



HHS Public Access

Author manuscript

Cochrane Database Syst Rev. Author manuscript; available in PMC 2017 April 07.

Published in final edited form as:

Cochrane Database Syst Rev. 2016 ; 2016(4) : . doi:10.1002/14651858.CD012139.

Spinal rehabilitative exercise or manual treatment for the prevention of tension-type headache in adults

Brent Leininger¹, Gert Brønfort¹, Mitchell Haas², John Schmitt³, Roni L Evans¹, Morris Levin⁴, Kristine Westrom⁵, and Charles H Goldsmith⁶

¹Integrative Health & Wellbeing Research Program, Center for Spirituality & Healing, University of Minnesota, Minneapolis, MN, USA

²University of Western States, Portland, OR, USA

³St Catherine University, Minneapolis, MN, USA

⁴Neurology, UCSF School of Medicine, San Francisco, California, USA

⁵Chaska, Minnesota, USA

⁶Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada

Abstract

This is the protocol for a review and there is no abstract. The objectives are as follows:

Contact address: Gert Brønfort, Integrative Health & Wellbeing Research Program, Center for Spirituality & Healing, University of Minnesota, 420 Delaware Street SE, Minneapolis, MN, 55455, USA. bronf003@umn.edu.

CONTRIBUTIONS OF AUTHORS

Gert Bronfort registered the title, drafted the protocol, will develop the search strategy, search for studies, select studies for inclusion, provide clinical guidance, complete and interpret the analysis, draft and finalize the review.

Roni Evans will select studies for inclusion, draft and finalize the review.

Charlie Goldsmith will provide methodological and statistical guidance, complete and interpret the analysis.

Mitchell Haas registered the title, drafted the protocol, will develop the search strategy, search for studies, select studies for inclusion, provide clinical guidance, complete and interpret the analysis, draft and finalize the review.

Brent Leininger registered the title, drafted the protocol, will develop the search strategy, search for and obtain copies of studies, extract and enter data into Review Manager, complete and interpret the analysis, draft and finalize the review.

Morris Levin will provide clinical guidance, complete and interpret the analysis, draft and finalize the review.

John Schmitt will extract data and provide clinical guidance.

Kristine Westrom will participate in drafting and finalizing the review.

All authors contributed to and approve this version of the protocol. Gert Bronfort, Mitchell Haas, Roni Evans and John Schmitt will be responsible for updating the review.

DECLARATIONS OF INTEREST

Four authors are researchers with chiropractic training, one is a physical therapy researcher, one is a statistician and two are medical doctors with research backgrounds. All members may have a potential special professional interest in the effectiveness of these interventions.

Gert Bronfort declares no financial conflicts of interest.

Roni Evans declares no financial conflicts of interest.

Mitchell Haas declares no financial conflicts of interest.

Charlie Goldsmith declares no financial conflicts of interest.

Brent Leininger declares no financial conflicts of interest.

Morris Levin declares no financial conflicts of interest. ML has received small honoraria for consulting with Depomed and Allergan who produce medications for migraine headache and related conditions. These companies did not fund this review.

John Schmitt declares no financial conflicts of interest.

Kristine Westrom declares no financial conflicts of interest.

Review authors who have been authors of clinical trials that may be included in the review will not be involved in decisions regarding the inclusion or 'Risk of bias' assessment of such trials, to minimize potential personal conflicts of interest.

To assess the short- and long-term effects of manual treatment and spinal rehabilitative exercise for the prevention of tension-type headache in adults.

BACKGROUND

Tension-type headache (TTH) is the second most common condition and the most common headache disorder worldwide, with global prevalence estimates ranging from 22% to 40% (Stovner 2007; Vos 2012; Global Burden of Disease Collaborators 2015). Although less disabling than migraine at the individual level, the societal burden of TTH is greater due to the higher prevalence and frequency of the condition (Stovner 2007). Across Europe, the total cost of TTH in adults is estimated at EUR21 billion, with indirect costs due to lost work productivity accounting for 92% of costs (Linde 2012). Most individuals (90%) utilize over-the-counter medications to manage the condition, with only 7% of patients reporting prescription medication use and less than 1% reporting prophylactic medication use (Lyngberg 2005). The use of medication for the management of TTH can be a concern due to potential side effects, including the development of medication overuse headache (Bendtsen 2010b). One quarter of medication overuse headache patients develop the condition in response to overtreatment of TTH (Jonsson 2011). Many TTH patients use physical treatments including spinal rehabilitative exercise (e.g. stretching, strengthening, motor control exercises) and manual treatments (e.g. spinal manipulation, mobilization, massage) as an alternative or adjunct to pharmacological management (Rasmussen 1992; Wells 2010).

Description of the condition

TTH typically presents as a bilateral, pressing or tightening headache pain that is mild to moderate in severity and not aggravated by physical activity (IHS 2013). TTH is distinguished from migraine by the absence of nausea/vomiting or the combination of light and sound sensitivity. TTH is classified by the International Headache Society (IHS) as either infrequent episodic, frequent episodic or chronic, based on frequency (< 1, 1 to 15, and > 15 days per month for 3 months, respectively) and may be associated with pericranial tenderness. Chronic TTH may be associated with mild nausea, light sensitivity (photophobia), or sound sensitivity (phonophobia) but any combination of these symptoms would lead to a diagnosis of migraine.

Description of the intervention

For the purpose of this review, manual treatment is defined as the therapeutic application of manual force and may include manipulation, mobilization, or massage primarily applied to the cervical spine and surrounding structures. Spinal manipulation is characterized by the application of a high-velocity, low-amplitude force resulting in motion slightly beyond the passive range of the targeted joint (Haldeman 2005; Evans 2010). Spinal mobilization is characterized by the application of low-velocity, variable-amplitude force resulting in motion within the passive range of the targeted joint (Haldeman 2005). Massage is characterized by the manipulation of muscles and other soft tissues of the body. Many distinct forms of massage therapy exist such as Swedish, structural, relaxation, myofascial or

connective tissue release, and cross-friction massage. Trigger point therapy is another form of massage involving direct manual pressure to taut bands of myofascial tissue or ‘trigger points’ proposed to be responsible for local and referred pain (Simons 2008). Exercise therapy is defined as planned or structured physical activity to improve or maintain components of physical fitness (Caspersen 1985). For the purpose of this review, we will focus on spinal rehabilitative exercises for the improvement of spinal muscle strength, endurance, flexibility or motor control of the cervical spine. Spinal rehabilitative exercises may be implemented in one-on-one or group sessions.

