RCT data were not extracted because pain was reported as the number of words describing the intensity or unpleasantness of the pain, and functional status was defined both gonometrically and as number of activities causing less than 50% pain [67]. In another study the LBP rating scale was used [55], which is a 130-point combination outcome measure, consisting of back and leg pain and a disability and physical impairment index and therefore, was not clear how this would translate to the other functional scales. Ten RCTs were identified, which examined functional status [11, 40, 42, 43, 49, 50, 53, 58, 67, 78], as measured by the RMDQ [42, 43, 58], HFAQ [40] and PDI [11, 50]. Four RCTs (all with a high risk of bias) examined recovery [15, 22, 46, 67], but the data were deemed unreliable in all but one study [15] due to for example, a large percentage of drop-outs [46], and also noted elsewhere [31] concerning the study by Thomas and Lundberg [67]. A summary of the effect estimates can be found in Table 3.

Versus no treatment or waiting list control Only one study (with a low risk of bias [11]) was identified for which

data could be extracted and interpreted. A large significant effect was observed at 8 weeks for pain relief in favour of acupuncture (MWD -24.10 , 95% CI -31.52 to -16.68) and for functional status (SMD -0.61 , 95% CI -0.90 to -0.33) (low quality evidence).

Versus sham, placebo or passive modalities In total, seven RCTs (5 with a low risk of bias [11, 15, 40, 42, 50]) were identified [11, 15, 40, 42, 46, 50, 57]. It was not possible to pool data due to an extreme outlier, which demonstrated a large significant and clinically relevant effect (MWD -42.30 , 95% CI -53.29 to -31.31 at 1 month follow-up) (low quality evidence) [42]. This was a small study ($n = 19$), which examined sham acupuncture in elderly subjects. In a subsequent post hoc analysis, it was decided to exclude this study, which allowed data to be pooled at all follow-up measurements. For pain relief, a small, significant, but not clinically relevant effect was observed at the short-term and intermediate follow-ups in favour of acupuncture (MWD -5.88 , 95% CI -11.20 to -0.55 ; -7.27 95% CI -12.66 to -1.89 ; -3.26 , 95% CI -6.28 to -0.23 , respectively) [moderate quality evidence (serious imprecision)]. Similarly, a significant, but small clinical effect was observed for functional status at the short-term and intermediate follow-up measurements (SMD -0.18 , 95% CI -0.32 to -0.04 ; -0.28 , 95% CI -0.41 to -0.16 ; -0.27 , 95% CI -0.40 to -0.15 , respectively) [low quality evidence (serious imprecision, inconsistent)]. In total, three RCTs examined recovery for the short-term and intermediate follow-up measurements [15, 40, 46]; however, one of these studies presented the data as a continuous outcome [40]. Only one of these studies was thought to have a low risk of bias, which demonstrated no significant effect at any of these follow-up measurements (low quality evidence) [15].

Acupuncture plus an intervention versus intervention alone In total, five RCTs (3 with a low risk of bias [50, 58, 79]) were identified, which examined the therapeutic effects of acupuncture in addition to another therapy (physiotherapy [50], standard medical care [38, 58, 78], and exercise [79]). A significant, but not clinically relevant effect was observed for pain relief at 1, 3 and 12 months, but not 6 months (for which there was no data) (MWD -9.80 , 95% CI -14.93 to -4.67 ; -16.91 , 95% CI -25.18 to -8.64 ; -14.00 , 95% CI -21.83 to -6.17 , respectively) [low quality evidence (indirect, imprecise)]. A strong, significant clinically relevant effect was observed for functional status at 1 and 3 months in favour of acupuncture (SMD -1.04 , 95% CI -1.46 to -0.61 ; -0.66 , 95% CI -0.74 to -0.58 , respectively) [low quality evidence (indirect, imprecise)]. A significant effect was observed for recovery at 3 months in one study with a high risk of bias

