HHS Public Access

Author manuscript

Cochrane Database Syst Rev. Author manuscript; available in PMC 2017 May 01.

Published in final edited form as:

Cochrane Database Syst Rev. 2016 May ; 2016(5): . doi:10.1002/14651858.CD012205.

Spinal rehabilitative exercise or manual treatment for the prevention of cervicogenic headache in adults

Mitchell Haas[1], **Gert Brønfort**[2], **Roni L Evans**[2], **Brent Leininger**[2], **John Schmitt**[3], **Morris Levin**[4], **Kristine Westrom**[5], and **Charles H Goldsmith**[6]

[1]University of Western States, Portland, OR, USA

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

[3]St Catherine University, Minneapolis, MN, USA

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

[5]Chaska, Minnesota, USA

[6]Faculty 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:

To quantify and compare the short- and long-term effects of manual treatment and spinal rehabilitative exercise for cervicogenic headache, classified according to the International

Contact address: Mitchell Haas, University of Western States, 2900 NE 132nd Avenue, Portland, OR, 97230, USA. mhaas@uws.edu.

CONTRIBUTIONS OF AUTHORS

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.
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.
Brent Leininger registered the title, drafted the protocol, will search for and obtain copies of studies, extract and enter data into Review Manager, 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.
Charlie Goldsmith will provide methodological and statistical guidance, complete and interpret the analysis.

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 a research background. All members may have a potential special professional interest in the effectiveness of these interventions.
Mitchell Haas: none known.
Gert Bronfort: none known.
Roni Evans: none known.
Brent Leininger: none known.
John Schmitt: none known.
Morris Levin: none known. 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.
Kristine Westrom: none known.
Charlie Goldsmith: none known.
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.

Headache Society's (IHS) diagnostic criteria, with an active or placebo/sham comparison or wait-list control.

BACKGROUND

Description of the condition

Cervicogenic headache (CGH) was recognized as a distinct classification of headache in 1988 by the International Headache Society (IHS) (IHS 1988). Point prevalence estimates for CGH range from 0.4% to4.6% (Nilsson 1995; Sjaastad 2000; Sjaastad 2008). However, there has been some disagreement about the definition of CGH (Pollmann 1997), and some have speculated that the prevalence may be as high as 15% to 20% of patients with a headache complaint (Haldeman 2001). CGH patients have been shown to have a substantial quality-of-life burden, with impairment comparable to patients with episodic tension-type headache and migraine without aura (van Suijlekom 2003). Treatment recommendations include surgery for specifically-identified conditions; injections into various neck structures; medications such as non-steroidal anti-inflammatory drugs (NSAIDs); neck exercises; manual therapy including spinal manipulation, mobilization, and massage; and multimodal approaches (Bronfort 2010; Fernandez-de-Las-Penas 2015; Gallagher 2007; Haldeman 2001; Racicki 2013). Non-pharmacological interventions such as manual therapy and exercise may have valuable risk/benefit profiles because of the risks associated with medications such as NSAIDs (Dabbs 1995; Wolfe 1999) and surgical procedures in general. Physical examination for screening for CGH has demonstrated good reliability and accuracy (Rubio-Ochoa 2016).

CGH has been characterized by pain that starts in the neck or occipital area and can move to other areas of the head (Bogduk 1992). Typically, the patient displays decreased range of cervical motion and palpable tenderness of the cervical paraspinal tissues (Sjaastad 1998).

We will define cervicogenic headache (CGH) according to current IHS criteria IHS 2013:

- **a.** "any headache fulfilling criterion C;
- **b.** clinical, laboratory and/or imaging evidence of a disorder or lesion within the cervical spine or soft tissues of the neck, known to be able to cause headache;
- **c.** evidence of causation demonstrated by at least two of the following:
  - **1.** headache has developed in temporal relation to the onset of the cervical disorder or appearance of the lesion;
  - **2.** headache has significantly improved or resolved in parallel with improvement in

or resolution of the cervical disorder or lesion;

3. cervical range of motion is reduced and headache is made significantly worse by provocative manoeuvres;

4. headache is abolished following diagnostic blockade of a cervical structure or its nerve supply.

d. not better accounted for by another ICHD-3 diagnosis." We note that criterion C2 must be excluded because such post hoc criteria are not compatible with randomized controlled trial designs. The IHS has developed and refined the definition of CGH over the years (IHS 1988; IHS 2004), but we recognize that earlier CGH definitions used in some randomized controlled trials are essentially compatible with the current definition and have negligible effects on headache characteristics that have been used for study eligibility criteria.