How the intervention might work

The etiology of TTH is not entirely clear, but there is emerging evidence that dysfunction of the muscles and other structures in the cervical spine might contribute to the disorder (Fernandez-de-las-Penas 2007a; Fernandez-de-las-Penas 2007b; Watson 2012; Ashina 2015). Central hypersensitivity of the nervous system is one possible model for chronic TTH, and ongoing pain generation from the muscles and other structures of the neck is a potential trigger for central sensitization (Fernandez-de-las-Penas 2007c; Bendtsen 2016). There is evidence that manual treatment and spinal rehabilitative exercise decrease nociceptive input from the cervical spine and surrounding musculature (Jull 2002; O’Leary 2007; Coronado 2012) which may explain how these interventions can prevent TTH. There is also developing evidence that manual treatment modulates pain centrally through spinal and supraspinal mechanisms (Bialosky 2009; Pickar 2012; Bialosky 2014). The supraspinal mechanisms may be directly related to mechanical input into the spine, as well as expectancy, placebo, and other nonspecific effects. The effect of manual treatment on central pain processing mechanisms may provide an additional pathway for the reduction of TTH.

Why it is important to do this review

TTH is frequently self-managed with over-the-counter medications, but many patients also use non-pharmacological interventions including manual treatment and spinal rehabilitative exercise (Rasmussen 1992; Wells 2010). This protocol is one of a series of planned new reviews that will serve to update the original review (Bronfort 2004) to provide a comprehensive evaluation of the evidence regarding the efficacy of manual treatment and spinal rehabilitative exercise for TTH. We will compare our findings with those of other systematic reviews published after our original review (Lenssinck 2004; Biondi 2005; Fernandez-de-Las-Penas 2006; Bronfort 2010; Bryans 2011; Posadzki 2012; Gil-Martinez 2013; Chaibi 2014; Clar 2014; Mesa-Jimenez 2015; Luedtke 2015). We will divide our previous review into three separate reviews, one for TTH, one for cervicogenic headache (in press), and one for migraine headache (Bronfort 2015). We will utilize risk of bias and quality of evidence assessments recommended by Cochrane for the updated systematic review (Higgins 2011).

OBJECTIVES

To assess the short- and long-term effects of manual treatment and spinal rehabilitative exercise for the prevention of tension-type headache in adults.

METHODS

Criteria for considering studies for this review

Types of studies—We will only include randomized controlled trials (RCTs). We will exclude quasi-randomized studies (e.g., treatment allocation by date of birth, hospital record number, or alternation). Study reports in any language will be included. We will not exclude RCTs on the basis of methodological quality. We will conduct separate analyses for studies using only manual treatment or spinal rehabilitative exercise and studies combining the interventions.

Types of participants—We will include studies reporting on individuals aged 18 years and older with episodic, frequent episodic, or chronic tension-type headache (TTH) with or without pericranial tenderness, classified according to the International Headache Societies (IHS) 2013 criteria (IHS 2013). Some studies are anticipated to pre-date or not utilize the IHS classification system. Two review authors, including one neurologist (ML), will determine if studies pre-dating or not utilizing the IHS classification system can be classified as TTH using reported data (e.g. inclusion/exclusion criteria, diagnostic criteria, baseline clinical characteristics) (McCroory 2005). Studies including individuals with probable TTH or a mixture of headache disorders (e.g. both TTH and migraine) will be included if TTH or probable TTH is the predominant primary headache. Trials may include mixed populations of frequent episodic and chronic TTH patients. We will contact the authors of trials with mixed populations to acquire data for the frequent episodic and chronic TTH population separately. If we can not acquire data stratified by frequent episodic and chronic TTH, the mixed trial(s) will be analyzed separately.

Types of interventions—We will include studies assessing the effect of one or more types of manual treatment or spinal rehabilitative exercise primarily applied to the cervical spine and surrounding structures. Manual treatment can consist of spinal manipulation, mobilization, or massage techniques. Spinal rehabilitative exercise can consist of strengthening, stretching, or motor control exercises (including proprioceptive exercises). Interventions can be used alone or in combination with other active treatments (e.g. general physical therapy) but the manual treatment or spinal rehabilitative exercise must be the primary therapy assessed in the study. We will analyze single intervention studies (e.g. manual treatment or spinal rehabilitative exercise alone) independently from studies including a combination of therapies. In general, acceptable comparison groups will include placebo, no treatment (e.g., wait-list control), and any other type of active intervention.

Types of outcome measures—The primary and secondary outcome measures will follow the recommendations of the 2010 IHS guidelines on controlled trials of drugs in tension-type headache (TTH) (Bendtsen 2010a).

Primary outcomes—We will use: 1) patient-rated headache frequency measured in number of TTH days, and 2) area under the headache curve (i.e. headache index) as the primary outcome measures, as recommended by the 2010 guidelines on controlled trials of drugs in TTH (Bendtsen 2010a).

Secondary outcomes—We will include the following secondary outcomes, if available: headache intensity, duration, headache disability (e.g. headache disability index (HDI)), analgesic use, quality of life, or other pain or disability patient-reported outcomes. We will report responder rates of patients achieving 30% and 50% improvement in the primary outcome measures along with number needed to treat for an additional beneficial outcome (NNTB), if available (Bendtsen 2010a). Data on costs, cost-effectiveness, and adverse events, if available, will also be reported in the review. For adverse events, we will describe the collection methods used (active/passive surveillance), the definition of adverse events, the proportion of people reporting adverse and serious adverse events, and the number of withdrawals from treatment due to adverse events.