Table 3 Summary effect estimates for acupuncture

Outcome or subgroup	Studies	Participants	Statistical method	Effect estimate
5 Acupuncture versus no treatment or waiting list control				
5.1 Pain	1	214	Mean difference (IV, Random, 95% CI)	-24.10 [-31.52, -16.68]
5.1.1 Pain at 1 month	0	0	Mean difference (IV, Random, 95% CI)	Not estimable
5.1.2 Pain at 3 months	1	214	Mean difference (IV, Random, 95% CI)	-24.10 [-31.52, -16.68]
5.1.3 Pain at 6 months	0	0	Mean difference (IV, Random, 95% CI)	Not estimable
5.1.4 Pain at 12 months	0	0	Mean Difference (IV, Random, 95% CI)	Not estimable
5.2 Disability	1	214	Standard mean difference (IV, Random, 95% CI)	-0.61 [-0.90, -0.33]
5.2.1 Disability at 1 month	0	0	Standard mean difference (IV, Random, 95% CI)	Not estimable
5.2.2 Disability at 3 months	1	214	Std. Mean Difference (IV, Random, 95% CI)	-0.61 [-0.90, -0.33]
5.2.3 Disability at 6 months	0	0	Standard mean difference (IV, Random, 95% CI)	Not estimable
5.2.4 Disability at 12 months	0	0	Standard mean difference (IV, Random, 95% CI)	Not estimable
5.3 Recovery	0	0	Risk ratio (M-H, Random, 95% CI)	Not estimable
6 Acupuncture versus sham/placebo/passive modalities				
6.1 Pain	6		Mean difference (IV, Random, 95% CI)	Subtotals only
6.1.1 Pain at 1 month	4	918	Mean difference (IV, Random, 95% CI)	-5.88 [-11.20, -0.55]
6.1.2 Pain at 3 months	4	1076	Mean difference (IV, Random, 95% CI)	-7.27 [-12.66, -1.89]
6.1.3 Pain at 6 months	3	989	Mean difference (IV, Random, 95% CI)	-3.26 [-6.28, -0.23]
6.1.4 Pain at 12 months	2	290	Mean difference (IV, Random, 95% CI)	-4.74 [-10.50, 1.02]
6.2 Disability	3		Standard mean difference (IV, Random, 95% CI)	Subtotals only
6.2.1 Disability at 1 month	1	745	Standard mean difference (IV, Random, 95% CI)	-0.18 [-0.32, -0.04]
6.2.2 Disability at 3 months	3	1044	Standard mean difference (IV, Random, 95% CI)	-0.28 [-0.41, -0.16]
6.2.3 Disability at 6 months	2	962	Standard mean difference (IV, Random, 95% CI)	-0.27 [-0.40, -0.15]
6.2.4 Disability at 12 months	2	290	Standard mean difference (IV, Random, 95% CI)	-0.22 [-0.46, 0.03]
6.3 Recovery	2		Risk ratio (M-H, Random, 95% CI)	Subtotals only
6.3.1 Recovery at 1 month	2	90	Risk ratio (M-H, Random, 95% CI)	1.92 [0.41, 9.09]
6.3.2 Recovery at 3 months	1	50	Risk ratio (M-H, Random, 95% CI)	3.53 [0.91, 13.62]
6.3.3 Recovery at 6 months	2	83	Risk ratio (M-H, Random, 95% CI)	1.82 [0.72, 4.61]
6.3.4 Recovery at 12 months	0	0	Risk ratio (M-H, Random, 95% CI)	Not estimable
7 Acupuncture + intervention versus intervention alone				
7.1 Pain	3		Mean difference (IV, Random, 95% CI)	Subtotals only
7.1.1 Pain at 1 month	2	99	Mean difference (IV, Random, 95% CI)	-9.80 [-14.93, -4.67]
7.1.2 Pain at 3 months	3	185	Mean difference (IV, Random, 95% CI)	-16.91 [-25.18, -8.64]
7.1.3 Pain at 6 months	0	0	Mean difference (IV, Random, 95% CI)	Not estimable
7.1.4 Pain at 12 months	1	86	Mean Difference (IV, Random, 95% CI)	-14.00 [-21.83, -6.17]
7.2 Disability	4		Standard mean difference (IV, Random, 95% CI)	Subtotals only
7.2.1 Disability at 1 month	2	99	Standard mean difference (IV, Random, 95% CI)	-1.04 [-1.46, -0.61]
7.2.2 Disability at 3 months	4	2824	Standard mean difference (IV, Random, 95% CI)	-0.66 [-0.74, -0.58]
7.2.3 Disability at 6 months	0	0	Standard mean difference (IV, Random, 95% CI)	Not estimable
7.2.4 Disability at 12 months	1	131	Standard mean difference (IV, Random, 95% CI)	-0.28 [-0.66, 0.09]
7.3 Recovery	1		Risk ratio (M-H, Random, 95% CI)	Subtotals only
7.3.1 at 1 month	0	0	Risk ratio (M-H, Random, 95% CI)	Not estimable
7.3.2 at 3 months	1	56	Risk ratio (M-H, Random, 95% CI)	5.90 [1.96, 17.70]
7.3.3 at 6 months	0	0	Risk ratio (M-H, Random, 95% CI)	Not estimable
7.3.4 at 12 months	0	0	Risk ratio (M-H, Random, 95% CI)	Not estimable
8 Acupuncture versus any other intervention				
8.1 Pain	1		Mean difference (IV, Random, 95% CI)	Subtotals only
8.1.1 Pain at 1 month	1	731	Mean difference (IV, Random, 95% CI)	-8.50 [-11.04, -5.96]
8.1.2 Pain at 3 months	1	734	Mean difference (IV, Random, 95% CI)	-9.40 [-12.13, -6.67]