Description of the intervention

Manual treatment and/or spinal rehabilitative exercises will be included. 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 (Bergmann 2011; Evans 2010; Haldeman 2005). 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; Bergmann 2011). 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

By definition, CGH is caused by a disorder/lesion of the cervical spine or adjacent soft tissues in the neck. The etiology of CGH pain is believed to originate from pain generation in the neck (Becker 2010), for example from spinal joints (Dwyer 1990), muscles (Schmidt-Hansen 2006), and nerve entrapment (Baron 2011). The definition of CGH includes dysfunction of the cervical spine commonly treated with the therapies under study in this review. Manual treatment and spinal rehabilitative exercise are performed to reduce pain and improve the biomechanical function of the spine including intersegmental and global range of motion, muscle strength, endurance and motor control (Benjamin 2009; Bergmann 2011; Boyling 2004; Jull 2008; O'Leary 2007). There is evidence that manual treatment and spinal rehabilitative exercise decrease nociceptive input from cervical spine structures (Coronado 2012; Jull 2002; O'Leary 2007). CGH usually presents with neck pain and stiffness (Bogduk 1992; Sjaastad 1998) and the reduction of nociceptive afferent input from the cervical spine and surrounding tissues to the trigeminal-cervical nucleus may explain how manual treatment and spinal rehabilitative exercise can improve CGH. Additionally, there is emerging evidence that manual treatment can modulate pain centrally through spinal and supraspinal mechanisms (Bialosky 2009; Bialosky 2014; Pickar 2012). The supraspinal mechanisms may be related directly to mechanical input into the spine, as well as expectancy, placebo, and other nonspecific effects. The effect of manual treatment on central mechanisms of pain processing may provide an additional pathway for the reduction of CGH symptoms.

Harm

It has long been suspected that stroke from cervical artery dissection can be caused by cervical spinal manipulation (Kawchuk 2008). Cervical artery dissection is a rare event and case-control studies have consistently found an association with cervical spine manipulation (Biller 2014; Church 2016). However, the best evidence suggests that the association is not causal. A case-control and case-cross-over study using 100 million person-years of data found no excess risk of stroke from chiropractic care over medical care and suggested the association was due to care-seeking for neck pain and headache, which precedes 80% of vertebrobasilar strokes (Cassidy 2008). It has been concluded that the clinical reports rarely contain useful information for assessing the association between cervical spinal manipulation, cervical artery dissection, and stroke (Wynd 2013), and that no causal relation can be established (Chung 2015; Church 2016).

Why it is important to do this review

CGH is treated both pharmacologically and non-pharmacologically, including by manual treatment and spinal rehabilitative exercise (Biondi 2005a; Wells 2010). This protocol is one of a series of planned new reviews that will serve to update our original review (Bronfort 2004), to provide a comprehensive evaluation of the evidence regarding the efficacy of manual treatment and spinal rehabilitative exercise for CGH. We will also conduct reviews of spinal rehabilitative exercise and manual treatment for migraine (Bronfort 2015) and for tension-type headache (protocol in press). We will compare our findings with other systematic reviews published after our original review (Biondi 2005b; Bronfort 2010;

Bryans 2011; Chaibi 2012; Clar 2014; Fernandez-de-las-Penas 2005; Vernon 1999). We will utilize 'Risk of bias' and quality of evidence assessments recommended by Cochrane for the updated systematic review (Higgins 2011).

OBJECTIVES

To quantify and compare the short- and long-term effects of manual treatment and spinal rehabilitative exercise for cervicogenic headache, classified according to the International Headache Society's (IHS) diagnostic criteria, with an active or placebo/sham comparison or wait-list control.

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 limit studies to those that isolate the effects of the target manual treatment, spinal rehabilitative exercise, or combination of both.

