Efficacy and acceptability of pharmacological and non-pharmacological interventions for non-specific chronic low back pain: a protocol for a systematic review and network meta-analysis

Trevor Thompson[1], Sofia Dias[2], Damian Poulter[1], Sharon Weldon[3,4], Lucy Marsh[1], Claire Rossato[1], Jae Il Shin[5], Joseph Firth[6,7], Nicola Veronese[8], Elena Dragioti[9], Brendon Stubbs[10], Marco Solmi[11], Christopher G Maher[12], Andrea Cipriani[13,14], John P.A. Ioannidis[15]

[1]School of Human Sciences, University of Greenwich, Park Row, London SE10 9LS, UK

[2]Centre for Reviews and Dissemination, University of York, York, UK

[3]School of Health Sciences, University of Greenwich, London SE9 2UG, UK;

[4]Barts Health NHS Trust, The Royal London Hospital, Whitechapel Rd, Whitechapel E1 1BB, UK

[5]Department of Pediatrics, Yonsei University College of Medicine, Seoul, Republic of Korea

[6]NICM Health Research Institute, Western Sydney University, Sydney, Australia

[7]Division of Psychology and Mental Health, University of Manchester, UK

[8]National Research Council, Neuroscience Institute, Aging Branch, Padova, Italy

[9]Pain and Rehabilitation Centre and Dept of Health, Medicine and Caring Sciences, Linköping University, Sweden

[10]King's College London and South London and Maudsley NHS Foundation Trust, UK

[11]Neurosciences Department, University of Padua, Padua, Italy

[12]Sydney School of Public Health, The University of Sydney, Australia

[13]Department of Psychiatry, University of Oxford, UK

[14]Oxford Health NHS Foundation Trust, Warneford Hospital, OX3 8AX, Oxford, UK

[15]Meta-Research Innovation at Stanford (METRICS) and Departments of Medicine, Health Research and Policy, Biomedical Science and Statistics, Stanford University, CA, USA

Email details: Trevor Thompson, t.thompson@gre.ac.uk, Sofia Dias, sofia.dias@york.ac.uk, Damian Poulter, d.r.poulter@gre.ac.uk, Sharon Weldon, s.m.weldon@gre.ac.uk, Lucy Marsh, l.b.marsh@gre.ac.uk, Claire Rossato, c.rossato@gre.ac.uk, Jae Il Shin, shinji@yuhs.ac, Joseph Firth, J.Firth@westernsydney.edu.au, Nicola Veronese, ilmannato@gmail.com, Elena Dragioti, elena.dragioti@liu.se, Brendon Stubbs, brendon.stubbs@kcl.ac.uk, Marco Solmi,

marco.solmi83@gmail.com, Christopher Maher, christopher.maher@sydney.edu.au, Andrea Cipriani andrea.cipriani@psych.ox.ac.uk, John Ioannidis, jioannid@stanford.edu

ABSTRACT

Background Despite the enormous financial and humanistic burden of chronic low back pain (CLBP), there is little consensus on what constitutes the best treatment options from a multitude of competing interventions. The objective of this network meta-analysis (NMA) is to determine the relative efficacy and acceptability of primary care treatments for non-specific CLBP, with the overarching aim of providing a comprehensive evidence base for informing treatment decisions. **Methods** We will perform a systematic search to identify randomized controlled trials of interventions endorsed in primary care guidelines for the treatment of non-specific CLBP in adults. Information sources searched will include major bibliographic databases (MEDLINE, Embase, CENTRAL, CINAHL, PsycINFO and LILACS) and clinical trial registries. Our primary outcomes will be patient-reported pain ratings and treatment acceptability (all-cause discontinuation), and secondary outcomes will be functional ability, quality of life and patient/physician ratings of overall improvement. A hierarchical Bayesian class-based NMA will be performed to determine the relative effects of different classes of pharmacological (NSAIDs, opioids, paracetamol, anti-depressants, muscle relaxants) and non-pharmacological (exercise, patient education, manual therapies, psychological therapy, multidisciplinary approaches, massage, acupuncture, mindfulness) interventions and individual treatments within a class (e.g. NSAIDs: diclofenac, ibuprofen, naproxen etc.). We will conduct risk of bias assessments and threshold analysis to assess the robustness of the findings to potential bias. We will compute the effect of different interventions relative to placebo/no treatment for both short and long term efficacy and

