2018

Paracetamol, NSAIDS and opioid analgesics for chronic low back pain: A network meta-analysis (Protocol)

Matthew K. Bagg
Andrew J. McLachlan
Christopher G. Maher
Steven J. Kamper
Christopher M. Williams

See next page for additional authors

Follow this and additional works at: https://researchonline.nd.edu.au/physiotherapy_article

Part of the Physical Therapy Commons, and the Physiotherapy Commons

This article was originally published as:

Original article available here:
https://doi.org/10.1002/14651858.CD013045

This article is posted on ResearchOnline@ND at
https://researchonline.nd.edu.au/physiotherapy_article/139. For more information, please contact researchonline@nd.edu.au.
This article first published in the Cochrane Database of Systematic Reviews, June 2018. Available:  https://doi.org/10.1002/14651858.CD013045

Paracetamol, NSAIDS and opioid analgesics for chronic low back pain: a network meta-analysis (Protocol)


www.cochranelibrary.com
**TABLE OF CONTENTS**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEADER</td>
<td>1</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>1</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>OBJECTIVES</td>
<td>3</td>
</tr>
<tr>
<td>METHODS</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1</td>
<td>4</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>10</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>10</td>
</tr>
<tr>
<td>ADDITIONAL TABLES</td>
<td>15</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>18</td>
</tr>
<tr>
<td>CONTRIBUTIONS OF AUTHORS</td>
<td>25</td>
</tr>
<tr>
<td>DECLARATIONS OF INTEREST</td>
<td>25</td>
</tr>
<tr>
<td>SOURCES OF SUPPORT</td>
<td>26</td>
</tr>
</tbody>
</table>

Paracetamol, NSAIDS and opioid analgesics for chronic low back pain: a network meta-analysis (Protocol)
**Paracetamol, NSAIDS and opioid analgesics for chronic low back pain: a network meta-analysis**

Matthew K Bagg\(^1,2,3\), Andrew J McLachlan\(^4,5\), Christopher G Maher\(^6\), Steven J Kamper\(^6\), Christopher M Williams\(^7,8\), Nicholas Henschke\(^9\), Benedict M Wand\(^10\), G L. Moseley\(^1,11\), Markus Hübscher\(^1\), Neil E O’Connell\(^12\), Maurits W van Tulder\(^13\), Adriani Nikolakopoulou\(^14\), James H McAuley\(^1,15\)

\(^1\)Neuroscience Research Australia, Sydney, Australia. \(^2\)Prince of Wales Clinical School, University of New South Wales, Sydney, Australia. \(^3\)New College Village, University of New South Wales, Sydney, Australia. \(^4\)Faculty of Pharmacy, University of Sydney, Sydney, Australia. \(^5\)School of Public Health, University of Sydney, Sydney, Australia. \(^6\)School of Physiotherapy, The University of Notre Dame Australia Fremantle, Perth, Australia. \(^7\)School of Medicine and Public Health, University of Newcastle, Callaghan, Australia. \(^8\)Hunter Medical Research Institute, New Lambton, Australia. \(^9\)Sydney Medical School, University of Sydney, Sydney, Australia. \(^10\)School of Physiotherapy, The University of Notre Dame Australia Fremantle, Perth, Australia. \(^11\)Sanson Institute for Health Research, University of South Australia, Adelaide, Australia. \(^12\)Health Economics Research Group, Institute of Environment, Health and Societies, Department of Clinical Sciences, Brunel University London, Uxbridge, UK. \(^13\)Department of Health Sciences, Faculty of Science and Amsterdam Movement Science institute, Vrije Universiteit, Amsterdam, Netherlands. \(^14\)Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland. \(^15\)School of Medical Sciences, University of New South Wales, Sydney, Australia

Contact address: James H McAuley, Neuroscience Research Australia, Sydney, NSW, Australia. j.mcAuley@neura.edu.au.

**Editorial group:** Cochrane Back and Neck Group.

**Publication status and date:** New, published in Issue 6, 2018.


Copyright © 2018 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

**A B S T R A C T**

This is a protocol for a Cochrane Review (Intervention). The objectives are as follows:

To answer the clinical question: ‘what analgesic medicine shall I prescribe this patient with chronic low back pain to reduce their pain?’.

The objectives are to determine the analgesic effects, safety, effect on function, and relative rank according to analgesic effect, safety and effect on function of a single course of opioid analgesics, NSAIDs or paracetamol or combinations of these medicines.

**Description of the condition**

Low back pain (LBP) presents an enormous global health problem. It is the leading cause of disability (GBD Collaborators 2017a).
and fourth-leading cause of disease burden amongst non-communicable health conditions (GBD Collaborators 2017b). In the preceding 20 years, the amount of disability caused by LBP has remained stable; however the disease burden has increased. Recent-onset (acute) LBP has a favourable initial prognosis (Menezes Costa 2012). However, only one-third of people are recovered 12 weeks from pain onset and this changes little at 12 months from pain onset (Itz 2013). LBP persisting longer than 12 weeks is termed chronic low back pain (Furlan 2015; Treede 2015). Technically, pain in the low back area is a symptom. Nonetheless, the usual use of the term LBP is to describe the condition of non-specific LBP (NSLBP). NSLBP is the most common form of LBP; commonly stated to be 90% of all cases (Koes 2006), and is so termed because there is no identifiable problem (in the back) that is causing the LBP symptoms (Maher 2017). Much less common conditions for which there is an identifiable problem include fracture, infection, malignancy (tumour), radiculopathy (sciatica involving problematic nerve function), spinal stenosis (narrowing of spinal openings) or inflammatory arthropathy (inflammatory joint problems) (Downie 2013; Enthoven 2016a; Hensche 2008; Koes 2006). The focus of this review is chronic non-specific LBP (CLBP).

**Description of the intervention**

Medicines prescribed to reduce pain intensity (analgesic medicines) are the most common form of treatment for CLBP (Gore 2012; Hart 2015; Ivanova 2011). Analgesic medicines work in various ways to reduce the intensity of pain but may also cause unwanted harmful effects. Clinical decisions to recommend particular pain-relieving medicines involve a balance between the anticipated reduction in pain and the possibility of side effects. This is particularly important when people have other relevant health conditions such as the inability to tolerate some medicines, or problems with kidney or liver function. Clinical use of these medicines varies across healthcare settings and geographical areas and is usually different to guideline recommendations (Amorin-Woods 2014; Bishop 2003; Cifuentes 2010; Gore 2012; Gouveia 2017; Hart 2015; Hoffmann 2013; Ivanova 2011; Mafi 2013; Ndlouvu 2014; Webster 2005; Webster 2007). In this review, we will focus on the three most commonly prescribed classes of analgesic medicines: paracetamol (acetaminophen); non-steroidal anti-inflammatory drugs (NSAIDs); and opioid analgesics.

**How the intervention might work**

The human body creates its own natural opioid molecules (endogenous opioids), for a variety of functions. The nervous system contains receptors for these opioids on the exterior of cells (Kieffer 2009). When opioid molecules bind to receptors, this causes changes within the cells. Opioid analgesics, as they are termed in this review, are medicines made synthetically that match or are chemically similar to endogenous opioids. When taken, opioid analgesics attach to the opioid receptors on nervous system cells, causing a reduction in perceived pain (Kieffer 2009; Rachinger-Adam 2011). Opioids have other effects as well, depending on the location and type of the receptor. One common side effect from taking opioids is constipation, because the medicine also binds to receptors in the digestive system and interferes with digestion (Rachinger-Adam 2011); other side effects are that people taking opioids may develop tolerance (they require larger doses for the same effect) or dependence on the medicine (they experience withdrawal symptoms if a dose is missed), or both (Deyo 2015).

NSAIDs and paracetamol interfere with the natural inflammatory processes in the body, in particular the production of chemicals called prostaglandins (Brune 2015). Prostaglandins are themselves involved in several processes. One of these is pain perception, which they contribute to by increasing inflammation. They also contribute to blood-clotting processes and support the lining of the stomach (Brune 2015). An enzyme called cyclooxygenase (COX) helps to create prostaglandins. NSAIDs lessen the effect of COX type-1 (COX-1) and COX type-2 (COX-2), which reduces prostaglandins and, in turn, reduces the perception of pain. The unavoidable reduction in prostaglandins elsewhere (e.g. the stomach) is associated with side effects (e.g. stomach ulcers) (Brune 2015). Some NSAIDs selectively inhibit COX-2, which is less present in the stomach: these ‘selective’ NSAIDs are associated with diminished digestive system side effects (Davies 2000; Sostres 2013; Whittle 2000). Both selective and non-selective NSAIDs have known cardiovascular risks (Baigent 2013; McGettigan 2011).