Types of participants—We will include studies reporting on individuals 18 and older with cervicogenic headache classified according to the International Headache Society's (IHS) 2013 criteria (IHS 2013). Some studies are anticipated to pre-date or not utilize the IHS classification system. Two review authors, including one who is a neurologist, will determine if studies pre-dating or not utilizing the IHS classification system can be classified as cervicogenic headache using reported data (e.g. inclusion/exclusion criteria, diagnostic criteria, baseline clinical characteristics) (McCrory 2005). We will also include studies with a focus on CGH as long as participants meet the IHS definition for CGH, and even if distinct other types of headache were included or if there is some ambiguity about headache type (e.g., overlap between definitions between CGH and certain types of tension-type headaches).

Types of interventions—Included studies must assess 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) for the spine. 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 for tension-type headache (Bendtsen 2010).

Primary outcomes: We will use:

1. patient-rated headache frequency measured in number of CGH 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 tension-type headache (Bendtsen 2010). This is consistent with our companion review for tension-type headache (Leininger 2016).

Secondary outcomes: We will include the following secondary outcomes, if available:

1. headache intensity;
2. duration;
3. headache disability (e.g. headache disability index (HDI));
4. analgesic use;
5. 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 measure along with number needed to treat (NNT), if available (Bendtsen 2010). Data on costs 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 three months after the initial study treatment. Long-term follow up will be defined as outcomes evaluated more than three months after onset of study therapy.

Search methods for identification of studies

Electronic searches—We will identify studies by a comprehensive computerized search of the following databases:

- Cochrane Central Register of Controlled Trials (*Cochrane Library*);
- MEDLINE (Ovid);
- EMBASE (Ovid);
- CINAHL (Ebsco);

- BIOSIS (ISI);
- ISI Web of Science: Science Citation Index;
- Dissertation Abstracts;
- MANTIS;
- Index to Chiropractic Literature;
- Chiropractic Research Archives/Abstracts Collection;
- Physiotherapy Evidence Database.

Medical subject headings (MeSH) or equivalent, and text-word terms, will be used. In MEDLINE, we will use a published search strategy for identifying RCTs (Lefebvre 2011). There will be no language restrictions. Searches will be tailored to individual databases. The search strategy for MEDLINE is in Appendix 1.

Searching other resources—We will check the citations in included publications for 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);
- Meta-Register of controlled trials (mRCT) (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 using Poisson regression (Kastner 2007).

Data collection and analysis

Selection of studies—We will not anonymize the studies in any way before assessment. 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. Two authors (MH, BL) will independently select trials to be included in the review based on the explicit inclusion criteria. We will resolve differences in the results of selection by discussion; a third review author will be consulted (GB) 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.

Data extraction and management—We will record explicit information about patient demographics, clinical characteristics, interventions, and outcome measures using standardized abstracting forms. Two non-blinded 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 3, 0 to 100 or frequency per 7, 14 or 28 days). We will normalize outcomes to a common 0 to 100 scale (also referred to as a percentage point scale) to facilitate analysis using mean differences (MDs), a more clinically-intuitive measure than standardized mean differences (SMDs). We will use SMDs for outcomes with conceptually different domains (e.g. disability measured with different individual domains). We will enter all original data on outcomes as normalized mean percentage point scores. We will attempt to standardize headache frequency outcomes to four weeks (28 days). We will enter data into Cochrane's statistical software, Review Manager 2014, to create normalized MD scores and SMD scores whenever possible. 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 include Cochrane's tool for assessing risk of bias (Higgins 2011). At least two authors (MH, BL) will independently assess the risk of bias in each included outcome per study. We will resolve differences in ratings by discussion or by consulting a third author (GB) 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.

We will not exclude outcomes from individual studies 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 for Systematic Reviews of Interventions* (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 for Systematic Reviews of Interventions* 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 (Higgins 2011; Jadad 1996):

- We will judge outcomes for individual studies scoring 3/5 on the Oxford scale as low risk of bias if the remaining domains assessed using the Cochrane 'Risk of bias' tool, 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;
- For 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, we will judge them as being at moderate risk of bias;
- We will judge all outcomes failing to meet the criteria for low or moderate risk of bias 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 and NNT or NNH (the number needed to harm). We will assess the clinical importance by following the guidance of 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. If the standard deviation for follow-up measurements is missing, the standard deviation for that measure at baseline will be used for subsequent follow-up measurements. When standard deviations are not reported, these will be estimated from the confidence intervals if possible. In the absence of these statistics, standard deviations 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 standard deviation 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, intervention, comparison group, outcomes, and follow-up time point. 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. 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), and comparison (i.e. active comparison, placebo/sham, wait-list) using a random-effects model in Review Manager (Review Manager 2014). If outcomes with low or moderate risk of bias are not available, we will pool outcomes with high risk of bias. If $I^2 \geq 50\%$, (Higgins 2002) we will acknowledge the difficulty of making inferences 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 (GRADEpro GDT 2015; Higgins 2011). We will use the GRADE approach for all outcomes, independently of the decision to pool trials for meta-analysis. Two authors (MH, BL) will independently assess the quality of the body of evidence for each outcome. We will resolve differences in ratings by discussion or by consulting a third author (GB) if disagreements cannot be resolved.