Table 3 continued

Outcome or subgroup	Studies	Participants	Statistical method	Effect estimate
8.1.3 Pain at 6 months	1	741	Mean difference (IV, Random, 95% CI)	−12.10 [−15.25, −8.95]
8.1.4 Pain at 12 months	0	0	Mean difference (IV, Random, 95% CI)	Not estimable
8.2 Disability	1		Standard mean difference (IV, Random, 95% CI)	Subtotals only
8.2.1 Disability at 1 month	1	731	Standard mean difference (IV, Random, 95% CI)	−0.53 [−0.67, −0.38]
8.2.2 Disability at 3 months	1	734	Standard mean difference (IV, Random, 95% CI)	−0.64 [−0.79, −0.49]
8.2.3 Disability at 6 months	1	741	Standard mean difference (IV, Random, 95% CI)	−0.76 [−0.91, −0.61]
8.2.4 Disability at 12 months	0	0	Standard mean difference (IV, Random, 95% CI)	Not estimable
8.3 Recovery	0	0	Risk ratio (M-H, Random, 95% CI)	Not estimable

in favour of acupuncture (RR 5.90; 95% CI 1.96–17.70) (very low quality evidence) [38]. This was a study which examined the effects of acupuncture in exclusively male subjects who had failed traditional medical or surgical therapy. The long-term follow-up measurement was highly variable and therefore, not presented.

Versus any other intervention Only one study (with a low risk of bias [40]) examined the effects of acupuncture versus another intervention, namely standard care, consisting of treatment by a physician or physiotherapist and comprised a “multimodal treatment program”. Patients in the acupuncture group were allowed “rescue medication” for acute episodes consisting of a short course of NSAIDs (no more than 2 days per week). A statistically, but not clinically relevant effect was observed for pain at the short-term and intermediate follow-ups (MWD −8.50, 95% CI −11.04 to −5.96; −9.40, 95% CI −12.13 to −6.67; −12.10, 95% CI −15.25 to −8.95, respectively) (low quality evidence). In contrast, a moderate statistically significant and clinically relevant effect was observed for functional status at the short-term and intermediate follow-ups in favour of acupuncture (SMD −0.53, 95% CI −0.67 to −0.38; −0.64, 95% CI −0.79 to −0.49; −0.76, 95% CI −0.91 to −0.61, respectively) (low quality evidence). Recovery was examined but analysed as a continuous variable.

Versus other types of acupuncture One small study ($n = 35$) (with a high risk of bias) examined the effect of standard acupuncture versus superficial (acupuncture sensation *de qi*) and deep needling at trigger points in patients 65 years of age and older (very low quality evidence) [43]. There were no significant differences between the groups at the end of either of two treatment phases. Another study ($n = 61$) (with a low risk of bias) examined the effect of auricular electroacupuncture to conventional manual auricular acupuncture (low quality evidence) [64]. Electroacupuncture resulted in significantly and clinically more pain relief and functional improvement at the short and intermediate phase compared with manual acupuncture.

Another study examined the effects of an ancient versus regular needling technique, but it was unclear whether the assessment was conducted immediately following treatment or at 2 months [24].

Herbal medicine

In total, seven RCTs (all with a low risk of bias and without a fatal flaw) were included in the analyses [17–21, 27, 45]. A summary of the effect estimates can be found in Table 4.

Versus no treatment or waiting list control No studies examined this contrast.

Versus sham, placebo or passive modalities In total, five RCTs examined this contrast, consisting of placebo tablets or plasters [17, 18, 21, 27, 45]. The principal outcome measure was perceived recovery at 1 month. Three of the studies defined recovery as the number of patients pain-free and who had not used Tramadol in the last week of the trial [17, 18, 21]. Data could not be pooled due to heterogeneity; therefore, in a post hoc analysis data were examined by the specific type of herbal medicine.

Harpagophytum versus placebo In total, two RCTs were identified, which demonstrated a non-significant effect in favour of the herbal medicine for short-term recovery (RR 3.02, 95% CI 0.51–17.91) [moderate quality evidence (indirect; imprecise)] [18, 21].

Capsicum plaster (cayenne pepper) versus placebo In total, two RCTs were identified, which demonstrated a small, significant effect in favour of the herbal medicine for short-term recovery (RR 1.59, 95% CI 1.37–1.85) [moderate quality evidence (limitations)] [27, 45].

Willow bark (salix alba) versus placebo One RCT was identified, which demonstrated a strong, significant effect in favour of the herbal medicine for short-term recovery (RR 6.75, 95% CI 2.49–18.28) (low quality evidence) [17].