acceptability. **Discussion** While many factors are important in selecting an appropriate intervention for an individual patient, evidence for the analgesic effects and acceptability of a treatment are key factors in guiding this selection. Thus, this NMA will provide an important source of evidence to inform treatment decisions and future clinical guidelines.
Keywords: *Low back pain; network meta-analysis; systematic review; protocol; randomized controlled trial*

Systematic review registration

PROSPERO registry number: CRD42019138115,

http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42019138115

1 Background

Low back pain is the leading cause of years lived with disability across the world (GBD, 2017). It is also the second most common reason reported by patients for visiting their family doctor (Finley et al., 2018) and has an estimated lifetime prevalence of 80% (World Health Organization, 2003). The most common type of low back pain by far is the non-specific type (Bardin et al., 2017), indicating the absence of an identifiable cause. While acute episodes of non-specific low back pain can improve markedly in the first 6 weeks, recent esimates suggest that pain can persist for over 12 weeks in 24%-61% of cases (Costa et al., 2012). This type of chronic low back pain (CLBP) carries an enormous economic burden both from direct (e.g. treatment) and indirect (e.g. lost work productivity) costs. In the UK, the cost to the NHS from low back pain exceeds £12 billion a year (NatCen Social Research, 2014), with the chronic form representing the largest proportion of these costs

(Buchbinder and Underwood, 2012). CLBP is also associated with impaired quality of life, mobility and daily function as well as social isolation, disability and depression (National Institute for Health and Care Excellence, 2016).

Because the underlying pathology of non-specific CLBP is by definition unidentified, treatment is largely focused on reducing pain symptoms, and a range of pharmacological and non-pharmacological intervention strategies are used in clinical practice (Maher et al., 2017). A recent review of international practice guidelines (Oliveira et al., 2018) found that while NSAIDs and exercise were commonly recommended, the endorsement of many other treatments including opioids, antidepressants, paracetamol, muscle relaxants, spinal manipulation and acupuncture varied considerably across guidelines. The apparent uncertainty over which pool of interventions constitute the most effective options for treating non-specific CLBP suggests the need for a stronger evidence base.

Network meta-analysis (NMA) provides a powerful means of assessing multiple competing interventions by synthesising data across a network of different treatments (Dias and Caldwell, 2019). By incorporating indirect evidence (where two treatments can be compared by assessing their performance relative to a common comparator such as placebo) the relative effects of two interventions can be evaluated even when no head-to-head trials are available. This cannot be achieved with standard pairwise meta-analysis and helps to establish a hierarchy of the best interventions for a particular condition. In addition, where there is both direct and indirect evidence, these can be combined using all the available evidence to compute the relative treatment effect.

The objective of this NMA is to assess the effectiveness and acceptability of interventions endorsed in primary care practice guidelines for the treatment of non-specific CLBP, with the aim of providing a comprehensive evidence base to inform treatment decisions. The project is called **S**tudy of **P**ain **I**nterventions using **N**etwork meta-**A**nalysis: **L**ow-back pain (SPINAL).

2 Methods/Design

This protocol conforms to PRISMA-P (Moher et al., 2015) recommendations (Additional File 1) and was developed based on guidelines for systematic reviews of back pain interventions from the Cochrane Back and Neck Group (Furlan et al., 2015). Eligibility criteria were developed using the PICOS framework and are reported in detail in the following sections and summarised briefly in Table 1.

Table 1. Summary of PICOS eligibility criteria (Section 2 lists detailed criteria).