Timing of outcome assessments—Short-term follow up will be defined as outcomes evaluated up to 3 months after randomization. Medium-term follow up will be defined as outcomes evaluated between 3 months and 1 year after randomization. Long-term follow up will be defined as outcomes evaluated 1 year or later after randomization. In the event that multiple outcome assessments are reported within a single study, the assessments closest to 3, 6, and 12 months after randomization will be used for the short-, medium-, and long-term analyses.

Search methods for identification of studies

Electronic searches—We will search the following databases without language restrictions.

- The Cochrane Central Register of Controlled Trials (CENTRAL) (via the *Cochrane Library*)
- MEDLINE (via Ovid)
- EMBASE (via Ovid)
- CINAHL (EBSCO)
- ISI Web of Science: Science Citation Index
- Dissertation Abstracts
- BIOSIS (ISI)
- MANTIS
- Index to Chiropractic Literature
- Chiropractic Research Archives/Abstracts Collection
- Physiotherapy Evidence Database (PEDro)
- LILACS

In MEDLINE, a published search strategy for identifying RCTs (Lefebvre 2011) will be applied in combination with specific search terms, without language restriction (Appendix 1). The full MED-LINE search strategy will be adapted for use in the other electronic

databases. If multiple publications reporting different aspects of the same trial are located, we will include all publications for assessment.

Searching other resources—We will perform forward and backward citation tracking of studies included in the review to try to identify additional studies that may qualify for this review. In addition, we will search the following trial registries for completed and ongoing studies.

- [Clinical Trials.gov \(www.clinicaltrials.gov\)](http://www.clinicaltrials.gov)
- [Meta-Register of Controlled Trials \(mRCT\) \(www.controlled-trials.com/mrct\)](http://www.controlled-trials.com/mrct)
- World Health Organization’s (WHO) International Clinical Trials Registry Platform (ICTRP) (apps.who.int/trialsearch/)

Horizon estimation will be attempted after the second and subsequent searches to estimate how many articles are missing with 95% confidence intervals (CIs) using Poisson regression (Kastner 2007).

Data collection and analysis

Selection of studies—Two authors (BL, GB) will independently select trials to be included in the review based on the explicit inclusion criteria. The authors are too familiar with the relevant RCTs to conduct blinded reviews. The two authors will resolve differences in the results of selection by discussion; a third review author will be consulted if disagreements cannot be resolved. Prior to resolution of disagreements, we will calculate agreement between authors. We will select articles initially on the basis of their abstracts; if a determination cannot be made based on the abstracts, we will retrieve the full articles for review. Authors responsible for the conduct of an RCT considered for inclusion will not participate in decisions regarding inclusion/exclusion or quality assessment of their trial. We will include a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart in the full review which will show the status of identified studies (Moher 2009) as recommended in Part 2, Section 11.2.1 of the *Cochrane Handbook* (Higgins 2011). We will include studies in the review irrespective of whether measured outcome data are reported in a ‘usable’ way.

Data extraction and management—We will record explicit information about patient demographics, type of tension -type headache (TTH), clinical characteristics, interventions, and outcome measures using standardized abstracting forms. Key intervention information as described by the template for intervention description and replication (TIDieR) checklist will be recorded (Hoffmann 2014). Two non-blinded review authors (BL, JS) will independently extract and record relevant data from each article. Similar headache outcome constructs (e.g. intensity, frequency) will likely be measured on different scales (e.g. intensity of 0 to 10, 0 to 100 or frequency per 7, 14 or 28 days). We will normalise outcomes to a common 0 to 100 scale (also referred to as a percentage point scale) to facilitate analysis using mean differences, a more clinically-intuitive measure than standardized mean differences. Headache intensity measured on a 0 to 3 scale will be considered separately.

Headache frequency outcomes will be converted to a 4 week (28 day) period. We will use standardized mean differences (SMDs) for outcomes measuring the same underlying construct using distinct items (e.g. disability measured with different individual questions). We will enter data into Review Manager 5 (RevMan 2014) to create normalized mean difference (MD) scores and SMD scores. We will attempt to contact authors if there is uncertainty about important aspects of methods or data in the published report.

Assessment of risk of bias in included studies—This review will transition from the quality assessment tool used in the previous review to the ‘Risk of bias’ assessment recommended by Cochrane. At least two authors (BL, GB) will independently assess the risk of bias of each included outcome per study. Differences in ratings will be resolved by discussion between the two review authors; a third author (MH) will be consulted if disagreements cannot be resolved. Prior to resolution of disagreements, we will calculate agreement between authors. We will assess the following seven domains for risk of bias:

1. random sequence generation;
2. concealment of treatment allocation;
3. blinding of participants and/or personnel (blinding of treatment providers is not possible for clinical trials investigating manual treatment or spinal rehabilitative exercise);
4. blinding of outcome assessment;
5. incomplete outcome data: withdrawal/drop-out rate and intention-to-treat analysis;
6. selective outcome reporting;
7. other bias: similar at baseline, similar co-interventions, acceptable compliance, similar timing of assessment.

Outcomes from individual studies will not be excluded from further analyses based on the results of ‘Risk of bias’ assessments. We will rate each domain as ‘low risk’, ‘high risk’, or ‘unclear risk’ based on the criteria recommended in the *Cochrane Handbook* (Higgins 2011) and outlined in Appendix 2. In addition, we will assess the quality of each outcome using the Oxford quality scale (Jadad 1996).

‘Risk of bias’ assessments can vary among authors of Cochrane reviews, and the *Cochrane Handbook* does not specify how review authors should summarize overall risk of bias for a particular outcome (Furlan 2009; Higgins 2011). We will use the following operational definitions when judging individual outcomes within studies for overall risk of bias.