Table 4 Summary effect estimates for herbal medicine

Outcome or subgroup	Studies	Participants	Statistical Method	Effect Estimate
9 Herbal medicine versus no treatment or waiting list control				
No studies were identified				
10 Herbal medicine versus sham/placebo/passive modalities				
10.1 Pain	0	0	Mean difference (IV, Random, 95% CI)	Not estimable
10.2 Disability	0	0	Standard mean difference (IV, Random, 95% CI)	Not estimable
10.3 Recovery	5		Risk ratio (M-H, Random, 95% CI)	Subtotals only
10.3.1 Recovery at 1 month	5	841	Risk ratio (M-H, Random, 95% CI)	1.92 [1.31, 2.81]
11 Herbal medicine + intervention versus intervention alone				
No studies were identified				
12 Herbal medicine versus any other intervention				
12.1 Pain	2	271	Mean difference (IV, Random, 95% CI)	−2.56 [−8.38, 3.25]
12.1.1 Pain at 1 month	2	271	Mean difference (IV, Random, 95% CI)	−2.56 [−8.38, 3.25]
12.2 Disability	0	0	Mean difference (IV, Random, 95% CI)	Not estimable
12.3 Recovery	2		Risk ratio (M-H, Random, 95% CI)	Subtotals only
12.3.1 Recovery at 1 month	2	271	Risk ratio (M-H, Random, 95% CI)	0.88 [0.70, 1.10]

Herbal medicine plus an intervention versus intervention alone No studies examined this contrast.

Versus any other intervention In total, two RCTs examined this contrast, which compared herbal medicine to COX-2 inhibitors (Rofecoxib) [19, 20]. The outcomes, pain and perceived recovery, were examined at 1 month only. No data were available for the intermediate or long-term follow-up. Data could not be pooled for the group due to heterogeneity; therefore, data were analysed per specific type of herbal medicine.

Willow bark (salix alba) versus Rofecoxib One RCT was identified, which demonstrated a small, significant effect in favour of the herbal medicine for pain relief (MWD) −6.00, 95% CI −11.45 to −0.55), but a non-significant effect in favour of Rofecoxib for recovery (RR 0.87, 95% CI (0.69 to 1.10) (low quality evidence) [19].

Harpagophytum versus Rofecoxib One RCT was identified, which demonstrated no significant effect for pain relief (MD 0.00, 95% CI −3.09 to 3.09) or for recovery (RR 1.00, 95% CI 0.38–2.61) (low quality evidence) [20].

Discussion

The lack of studies with a low risk of bias, especially in regard to SMT precludes any strong conclusions; however, the principal findings, which are based upon low- to very low-quality evidence, suggest that SMT does not provide a more clinically beneficial effect compared with sham,

passive modalities or any other intervention for treatment of chronic LBP. There is evidence, however, that acupuncture provides a short-term clinically relevant effect when compared with a waiting list control or when acupuncture is added to another intervention. Although there are some good results for individual herbal medicines in short-term individual trials, the lack of homogeneity across studies did not allow for a pooled estimate of the effect. In general, these results are in agreement with other recent systematic reviews on SMT [2, 25], but in contrast with others [12, 51, 74]. These results are also in agreement with recent reviews on acupuncture [1, 31, 54] and herbal medicine [32].

Perhaps some criticism might be levelled at our definition of chronic LBP. To our knowledge, this is the first review to exclusively examine subjects suffering LBP for more than 3 months. Previous reviews, especially for spinal manipulation, have included studies with a mixed population, e.g. sub-acute and chronic, in many cases, limiting the analysis to those studies in which the majority had chronic LBP [3, 12]. Furthermore, data for exclusively chronic LBP were not presented. This decision had definite consequences for the number of studies included in this review because a large majority of studies investigating SMT, and to a lesser extent acupuncture, examined mixed populations. This decision was not taken lightly. In our estimation, the inclusion of studies investigating exclusively chronic LBP allows for a better estimation of the effect of these interventions. However, this also meant excluding some studies with a low risk of bias, which have demonstrated a positive effect of CAM [69]. At best and based upon the findings from this review, CAM therapies

tend to have a short-term effect; therefore, studies which investigated a mixed population might overestimate these effects because patients with a shorter duration of LBP might be more likely to respond to treatment. This is consistent with the observation that longer duration of pain is a consistent predictor of a worse outcome for numerous musculoskeletal complaints [6, 7, 44, 47, 65, 71–73].