	Inclusion criteria	**Exclusion criteria**
Population	Adults (>=18yrs) with non-specific CLBP	Patient baseline pain < 4/10; radicular pain or LBP with a known cause; LBP < 12 weeks
Intervention	Primary care interventions for CLBP	Surgical or invasive interventional procedures
Comparison	A different eligible intervention or a control (placebo/sham or no intervention)	
Outcome	Pain ratings or acceptability (all cause discontinuation)	
Study type	Randomized clinical trials	

2.1 Population

Inclusion criteria. We will include studies of adults (>=18 years) with non-specific CLBP. This is typically defined as pain without a specific known cause or pathology that persists for 12 or more weeks and that occurs below the costal margin and above the inferior gluteal folds.

Studies that simply describe low back pain as non-specific or chronic without providing detail of how this was determined will be included, provided this designation does not conflict with information elsewhere in the text (e.g. where a specific cause of LBP such as infection, cancer or fracture is listed, or where there is an obvious non-chronic symptom duration). Where it cannot be reliably determined whether LBP is specific or non-specific, we will assume non-specific as this represents the vast majority of LBP cases (Oliveira et al., 2018). Where LBP duration cannot be reliably determined, we will assume LBP is acute and exclude the study as it seems likely that any chronicity would have been referred to in the text; but we will document such studies and include them as part of a sensitivity analysis if there are >5 such studies.

Exclusion criteria. We will exclude studies of LBP patients with radicular pain, e.g. sciatica (or where >10% of participants have radicular symptoms in mixed samples of patients with and without radicular pain). Radicular symptoms are typically a result of spinal nerve compromise, and represent a population that may require different treatment options and who are commonly differentiated in treatment guidelines (Oliveira et al., 2018). To help ensure a consistent patient population, we will exclude studies with a minimum baseline threshold for individual patient eligibility that is below 4 on a 0-10 rating, unless separate data are available for participants with baseline pain of 4 or above. We chose a threshold of

4 or above as this represents a common and established individual patient entry criterion and will ensure a homogenous sample of patients with pain of at least a moderate, clinically meaningful level (Boonstra et al., 2016) who are the most likely to seek treatment. If a trial does not specify individual baseline pain as an entry criterion, we will calculate z-scores from the sample mean baseline pain using the formula z = (Mean Baseline Pain– 4.0)/SD and retain only trials where z > -1, indicating approximately 85% of patients reporting a baseline pain of 4 or more.

Whenever we encounter trials that include both eligible and ineligible patients, we will try to determine whether data on the eligible subset can be extracted separately (e.g., in trials including both children and adults, separate the adults; in trials including both patients with and without sciatica, separate those without sciatica; in trials with baseline pain both <4 and >=4, separate those with >=4 pain; and in trials with LBP duration both below and above 12 weeks, separate those with LBP>=12 weeks). If the data for the eligible subset are not available from the published papers and cannot be obtained from the authors, the entire trial will be included, if the percentage of eligible patients is expected to be more than 85% (as exemplified for the baseline pain criterion above).

2.2 Interventions

We will include interventions for the treatment of CLBP in primary care that are endorsed by any of the 15 clinical practice guidelines reviewed by Oliveira et al. (2018), with the exception of herbal medicine as this is endorsed by only one guideline (and recommended against in one other guideline) and is often studied in trials of very low quality (Gagnier et al., 2016). Our rationale for focusing on treatments only included in practice guidelines is

that these represent the pool of intervention strategies more likely to be adopted in clinical practice and because their presence in guidelines usually indicates a higher quality evidence base (Oliveira et al., 2018). Surgical and interventional pain management (e.g. spinal injections, radiofrequency denervation, deep brain and spinal cord stimulation (Morlion, 2013)) will be excluded as these are invasive procedures that are recommended for low back only as next-line treatment in secondary or tertiary care for severe or refractory LBP where conservative primary care treatments have failed, and are not recommended in any guidelines when LBP is chronic and non-specific (Oliveira et al., 2018).