Paracetamol also inhibits the COX-1 and COX-2 enzymes (Graham 2013; Jozwiak-Bebenista 2014), particularly COX-2 (Graham 2013; Hinz 2008; Hinz 2012). Paracetamol is generally a safe medicine to take. However, overdosing (exceeding the recommended daily dose) is associated with an increased risk of liver problems and death (Daly 2008; Graham 2013; Sheen 2002).

**Why it is important to do this review**

There is uncertainty about the comparative effectiveness of most analgesic medicines for CLBP. This is problematic given their wide use (Gore 2012; Hart 2015; Ivanova 2011), side effects (Baigent 2013; Deyo 2015; Humphreys 2017; Martell 2007; McGettigan 2011; Shaheed 2016), and the variety of available formulations. Evidence of comparative effectiveness - the analgesic effect and safety of a medicine in comparison to other medicines that may be prescribed for the same condition - is the information required most by clinicians. This information is presently insufficiently described in the literature. Systematic reviews of these medicines typically include limited data on comparative effectiveness (Chaparro 2014; Chou 2017; Enthoven 2016b; Saragiotto 2016b), and
focus on effects compared to placebo (Chaparro 2014; Chou 2007; Chung 2013; Enthoven 2016b; Furlan 2006; Furlan 2011; Kuipers 2011; Machado 2017; Shaheed 2016; White 2011a), or aspects of clinical trial design (Furlan 2006; Furlan 2011). While many factors influence prescribing decisions (Bhamb 2006; Cabana 1999; Darlow 2012; Perrot 2008; Schers 2001), the lack of evidence of comparative effectiveness for many analgesics means that prescribing is less evidence-informed and more reliant on other factors such as clinical experience and patient preferences. Network meta-analysis (NMA) can provide the information on comparative effectiveness that is needed to make informed clinical decisions. NMA is a simultaneous comparison of all competing effects (Salanti 2012). The results of NMA are expressed as the effect of a treatment, on a particular outcome, compared to every other competing treatment. This review will provide a comprehensive evaluation of the comparative effectiveness of analgesic medicines for CLBP and will express these data in formats that are accessible to clinicians.

**Description of network meta-analysis**

Network meta-analysis (NMA) is an extension of pairwise meta-analysis that compares more than two treatments (Lu 2004). NMA uses networks to represent the evidence base. Each competing treatment is represented as a node (vertex) and each comparison between treatments is represented by lines (edges) connecting the treatments (Jansen 2011). As with pairwise meta-analysis, it is important that an NMA is conducted within a rigorous systematic review process to ensure that all relevant information is included. Information that is missing from the network, for example clinical trials of a particular treatment comparison, may lead to bias in the results (Mills 2013). A network with no missing information will likely still contain a mix of solid and blank (or absent) edges. Solid edges are termed ‘direct’ comparisons because these have been compared in clinical trials. Absent edges are termed ‘indirect’ comparisons - there have been no clinical trials of these comparisons (Bucher 1997; Jansen 2011).

NMA uses the ‘direct’ clinical trial data to ‘indirectly’ estimate the effects for the missing comparisons (blank edges) (Caldwell 2005; Cipriani 2009; Jansen 2011; Jansen 2013; Mills 2013; Song 2003; Sutton 2008). In this way, NMA can be said to fill in the gaps in the evidence base. This process involves fitting a single statistical model to the network, which simultaneously combines all the information from the clinical trials (the direct comparisons) together with the information about the network structure (in particular, the missing comparisons) (Efthimiou 2016; Krahn 2013; Salanti 2012).

The transitivity assumption underpins the validity of NMA. The assumption holds when the characteristics of the participants and trials included in the network are sufficiently similar across all of the direct comparisons in the network (Jansen 2013). From a researcher’s perspective, transitivity means that if all treatments in the network were included in a single mega-trial, there is an equal likelihood that any one participant may be randomised to any of those treatments (Salanti 2012). From a participant’s perspective, transitivity means that if they were a participant in any one of the trials in the network, they would have an equal chance of being a participant in any of the other trials in the network. The results from NMA may be portrayed in formats that are useful for clinical decision-making. The results for each treatment are displayed in a league table, which may display results for two outcomes (e.g. analgesic effect and safety). This is helpful for clinicians, because they can see the evidence for a treatment’s benefits alongside its risks. The league table displays all of the results, which means that clinicians have all of the required information in one place. These results can also be used to produce rankings of the effect of treatments on a particular outcome (Dominici 1999; Hoaglin 2011; Jansen 2011; Sutton 2008; van der Valk 2009).

**OBJECTIVES**

To answer the clinical question: ‘what analgesic medicine shall I prescribe this patient with chronic low back pain to reduce their pain?’.

The objectives are to determine the analgesic effects, safety, effect on function, and relative rank according to analgesic effect, safety and effect on function of a single course of opioid analgesics, NSAIDs or paracetamol or combinations of these medicines.

**METHODS**

**Criteria for considering studies for this review**

**Types of studies**

We will include conventional and enriched-design parallel group RCTs and the first phase of cross-over RCTs. We will exclude cluster RCTs. Studies must compare at least two of the interventions of interest.

**Types of participants**

We will include studies of people with chronic, non-specific LBP defined as a primary area of pain between the twelfth rib and gluteal fold, with or without associated leg pain, for which the same episode has persisted for longer than three months (Furlan 2015; Treede 2015). We will exclude studies where subjects have leg pain that meets the definition of sciatica used by Koes 2007 or where low back pain is caused by pathological entities such as infection, neoplasm, metastasis, osteoporosis, inflammatory disease.
or fractures. We will include studies that include participants with conditions other than CLBP (e.g. sciatica or ankylosing spondylitis) if separate data is reported for the CLBP participants or may be obtained from study authors.

**Types of interventions**

We will include studies of analgesic medicines prescribed for the intended purpose of reducing pain intensity for someone with CLBP, defined as paracetamol (acetaminophen), NSAIDs and opioid analgesics. The interventions of interest are listed in Appendix 1 and include placebo and no treatment. Each medicine is a separate intervention (node) in the analysis. We will not combine different medicines into the same node. The NMA will compare every medicine to every other medicine. If we identify NSAIDs or opioid analgesics that are inadvertently not listed in Appendix 1, we will consider them as eligible and we will include them in the network after assessing their comparability with the pre-specified set of competing interventions. We will report the findings for these interventions in the results and the conclusions of the review.

We will include analgesic medicines delivered as mono or combination therapy (delivery of one or more agents at once for the same intended therapeutic effect, in single or multiple formulations) via all systemic routes of administration. We will exclude local therapeutic injections or topical applications and medicines or dosages that are not currently licensed for human use. We will include trials that contain non-pharmacological co-interventions (intended by the trial investigators) in one or more of the intervention arms. Such co-interventions will be categorised as i) physical therapies, or ii) psychological therapies, or iii) combinations of i) and ii), for the evaluation of transitivity.

We assume that all analgesic medicines, defined herein, are directly comparable treatments. In other words, we assume that the distribution of important characteristics (effect modifiers) is the same across all treatment comparisons (Salanti 2012). The network diagram in Figure 1 displays all possible pairwise comparisons between opioid analgesics. We have not included NSAIDs or paracetamol in this diagram for the sake of visual simplicity. All medicines are categorised according to the World Health Organization (WHO) Anatomical Therapeutic Chemical (ATC) system. The ATC codes are listed in Appendix 1.

**Figure 1.** All theoretically possible comparisons between the opioid analgesic interventions of interest. NSAIDs and paracetamol are not included for the sake of simplicity.
Network nodes will be further defined using the relevant licensed dosing range (LDR) from the US Food and Drug Administration (FDA), the Australian Therapeutic Goods Administration (TGA) or the European Medicines Agency (EMA). For most medicines, this is equivalent to the standard dosing range (SDR). For older medicines, the SDR may be smaller than the LDR. The network will include these doses as separate nodes. Thus, for drug A there may be a single LDR node, whereas for drug B there may be a ‘less than SDR’ node, an SDR node and a ‘greater than SDR’ node. We will use the standard dosing range from the Australian Medicines Handbook (AMH) and the Prescriber’s Digital Reference or MIMS when the medicine is not listed in the AMH (AMH 2017; MIMS 2017; PDR 2017). The placebo node is defined as any drug intervention that does not contain an active ingredient. The ‘no treatment’ node is defined as any trial arm that contains no investigator-intended treatment, and includes trial arms that comprise continuation of usual care or being placed on a wait-list.