Principal limitations include the relatively few studies with a low risk of bias, which was most notable for SMT. By comparison, we identified four times as many studies for acupuncture, while all the studies for herbal medicine had a low risk of bias. Nevertheless, despite the fact that the quality of studies has improved in the past decade, less than half of all studies included in this review were considered to meet the minimum requirement. Another important limitation was the type of data used, that is, we examined the mean and standard deviations for the individual interventions at the various time measurements without regard to (small) baseline differences between the different interventions. Optimally, it would have been best to examine within mean changes, but in many cases, these data were not available. It is unclear what this difference in the analysis would have meant to the overall results and conclusions; however, we did “eyeball” these differences before extracting the data. It is also open for debate whether our definition of a fatal flaw is suitable and appropriate and whether these studies should be excluded from the meta-analyses. This exclusion was added as a buffer in order to prevent a biased assessment. Although a sensitivity analysis could have been conducted in order to determine to what extent these studies had on the overall effect, sensitivity analyses are by definition post hoc and therefore, also subject to potential bias. Based upon our assessment, two of the three studies found to have a fatal flaw, also had a high risk of bias [16, 36] and presented data as a median and interquartile range, presumably due to the skewness of the study data [36, 60]. Whether this definition finds universal recognition remains to be discussed and is a subject for future study. In addition, there was a limited amount of data for the primary outcome variables. This is particularly cogent for recovery. As a result, our analyses were limited to pain and functional status. The limited amount of data also made it difficult to discern whether different types of SMT (e.g. high-velocity low amplitude techniques, such as used by most chiropractors, or mobilization techniques, such as used by manual therapists or osteopaths), different types of acupuncture (e.g. electroacupuncture or manually-stimulated acupuncture, or Western vs. Chinese acupuncture) or dosages of herbal medicine are more effective than another. Finally, studies were excluded if the unique contribution of CAM could not be discerned. This means pragmatic studies, which include multiple therapies and might more

closely resemble clinical treatment, were excluded. The effect of one such recent study with a low risk of bias, which included SMT among other interventions (i.e. exercise) in one of the treatment arms, demonstrated a small significant effect on pain relief and a moderate effect on functional improvement compared with a physician consultation alone [61]. While these studies are vital to our understanding, they risk being excluded from future systematic reviews.

A major strength of this review is an overview of the effects of CAM for the treatment of chronic LBP. Ideally, we would have liked to have compared the different CAM interventions with one another; however, these studies are lacking. In fact, only one study directly examined the effect of SMT versus acupuncture and versus medication (NSAIDs); however, this study had a fatal flaw [60]. Alternatively, we wanted to conduct indirect comparisons with the various CAM therapies; however, there was too much heterogeneity between studies to make for a useful analysis.

While this review suggests a lack of benefit associated with SMT for chronic LBP, this is not necessarily an indictment of its efficacy. A recent study, which examined the effects of manipulation in addition to other interventions in patients with acute and subacute LBP, concluded that outcomes are improved when patients are matched to an intervention according to their signs and symptoms as opposed to being randomly assigned to a particular therapy [8]. Perhaps this indicates the need for a new generation of clinical trials, in which sub-grouping of the participants is taken into account.

Authors' conclusions

In conclusion, overall, the lack of studies with a low risk of bias precludes any strong recommendations, particularly with regard to SMT. Randomized trials with a low risk of bias and adequate sample sizes are direly needed.

Recommendations

Future efforts should focus on improving the quality of studies by considering the quality items discussed in this report. Furthermore, researchers should better define non-specific LBP so that the studied populations are as homogenous as possible and consider classification schemes in which subjects are randomized to individual therapies for which they are likely to respond.

Acknowledgments This work was funded by the Dutch National Health Insurance Council.