We will judge outcomes for individual studies scoring 3/5 on the Oxford scale (Jadad 1996) as low risk of bias if the remaining domains recommended in the *Handbook*, but not assessed by the Oxford scale (i.e. allocation concealment, selective outcome reporting, other risk of bias) have no more than one rating of unclear risk of bias. Outcomes scoring 3/5 on the Oxford scale, but with either high risk of bias in one domain or unclear risk of bias in two domains not assessed by the Oxford scale will be considered moderate risk of bias. All

outcomes failing to meet the criteria for low or moderate risk of bias will be characterized as high risk of bias.

Measures of treatment effect—We will use mean differences (MDs) and standardized mean differences (SMDs) as the effect measures for continuous outcomes. We will compute SMDs as described by Cohen (Cohen 1988) and Glass (Glass 1981): difference in treatment and control group means divided by the pooled standard deviation. Correction for SMD estimate bias, associated with small sample sizes ($n < 50$), will be accomplished using the method described by Hedges and Olkin (Hedges 1985). For dichotomous outcomes, we will calculate risk differences (RDs) and number needed to treat for an additional beneficial outcome (NNTB). We will assess the clinical importance by following the guidance by the IMMPACT group (Dworkin 2009). Determination of the clinical importance of between-group mean differences has not been well-standardized; however, we will facilitate interpretation by considering many factors in aggregate, including the magnitude of group differences, responder analyses, (e.g. 50% improvement), type of comparison, durability of treatment effect, intervention safety and tolerance, cost, and patients' ability to adhere to treatment.

Unit of analysis issues—For studies that include three or more interventions, we will combine the comparison groups in the meta-analysis to allow for one 'pair-wise' comparison, if clinically possible. This will prevent double counting the participants in the manual treatment or spinal rehabilitative exercise group. Cross-over studies will be included in the review; however, only data from the first period of the trial prior to the cross-over will be analyzed. For cluster randomized trials, we will include the direct treatment effect estimate if the trial author(s) properly accounted for the clustered design within the analysis. If the clustered design is not appropriately accounted for in the analysis, we will treat the individual clusters as the unit of analysis.

Dealing with missing data—We will contact the corresponding author of clinical trials with unclear reporting of trial methodology or results, for additional information. If we are unable to secure additional information pertaining to study results, we will use the following strategy for dealing with missing data. Where data are reported in a graph and not in a table, we will estimate the means and standard deviations (SDs). If the SD for follow-up measurements is missing, the SD for that measure at baseline will be used for subsequent follow-up measurements. When SDs are not reported, these will be estimated from the confidence intervals if possible. In the absence of these statistics, SDs will be calculated from T scores, P values, and F values, provided sample sizes are given (Higgins 2011). Finally, if no measure of variation is reported anywhere in the text, the SD will be estimated based upon other studies with a similar population and risk of bias.

Assessment of heterogeneity—Prior to calculation of a pooled effect measure, we will assess the reasonableness of pooling on clinical grounds. The possible sources of clinical heterogeneity to be considered are: patient population (i.e. frequent episodic, chronic, or mixed TTH), intervention (types of manual treatment or supervised rehabilitative exercises), comparison group (i.e. active comparison, placebo/sham, wait list), outcomes (e.g.

frequency, intensity, area under the headache curve), and follow-up time point (i.e. short-term, medium-term, long-term). If pooling seems appropriate on clinical grounds, we will then test for statistical heterogeneity across studies using Chi^2 and the I^2 statistic (proportion of variation between studies due to heterogeneity).

Assessment of reporting biases—We will use funnel plots as one tool for assessing potential publication bias when 10 or more studies are included in an analysis. Publication bias is just one possible explanation for funnel plot asymmetry (Higgins 2011). Selective outcome reporting, publication location, poor methodological design, and heterogeneity of treatment effects are other possible reasons for funnel plot asymmetry (Higgins 2011). We will also consider the number and size of published clinical trials in addition to evidence of smaller treatment effects in unpublished trials located by our search strategy.

Data synthesis—For the main analyses, we will pool outcomes with low or moderate risk of bias by type of intervention (i.e. manual therapy, exercise therapy, combined manual and exercise therapy), tension-type headache (TTH) frequency (frequent episodic, chronic, mixed), and comparison (i.e. active comparison, placebo/sham, wait list) using random-effects models in Review Manager, version 5.3 (RevMan 2014). If outcomes with low or moderate risk of bias are not available, we will pool outcomes with high risk of bias. If I^2 50% (Higgins 2002), we will acknowledge the difficulty of making inference from pooled estimates and emphasize individual trial results using best-evidence synthesis methodology. We will assess the quality of the body of the evidence and synthesize the findings of multiple studies using the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach (Higgins 2011; GRADEpro GDT 2015). We will use the GRADE approach for all outcomes, independent of the decision to pool trials for meta-analysis. Domains that may decrease the quality of the evidence are:

1. the study design;
2. risk of bias;
3. consistency of results;
4. directness (generalizability);
5. precision (sufficient data);
6. publication bias.

Domains that may increase the quality of the evidence are:

1. large magnitude of effect;
2. all residual confounding would have reduced the observed effect (true effect underestimated);
3. a dose-response gradient is evident.

High-quality evidence is defined as outcomes from RCTs with low risk of bias that provide consistent, direct, and precise results for the outcome. The quality of the evidence will be reduced by one level for each of the six domains not met, or increased by one level for each

of three factors. If only studies with high risk of bias are present for a given outcome, the quality of evidence will decrease by two levels for the ‘Risk of bias’ domain. Further details are provided in Appendix 3.

- **High-quality evidence:** Further research is very unlikely to change our confidence in the estimate of effect. There are consistent findings among 75% of RCTs with low risk of bias that are generalizable to the population in question. There are sufficient data, with narrow confidence intervals. There are no known or suspected reporting biases. (All of the domains are met.)
- **Moderate-quality evidence:** Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate. (One of the domains is not met.)
- **Low quality evidence:** Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. (Two of the domains are not met or only high risk of bias studies are included.)
- **Very low quality evidence:** We are very uncertain about the estimate. (Three of the domains are not met.)