Both single and combined treatments are considered eligible and medications may be fixed or flexibly dosed. For medications approved for pain, we will include only trials that use licenced dosing ranges based on European Medicines Agency guidelines. Where a drug is used off-label and no dosing guidelines exist for pain management, we will include all such trials but perform sensitivity analysis removing studies using dosages outside the approved dosing range for that drug's approved indication.

2.2.1 *Classification of interventions*

Treatments will be grouped into intervention classes to allow us to compare the relative effects of intervention classes as well as individual treatments within a class, using a Bayesian hierarchical class-based NMA model (Dias et al., 2018; Dominici et al., 1999). Grouping individual treatments into meaningful classes maximises statistical power and provides a simpler and more interpretable framework on which to ultimately inform treatment decisions (comparing each individual treatment with every other for 40 treatments, for example, would result in 780 potential comparisons). We will also perform

separate analysis of pharmacological and non- pharmacological networks as described in section 3.2.

Initial classifications were informed by key reviews of treatment guidelines for CLBP interventions (Chou et al., 2017; Oliveira et al., 2018; van Tulder and Koes, 2013; Maher et al., 2017; Foster et al., 2018; National Institute for Health and Care Excellence, 2016) and then circulated to seven members of the Lancet Low Back Pain Series Working Group (not previously known to the lead author) for evaluation and comment. We received responses from five members (see Acknowledgements section) and subsequent refinements were made resulting in a final set of classifications (Table 2). Classifications are differentiated primarily by mechanisms of action, although when putative mechanisms were unclear (e.g. acupuncture) or there was uncertainty over the most appropriate classification, that treatment was listed in its own class.

A non-exhaustive list of examples of the most common interventions that comprise each class are given in Table 2. Pharmacological interventions returned by searches that are not listed in Table 2 will be classified based on MeSH and emtree headings and non-pharmacological interventions will be classified after discussion with the review team prior to analysis with rationale for these classifications documented in the final report.

In the absence of any definitive criteria for differentiating 'weak' vs. 'strong' opioids we followed the classifications used by Whittle et al. (2011) where strong opioids are generally those with higher rates of conversion to morphine. For topical pharmacological agents, while the agents used (e.g. ibuprofen) are also often present in other classes, we

197 nevertheless assessed this as a distinct class given the potential benefits of topical relative to
198 systemic administration. We defined exercise therapy as 'a series of specific movements
199 with the aim of training or developing the body by a routine practice or as physical training
200 to promote good physical health' (Abenhaim et al., 2000). Although there are numerous
201 meaningful ways to categorise exercise types, we decided on two basic classifications of
202 non-specific and mind-body type approaches. However, if excessive heterogeneity is
203 observed within each exercise type relative to other classes, we will explore sources of
204 possible heterogeneity based on pre-defined exercise characteristics identified by Hayden et
205 al. (2005) as potentially important to efficacy (including dose/intensity, supervised vs. non-
206 supervised, delivery type and design), and consider reclassification if necessary. Finally, as no
207 consensus could be reached on the classification of McKenzie therapy, we provisionally
208 classified this as education as the approach invokes components of several treatments, but
209 we will explore the impact of this decision in a sensitivity analysis.

Table 2. Intervention classes and individual treatments (generic drug names given for pharmacological agents)

CLASS	**Examples of Individual treatments**
Pharmacological	
Antidepressants: SNRI	duloxetine, desvenlafaxine, levomilnacipran, venlafaxine, milnacipran
Antidepressants: SSRI	fluoxetine, fluvoxamine, paroxetine, escitalopram, citalopram, sertraline, vilazodone
Antidepressants: tricyclic	amitriptyline, amoxapine, desipramine, imipramine, doxepin, clomipramine, trimipramine, protriptyline, imipramine, nortriptyline, doxepin, nortriptyline