Conflict of interest statement None.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Ammendolia C, Furlan AD, Imamura M, Irvin E, van Tulder M (2008) Evidence-informed management of chronic low back pain with needle acupuncture. *Spine J* 8(1):160–172
- Assendelft WJ, Morton SC, Yu EI, Suttrop MJ, Shekelle PG (2003) Spinal manipulative therapy for low back pain. A meta-analysis of effectiveness relative to other therapies. *Ann Intern Med* 138(11):871–881
- Assendelft WJ, Morton SC, Yu EI, Suttrop MJ, Shekelle PG (2004) Spinal manipulative therapy for low back pain. *Cochrane Database Syst Rev*(1):CD000447
- Atkins D, Best D, Briss PA et al (2004) Grading quality of evidence and strength of recommendations. *BMJ* 328(7454):1490
- Bombardier C (2000) Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine* 25(24):3100–3103
- Bot SD, van der Waal JM, Terwee CB et al (2005) Course and prognosis of elbow complaints: a cohort study in general practice. *Ann Rheum Dis* 64(9):1331–1336
- Bot SD, van der Waal JM, Terwee CB et al (2005) Predictors of outcome in neck and shoulder symptoms: a cohort study in general practice. *Spine* 30(16):E459–E470
- Brennan GP, Fritz JM, Hunter SJ et al (2006) Identifying subgroups of patients with acute/subacute “nonspecific” low back pain: results of a randomized clinical trial. *Spine* 31(6):623–631
- Brinkhaus B, Becker-Witt C, Jena S et al (2003) Acupuncture Randomized Trials (ART) in patients with chronic low back pain and osteoarthritis of the knee—design and protocols. *Forsch Komplementarmed Klass Naturheilkd* 10(4):185–191
- Brinkhaus B, Witt CM, Jena S et al (2006) Interventions and physician characteristics in a randomized multicenter trial of acupuncture in patients with low-back pain 36. *J Altern Complement Med* 12(7):649–657
- Brinkhaus B, Witt CM, Jena S et al (2006) Acupuncture in patients with chronic low back pain: a randomized controlled trial. *Arch Intern Med* 166(4):450–457
- Bronfort G, Haas M, Evans R, Kawchuk G, Dagenais S (2008) Evidence-informed management of chronic low back pain with spinal manipulation and mobilization. *Spine J* 8(1):213–225
- Cambron JA, Gudavalli MR, Hedeker D et al (2006) One-year follow-up of a randomized clinical trial comparing flexion distraction with an exercise program for chronic low-back pain. *J Altern Complement Med* 12(7):659–668
- Cambron JA, Gudavalli MR, McGregor M et al (2006) Amount of health care and self-care following a randomized clinical trial comparing flexion-distraction with exercise program for chronic low back pain. *Chiropr Osteopat* 14:19
- Carlsson CP, Sjolund BH (2001) Acupuncture for chronic low back pain: a randomized placebo-controlled study with long-term follow-up. *Clin J Pain* 17(4):296–305
- Chown M, Whittamore L, Rush M et al (2008) A prospective study of patients with chronic back pain randomised to group exercise, physiotherapy or osteopathy. *Physiotherapy* 94(1):21–28
- Chrubasik S, Eisenberg E, Balan E et al (2000) Treatment of low back pain exacerbations with willow bark extract: a randomized double-blind study. *Am J Med* 109(1):9–14
- Chrubasik S, Junck H, Breitschwerdt H, Conrath C, Zappe H (1999) Effectiveness of *Harpagophytum* extract WS 1531 in the treatment of exacerbation of low back pain: a randomized, placebo-controlled, double-blind study. *Eur J Anaesthesiol* 16(2): 118–129
- Chrubasik S, Kunzel O, Model A, Conrath C, Black A (2001) Treatment of low back pain with a herbal or synthetic anti-rheumatic: a randomized controlled study. Willow bark extract for low back pain. *Rheumatology (Oxford)* 40(12):1388–1393
- Chrubasik S, Model A, Black A, Pollak S (2003) A randomized double-blind pilot study comparing Dolotefin and Vioxx in the treatment of low back pain. *Rheumatology (Oxford)* 42(1):141–148
- Chrubasik S, Zimpfer C, Schtt U, Ziegler R (1996) Effectiveness of *Harpagophytum procumbens* in treatment of acute low back pain. *Phytomedicine* 3:1–10
- Coan RM, Wong G, Ku SL et al (1980) The acupuncture treatment of low back pain: a randomized controlled study. *Am J Chin Med* 8(1–2):181–189
- Cohen J (1988) *Statistical power analysis for the behavioural sciences*. Lawrence Erlbaum, New Jersey
- Ding YD (1998) Fly-probing-acupoint manipulation as a main treatment for lumbago. *Shanghai J Acupunct Moxib* 17(5):25–26
- Ferreira ML, Ferreira PH, Latimer J, Herbert R, Maher CG (2002) Does spinal manipulative therapy help people with chronic low back pain? *Aust J Physiother* 48(4):277–284
- Ferreira ML, Ferreira PH, Latimer J et al (2007) Comparison of general exercise, motor control exercise and spinal manipulative therapy for chronic low back pain: a randomized trial. *Pain* 131(1–2):31–37
- Frerick H, Keitel W, Kuhn U et al (2003) Topical treatment of chronic low back pain with a capsicum plaster. *Pain* 106(1–2):59–64
- Frymoyer JW (1992) Can low back pain disability be prevented? *Baillieres Clin Rheumatol* 6(3):595–606
- Frymoyer JW, Cats-Baril WL (1991) An overview of the incidences and costs of low back pain. *Orthop Clin North Am* 22(2):263–271
- Furlan AD, Pennick V, Bombardier C, Tulder Mv, on behalf of the Editorial Board of the Cochrane Back Review Group (2008) Updated method guidelines for systematic reviews in the Cochrane Back Review Group. Cochrane Back Review Group Editorial Board
- Furlan AD, van Tulder MW, Cherkin DC et al (2005) Acupuncture and dry-needling for low back pain. *Cochrane Database Syst Rev*(1):CD001351
- Gagnier JJ, van Tulder M, Berman B, Bombardier C (2006) Herbal medicine for low back pain. *Cochrane Database Syst Rev* (2):CD004504
- Giles LG, Muller R (2003) Chronic spinal pain: a randomized clinical trial comparing medication, acupuncture, and spinal manipulation. *Spine* 28(14):1490–1502
- Giles LG, Muller R (1999) Chronic spinal pain syndromes: a clinical pilot trial comparing acupuncture, a nonsteroidal anti-inflammatory drug, and spinal manipulation. *J Manipulative Physiol Ther* 22(6):376–381
- Goldby LJ, Moore AP, Doust J, Trew ME (2006) A randomized controlled trial investigating the efficiency of musculoskeletal physiotherapy on chronic low back disorder. *Spine* 31(10):1083–1093
- Grant DJ, Bishop-Miller J, Winchester DM, Anderson M, Faulkner S (1999) A randomized comparative trial of