**Types of outcome measures**

**Primary outcomes**

The co-primary outcomes are pain and safety.

- Pain is defined as pain intensity, measured at the time point closest to the end of treatment. Pain intensity may be measured with a continuous self-report scale (e.g. visual analogue scale (VAS) or numeric rating scale (NRS)), a rating scale within a composite measure of pain (e.g. McGill Pain Questionnaire), or an ordinal scale with greater than six levels (we will consider such ordinal scales to exhibit continuous properties). We will not exclude studies that use other measurement tools. We will estimate the relative ranking of the competing interventions according to their effect on pain intensity at the end of treatment.

- Safety is defined as the proportion of participants who experience a serious adverse effect during the treatment period. Adverse effects are described broadly as any of ‘adverse event’, ‘adverse drug reaction’, ‘side effect’, ‘toxic effect’ or ‘complication’ that are associated with the medicine under investigation. A serious adverse effect is that which causes a reduction in dose or cessation of treatment. No change or an increase in pain is not considered an adverse effect. We will estimate the relative ranking of the competing interventions according to their effect on safety.

We note that randomised controlled trials have limitations in the evaluation of medicine safety, particularly enriched designs (Furlan 2011). The study size, duration of treatment, length of follow-up and inclusion criteria are usually not appropriate to detect rare events or evaluate long-term adverse effects (Brewer 1999; Sills 1986; Sutton 2002). Data on adverse effects may be available from multiple sources (Sutton 2002), including observational cohort designs, which are not included in this review. The Cochrane Back and Neck Group is currently conducting a review of harms of opioids for chronic non-cancer pain (Furlan 2014), and other reviews of harms of NSAIDs have been published (for example, Baigent 2013).

**Secondary outcomes**

- Pain, defined as pain intensity, measured at 1-month post randomisation, provided the individual study treatment is complete. The outcome is otherwise defined as above.

- Functional ability, defined as low back specific function, measured at end of treatment. Functional ability may be measured with a continuous, self-report scale (e.g. Roland Morris Disability Questionnaire (RMDQ) or Oswestry Disability Index (ODI)), a rating scale within a composite measure (e.g. SF-36), or an ordinal scale with greater than six levels (we will consider such ordinal scales to exhibit continuous properties). We will not exclude studies that use other measurement tools. We will estimate the relative ranking of the competing interventions according to their effect on functional ability at the end of treatment.

**Search methods for identification of studies**

**Electronic searches**

We will search for all possible comparisons formed by the interventions of interest. We will search the following electronic databases from inception to current.

- Cochrane Back and Neck Group’s Trials Register (through CENTRAL).
- Cochrane Central Register of Controlled Trials (CENTRAL) in the Cochrane Library, current issue.
- MEDLINE (Ovid) (1946 to current).
- Embase (Ovid) (1980 to current).
- CINAHL (EBSCO) (1982 to current).
- ClinicalTrials.gov (ClinicalTrials.gov/ct2/home).
- WHO International Clinical Trial Registry Platform (WHO ICTRP) (apps.who.int/trialsearch/Default.aspx).

We have developed the search strategies specifically for this review, using the recommended back pain search terms (Furlan 2015), and specific terms for the interventions of interest and their combinations. The search strategy for MEDLINE is listed in Appendix 2.
Searching other resources

We will check the reference lists of eligible studies and previous systematic reviews to identify additional studies. We will contact the corresponding authors of the most recent systematic reviews on opioid analgesics, NSAIDs and paracetamol to enquire if they are aware of any additional studies in the area of their review. We will search for the protocols and registrations of all included studies.

Data collection and analysis

Selection of studies

Two reviewers from a panel of six (MKB, SK, BMW, AM, NH, CW) will independently screen identified studies for eligibility in two independent stages: i) title and abstract; ii) full text. Review authors will not be involved in the decision to include any study with which they have been involved. The reviewers will resolve disagreements through discussion, with arbitration by a third reviewer (JHM) if required. We will contact study authors up to three times to obtain additional information to determine study eligibility. We will place the study in ‘Studies awaiting assessment’ for this iteration of the review if no reply is received within six weeks. We will obtain translations of articles written in languages that cannot be read by the authors through colleagues or Cochrane resources such as Task Exchange. We will summarise the search process using a PRISMA flow diagram (Liberati 2009).

Data extraction and management

Two reviewers from a panel of six (MKB, BMW, MH, NOC, NH, CW) will independently extract data from included studies. Review authors will not extract data from any study in which they have had any involvement. We will use standardised, piloted, data extraction forms. Required data will be taken from previous Cochrane or non-Cochrane Reviews (conducted by the authors) when possible. We will scan these data for implausible values; and if found we will refer to the original article. The colleagues providing the translation of studies written in other languages will also extract data from these studies. The reviewers will resolve disagreements through discussion, with arbitration by a third reviewer (JHM) if required. We will contact study authors up to three times to request additional data. We will consider the data unobtainable for this iteration of the review if no reply is received within six weeks. We will only extract data for the interventions of interest from multi-arm studies of many interventions. We will preference published over unpublished data in situations where data sources conflict, because published data have been through peer review.

Study methods

We will extract data on the trial design. This will include the number, country and setting of trial sites, sample size and total duration of the trial.

Participants

We will extract data on the individual study sample. This will include the diagnosis, duration of LBP episode, age, male/female ratio and co-morbidities, including alternate sites of pain. We will extract arm-level pain intensity at baseline (mean (SD)).

Interventions

We will extract data on the interventions. This will include the duration of the intervention, including duration of any washout, run-in or titration period; dosage regimen, including titration; and routes of administration and usage of rescue medication.

Outcomes

We will extract the type and dimensions of the scale or measure used to assess pain or functional ability and the time from randomisation at which the end of treatment data were obtained in the individual trials. We will extract the definition of ‘adverse effect’ and ‘serious adverse effect’ used in each study.

Results

We will extract the number of participants allocated to each intervention group and the proportion of adherence to the intervention, including the definition of adherence. We will extract the proportion of participants in each group who discontinued treatment due to an adverse effect. We will extract from each trial the event rates and descriptions of all reported adverse effects and serious adverse effects. We will extract the outcome score (preferred) for pain and functional ability or the change in outcome from baseline and the accompanying measure of variance (or available statistic to estimate these values) for each group at the time point closest to the end of treatment. We will also extract these data for pain at the time point closest to 1-month post-randomisation, provided study treatment is complete. We will extract the between-group differences in scores and the accompanying measure of variance if group-level data are not available. We will select and extract data from a single outcome measure, in studies with more than one relevant outcome measure for pain, in the following order: 100 mm VAS; 10 cm VAS; 11-point NRS; rating scale for pain intensity from a composite measure of pain; ordinal scale with more than six levels. We will select and extract data from a single outcome measure, in studies with more than one relevant outcome measure for functional ability, in the following order: ODI; RMDQ; rating scale for functional ability from a composite measure; ordinal scale with more than six levels.

Paracetamol, NSAIDs and opioid analgesics for chronic low back pain: a network meta-analysis (Protocol)

Copyright © 2018 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.
Assessment of risk of bias in included studies

Two reviewers from a panel of six (MKB, CW, MH, NH, SK, NOC) will independently appraise outcome-level risk of bias for the domains of selection, performance, attrition, detection, reporting and other sources of bias, using the Cochrane ‘Risk of bias’ tool, version 5.1.0 (Higgins 2011), and recommendations of Furlan 2015 (Table 1; Table 2). Review authors will not appraise risk of bias for any study in which they have had any involvement. We will pilot-test the ‘Risk of bias’ assessment procedure on a small number of articles. ‘Risk of bias’ assessments will be taken from previous Cochrane Reviews or from previous reviews conducted by the authors when possible. We will add any missing outcome-level assessments.

We will determine single outcome-level risk of bias ratings for each study, using an adaptation of the criteria in Furukawa 2016 for the GRADE evaluation. Outcomes are at low overall risk when three or fewer domains are rated ‘unclear’ risk and no domains are rated ‘high’; moderate overall risk if a single domain is rated as ‘high’ risk of bias, or no domain is rated as ‘high’ risk but four or more are rated as ‘unclear’; and high overall risk in all other instances. The reviewers will resolve disagreements through discussion, with arbitration by a third reviewer (JHM) if required.

Measures of treatment effect

Relative treatment effects

We will convert all outcome data for pain and functional ability to common 0 to 100 point scales (mean (SD)). We will estimate the relative treatment effects of the competing interventions on pain and functional ability using weighted mean differences (WMD), with 95% confidence intervals (95% CIs); and on safety using risk ratios (RRs), with 95% CIs. We will present the results from NMA as WMDs (pain and functional ability) and RRs (safety), with 95% CIs, for each intervention compared to the reference (placebo).