NSAIDs	ibuprofen, naproxen, sulindac, ketoprofen, tolmetin, etodolac, fenoprofen, diclofenac, flurbiprofen, piroxicam, ketorolac, indomethacin, meloxicam, nabumetone, oxaprozin mefenamic acid, diflunisal, fenoprofen
Opioids (strong)	morphine, hydromorphone, oxycodone, fentanyl, methadone, buprenorphine, diamorphine, tapentadol
Opioids (weak)	codeine, hydrocodone, tramadol, pentazocine, tilidine
Muscle relaxants: benzodiazepines	diazepam, estazolam, quazepam, alprazolam, chlordiazepoxide, clorazepate, lorazepam, flurazepam, clonazepam, temazepam, midazolam
Muscle relaxants: skeletal	flupirtin, orphenadrine, dantrolene, carisoprodol, tizanidine, incobotulinumtoxinA, cyclobenzaprine, metaxalone, baclofen, methocarbamol, chlorzoxazone
Paracetamol	
Topical agents (non-opioid)	diclofenac, capsaicin, lidocaine
Non-pharmacological treatments	
Acupuncture	acupuncture, dry needling
Exercise: non-specific	Walking, swimming, running, stretching, aerobics
Exercise: mind-body and bodily awareness	yoga, tai chi, Pilates, motor control exercise, alexander technique
Manual therapy: spinal manipulation	high velocity thrust techniques at or near the end of the passive or physiologic range of motion
Manual therapy: spinal mobilization	low-grade velocity movement techniques within the patient's range of motion and control
Massage	soft tissue massage, acupressure
Mindfulness	mindfulness, mindfulness-based stress reduction
Multidisciplinary approaches	packages that include coordinated delivery of interventions from across different disciplinary practices/clinics (which typically consist of physical and psychological therapy, e.g. education + physiotherapy + exercise + counselling)
Patient education: basic	back school (e.g. instruction on anatomy and function of the back), brief educational intervention, advice on importance of staying active, reassurance, McKenzie therapy

Patient education: pain neuroscience	educational sessions that describe the neurobiology and neurophysiology of pain by the nervous system
Psychological therapy	CBT, operant therapy, behavioural therapy, self-regulatory therapy

2.3 Comparator

A different eligible individual treatment or a control condition (placebo/sham or no-intervention).

2.4 Outcomes

2.4.1 *Primary outcomes*

(1) Pain intensity, assessed with an established rating scale (e.g. 0-10 numerical rating scale or VAS) at specific time periods defined below

(2) Acceptability, defined as (one minus) the proportion of patients who discontinued treatment during the trial for any reason

2.4.1.1 Assessment Timing

The effects of different interventions on pain will be evaluated within the following, distinct assessment windows: immediate (≤2 weeks post-randomisation), short-term (>2 weeks to ≤3 months), medium-term (>3 months to <12 months), long-term (>12 months). These time windows were selected based on a sample of 24 eligible articles from provisional searches. If these divisions fail to sensitively reflect the pattern of assessment timings used across studies, we may reclassify these windows prior to analysis to reflect trial practices.

As many pharmacological interventions may be more likely to be trialled for immediate and short-term outcomes, and certain non-pharmacological treatment (e.g. exercise) trials may be more likely to include long-term outcomes, separate analyses in each time window ensures that the relative efficacies of competing interventions will be evaluated in time windows appropriate for how those interventions are used. When pain ratings have been collected by the study authors at multiple time points *within* a time window, we will use the time point closest to the median for the immediate and short-term windows and the longest follow-up for the long term follow-up window. If data are not reported at these time points (but are reported for other time points), we will make every possible attempt to retrieve these data to reduce the possibility of exaggerated treatment effects from selective reporting of the largest effects (Page et al., 2014). If we are unable to retrieve the preferred data, we will use outcomes at the next closest time point but conduct sensitivity analysis excluding these studies.

2.4.1.2 Effect sizes

Odds ratios will be computed for acceptability.If sufficient data are available, odds ratios for pain will also be computed contrasting the number of treatment responders across two interventions (or an intervention and control). A responder will be defined as a patient who demonstrates >=30% and >=50% reduction from baseline pain rating (we will examine both thresholds separately) reflecting 'moderate' and 'substantial' clinically important improvement according to IMMPACT recommendations (Dworkin et al., 2009). When a study does not report treatment response rate, we will impute these from continuous pain ratings with an established conversion formula (Furukawa et al., 2005; Samara et al., 2013),

unless an excessive number of imputations are required given that this imputation assumes a normal distribution which is usually untestable.