- acupuncture versus transcutaneous electrical nerve stimulation for chronic back pain in the elderly. *Pain* 82(1):9–13
37. Gudavalli MR, Cambron JA, McGregor M et al (2006) A randomized clinical trial and subgroup analysis to compare flexion-distraction with active exercise for chronic low back pain. *Eur Spine J* 15(7):1070–1082
 38. Gunn CC, Milbrandt WE, Little AS, Mason KE (1980) Dry needling of muscle motor points for chronic low-back pain: a randomized clinical trial with long-term follow-up. *Spine* 5(3):279–291
 39. Haake M, Muller HH, Schade-Brittinger C et al (2003) The German multicenter, randomized, partially blinded, prospective trial of acupuncture for chronic low-back pain: a preliminary report on the rationale and design of the trial. *J Altern Complement Med* 9(5):763–770
 40. Haake M, Muller H-H, Schade-Brittinger C et al (2007) German Acupuncture Trials (GERAC) for chronic low back pain: randomized, multicenter, blinded, parallel-group trial with 3 groups. *Arch Intern Med* 167(17):1892–1898
 41. Haase I, Schwarz A, Burger A, Kladny B (2001) Comparison of Hannover Functional Ability Questionnaire (FFbH) and the SF-36 subscale “Physical Functioning”. *Rehabilitation (Stuttg)* 40(1):40–42
 42. Itoh K, Katsumi Y, Hirota S, Kitakoji H (2006) Effects of trigger point acupuncture on chronic low back pain in elderly patients—a sham-controlled randomised trial. *Acupuncture in Medicine* 24(1):5–12
 43. Itoh K, Katsumi Y, Kitakoji H (2004) Trigger point acupuncture treatment of chronic low back pain in elderly patients—a blinded RCT. *Acupunct Med* 22(4):170–177
 44. Jellema P, van der Horst HE, Vlaeyen JW et al (2006) Predictors of outcome in patients with (sub)acute low back pain differ across treatment groups. *Spine* 31(15):1699–1705
 45. Keitel W, Frerick H, Kuhn U et al (2001) Capsicum pain plaster in chronic non-specific low back pain. *Arzneimittelforschung* 51(11):896–903
 46. Kerr DP, Walsh DM, Baxter D (2003) Acupuncture in the management of chronic low back pain: a blinded randomized controlled trial. *Clin J Pain* 19(6):364–370
 47. Kuijpers T, van der Windt DA, Boeke AJ et al (2006) Clinical prediction rules for the prognosis of shoulder pain in general practice. *Pain* 120(3):276–285
 48. Lautenschlager J, Mau W, Kohlmann T et al (1997) Comparative evaluation of a German version of the Health Assessment Questionnaire and the Hannover Functional Capacity Questionnaire. *Z Rheumatol* 56(3):144–155
 49. Lehmann TR, Russell DW, Spratt KF et al (1986) Efficacy of electroacupuncture and TENS in the rehabilitation of chronic low back pain patients. *Pain* 26(3):277–290
 50. Leibing E, Leonhardt U, Koster G et al (2002) Acupuncture treatment of chronic low-back pain: a randomized, blinded, placebo-controlled trial with 9-month follow-up. *Pain* 96(1–2):189–196
 51. Licciardone JC, Brimhall AK, King LN (2005) Osteopathic manipulative treatment for low back pain: a systematic review and meta-analysis of randomized controlled trials. *BMC Musculoskelet Disord* 6:43
 52. Licciardone JC, Stoll ST, Fulda KG et al (2003) Osteopathic manipulative treatment for chronic low back pain: a randomized controlled trial. *Spine* 28(13):1355–1362
 53. MacDonald AJR, MacRae KD, Master BR, Rubin AP (1983) Superficial acupuncture in the relief of chronic low back pain. *Ann R Coll Surg Engl* 65(1):44–46
 54. Manheimer E, White A, Berman B, Forys K, Ernst E (2005) Meta-analysis: acupuncture for low back pain. *Ann Intern Med* 142(8):651–663
 55. Manniche C, Asmussen K, Lauritsen B et al (1994) Low Back Pain Rating scale: validation of a tool for assessment of low back pain. *Pain* 57(3):317–326
 56. Mendelson G, Kidson MA, Loh ST et al (1978) Acupuncture analgesia for chronic low back pain. *Clin Exp Neurol* 15:182–185