We will interpret the meaningfulness of the effect of medicines on pain using the threshold of 10 points on a 0 to 100 point scale (Chou 2017). We will use the 95% CIs of the risk ratio for safety to judge meaningfulness on this outcome.

Relative treatment ranking

We will obtain hierarchies (rankings) of the effect of all interventions on pain and on safety using two measures: the probability to be at each possible rank; and the surface under the cumulative ranking curve (SUCRA) (Salanti 2011). We will present these hierarchies using rankograms, cumulative probability plots and clustered ranking graphs (Chaimani 2013; Salanti 2011).

Unit of analysis issues

Studies with multiple treatment groups (multi-arm trials)

We will fit the network meta-analysis models using multivariate meta-analysis in Stata (Higgins 2012; StataCorp 2015; White 2009; White 2011b; White 2012; White 2015), which accounts for the presence of multi-arm trials. We will combine intervention arms from the same trial that are within the same licensed dosing range, using the formulae in Table 7.7a of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011).

Dealing with missing data

We will contact study authors to provide missing data. We will impute missing data that are required for meta-analysis (e.g. standard deviations), if they are not obtainable for this iteration of the review, using previously published methods (Wan 2014), and those described in Sections 7.7 and 16.1 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011).

Assessment of heterogeneity

Assessment of heterogeneity within treatment comparisons

We will use clinical and methodological judgement to assess the presence of clinical and methodological heterogeneity within each pairwise comparison by comparing the trial and study population characteristics across all eligible trials. A judgement of excessive clinical or methodological heterogeneity within a particular pairwise comparison will prevent meta-analysis.

Assessment of transitivity across the network

Transitivity is the fundamental assumption underpinning the validity of network meta-analysis. It is a statistically untestable assumption, relying on a subjective assessment of clinical and methodological heterogeneity across the network. Transitivity assumes that there is no imbalance in the distribution of treatment effect modifiers across each of the comparisons in the network (Jansen 2013). Any imbalance threatens the transitivity assumption. We will use several methods to evaluate the transitivity assumption. We consider the following clinical and methodological factors to be potential treatment effect modifiers, in the absence of robust evidence to date for effect modification in LBP trials (Saragiotto 2016a).

- Baseline pain intensity
- Intended co-interventions, categorised as i) physical therapies ii) psychological therapies, iii) combinations of i) and ii)
- Sample size (Dechartres 2016)
- Enriched design
We will visually assess the distributions of these effect modifiers across all treatment comparisons in the network, using weighted network plots in Stata (Chaimani 2013; Salanti 2012). We anticipate that insufficient reporting of effect modifiers in individual studies and pairwise comparisons containing few studies will make assessment of the distribution of effect modifiers difficult (Cipriani 2013). We will interpret a clearly dissimilar distribution of an effect modifier across trials as a threat to transitivity. We will proceed with NMA in the case of minor dissimilarities; and explore the influence of the effect modifier on inconsistency/heterogeneity using network meta-regression or subgroup analyses (or both). We will consider excluding network nodes in the case of considerable dissimilarity and will consider not proceeding with NMA if intransitivity persists.

An equivalent expression of transitivity is that treatments missing from each trial in the network are missing at random (Lu 2006), which implies that there is no preference for a particular (set of) comparison treatment/s (Salanti 2012). We will calculate network geometry metrics for diversity and co-occurrence, using EcoSim Professional v1.2d (Entsminger 2014), to assess the presence of comparator preference bias. We will interpret a probability of interspecific encounter (PIE) index less than 0.75 and a statistically significant C-score (P < 0.10) as indicative of limited diversity and likely co-occurrence, respectively (Hurlbert 1971; Salanti 2008). We will construct network plots and weight the edges by the number of checkerboard units for each comparison if limited diversity or likely co-occurrence are identified. We will manually scrutinise comparisons with a large number of checkerboard units to determine whether this co-occurrence is justifiable (Salanti 2008). We will exclude network nodes (treatments) forming comparisons for which there is non-justifiable co-occurrence.

Assessment of reporting biases

We will assess small study effects in pairwise comparisons, using contour enhanced funnel plots (Peters 2008), when there are at least 10 studies available. Various factors contribute to the association between study effect size and funnel plot asymmetry. Contour-enhanced funnel plots assist interpretation of asymmetry that is due to publication bias and not other factors, such as lesser methodological quality. The contour lines imposed on the plot indicate levels of statistical significance. We will interpret an absence of studies in areas of non-significance as suggestive of publication bias for that pairwise comparison.

Data synthesis

Methods for direct treatment comparisons

We will perform pairwise random-effects meta-analyses in Stata or Review Manager 5 for each direct comparison for which there are at least two studies available (StataCorp 2015; Review Manager 2014).

Methods for indirect and mixed comparisons

We aim to perform network meta-analysis within a frequentist framework in Stata (Higgins 2012; White 2009; White 2011b; White 2012; White 2015). The multivariate meta-analysis models incorporate random effects for heterogeneity and for inconsistency. We will not stratify study inclusion to the analysis by risk of bias.

Assessment of statistical heterogeneity

Assumptions when estimating the heterogeneity

We will assume that the heterogeneity variance is different for each direct comparison in standard pairwise meta-analyses. We will assume that the heterogeneity variance is the same across the different comparisons in network meta-analysis.

Measures and tests for heterogeneity

We will test for the presence of statistical heterogeneity (variance in true effects) within each pairwise comparison using the Q statistic, with alpha less than 0.10 as we anticipate a small number of trials per comparison. We will calculate 95% prediction intervals for the pooled effects and interpret prediction intervals spanning greater than 15 points (on a 0 to 100 scale) on either side of the pooled effect as indicative of important heterogeneity. We will visually inspect the distribution of effect sizes in the forest plots and calculate the I² value to indicate the proportion of observed variance that is due to heterogeneity (Borenstein 2009). We will interpret I² greater than 50% as indicative of important heterogeneity. We will use the heterogeneity variances from the NMA models as measures of total network heterogeneity.

Assessment of statistical inconsistency

Assessment of inconsistency is only possible if there are closed loops of evidence in the network (all comparisons have direct evidence). We will rely on the assessment of transitivity to infer the presence of inconsistency if there are no closed loops. Inconsistency may not be detected for two reasons. Firstly, tests to detect inconsistency have low power (Higgins 2012; Krahn 2013; Song 2012; Veroniki 2014). Secondly, heterogeneity and inconsistency are intertwined. In situations where there is large heterogeneity, this may mask the presence of inconsistency (Song 2012; Veroniki 2013; Veroniki 2014).
Global approaches for evaluating consistency

We will evaluate consistency across the entire network using the ‘design-by-treatment’ interaction model (Higgins 2012; White 2012); and infer the presence of inconsistency based on P < 0.10.

Local approaches for evaluating consistency

We will evaluate consistency in closed loops with the loop-specific approach (Bucher 1997) and node-splitting (Dias 2010), using a threshold of P < 0.10 for either approach.

Strategy for investigating the sources of inconsistency

We will employ a staged approach to investigate any significant inconsistency that we encounter (Cipriani 2013; Salanti 2012). In the first instance, we will check for data extraction errors in the comparisons identified as inconsistent and those with important heterogeneity. Secondly, we will test whether the inconsistency may be explained using pre-specified covariates in network meta-regression and subgroup analyses, provided sufficient studies are available. Lastly, if there remains significant unexplained inconsistency, we will consider not proceeding with NMA. This judgement will involve the clinical and methodological evaluation of transitivity, the approaches to identify inconsistency and the knowledge that small amounts of inconsistency may be due to chance (Veroniki 2013).

Confidence in cumulative evidence

We will use the approach described by Salanti 2014 to construct judgements of confidence in each of the pairwise effects derived from NMA for pain and safety at end of treatment and confidence in the ranking of treatments with respect to their effect on pain and safety at end of treatment. We will likely use the Confidence in Network Meta-Analysis (CINeMA) web application for this process (ISPM 2017). We will consider the five GRADE domains: risk of bias; indirectness; inconsistency; imprecision; and publication bias. Initial judgements of confidence will be ‘high’, because the data come from randomised controlled trials, which are the ideal study design for the research question. The procedures for evaluation and downgrading of judgements differ slightly between those for confidence in the pairwise effects and confidence in the treatment rankings. These are described in Appendix 3.