As odds ratios can be difficult to interpret for many people, we will also present additional statistics generally perceived as more intuitive. Specifically, we will calculate risk ratios, absolute risk differences and numbers needed to treat for primary outcomes, by back transformation of the odds ratios. The baseline risk value needed for this transformation will be estimated from random-effects meta-analysis of risk from the placebo arm of placebo-controlled trials. For this purpose, we will use a subset of trials (Dias et al., 2018) judged to be representative of the overall population of chronic low back pain patients based on expert clinical input of the review team.

For pain, we will also calculate effect size as the mean difference in pain ratings across treatments, as these are expected to be reported in nearly all studies. If pain ratings are not reported on the usual 0-10 scale, they will be normalised to this scale. We will use post-treatment scores to compute effect size, unless only change from baseline scores are reported in which case we will use these. Effect sizes using either method can be legitimately pooled (da Costa et al., 2013), and both produce the same effect size when study pre-treatment scores are equal across groups (as would be expected here given only randomised designs are eligible). Where we do use change from baseline scores and standard deviation(s) needed for effect size computations are not reported, they will be computed in the following priority order. First, using standard formula (Borenstein et al., 2009) based on the change score variance and the study pre-post correlation (or if

unavailable, the average pre-post correlation across studies that report it). Second, using the average standard deviation based on studies that report it.

2.4.2 *Secondary outcomes*

Based on recommendations for a core outcome set (COS) in non-specific low back pain (Chiarotto et al., 2018) we also included the following outcomes and associated recommended assessment measures:

(1) Physical functioning (PF), assessed with the Oswestry Disability Index 2.1a or Roland-Morris Disability Questionnaire (the two recommended COS measures and the most commonly used in trials). If a study does not employ either scale, we will include any of the following: Quebec Back Pain Disability Scale, BPI-PI, MPI-PI, SF-36-PF, PROMIS-PF, CLBPDQ, LBPRS-DI, ODI 1.0 as there is evidence of their validity as assessments of PF (Chiarotto et al., 2018)

(2) Health-related quality of life, assessed with the Short-Form Health Survey (SF-12/ SF-36) or PROMIS-GH-10.

(3) Patient or physician ratings of overall improvement.

As all secondary outcomes are assessed on a continuous measure, we will use the mean difference as the effect size. If an outcome is assessed by multiple different scales we will use the most common scale and convert scores from any other scales to the same metric if an established mapping algorithm exists. If this results in a low number of available studies for (e.g. <60% of the total studies reporting that outcome), to maximise data inclusion we will standardize all scales for that outcome and use the *standardized* mean difference,

297 provided that an inspection of the domain of the scales suggests the scales can be
298 meaningfully combined. We will conduct sensitivity analysis In all instances where scales
299 have been combined.
300

301 2.4.3 *Outcomes with missing data*

302 Where missing participant data is present, studies may report analysis on only the subset of
303 patients who adhered to the intervention (per-protocol) or on all participants who were
304 assigned to the intervention at the start of the trial (intention-to-treat) after missing data
305 has been imputed (e.g. using last observation carried forward). If both per-protocol and
306 intention-to-treat analyses are reported, we will prioritise intention-to-treat data (Sterne et
307 al., in press). In all instances, we will report whether analysis was conducted on data that
308 were complete, complete after imputation or incomplete, and we will examine and report
309 any material differences in results across these types. When primary outcomes are missing,
310 an effort will be made to contact authors to obtain data.

311 2.4.4 *Study Designs*

312 Only randomised controlled trials comparing an active intervention with another eligible
313 intervention or control will be included. Randomisation can be at the individual or group
314 level and both parallel group and crossover designs will be included. For crossover designs,
315 only data from the first trial period will be extracted to eliminate any possibility of carryover
316 effects.