 57. Mendelson G, Selwood TS, Kranz H et al (1983) Acupuncture treatment of chronic back pain. A double-blind placebo-controlled trial. *Am J Med* 74(1):49–55
 58. Meng CF, Wang D, Ngeow J et al (2003) Acupuncture for chronic low back pain in older patients: a randomized, controlled trial. *Rheumatology* 42(12):1508–1517
 59. Mohseni-Bandpei MA, Critchley J, Staunton T, Richardson B (2006) A prospective randomised controlled trial of spinal manipulation, ultrasound in the treatment of chronic low back pain. *Physiotherapy* 92(1):34–42
 60. Muller R, Giles LG (2005) Long-term follow-up of a randomized clinical trial assessing the efficacy of medication, acupuncture, and spinal manipulation for chronic mechanical spinal pain syndromes. *J Manipulative Physiol Ther* 28(1):3–11
 61. Niemisto L, Rissanen P, Sarna S et al (2005) Cost-effectiveness of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain: a prospective randomized trial with 2-year follow-up. *Spine* 30(10):1109–1115
 62. Ongley MJ, Klein RG, Dorman TA, Eek BC, Hubert LJ (1987) A new approach to the treatment of chronic low back pain. *Lancet* 2(8551):143–146
 63. Roland M, Morris R (1983) A study of the natural history of back pain—part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 8(2):141–144
 64. Sator-Katzenschlager SM, Scharbert G, Kozek-Langenecker SA et al (2004) The short- and long-term benefit in chronic low back pain through adjuvant electrical versus manual auricular acupuncture. *Anesth Analg* 98(5):1359–1364
 65. Smidt N, Lewis M, van der Windt DA et al (2006) Lateral epicondylitis in general practice: course and prognostic indicators of outcome. *J Rheumatol* 33(10):2053–2059
 66. Tait RC, Chibnall JT, Krause S (1990) The Pain Disability Index: psychometric properties. *Pain* 40(2):171–182
 67. Thomas M, Lundberg T (1994) Importance of modes of acupuncture in the treatment of chronic nociceptive low back pain. *Acta Anaesthesiol Scand* 38(1):63–69
 68. Timm KE (1994) A randomized-control study of active and passive treatments for chronic low back pain following L5 laminectomy. *J Orthop Sports Phys Ther* 20(6):276–286
 69. UK BEAM Trial Team (2004) United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. *BMJ* 329(7479):1377
 70. van der Roer N, Ostelo RW, Bekkering GE, van Tulder MW, de Vet HC (2006) Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. *Spine* 31(5):578–582
 71. van der Waal JM, Bot SD, Terwee CB et al (2006) The course and prognosis of hip complaints in general practice. *Ann Behav Med* 31(3):297–308
 72. van der Waal JM, Bot SD, Terwee CB et al (2005) Course and prognosis of knee complaints in general practice. *Arthritis Rheum* 53(6):920–930
 73. van der Windt DA, Kuijpers T, Jellema P, van der Heijden GJ, Bouter LM (2007) Do psychological factors predict outcome in both low-back pain and shoulder pain? *Ann Rheum Dis* 66(3):313–319
 74. van Tulder MW, Koes B, Malmivaara A (2006) Outcome of non-invasive treatment modalities on back pain: an evidence-based review. *Eur Spine J* 15(Suppl 1):S64–S81

75. Von Korff M, Ormel J, Keefe FJ, Dworkin SF (1992) Grading the severity of chronic pain. *Pain* 50(2):133–149
76. Waddell G (1996) Low back pain: a twentieth century health care enigma. *Spine* 21(24):2820–2825
77. Wilkey A, Gregory M, Byfield D, McCarthy PW (2008) A comparison between chiropractic management and pain clinic management for chronic low-back pain in a national health service outpatient clinic. *J Altern Complement Med* 14(5):465–473
78. Witt CM, Jena S, Selim D et al (2006) Pragmatic randomized trial evaluating the clinical and economic effectiveness of acupuncture for chronic low back pain. *Am J Epidemiol* 164(5):487–496
79. Yeung CK, Leung MC, Chow DH (2003) The use of electroacupuncture in conjunction with exercise for the treatment of chronic low-back pain. *J Altern Complement Med* 9(4):479–490