We also will consider adverse events and costs to place the results into a larger clinical context.

Subgroup analysis and investigation of heterogeneity—Subgroup analyses are planned to assess the influence of type of manual therapy (manipulation, mobilization, massage, mixed), and exercise therapy (strengthening, stretching, motor control, mixed) on the overall results. In addition, the impact of the presence of pericranial tenderness will also be assessed in a subgroup analysis. We will emphasize the subgroup analyses in the narrative because of their different interpretations.

Sensitivity analysis—We will include outcomes with a high risk of bias in the main analyses as a sensitivity analysis. In addition, we will reclassify outcomes originally rated as low risk of bias, but containing domains rated as unclear, as moderate risk of bias. Data synthesis will then be repeated and the new quality of evidence ratings will be compared with the original ones.

‘Summary of findings’ table—We will present results for primary outcomes using ‘Summary of findings’ tables from the GRADE system (GRADEpro GDT 2015). Strength of evidence findings for secondary outcomes will be reported within forest plots. We will report the number of studies and participants addressing each outcome, the magnitude of treatment effect, the overall quality, and reasons for up-or downgrading the evidence.

Acknowledgments

Cochrane Review Group funding acknowledgement: The National Institute for Health Research (NIHR) is the largest single funder of the Cochrane PaPaS Group. Disclaimer: The views and opinions expressed herein are those

of the authors and do not necessarily reflect those of the NIHR, National Health Service (NHS) or the Department of Health.

SOURCES OF SUPPORT

Internal sources

- Center for Spirituality & Healing, University of Minnesota, USA.
Salary support for Gert Bronfort, Roni Evans, and Brent Leininger
- Simon Fraser University, Canada.
Salary support for Charles Goldsmith
- University of Western States, USA.
Salary support for Mitchell Haas
- UCSF School of Medicine, USA.
Salary support for Morris Levin
- Doctor of Physical Therapy Program, St Catherine's University, USA.
Salary support for John Schmitt

External sources

- National Center for Complementary and Integrative Health, National Institute of Health, USA.

Salary support for Brent Leininger through a post-doctoral fellowship training grant (#F32AT007507) and career development award (#K01AT008965). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

Additional references

- Ashina S, Bendtsen L, Lyngberg A, Lipton RB, Hajjiyeva N, Jensen R. Prevalence of neck pain in migraine and tension-type headache: a population study. *Cephalalgia*. 2015 Mar; 35(3):211–219. [PubMed: 24853166]
- Bendtsen L, Bigal ME, Cerbo R, Diener HC, et al. Guidelines for controlled trials of drugs in tension-type headache: second edition. *Cephalalgia*. 2010a; 30(1):1–16. [PubMed: 19614696]
- Bendtsen L, Evers S, Linde M, Mitsikostas DD, Sandrini G, Schoenen J. EFNS guideline on the treatment of tension-type headache - report of an EFNS task force. *European Journal of Neurology*. 2010b; 17:1318–1325. [PubMed: 20482606]
- Bendtsen L, Ashina S, Moore A, Steiner T. Muscles and their role in episodic tension-type headache: implications for treatment. *European Journal of Pain*. 2016; 20(2):166–175. [PubMed: 26147739]
- Bialosky JE, Bishop MD, Price DD, Robinson ME, George SZ. The mechanisms of manual therapy in the treatment of musculoskeletal pain: a comprehensive model. *Manual Therapy*. 2009 Oct; 14(5): 531–538. [PubMed: 19027342]
- Bialosky JE, George SZ, Horn ME, Price DD, Staud R, Robinson ME. Spinal manipulative therapy-specific changes in pain sensitivity in individuals with low back pain. *Journal of Pain*. 2014 Feb; 15(2):136–148. [PubMed: 24361109]
- Biondi DM. Physical treatments for headache: a structured review. *Headache*. 2005; 45(6):738–746. [PubMed: 15953306]
- Bronfort G, Haas M, Evans R, Leininger B, Triano J. Effectiveness of manual therapies: the UK evidence report. *Chiropractic & Osteopathy*. 2010; 18:3. [PubMed: 20184717]
- Bronfort G, Evans RL, Goldsmith CH, Haas M, Leininger B, Levin M, et al. Spinal rehabilitative exercise and manual treatment for the prevention of migraine attacks in adults. *Cochrane Database of Systematic Reviews*. 2015; (8)