2.4.5 *Language*

No language restrictions will be initially applied, although studies for which adequate translation cannot be obtained will be considered potentially eligible and described in the final report but will not be included in the meta-analysis.

2.5 Information sources

We will search for published RCTs indexed in the following databases by the final search date: MEDLINE (1946-), MEDLINE In-Process, EMBASE (1974-), CENTRAL, CINAHL (1937-), LILACS (1982-) and PsycINFO (1967-). We will also search for published, unpublished and ongoing trials in clinical trial registries ClinicalTrials.gov and WHO International Clinical Trials Registry Platform (ICTRP). We will complement published data with results reported in these trial registries. We will additionally search the websites of drug regulatory bodies of the FDA (USA), MHRA (UK) and EMA (Europe). It is important to include unpublished data, since the well-known bias towards publication of significant findings can, when relying on published literature alone, lead to an overestimation of treatment effects and an underestimation of adverse effects (Dwan et al., 2013). The search strategy will be augmented through hand searching of relevant reviews and of the reference lists of included articles for additional studies.

For unpublished clinical trials, if a study is listed as ongoing and >=1 year has elapsed since registration, we will attempt to establish whether the listed trial status is current. If it emerges that such trials have in fact been completed or terminated, we will attempt to obtain data from: (a) the trial registry, (b) study authors, (c) drug regulatory agency websites, and (d) OpenTrials (which while still in its preliminary stages can provide a wide

range of unpublished evidence including regulatory documents, clinical study reports and protocols). Where possible, the same sources will be approached when a trial has been published but key primary outcomes are not reported or reported only partially in the journal publication.

2.6 Search strategy

The search strategy was informed by PICOS criteria and will be comprised of three groups of terms relating to (1) randomized trials, (2) CLBP and (3) interventions. Search terms will be combined with a Boolean "AND" and consist of both controlled subject headings (where provided by the database) and free-text keywords in titles and abstracts.

Randomized trials will be identified using highly sensitive search filters validated for each database (Eady et al., 2008; Glanville et al., 2019; Manríquez, 2008; Wong et al., 2006) and CLBP studies identified using search terms suggested by Furlan et al (2015). For identifying treatments, we will employ subject headings for intervention trials and an extensive list of keywords for specific interventions from clinical practice guidelines (Foster et al., 2018; National Institute for Health and Care Excellence, 2016; Oliveira et al., 2018) and relevant Cochrane Reviews (https://back.cochrane.org/our-reviews).

Search strings were reviewed and approved by a healthcare information specialist at the University of Greenwich (see Additional File 2 for the draft MEDLINE example).

2.7 Study selection

Records returned by initial searches will be screened for relevancy in two stages. First, the titles and abstracts of each record will be independently screened by two members of the review team, who will exclude studies not meeting eligibility criteria. The online software Rayyan (Ouzzani et al., 2016) will be used to facilitate first stage screening by highlighting keywords relating to inclusion and exclusion criteria. Second, the full-text of the remaining articles will be screened by the same two reviewers, who will retain for inclusion in the NMA only those that meet eligibility criteria. Disagreements at any stage will be resolved through discussion or, if not resolved, with a third member of the review team.

2.8 Data Extraction

Data from each study will be extracted by one member of the review team and checked for accuracy by a senior member of the review team, with sets of studies distributed across a pool of reviewers. We will use a standardized excel coding form adapted from our previous work, with explanatory notes provided on how coding should be performed for each variable to ensure consistency across coders. If there are missing methods data or missing outcome data, the corresponding author will be contacted via e-mail with one additional reminder email sent within 3 weeks if no response is received. Subsequently, other authors will be contacted. If no response is received before analysis is conducted, the study will be excluded from the NMA but the basic study findings will be described in a separate section of the final report. When data are identified as being published across multiple sources we will prioritise extraction from the most complete data sources. Where these sources include both published and unpublished data, we will extract both but prioritise published data in

the analysis as this has been subject to peer-review, but conduct sensitivity analysis including both published and unpublished data.