Domains that may decrease the quality of the evidence are:

1. study design;
2. risk of bias;
3. consistency of results;

4. directness (generalizability);
5. precision (sufficient data); and
6. publication bias.

Quality of evidence will be decreased if there is a serious (−1 category) or very serious (−2) limitation to study quality (design, risk of bias), important inconsistency (−1), some (−1) or major (−2) uncertainty about directness, imprecise or sparse data (−1), or high probability of reporting bias (−1).

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); and
3. a dose-response gradient is evident.

Quality of evidence will be increased one category for each of these three domains.

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. We will assess the level of quality of evidence as follows:

- 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: 1) manual therapy (manipulation, mobilization, massage, mixed); 2) exercise therapy (strengthening, stretching, motor control, mixed); and 3) frequency (< 15 and ≥ 15 headaches per month) on the overall results. 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 all outcomes using 'Summary of findings' tables from the GRADE system (GRADEpro GDT 2015). 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 down-grading 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

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

External sources

- National Center for Complementary and Integrative Health, National Institutes of Health, USA.
  Provided salary support for Brent Leininger through a post-doctoral fellowship training grant (#F32AT007507). 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

Baron EP, Cerian N, Tepper SJ. Role of greater occipital nerve blockst injections for patients with dizziness and headache. Neurologist. 2011; 17(6):312–317. [PubMed: 22045281]

Becker WJ. Cervicogenic headache: evidence that the neck is a pain generator. Headache. 2010; 50(4): 699–705. [PubMed: 20456156]

Bendtsen L, Bigal ME, Cerbo R, Diener HC, Holroyd K, Lample C, et al. Guidelines for controlled trials of drugs in tension-type headache: second edition. Cephalalgia. 2010; 30(1):1–16. [PubMed: 19614696]

Benjamin, P.; Tappan, F. Tappan's Handbook of Healing Massage Techniques: Classic, Holistic, and Emerging Methods. Boston: Pearson Education; 2009.

Bergmann, TF.; Peterson, DH. Chiropractic Technique: Principles and Procedures. 3rd. St. Louis: Mosby; 2011.

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; 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; 15(2): 136–148. [PubMed: 24361109]

Biller J, Sacco RL, Albuquerque FC, Demaerschalk BM, Fayad P, Long PH, et al. Cervical arterial dissections and association with cervical manipulative therapy: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2014; 45(10):3155–3174. [PubMed: 25104849]

Biondi DM. Cervicogenic headache: a review of diagnostic and treatment strategies. Journal of the American Osteopathic Association. 2005; 105(4 Suppl 2):16S–22S. [PubMed: 15928349]

Biondi DM. Physical treatments for headache: a structured review. Headache. 2005; 45(6):738–746. [PubMed: 15953306]

Bogduk N. The anatomical basis for cervicogenic headache. Journal of Manipulative and Physiological Therapeutics. 1992; 15:67–69. [PubMed: 1740655]

Boyling, JD.; Jull, GA. Grieve's Modern Manual Therapy. 3rd. London: Churchill Livingstone; 2004.

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]

Brønfort 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 Reports. 1985; 100(2):126–131. [PubMed: 3920711]

Cassidy JD, Boyle E, Cote P, He Y, Hogg-Johnson S, Silver FL, Bondy SJ. Risk of vertebrobasilar stroke and chiropractic care: results of a population-based case-control and case-crossover study. Spine. 2008; 33:S176–S183. [PubMed: 18204390]

Chaibi A, Russell MB. Manual therapies for cervicogenic headache: a systematic review. Journal of Headache and Pain. 2012; 13(5):351–359. [PubMed: 22460941]

Chung CL, Cote P, Stern P, L'Esperance G. The association between cervical spine manipulation and carotid artery dissection: a systematic review of the literature. Journal of Manipulative and Physiological Therapy. 2015; 38:672–676.