- Bryans R, Descarreaux M, Duranleau M, Marcoux H, Potter B, Ruegg R, et al. Evidence-based guidelines for the chiropractic treatment of adults with headache. *Journal of Manipulative and Physiological Therapeutics*. 2011; 34(5):274–289. [PubMed: 21640251]
- Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Report*. 1985; 100(2):126–131.
- Chaibi A, Russell MB. Manual therapies for primary chronic headaches: a systematic review of randomized controlled trials. *The Journal of Headache and Pain*. 2014; 15:67. [PubMed: 25278005]
- Clar C, Tsertsvadze A, Court R, Hundt GL, Clarke A, Sutcliffe P. Clinical effectiveness of manual therapy for the management of musculoskeletal and non-musculoskeletal conditions: systematic review and update of UK evidence report. *Chiropractic & Manual Therapies*. 2014; 22(1):12. [PubMed: 24679336]
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1988.
- Coronado RA, Gay CW, Bialosky JE, Carnaby GD, Bishop MD, George SZ. Changes in pain sensitivity following spinal manipulation: a systematic review and meta-analysis. *Journal of Electromyography and Kinesiology*. 2012 Oct; 22(5):752–767. [PubMed: 22296867]
- Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2009; 146:238–244. [PubMed: 19836888]
- Evans DW, Lucas N. What is ‘manipulation’? A reappraisal. *Manual Therapy*. 2010; 15(3):286–291. [PubMed: 20080431]
- Fernandez-de-Las-Penas C, Alonso-Blanco C, Cuadrado ML, Miangolarra JC, Barriga FJ, Pareja JA. Are manual therapies effective in reducing pain from tension-type headache?: a systematic review. *The Clinical Journal of Pain*. 2006; 22(3):278–285. [PUBMED: 16514329]. [PubMed: 16514329]
- Fernandez-de-las-Penas C, Bueno A, Ferrando J, Elliott JM, Cuadrado ML, Pareja JA. Magnetic resonance imaging study of the morphometry of cervical extensor muscles in chronic tension-type headache. *Cephalalgia*. 2007a; 27:355–362. [PubMed: 17376113]
- Fernandez-de-las-Penas C, Cuadrado ML, Pareja JA. Myofascial trigger points, neck mobility, and forward head posture in episodic tension-type headache. *Headache*. 2007b; 47:662–672. [PubMed: 17501847]
- Fernandez-de-las-Penas C, Cuadrado ML, Arendt-Nielsen L, Simons DG, Pareja JA. Myofascial trigger points and sensitization: an updated pain model for tension-type headache. *Cephalalgia*. 2007c May; 27(5):383–393. [PubMed: 17359516]
- Furlan AD, Pennick V, Bombardier C, van Tulder M. Editorial Board of Cochrane Back Review Group. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine*. 2009; 34(18):1929–1941. [PubMed: 19680101]
- Gil-Martinez A, Kindelan-Calvo P, Agudo-Carmona D, Munoz-Plata R, Lopez-de-Uralde-Villanueva I, La Touche R. Therapeutic exercise as treatment for migraine and tension-type headaches: a systematic review of randomised clinical trials [Ejercicio terapeutico como tratamiento de las migranas y cefaleas tensionales: revision sistematica de ensayos clinicos aleatorizados.]. *Revista de neurologia*. 2013; 57(10):433–443. [PubMed: 24203665]
- Glass, GV.; McGaw, B.; Smith, ML. *Meta-analysis in Social Research*. Beverly Hills, CA: Sage Publications; 1981.
- Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* Aug. 2015; 386(9995):743–800.
- Brozek, J.; Oxman, A.; Schünemann, H. *GRADEpro Guideline Development Tool*. McMaster University (developed by Evidence Prime, Inc.); 2015.
- Haldeman, S.; Dagenais, S.; Budgell, B.; Grunnet-Nilsson, N.; Hooper, P.; Meeker, W., et al. *Principles and Practice of Chiropractic*. New York: McGraw-Hill; 2005.
- Hedges, LV.; Olkin, I. *Statistical Methods for Meta-analysis*. Orlando, FL: Academic Press; 1985.

- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*. 2002; 21:1539–1558. [PubMed: 12111919]
- Higgins, JPT.; Green, S., editors. *The Cochrane Collaboration*. 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. Available from www.cochrane-handbook.org
- Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014; 348:g1687. [PubMed: 24609605]
- Headache Classification Committee of the International Headache Society. *The International Classification of Headache Disorders, 3rd edition (beta version)*. *Cephalalgia*. 2013; 33(9):629–808. [PubMed: 23771276]
- Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials*. 1996 Feb; 17(1):1–12. [PubMed: 8721797]
- Jonsson P, Hedenrud T, Linde M. Epidemiology of medication overuse headache in the general Swedish population. *Cephalalgia*. 2011; 31:1015–1022. [PubMed: 21628444]
- Jull G, Trott P, Potter H, Zito G, Niere K, Shirley D, et al. A randomized controlled trial of exercise and manipulative therapy for cervicogenic headache. *Spine*. 2002 Sep; 27(17):1835–1843. [PubMed: 12221344]
- Kastner M, Straus S, Goldsmith CH. Estimating the Horizon of articles to decide when to stop searching in systematic reviews: an example using a systematic review of RCTs evaluating osteoporosis clinical decision support tools. *AMIA Annual Symposium Proceedings*. 2007; 11:389–393.
- Lefebvre, C.; Manheimer, E.; Glanville, J. Chapter 6: Searching for studies. In: Higgins, JPT.; Green, S., editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0* [updated March 2011]. *The Cochrane Collaboration*; 2011. Available from www.cochrane-handbook.org
- Lenz ML, Damen L, Verhagen AP, Berger MY, Passchier J, Koes BW. The effectiveness of physiotherapy and manipulation in patients with tension-type headache: a systematic review. *Pain*. 2004; 112(3):381–388. [PUBMED: 15561394]. [PubMed: 15561394]
- Linde M, Gustavsson A, Stovner LJ, Steiner TJ, Barré J, Katsarava Z, et al. The cost of headache disorders in Europe: the EuroLight project. *European Journal of Neurology*. 2012; 19(5):703–711. [PubMed: 22136117]
- Luedtke K, Allers A, Schulte LH, May A. Efficacy of interventions used by physiotherapists for patients with headache and migraine-systematic review and meta-analysis. *Cephalalgia*. 2015:1–19. [PUBMED: 26229071]. [PubMed: 25976401]
- Lynberg AC, Rasmussen BK, Jorgensen T, Jensen R. Secular changes in health care utilization and work absence for migraine and tension-type headache: a population based study. *European Journal of Epidemiology*. 2005; 20:1007–1014. [PubMed: 16331432]
- McCorry DC, Gray RN, Tfelt-Hansen P, Steiner TJ, Taylor FR. Methodological issues in systematic reviews of headache trials: adapting historical diagnostic classifications and outcome measures to present-day standards. *Headache*. 2005 May; 45(5):459–465. [PubMed: 15953262]
- Mesa-Jimenez JA, Lozano-Lopez C, Angulo-Diaz-Parreno S, Rodriguez-Fernandez AL, De-la-Hoz-Aizpurua JL, Fernandez-de-Las-Penas C. Multimodal manual therapy vs. pharmacological care for management of tension type headache: a meta-analysis of randomized trials. *Cephalalgia*. 2015; 35(14):1323–1332. [PUBMED: 25748428]. [PubMed: 25748428]
- Moher D, Liberati A, Tetzlaff J, Altman DG. the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*. 2009; 6(7):e1000097. [PubMed: 19621072]
- O’Leary S, Falla D, Hodges PW, Jull G, Vicenzino B. Specific therapeutic exercise of the neck induces immediate local hypoalgesia. *Journal of Pain*. 2007 Nov; 8(11):832–839. [PubMed: 17644487]
- Pickar JG, Bolton PS. Spinal manipulative therapy and somatosensory activation. *Journal of Electromyography and Kinesiology*. 2012 Oct; 22(5):785–794. [PubMed: 22349622]