Summary of findings

We will present ‘Summary of findings’ tables for pain and safety at end of treatment. The tables will be adapted from the current template used in Cochrane Reviews and will contain, at a minimum, the treatment effect, assumed risk (for safety only), quality of evidence (GRADE) and number of participants for each comparison in the network. An example ‘Summary of findings’ table is shown for a single comparison in Table 3. Development of methods for presentation of results of NMA is ongoing (e.g. Tan 2014; Veroniki 2016), and we anticipate the summary of findings in the finished review to incorporate these methodological advances where appropriate.

Subgroup analysis and investigation of heterogeneity

We will perform network meta-regression or subgroup analyses if sufficient studies are available, by investigating baseline pain intensity, sample size and enriched design as possible sources of inconsistency or heterogeneity. We may also investigate clinical or methodological factors identified during the review process that may threaten transitivity (Jansen 2012), as sources of inconsistency or heterogeneity, or both. We will assume for each network meta-regression model that the effects of the covariates are common (the same) across all comparisons in the network (Chaimani 2012; Dias 2013). This strong assumption is likely to make best use of the available data (Dias 2013; Efthimiou 2016). We specify the following assumptions about the direction of effect for each covariate.

- Baseline pain intensity (continuous): increasing magnitudes of the covariate increases the effect size between the intervention and comparator (compared to trials in which the covariate is less).
- Sample size (continuous): increasing magnitudes of the covariate reduces the effect size between the intervention and comparator.
- Enriched design (binary): the presence of the covariate increases the effect size for pain and decreases the effect size for safety, between the intervention and comparator (compared to trials in which the covariate is absent).

We plan to conduct a subgroup analysis if intended co-interventions are identified as threats to transitivity. We will group trials that test drug interventions alone separately from trials that include intended co-interventions, provided that sufficient data are available. We will assess small-study effects in pairwise comparisons using conventional funnel plots and across the entire network using comparison-adjusted funnel plots (Chaimani 2013).

Sensitivity analysis

We will conduct sensitivity analyses of pain and safety in which studies at high risk of bias are excluded, provided that the original network structure remains the same. We plan to investigate baseline pain intensity, sample size and enriched design in sensitivity analyses if sufficient data are not available for network meta-regression or subgroup analysis. We will dichotomise baseline pain intensity and sample size and exclude trials with baseline pain intensity higher than 70/100 (VAS) and sample size less than 50, respectively. We will also conduct sensitivity analyses of the analyses for pain at end of treatment and safety if we impute missing data for either of these outcomes.
ACKNOWLEDGEMENTS

Prof Georgia Salanti and A/Prof Andrea Cipriani for assistance with methodological development.

Dr Anna Chaimani for providing Stata code.

Dr Michel Rossignol, Dr Sarah Nevitt, Prof Roger Chou and Marilyn Walsh for their constructive feedback on the manuscript.

REFERENCES

Additional references

AMH 2017
Australian Medicines Handbook Pty Ltd.
Australian Medicines Handbook 2017 (online).

Amorin-Woods 2014

Baigent 2013

Bhamb 2006

Bishop 2003

Borenstein 2009

Brewer 1999
Brewer T, Colditz GA. Postmarketing surveillance and adverse drug reactions. JAMA 1999;281(9):824.

Brun 2015

Bucher 1997

Cabana 1999

Caldwell 2005

Chaimani 2012

Chaimani 2013

Chaparro 2014

Chou 2007

Chou 2017

Chung 2013
Cipriani 2013

Daly 2008

Darlow 2012

Davies 2000

Dechartres 2016

Deyo 2015

Dias 2010

Dias 2013

Dominici 1999

Downie 2013

Efthimiou 2016

Enthoven 2016a

Enthoven 2016b

Entsminger 2014 [Computer program]

Furlan 2006

Furlan 2011

Furlan 2014

Furlan 2015

Furukawa 2016

GBD Collaborators 2017a
GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases.

**GBD Collaborators 2017b**


**Gore 2012**


**Gouveia 2017**


**Graham 2013**


**Hart 2015**


**Hinz 2008**


**Hinz 2012**


**Hoaglin 2011**


**Hoffmann 2013**


**Humphreys 2017**


**Hurlbert 1971**


**ISPM 2017**


**Itz 2013**


**Ivanova 2011**


**Jansen 2011**


**Jansen 2012**


**Jansen 2013**


**Jozwiak-Bebenista 2014**

inconsistency in network meta-analyses. 


Liberati 2009

Lu 2004

Lu 2006

Machado 2017

Mafi 2013

Maher 2017

Martell 2007

McGettigan 2011

Menezes Costa 2012

Mills 2013

MIMS 2017
Haymarket Media Group Ltd. MIMS Online. www.mims.co.uk/ accessed prior to 11 December 2017.

Ndlovu 2014

PDR 2017

Perrot 2008

Peters 2008

Rachinger-Adam 2011

Review Manager 2014 [Computer program]

Salanti 2008

Salanti 2011
Salanti 2012

Salanti 2014

Saragiotto 2016a

Saragiotto 2016b

Schers 2001

Shaheed 2016

Sheen 2002

Sills 1986

Song 2003

Song 2012

Sostres 2013

StataCorp 2015 [Computer program]

Sutton 2002

Sutton 2008

Tan 2014

Treede 2015

van der Valk 2009

Veroniki 2013

Veroniki 2014

Veroniki 2016

Wan 2014

Webster 2005
**Webster 2007**

**White 2009**

**White 2011a**

**White 2011b**

**White 2012**

**White 2015**

**Whittle 2000**

* Indicates the major publication for the study

---

**ADDITIONAL TABLES**

**Table 1. Sources of risk of bias**

<table>
<thead>
<tr>
<th>Bias Domain</th>
<th>Source of Bias</th>
<th>Possible Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection</td>
<td>(1) Was the method of randomization adequate?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Selection</td>
<td>(2) Was the treatment allocation concealed?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Performance</td>
<td>(3) Was the patient blinded to the intervention?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Performance</td>
<td>(4) Was the care provider blinded to the intervention?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Detection</td>
<td>(5) Was the outcome assessor blinded to the intervention?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Attrition</td>
<td>(6) Was the drop-out rate described and acceptable?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Attrition</td>
<td>(7) Were all randomized participants analyzed in the group to which they were allocated?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Reporting</td>
<td>(8) Are reports of the study free of suggestion of selective outcome reporting?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Selection</td>
<td>(9) Were the groups similar at baseline regarding the most important prognostic indicators?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Performance</td>
<td>(10) Were cointerventions avoided or similar?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Performance</td>
<td>(11) Was the compliance acceptable in all groups?</td>
<td>Yes/No/Unsure</td>
</tr>
<tr>
<td>Detection</td>
<td>(12) Was the timing of the outcome assessment similar in all groups?</td>
<td>Yes/No/Unsure</td>
</tr>
</tbody>
</table>
Table 1. Sources of risk of bias (Continued)

<table>
<thead>
<tr>
<th>Other</th>
<th>(13) Are other sources of potential bias unlikely?</th>
<th>Yes/No/Unsure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furlan 2015</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Criteria for a judgment of “yes” for the sources of risk of bias

1. A random (unpredictable) assignment sequence. Examples of adequate methods are coin toss (for studies with 2 groups), rolling a dice (for studies with 2 or more groups), drawing of balls of different colours, drawing of ballots with the study group labels from a dark bag, computer-generated random sequence, preordered sealed envelopes, sequentially-ordered vials, telephone call to a central office, and preordered list of treatment assignments. Examples of inadequate methods are: alternation, birth date, social insurance/security number, date in which they are invited to participate in the study, and hospital registration number.

2. Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.

3. Index and control groups are indistinguishable for the patients or if the success of blinding was tested among the patients and it was successful.

4. Index and control groups are indistinguishable for the care providers or if the success of blinding was tested among the care providers and it was successful.

5. Adequacy of blinding should be assessed for each primary outcome separately. This item should be scored “yes” if the success of blinding was tested among the outcome assessors and it was successful or:
   - for patient-reported outcomes in which the patient is the outcome assessor (e.g., pain, disability): the blinding procedure is adequate for outcome assessors if participant blinding is scored “yes”.
   - for outcome criteria assessed during scheduled visit and that supposes a contact between participants and outcome assessors (e.g., clinical examination): the blinding procedure is adequate if patients are blinded, and the treatment or adverse effects of the treatment cannot be noticed during clinical examination.
   - for outcome criteria that do not suppose a contact with participants (e.g., radiography, magnetic resonance imaging): the blinding procedure is adequate if the treatment or adverse effects of the treatment cannot be noticed when assessing the main outcome.
   - for outcome criteria that are clinical or therapeutic events that will be determined by the interaction between patients and care providers (e.g., cointerventions, hospitalization length, treatment failure), in which the care provider is the outcome assessor: the blinding procedure is adequate for outcome assessors if item “4” (caregivers) is scored “yes”.
   - for outcome criteria that are assessed from data of the medical forms: the blinding procedure is adequate if the treatment or adverse effects of the treatment cannot be noticed on the extracted data.