Church EW, Sieg EP, Zalatimo O, Hussain NS, Glantz M, Harbaugh RE. Systematic review and meta-analysis of chiropractic care and cervical artery dissection: no evidence for causation. Cureus. 2016; 8(2):e498. [PubMed: 27014532]

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; 22(5):752–767. [PubMed: 22296867]

Dabbs V, Lauretti WJ. A risk assessment of cervical manipulation vs NSAIDS for the treatment of neck pain. Journal of Manipulative and Physiological Therapeutics. 1995; 18(8):530–536. [PubMed: 8583176]

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]

Dwyer A, Aprill C, Bogduk N. Cervical Zygapophyseal joint pain patterns I: a study in normal volunteers. Spine. 1990; 15(6):453–457. [PubMed: 2402682]

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, Pareja JA. Spinal manipulative therapy in the management of cervicogenic headache. Headache. 2005; 45:1260–1263. [PubMed: 16178960]

Fernandez-de-Las-Penas C, Cuadrado ML. Physical therapy for headaches. Cephalalgia. 2015 Dec 9. pii: 0333102415596445. [Epub ahead of print].

Furlan AD, Pennick V, Bombardier C, van Tulder M. Editorial Board 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]

Gallagher RM. Cervicogenic headache. Expert Review of Neurotherapeutics. 2007; 7(10):1279–1283. [PubMed: 17939766]

Glass, GV.; McGaw, B.; Smith, ML. Meta-analysis in Social Research. Beverly Hills, CA: Sage Publications; 1981.

Brozek, J.; Oxman, A.; Schünemann, H. GRADEpro Guideline Development Tool [Software]. McMaster University (developed by Evidence Prime, Inc.); 2008.

Haldeman S, Dagenais S. Cervicogenic headaches: a critical review. Spine Journal. 2001; (1):31–46. [PubMed: 14588366]

Haldeman, S.; Dagenais, S.; Budgell, B.; Grunnet-Nilsson, N.; Hooper, P.; Meeker, W., et al., editors. 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. Cochrane Handbook for Systematic Reviews of Interventions. Vol. 5.1.0. The Cochrane Collaboration; 2009. Available from www.cochrane-handbook.org. [Updated 2011]

Headache Classification Committee of the International Headache Society. Classification and Diagnostic Criteria for Headache Disorders, Cranial Neuralgias, and Facial Pain. Cephalagia. 1988; 8(Suppl 7):1–96.

Headache Classification Committee of the International Headache Society. The international classification of headache disorders, 2nd edition. Cephalalgia. 2004; 24(Suppl 1):9–160. [PubMed: 14979299]

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; 17(1):1–12. [PubMed: 8721797]

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; 27(17):1835–1843. [PubMed: 12221344]

Jull, G.; Sterling, M.; Falla, D.; Treleaven, J.; O'Leary, S. Whiplash, Headache, and Neck Pain: Research-Based Directions for Physical Therapies. Edinburgh, UK: Elsevier; 2008.

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.

Kawchuk GN, Jhangri GS, Hurwitz EL, Wynd S, Haldeman S, Hill MD. The relation between the spatial distribution of vertebral artery compromise and exposure to cervical manipulation. Journal of Neurology. 2008; 3:371–377.

Lefebvre, C.; Manheimer, E.; Glanville, J. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration: Higgins JPT, Green S; 2011. Chapter 6: Searching for studies. Available from www.cochrane-handbook.org

Leininger B, Brønfort G, Haas M, Schmitt J, Evans RL, Levin M, et al. Spinal rehabilitative exercise or manual treatment for the prevention of tension-type headache in adults. Cochrane Database of Systematic Reviews. 2016; (4)

McCrory 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]

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]

Nilsson N. The prevalence of cervicogenic headache in a random population sample of 20–59 year olds. Spine. 1995; 20:1884–1888. [PubMed: 8560336]

O'Leary S, Falla D, Jull G. Recent advances in therapeutic exercise for the neck: implications for patients with head and neck pain. Australian Endodontic Journal. 2003; 29(3):138–142. [PubMed: 14700399]

Pickar JG, Bolton PS. Spinal manipulative therapy and somatosensory activation. Journal of Electromyography and Kinesiology. 2012; 22(5):785–794. [PubMed: 22349622]

Pollmann W, Keidel M, Pfaffenrath V. Headache and the cervical spine: a critical review. Cephalalgia. 1997; 17:801–816. [PubMed: 9453267]

Racicki S, Gerwin S, Diclaudio S, Reinmann S, Donaldson M. Conservative physical therapy management for the treatment of cervicogenic headache: a systematic review. Journal of Manipulative and Physiological Therapeutics. 2013; 21(2):113–124.

The Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager (RevMan). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.

Rubio-Ochoa J, Benitez-Martinez J, Lluch E, Santacruz-Zaragoza S, Gomez-Contreras P, Cook CE. Physical examination tests for screening and diagnosis of cervicogenic headache: a systematic review. Manual Therapy. 2016; 21:35–40. [PubMed: 26423982]

Schmidt-Hansen PT, Svensson P, Jensen TS, Graven-Nielsen T, Bach FW. Patterns of experimentally induced pain in pericranial muscles. Cephalalgia. 2006; 26(5):568–577. [PubMed: 16674766]

Simons DG. New views of myofascial trigger points: etiology and diagnosis. Archives of Physical Medicine and Rehabilitation. 2008; 89(1):157–159. [PubMed: 18164347]

Sjaastad O, Fredriksson TA, Pfaffenrath V. Cervicogenic headache: diagnostic criteria. Headache. 1998; 38:442–445. [PubMed: 9664748]

Sjaastad O, Fredriksson TA. Cervicogenic headache: criteria, classification and epidemiology. Clinical and Experimental Rheumatology. 2000; 18(2 Suppl 19):S3–S6. [PubMed: 10824278]

Sjaastad O, Bakketeig LS. Prevalence of cervicogenic headache: Vaga study of headache epidemiology. Acta Neurologica Scandinavica. 2008; 117:173–180. [PubMed: 18031563]

van Suijlekom HA, Lame I, Stomp-van den Berg SG, Kessels AG, Weber WE. Quality of life of patients with cervicogenic headache: a comparison with control subjects and patients with migraine or tension-type headache. Headache. 2003; 43(10):1034–1041. [PubMed: 14629238]

Vernon H, McDermaid CS, Hagino C. Systematic review of randomized clinical trials of complementary/alternative therapies in the treatment of tension-type and cervicogenic headache. Complementary Therapies in Medicine. 1999; 7:142–155. [PubMed: 10581824]

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]

Wolfe MM, Lichtenstein DR, Singh G. Gastrointestinal toxicity of non-steroidal anti-inflammatory drugs. New England Journal of Medicine. 1999; 340(24):1888–1899. [PubMed: 10369853]

Wynd S, Estaway M, Vohra S, Kawchuk G. The quality of reports on cervical arterial dissection following cervical spinal manipulation. PLoS One. 2013; 8:e59170. [PubMed: 23527121]

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, Bouter L. Non-invasive physical treatments for chronic/recurrent headache. Cochrane Database of Systematic Reviews. 2004; (3)

Appendix 1. MEDLINE search strategy

MEDLINE (Ovid)

1.	exp Headache Disorders/
2.	Headache/
3.	(cervicogenic* 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:

1.	mp=protocol supplementary concept, rare disease supplementary concept, title, original title, abstract, name of substance word, subject heading word, unique identifier
2.	pt=publication type
3.	ab=abstract
4.	sh=subject heading
5.	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; • Shuffling cards or envelopes; • Throwing dice; • Drawing of lots; • Minimization*. *Minimization may be implemented without a random element, and this is considered to be equivalent to being random

Criteria for the judgement of 'High risk' of bias.	The investigators describe a non-random component in the sequence generation process. Usually, the description would involve some systematic, non-random approach, for example: • 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. 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: • 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.	Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: • 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.	Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on: • 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; • 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.	Any one of the following: • 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.	Any one of the following: • 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.	Any one of the following: • 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.	Any one of the following: • 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.	Any one of the following: • 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.	Any one of the following: • Insufficient information to permit judgement of 'Low risk' or 'High risk'; • 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).
Criteria for the judgement of 'High risk' of bias.	Any one of the following: • 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.	There is at least one important risk of bias. For example, the study: • 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.	There may be a risk of bias, but there is either: • 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.