- Posadzki P, Ernst E. Spinal manipulations for tension-type headaches: a systematic review of randomized controlled trials. *Complementary Therapies in Medicine*. 2012; 20(4):232–239. [PUBMED: 22579436]. [PubMed: 22579436]
- Rasmussen BK, Jensen R, Olesen J. Impact of headache on sickness absence and utilisation of medical services: a Danish population study. *Journal of Epidemiology & Community Health*. 1992; 46(4): 443–446. [PubMed: 1431724]
- The Nordic Cochrane Centre. The Cochrane Collaboration. Review Manager. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.
- Simons DG. New views of myofascial trigger points: etiology and diagnosis. *Archives of Physical Medicine and Rehabilitation*. 2008; 89(1):157–159. [PubMed: 18164347]
- Stovner LJ, Hagen K, Jensen R, Katsarava Z, Lipton R, Scher A, et al. The global burden of headache: a documentation of headache prevalence and disability worldwide. *Cephalalgia*. 2007; 27(3):193–210. [PubMed: 17381554]
- Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012 Dec; 380(9859):2163–2196. [PubMed: 23245607]
- Watson DH, Drummond PD. Head pain referral during examination of the neck in migraine and tension-type headache. *Headache*. 2012 Sep; 52(8):1226–1235. [PubMed: 22607581]
- Wells RE, Phillips RS, Schachter SC, McCarthy EP. Complementary and alternative medicine use among US adults with common neurological conditions. *Journal of Neurology*. 2010; 257(11): 1822–1831. [PubMed: 20535493]

References to other published versions of this review

* Indicates the major publication for the study

Bronfort G, Nilsson N, Haas M, Evans R, Goldsmith CH, Assendelft WJ, et al. Non-invasive physical treatments for chronic/recurrent headache. *Cochrane Database of Systematic Reviews*. 2004; (3)

Appendix 1. MEDLINE Ovid search strategy

1. exp Headache Disorders/
2. Headache/
3. (migrain* or headache* or cephalgi* or cephalalgi*).mp.
4. 1 or 2 or 3
5. exp Exercise/
6. exp Physical Therapy Modalities/
7. exp Physical Medicine/
8. Osteopathic Medicine/
9. Chiropractic/
10. (manual* adj5 (treat* or therap*)).mp.
11. ((spine or spinal) adj5 (manipulat* or mobili*)).mp.
12. (trigger point adj5 therap*).mp.

13. (exercis* or strength* or aerobic* or yoga or pilates or tai chi or tai ji or stretch* or danc*).mp.
14. (massag* or reflexology or physiotherap* or physical therap* or acupressure or osteopath* or chiropract* or shiatsu or kinesiology or ((cranio sacral or craniosacral) and therap*).mp.
15. 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
16. randomized controlled trial.pt.
17. controlled clinical trial.pt.
18. randomized.ab.
19. placebo.ab.
20. drug therapy.fs.
21. randomly.ab.
22. trial.ab.
23. groups.ab.
24. 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23
25. exp animals/ not humans.sh.
26. 24 not 25
27. 4 and 15 and 26

key:

mp=protocol supplementary concept, rare disease supplementary concept, title, original title, abstract, name of substance word, subject

heading word, unique identifier

pt=publication type

ab=abstract

sh=subject heading

ti=title

Appendix 2. Operational criteria for risk of bias assessment

RANDOM SEQUENCE GENERATION

Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence.

Criteria for a judgement of 'Low risk' of bias.

The investigators describe a random component in the sequence generation process such as:

- Referring to a random number table;
- Using a computer random number generator;
- Coin tossing;

RANDOM SEQUENCE GENERATION	
Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence.	
	<ul style="list-style-type: none"> • Shuffling cards or envelopes; • Throwing dice; • Drawing of lots; • Minimization*. <p>*Minimization may be implemented without a random element, and this is considered to be equivalent to being random</p>
Criteria for the judgement of 'High risk' of bias.	<p>The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic, non-random approach, for example:</p> <ul style="list-style-type: none"> • Sequence generated by odd or even date of birth; • Sequence generated by some rule based on date (or day) of admission; • Sequence generated by some rule based on hospital or clinic record number. <p>Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgement or some method of non-random categorization of participants, for example:</p> <ul style="list-style-type: none"> • Allocation by judgement of the clinician; • Allocation by preference of the participant; • Allocation based on the results of a laboratory test or a series of tests; • Allocation by availability of the intervention.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information about the sequence generation process to permit judgement of 'Low risk' or 'High risk'
ALLOCATION CONCEALMENT	
Selection bias (biased allocation to interventions) due to inadequate concealment of allocations prior to assignment.	
Criteria for a judgement of 'Low risk' of bias.	<p>Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:</p> <ul style="list-style-type: none"> • Central allocation (including telephone, web-based and pharmacy-controlled randomization); • Sequentially numbered drug containers of identical appearance; • Sequentially numbered, opaque, sealed envelopes.
Criteria for the judgement of 'High risk' of bias.	<p>Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on:</p> <ul style="list-style-type: none"> • Using an open random allocation schedule (e.g. a list of random numbers); • Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or nonopaque or not sequentially numbered); • Alternation or rotation; • Date of birth; • Case record number;

RANDOM SEQUENCE GENERATION**Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence.**

	<ul style="list-style-type: none"> Any other explicitly unconcealed procedure.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information to permit judgement of 'Low risk or 'High risk. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement - for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed

BLINDING OF PARTICIPANTS AND PERSONNEL**Performance bias due to knowledge of the allocated interventions by participants and personnel during the study.**

Criteria for a judgement of 'Low risk' of bias.	<p>Any one of the following:</p> <ul style="list-style-type: none"> No blinding or incomplete blinding, but the review authors judge that the outcome is not likely to be influenced by lack of blinding; Blinding of participants and key study personnel ensured, and unlikely that the blinding could have been broken.
Criteria for the judgement of 'High risk' of bias.	<p>Any one of the following:</p> <ul style="list-style-type: none"> No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding; Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding.
Criteria for the judgement of 'Unclear risk' of bias.	<p>Any one of the following:</p> <ul style="list-style-type: none"> Insufficient information to permit judgement of 'Low risk or 'High risk'; The study did not address this outcome.