6. The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of withdrawals and drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up and does not lead to substantial bias a “yes” is scored. (N.B. these percentages are arbitrary, not supported by literature).

7. All randomized patients are reported/analyzed in the group they were allocated to by randomization for the most important moments of effect measurement (minus missing values) irrespective of noncompliance and cointerventions.
Table 2. Criteria for a judgment of “yes” for the sources of risk of bias (Continued)

8 All the results from all prespecified outcomes have been adequately reported in the published report of the trial. This information is either obtained by comparing the protocol and the report, or in the absence of the protocol, assessing that the published report includes enough information to make this judgment

9 Groups have to be similar at baseline regarding demographic factors, duration and severity of complaints, percentage of patients with neurological symptoms, and value of main outcome measure(s)

10 If there were no cointerventions or they were similar between the index and control groups

11 The reviewer determines if the compliance with the interventions is acceptable, based on the reported intensity, duration, number and frequency of sessions for both the index intervention and control intervention(s). For example, physiotherapy treatment is usually administered for several sessions; therefore it is necessary to assess how many sessions each patient attended. For single-session interventions (e.g. surgery), this item is irrelevant

12 Timing of outcome assessment should be identical for all intervention groups and for all primary outcome measures

13 Other types of biases.
- Industry-sponsored trials. The conflict of interest (COI) statement should explicitly state that the researchers have had full possession of the trial process from planning to reporting without funders with potential COI having any possibility to interfere in the process. If, for example, the statistical analyses have been done by a funder with a potential COI, usually “unsure” is scored

Furlan 2015

Table 3. Summary of findings

Paracetamol, NSAIDs and opioid analgesics for chronic low back pain: network meta-analysis

**Patient or population:** non-specific chronic low back pain  
**Settings:** primary and secondary care, (additional detail unknown at protocol stage)  
**Intervention:** paracetamol, NSAIDs or opioid analgesics  
**Comparison:** intervention A vs intervention B, of X possible comparisons (unknown at protocol stage)

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Outcome type</th>
<th>Outcome measure</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>Continuous</td>
<td>Pain intensity on 0 to 100 mm VAS</td>
<td>Measured at end of treatment</td>
</tr>
<tr>
<td>Safety</td>
<td>Dichotomous</td>
<td>Proportion of participants that experience a serious adverse event during treatment</td>
<td></td>
</tr>
</tbody>
</table>

GRADE Working Group grades of evidence  
**High quality:** further research is very unlikely to change our confidence in the estimate of effect.  
**Moderate quality:** further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.  
**Low quality:** 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.
Table 3. Summary of findings  (Continued)

Very low quality: we are very uncertain about the estimate.

### APPENDICES

**Appendix 1. Interventions of interest**

<table>
<thead>
<tr>
<th>Drug name</th>
<th>ATC code</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>morphine</td>
<td>N02AA01</td>
<td></td>
</tr>
<tr>
<td>morphine combinations</td>
<td>N02AA51</td>
<td></td>
</tr>
<tr>
<td>hydromorphone</td>
<td>N02AA03</td>
<td></td>
</tr>
<tr>
<td>oxymorphone</td>
<td>N02AA</td>
<td></td>
</tr>
<tr>
<td>nucromorphine</td>
<td>N02AA04</td>
<td></td>
</tr>
<tr>
<td>oxycodone</td>
<td>N02AA05</td>
<td></td>
</tr>
<tr>
<td>oxycodone and naloxone</td>
<td>N02AA55</td>
<td></td>
</tr>
<tr>
<td>oxycodone and paracetamol</td>
<td>N02AJ17</td>
<td></td>
</tr>
<tr>
<td>oxycodone and acetylsalicylic acid</td>
<td>N02AJ18</td>
<td></td>
</tr>
<tr>
<td>oxycodone and ibuprofen</td>
<td>N02AJ19</td>
<td></td>
</tr>
<tr>
<td>codeine</td>
<td>R05DA04</td>
<td></td>
</tr>
<tr>
<td>codeine, combinations excl. psycholeptics</td>
<td>N02AA59</td>
<td>where these are NSAIDs, as defined herein, or Paracetamol</td>
</tr>
<tr>
<td>codeine and paracetamol</td>
<td>N02AJ06</td>
<td></td>
</tr>
<tr>
<td>codeine and acetylsalicylic acid</td>
<td>N02AJ07</td>
<td></td>
</tr>
<tr>
<td>codeine and ibuprofen</td>
<td>N02AJ08</td>
<td></td>
</tr>
<tr>
<td>codeine and other non-opioid analgesics</td>
<td>N02AJ09</td>
<td>where these are NSAIDs, as defined herein, or Paracetamol</td>
</tr>
<tr>
<td>Drug</td>
<td>Code</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>dihydrocodeine</td>
<td>N02AA08</td>
<td></td>
</tr>
<tr>
<td>dihydrocodeine combinations</td>
<td>N02AA58</td>
<td></td>
</tr>
<tr>
<td>dihydrocodeine and paracetamol</td>
<td>N02AJ01</td>
<td></td>
</tr>
<tr>
<td>dihydrocodeine and acetylsalicylic acid</td>
<td>N02AJ02</td>
<td></td>
</tr>
<tr>
<td>dihydrocodeine and other non-opioid analgesics</td>
<td>N02AJ03</td>
<td></td>
</tr>
<tr>
<td>papaveretum</td>
<td>N02AA10</td>
<td></td>
</tr>
<tr>
<td>buprenorphine</td>
<td>N02AE01</td>
<td></td>
</tr>
<tr>
<td>tilidine</td>
<td>N02AX01</td>
<td></td>
</tr>
<tr>
<td>dezocine</td>
<td>N02AX03</td>
<td></td>
</tr>
<tr>
<td>meptazinol</td>
<td>N02AX05</td>
<td></td>
</tr>
<tr>
<td>tapentadol</td>
<td>N02AX06</td>
<td></td>
</tr>
<tr>
<td>tramadol</td>
<td>N02AX02</td>
<td></td>
</tr>
<tr>
<td>tramadol and paracetamol</td>
<td>N02AJ13</td>
<td></td>
</tr>
<tr>
<td>tramadol and dexamfetadiprofen</td>
<td>N02AJ14</td>
<td></td>
</tr>
<tr>
<td>tramadol and other non-opioid analgesics</td>
<td>N02AJ15</td>
<td></td>
</tr>
<tr>
<td>butaphenyl</td>
<td>N02AF01</td>
<td></td>
</tr>
<tr>
<td>nalbuphine</td>
<td>N02AF02</td>
<td></td>
</tr>
<tr>
<td>ketobemidone</td>
<td>N02AB01</td>
<td></td>
</tr>
<tr>
<td>pethidine</td>
<td>N02AB02</td>
<td></td>
</tr>
<tr>
<td>pethidine, combinations excl. psycholeptics</td>
<td>N02AB52</td>
<td></td>
</tr>
<tr>
<td>fentanyl</td>
<td>N02AB03</td>
<td></td>
</tr>
<tr>
<td>dextromoramide</td>
<td>N02AC01</td>
<td></td>
</tr>
<tr>
<td>piritramide</td>
<td>N02AC03</td>
<td></td>
</tr>
<tr>
<td>Drug Name</td>
<td>ATC Code</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>dextropropoxyphene</td>
<td>N02AC04</td>
<td></td>
</tr>
<tr>
<td>bezitramide</td>
<td>N02AC05</td>
<td></td>
</tr>
<tr>
<td>methadone</td>
<td>N07BC02</td>
<td></td>
</tr>
<tr>
<td>methadone, combinations excl. psycholeptics</td>
<td>N02AC52</td>
<td>where these are NSAIDs, as defined herein, or Paracetamol</td>
</tr>
<tr>
<td>dextropropoxyphene, combinations excl. psycholeptics</td>
<td>N02AC54</td>
<td>where these are NSAIDs, as defined herein, or Paracetamol</td>
</tr>
<tr>
<td>pentazocine</td>
<td>N02AD01</td>
<td></td>
</tr>
<tr>
<td>phenazocine</td>
<td>N02AD02</td>
<td></td>
</tr>
<tr>
<td>phenylbutazone</td>
<td>M01AA01</td>
<td></td>
</tr>
<tr>
<td>mofebutazone</td>
<td>M01AA02</td>
<td></td>
</tr>
<tr>
<td>oxyphenbutazone</td>
<td>M01AA03</td>
<td></td>
</tr>
<tr>
<td>piroxicam</td>
<td>M01AC01</td>
<td></td>
</tr>
<tr>
<td>lornoxicam</td>
<td>M01AC02</td>
<td></td>
</tr>
<tr>
<td>meloxicam</td>
<td>M01AC06</td>
<td></td>
</tr>
<tr>
<td>ibuprofen</td>
<td>M01AE01</td>
<td></td>
</tr>
<tr>
<td>naproxen</td>
<td>M01AE02</td>
<td></td>
</tr>
<tr>
<td>ketoprofen</td>
<td>M01AE03</td>
<td></td>
</tr>
<tr>
<td>fenoprofen</td>
<td>M01AE04</td>
<td></td>
</tr>
<tr>
<td>flurbiprofen</td>
<td>M01AE09</td>
<td></td>
</tr>
<tr>
<td>tiaprofenic acid</td>
<td>M01AE11</td>
<td></td>
</tr>
<tr>
<td>oxaprozin</td>
<td>M01AE12</td>
<td></td>
</tr>
<tr>
<td>dexibuprofen</td>
<td>M01AE14</td>
<td></td>
</tr>
<tr>
<td>dexketoprofen</td>
<td>M01AE17</td>
<td></td>
</tr>
<tr>
<td>mefenamic acid</td>
<td>M01AG01</td>
<td></td>
</tr>
<tr>
<td>Drug Name</td>
<td>ATC Code</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>tolfenamic acid</td>
<td>M01AG02</td>
<td></td>
</tr>
<tr>
<td>meclofenamic acid</td>
<td>M01AG04</td>
<td></td>
</tr>
<tr>
<td>indomethacin</td>
<td>M01AB01</td>
<td></td>
</tr>
<tr>
<td>sulindac</td>
<td>M01AB02</td>
<td></td>
</tr>
<tr>
<td>tolmexitin</td>
<td>M01AB03</td>
<td></td>
</tr>
<tr>
<td>zomepirac</td>
<td>M01AB04</td>
<td></td>
</tr>
<tr>
<td>diclofenac</td>
<td>M01AB05</td>
<td></td>
</tr>
<tr>
<td>alclofenac</td>
<td>M01AB06</td>
<td></td>
</tr>
<tr>
<td>etodolac</td>
<td>M01AB08</td>
<td></td>
</tr>
<tr>
<td>ketorolac</td>
<td>M01AB15</td>
<td></td>
</tr>
<tr>
<td>aceclofenac</td>
<td>M01AB16</td>
<td></td>
</tr>
<tr>
<td>bufexamac</td>
<td>M01AB17</td>
<td></td>
</tr>
<tr>
<td>celecoxib</td>
<td>M01AH01</td>
<td></td>
</tr>
<tr>
<td>valdecoxib</td>
<td>M01AH03</td>
<td></td>
</tr>
<tr>
<td>parecoxib</td>
<td>M01AH04</td>
<td></td>
</tr>
<tr>
<td>etoricoxib</td>
<td>M01AH05</td>
<td></td>
</tr>
<tr>
<td>nabumetone</td>
<td>M01AX01</td>
<td></td>
</tr>
<tr>
<td>glucosamine</td>
<td>M01AX05</td>
<td></td>
</tr>
<tr>
<td>GAG polysulfate</td>
<td>M01AX12</td>
<td></td>
</tr>
<tr>
<td>nimesulide</td>
<td>M01AX17</td>
<td></td>
</tr>
<tr>
<td>chondroitin sulphate</td>
<td>M01AX25</td>
<td></td>
</tr>
<tr>
<td>paracetamol (acetaminophen)</td>
<td>N02BE01</td>
<td></td>
</tr>
<tr>
<td>acetylsalicylic acid (aspirin)</td>
<td>N02BA01</td>
<td></td>
</tr>
<tr>
<td>acetylsalicylic acid, combinations excl. psycholeptics</td>
<td>M01BA03</td>
<td>where these are NSAIDs, as defined herein, or Paracetamol</td>
</tr>
</tbody>
</table>
Appendix 2. Search strategies

Search Strategy for MEDLINE (Ovid):

**Part A: Generic search for randomized controlled trials**
1. randomized controlled trial.pt.
2. controlled clinical trial.pt.
4. clinical trial.pt.
5. random*.ab,ti.
6. placebo.ab,ti.
7. drug therapy.fs.
8. trial.ab,ti.
9. groups.ab,ti.
10. or/1-9
11. (animals not (humans and animals)).sh.
12. 10 not 11

**Part B: Specific search for low back, sacrum and coccyx problems**
13. dorsalgia.ti,ab.
14. exp Back Pain/ 
15. backache.ti,ab.
16. ((lumb* or back) adj pain).ti,ab.
17. coccydynia.ti,ab.
18. sciatica.ti,ab.
19. spondylitis.ti,ab.
20. lumbago.ti,ab.
21. or/13-20

**Part C: Specific search for other spinal disorders**
22. Coccyx.sh
23. Lumbar Vertebrae.sh
24. Intervertebral disc.sh
25. Sacrum.sh
26. Intervertebral disc degeneration.sh
27. (disc adj degenerat*).ti,ab.
28. (disc adj prolapse*).ti,ab.
29. (disc adj herniat*).ti,ab.
30. spinal fusion.sh.
31. (facet adj joint*).ti,ab.
32. Intervertebral Disc Displacement.sh.
33. postlaminectomy.ti,ab.
34. or/22-33

**Part D: Specific search for interventions of interest**
35. morphine.sh or morphine.tw,kf.
36. hydromorphone.sh or hydromorphone.tw,kf.
37. oxycodone.sh or oxycodone.tw,kf.
38. oxymorphone.sh or oxymorphone.tw,kf.
39. nicomorphine.tw,kf.
40. codeine.sh or codeine.tw,kf.
41. exp naloxone/ or naloxone.tw,kf.
42. dihydrocodeine.tw,kf.
43. papaveratum.tw,kf.
44. butorphanol.tw,kf.
45. tilidine.sh or tilidine.tw,kf.
46. dezocine.tw,kf.
47. meptazinol.sh or meptazinol.tw,kf.
48. tapentadol.tw,kf.
49. tramadol.sh or tramadol.tw,kf.
50. butaphenial.tw,kf.
51. nalbuphine.sh or nalbuphine.tw,kf.
52. ketobemidone.tw,kf.
53. meperidine.sh or meperidine.tw,kf. or pethidine.tw,kf.
54. fentanyl.sh or fentanyl.tw,kf.
55. dextromoramide.sh or dextromoramide.tw,kf.
56. piritramide.sh or piritramide.tw,kf.
57. dextropropoxyphene.sh or dextropropoxyphene.tw,kf.
58. bezitramide.tw,kf.
59. methadone.sh or methadone.tw,kf.
60. pentazocine.sh or pentazocine.tw,kf.
61. phenazocine.sh or phenazocine.tw,kf.
62. phenylbutazone.sh or phenylbutazone.tw,kf.
63. mofebutazone.tw,kf.
64. oxyphenbutazone.sh or oxyphenbutazone.tw,kf.
65. meloxicam.tw,kf.
66. piroxicam.sh or piroxicam.tw,kf.
67. lornoxicam.tw,kf.
68. ibuprofen.sh or ibuprofen.tw,kf.
69. naproxen.sh or naproxen.tw,kf.
70. ketoprofen.sh or ketoprofen.tw,kf.
71. fenoprofen.sh or fenoprofen.tw,kf.
72. flurbiprofen.sh or flurbiprofen.tw,kf.
73. tiaprofenic acid.tw,kf.
74. oxaprozin.tw,kf.
75. dexibuprofen.tw,kf.
76. desketoprofen.tw,kf.
77. mefenamic acid.sh or mefenamic acid.tw,kf.
78. tolfenamic acid.tw,kf.
79. meclofenamic acid.sh or meclofenamic acid.tw,kf.
80. exp indomethacin/ or indometacin.tw,kf.
81. sulindac.sh or sulindac.tw,kf.
82. tolmetin.sh or tolmetin.tw,kf.
83. zomepirac.tw,kf.
84. diclofenac.sh or diclofenac.tw,kf.
85. alclofenac.tw,kf.
86. etodolac.sh or etodolac.tw,kf.
87. aceclofenac.tw,kf.
88. bufexamac.sh or bufexamac.tw,kf.
89. celecoxib.sh or celecoxib.tw,kf.
90. valdecoxib.tw,kf.
91. etoricoxib.tw,kf.
92. nabumetone.tw,kf.
93. exp glucosamine/ or glucosamine.tw,kf.
94. glucosaminoglycan polysulfate.tw,kf.
95. nimesulide.tw,kf.
96. chondroitin sulfate.tw,kf.
97. acetaminophen.sh or acetaminophen.tw,kf. or paracetamol.tw,kf.
98. aspirin.sh or aspirin.tw,kf. or acetylsalicylic acid.tw,kf.
99. or/35-98 (all interventions of interest)
Results
100. 21 or 34 (all back pain)
101. 99 and 100 (all back pain and all interventions of interest)
102. 12 and 101 (all RCTs of interventions of interest in back pain)

Appendix 3. The GRADE approach to evidence synthesis

The quality of evidence will be categorized as follows:

- **High (◆◆◆):** further research is very unlikely to change the confidence in the estimate of effect.
- **Moderate (◆◆):** further research is likely to have an important impact in the confidence in the estimate of effect.
- **Low (◆):** 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.
- **Very Low (◆):** any estimate of effect is very uncertain.