BLINDING OF OUTCOME ASSESSMENT**Detection bias due to knowledge of the allocated interventions by outcome assessors.**

Criteria for a judgement of 'Low risk' of bias.	<p>Any one of the following:</p> <ul style="list-style-type: none"> No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding; Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken.
Criteria for the judgement of 'High risk' of bias.	<p>Any one of the following:</p> <ul style="list-style-type: none"> No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding; Blinding of outcome assessment, but likely that the blinding could have been broken, and the outcome measurement is likely to be influenced by lack of blinding.
Criteria for the judgement of 'Unclear risk' of bias.	<p>Any one of the following:</p> <ul style="list-style-type: none"> Insufficient information to permit judgement of 'Low risk or 'High risk';

RANDOM SEQUENCE GENERATION**Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence.**

- The study did not address this outcome.

INCOMPLETE OUTCOME DATA**Attrition bias due to amount, nature or handling of incomplete outcome data.**

Criteria for a judgement of 'Low risk of bias.

Any one of the following:

- No missing outcome data;
- Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias);
- Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups;
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate;
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size;
- Missing data have been imputed using appropriate methods.

Criteria for the judgement of 'High risk' of bias.

Any one of the following:

- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups;
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate;
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size;
- 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization;
- Potentially inappropriate application of simple imputation.

Criteria for the judgement of 'Unclear risk' of bias.

Any one of the following:

- Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g. number randomized not stated, no reasons for missing data provided);
- The study did not address this outcome.

SELECTIVE REPORTING**Reporting bias due to selective outcome reporting.**

Criteria for a judgement of 'Low risk of bias.

Any of the following:

- The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way;
- The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon).

RANDOM SEQUENCE GENERATION	
Selection bias (biased allocation to interventions) due to inadequate generation of a randomised sequence.	
Criteria for the judgement of 'High risk' of bias.	<p>Any one of the following:</p> <ul style="list-style-type: none"> • Not all of the study's pre-specified primary outcomes have been reported; • One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; • One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); • One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; • The study report fails to include results for a key outcome that would be expected to have been reported for such a study.
Criteria for the judgement of 'Unclear risk' of bias.	Insufficient information to permit judgement of 'Low risk or 'High risk'. It is likely that the majority of studies will fall into this category
OTHER BIAS	
Bias due to problems not covered elsewhere in the table.	
Criteria for a judgement of 'Low risk' of bias.	The study appears to be free of other sources of bias.
Criteria for the judgement of 'High risk' of bias.	<p>There is at least one important risk of bias. For example, the study:</p> <ul style="list-style-type: none"> • Had a potential source of bias related to the specific study design used; or • Has been claimed to have been fraudulent; or • Had some other problem.
Criteria for the judgement of 'Unclear risk' of bias.	<p>There may be a risk of bias, but there is either:</p> <ul style="list-style-type: none"> • Insufficient information to assess whether an important risk of bias exists; or • Insufficient rationale or evidence that an identified problem will introduce bias.

Appendix 3. GRADE criteria

Factors that may decrease the overall quality of the evidence

1. Study design
 - Downgrade 2 levels - Observational study design (Note: Only RCTs are included within the review, so the evidence will not be downgraded for study design).
2. Risk of bias
 - Downgrade 1 level - Less than 75% of information is from studies with low risk of bias.

- Downgrade 2 levels - Most information (75%) is from studies with high risk of bias.
- 3. Inconsistency**
- Downgrade 1 level - Substantial amount of heterogeneity which impact the interpretation of results (e.g. wide variability of point estimates across studies; minimal or no overlap of confidence intervals; large I^2). Heterogeneity is acceptable if it is due to variability in the size of treatment benefits across studies.
- 4. Indirectness**
- Downgrade 1 level - Indirect evidence from a single population, intervention, comparison, or outcome measure not specified in the review (Note: Populations, interventions, comparison interventions, and outcome measures not specified in the inclusion criteria for the review will be excluded, so the evidence will not be downgraded for indirectness)
 - Downgrade 2 levels - Indirect evidence from more than one population, intervention, comparison, or outcome measure not specified in the review (Note: Populations, interventions, comparison interventions, and outcome measures not specified in the inclusion criteria for the review will be excluded, so the evidence will not be downgraded for indirectness).
- 5. Imprecision**
- Downgrade 1 level - The confidence interval includes evidence for and against treatment benefit and the optimal information size was not met. For the analysis of a continuous outcome, using an α level of 0.05, a β of 0.20, and an effect size of 0.20 standard deviations, a total sample size of approximately 400 is required. For the analysis of a dichotomous outcome, using an α level of 0.05, a β of 0.20, a control group probability of 0.45, and a risk difference of 0.10, a total sample size of approximately 800 is required.
- 6. Publication bias**
- Downgrade 1 level - Publication bias is suspected from asymmetry in funnel plot analyses, the number of small published trials with large treatment effects, or evidence of smaller treatment effects in unpublished studies.

Factors that may increase the overall quality of the evidence

(Note: These factors were incorporated into GRADE for observational study designs and will not be used to increase the quality of the evidence from RCTs)

1. Large magnitude of effect;
2. All residual confounding would have reduced the observed effect (true effect underestimated);
3. A dose-response gradient is evident.