Initial judgements of confidence will be ‘high’, because the data come from randomised controlled trials. We will consider the five GRADE domains: risk of bias, indirectness, inconsistency, imprecision and publication bias. The evidence available under each of these domains will be graded according to slightly different procedures for confidence in the pairwise effects and confidence in the treatment rankings.

Judgement of confidence in the pairwise effects

1) **Risk of bias.** We will rate each direct comparison as at low, moderate or high risk of bias, assigning scores of 0, −1 and −2, respectively. We will consider direct comparisons at high risk of bias when > 25% of participants in the comparison are from studies at high overall risk of bias, low risk when > 50% of participants are from studies at low overall risk and moderate risk in all other instances. We will use the percentage contributions of all direct comparisons to each pairwise effect (from the contribution matrix) to construct weighted averages of these scores. Then we will base judgements of the confidence to be placed in each pairwise effect on a weighted average of the risk of bias of direct comparisons feeding into it. We will downgrade one level for −1.5 < score < −0.5 and two levels for scores < −1.51, although this may change in practice if there is clear imbalance in the contribution of evidence that would render such action inappropriate.

2) **Indirectness.** We will evaluate the populations, interventions and outcomes in the studies contributing to each direct comparison for congruence with those specified for this review. We will consider downgrading a pairwise effect one level for indirectness if direct comparisons with important imbalances in populations, treatments or outcomes contribute the majority of information to it. In addition, we will downgrade all comparisons one additional level if there were concerns over the transitivity assumption identified prior to the analysis.

3) **Inconsistency.** We will evaluate each direct comparison for consistency in the direction and magnitude of the effect sizes from individual trials, considering the width of the prediction interval and magnitude of the heterogeneity parameter. We will downgrade direct comparisons one level if important heterogeneity is identified. Additionally, we will downgrade one level any direct comparisons that are implicated in loops with important incoherence or where there is a discrepancy between direct and indirect evidence.

4) **Imprecision.** We will downgrade due to imprecision based on consideration of the width of the confidence intervals. We will downgrade one level when these span either the null or the threshold for a clinically meaningful effect on pain intensity (10 points on a 0 to 100 scale (Chou 2017)) and two levels when the interval spans both. We will not consider sample sizes as there are no established criteria for this evaluation in the NMA context (Salanti 2014).

5) **Publication bias.** We will consider the likelihood of publication bias for each direct comparison, using both epidemiological judgement and the contour-enhanced funnel plots. We will downgrade one level if we consider it likely that studies have been conducted and not published.

Judgement of confidence in the treatment rankings

The difference between the judgement of confidence in the treatment rankings and that for the pairwise effect sizes is that this is a single judgement for the ranking evidence as a whole.
1) Risk of bias. We will use the scores of 0, −1, or −2 as previously for the judgements of overall risk of bias for each direct comparison and construct a weighted average for the entire network using the percentage contributions of each direct comparison to the entire network. We will downgrade one level for −1.5 < score < −0.5 and two levels for scores < −1.51, although this may change in practice if there is clear imbalance in the contribution of evidence that would render such action inappropriate.

2) Indirectness. We will use the above evaluation of congruence between study populations, interventions and outcomes. We will evaluate the contribution to the entire network by comparisons that are judged to exhibit indirectness. We will consider downgrading one level if there is an important contribution from one/more of these comparisons. In addition, we will downgrade all comparisons one additional level if there were concerns over the transitivity assumption identified prior to the analysis.

3) Inconsistency. We will consider the magnitude of the heterogeneity parameter from the NMA model and the result of the Chi² test for global inconsistency, acknowledging that we may fail to detect important global inconsistency due to the low power of such a test and the presence of large heterogeneity (Higgins 2012; Krahn 2013; Song 2012; Veroniki 2013; Veroniki 2014). We will downgrade one level if either heterogeneity or inconsistency are present and two levels if both are present.

4) Imprecision. Imprecision in ranking may be conceptualised as uncertainty in the probability that a treatment achieves a certain rank (Salanti 2014). We will evaluate the precision of the treatment rankings by examining the SUCRA values that are used to calculate these probabilities. We will deem rankings imprecise when there are similar probabilities for two/more treatment to be ranked at the same level. We will downgrade one level in this situation.

5) Publication bias. We will evaluate the likelihood of publication bias across the network, using epidemiological judgement, evidence of publication bias in direct comparisons and the comparison-adjusted funnel plot, acknowledging that plot asymmetry may due to real factors other than publication bias (Salanti 2014). We will downgrade one level if there is evidence of publication bias or small study effects.

**Contributions of Authors**

MKB provided NMA methodology expertise and wrote the protocol.

AJM provided pharmacological and systematic review methodology expertise.

AN provided NMA methodology and statistical analysis expertise.

CGM, SK, CMW, NH, GLM, BMW, MH, NOC and MvT provided systematic review methodology and clinical area expertise.

JHM is the guarantor and provided clinical area, systematic review and NMA methodology expertise.

All authors read, contributed to and approved the final version of the manuscript.

**Declarations of Interest**

CGM and AJM are conducting an investigator-initiated, NHMRC-funded trial of opioids for acute LBP (OPAL). The OPAL trial has no funding from pharmaceutical companies. They have both received supplementary research funding from Pfizer and GlaxoSmithKline for two investigator-initiated, NHMRC-funded trials of paracetamol and pregabalin for low back pain (PACE & PRECISE).

GLM has received lecturing fees from Kaiser Permanente for their Physical Therapy Residency programme, and consultancy fees from workers’ compensation boards in Australia, North America and Europe for advice on policy matters concerning acute back pain and persistent musculoskeletal pain states. GLM has received consultancy fees from Port Adelaide Football Club for professional services regarding pain and performance. GLM has received travel and consultancy fees from Arsenal Football Club for professional services regarding pain and performance. GLM has received travel support and expenses from the International Olympic Committee for committee roles focusing on pain management in elite sport. GLM receives guest lecturing fees from Noigroup Australasia and has received guest lecturing fees from Agile Physiotherapy (USA) and Results Physiotherapy (USA), for professional development courses. GLM receives speaker’s fees for lectures on pain and rehabilitation and royalties for books on pain and rehabilitation. GLM has received honoraria and travel support from Pfizer for two lectures and for participation in the ASIA Masterclass on Neuropathic Pain. These contributions did not mention drugs. GLM has also received travel support from Grünenthal to attend a European CRPS minimum standards of care summit. This contribution did not involve any drugs.

MKB has received travel support from the Chiropractor’s Association of Australia for speaking engagements unrelated to this topic.
MvT and CGM are on the Editorial Board of Cochrane Back and Neck. MvT was Co-ordinating Editor until September 2017. Editors are required to conduct at least one Cochrane Review. This requirement ensures that they are aware of the processes and commitment needed to conduct reviews. None of the Editors are first authors. Any Editor who is a review author is excluded from editorial decisions on the review in which they are contributors.

BMW, MH, SK, NH, NOC, CMW, AN and JHM have no declarations.

This research project did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

SOURCES OF SUPPORT

Internal sources

- No sources of support supplied

External sources

- National Health and Medical Research Council, Australia.
  Research Fellowship (CMW, SK), Principal Research Fellowship (RH, CGM, GLM)
  Australian Government Research Training Program Scholarship, Australia.
  (MKB)
  UNSW Research Excellence Award, Australia.
  (MKB)
  Neuroscience Research Australia PhD Candidature Supplementary Scholarship, Australia.
  (MKB)