When available study data do not allow computation of effect sizes using standard formula (e.g. based on means and SDs) we will: (a) extract other statistics (e.g. *F*, *p*, *t* etc) that allow effect sizes to be computed using alternative formula (Cooper et al., 2009), (b) contact study authors for data, (c) for missing SDs, used the pooled SD from other studies (Furukawa et al., 2006) or external data. Finally, where a pain rating scale assesses not only average pain, but least and worst pain over the previous period (as in the Brief Pain Inventory), we will use only average pain ratings.

2.9 Data items

Study Information extracted will include: (1) study identifiers (e.g., title, authors, publication date); (2) study characteristics (e.g., trial design, source of financial support, trial size, study location); (3) participant characteristics (e.g. mean sample age, male/female ratio, SES, pain duration, severity, and current or previous treatments); (4) intervention details (e.g. type and class of treatment, intervention details, duration, dosage, delivery method); (5) outcome data (including assessment used, timing, missing data details).

2.10 Robustness of findings and risk of bias

Risk of bias will be assessed for all studies using the revised Cochrane Risk of Bias (RoB) tool (RoB 2.0 Sterne et al., in press). Assessments will be carried out independently by two reviewers, with any disagreement resolved by discussion or, if needed, consultation with a

third reviewer. We will also collect additional measures of bias (see section 3.3) and examine their potential influence in meta-regression.

We will conduct threshold analysis (Phillippo et al., 2019; Caldwell et al., 2016) to quantify the level of bias that would have to be present in the estimated treatment effect to have resulted in a major change in treatment ranking (such as a change in the order of the highest ranked interventions). If the magnitude of such potential bias is implausible, then conclusions on the 'best' treatments are more robust. If the level of bias needed to overturn treatment decisions *is* plausible, then we will closely examine RoB scores for that treatment as well as relevant external work to determine whether such bias is likely to be present to help evaluate our confidence in the findings.

An alternate method for assessing robustness is Salanti's (Salanti et al., 2014) GRADE for NMA extension, implemented using the CINeMA web application. This estimates overall RoB for a treatment comparison by aggregating individual study RoB scores after weighting each score based on a study's contribution to the overall treatment effect size. For the proposed NMA, however, we chose threshold analysis as we will employ a Bayesian analysis (CINeMA currently applies frequentist weights), and threshold analysis is more suited to directly informing treatment decisions (Phillippo et al., 2019).

3 Data synthesis and analysis

We will provide a descriptive table summarising the key characteristics of each eligible study, including interventions, patient populations and trial characteristics. A network

diagram will show which intervention classes were compared, with larger network nodes indicating a greater number of patients and thicker connecting lines between nodes indicating a greater number of trials.

3.1 Consistency assumption

A key assumption of NMA is that each participant should be equally likely to have received any of the treatments in the network. If this assumption holds, a key consequence is that there should be no systematic differences in effect modifiers (such as important patient characteristics) across different sets of treatment comparisons that might otherwise explain apparent intervention differences (Cipriani et al., 2013).

As described in section 2.1, we will ensure similarity by restricting patient populations to those with non-specific LBP that is chronic only and who report a moderate or greater level of pain. We will also qualitatively assess the clinical similarity of populations across different treatment comparisons on potentially important factors such as age, sex, baseline pain severity and CLBP duration (Gurung et al., 2015; Beneciuk et al., 2017; Mallen et al., 2007), and present this in a summary table. Statistical tests of consistency we will employ are described in section 3.2.2 and 3.2.3. One common concern with comparing pharmacological and non-pharmacological interventions in general, is that one class of intervention is administered as a first-line treatment and the other is given to treatment resistant cases for whom previous interventions have failed. Because we are examining chronic LBP, however, treatment failure would have been likely for all patients during the acute phase of their LBP in order for chronic LBP to develop.

3.2 Network meta-analysis

We will conduct a Bayesian NMA to estimate relative treatment effects based on a synthesis of direct (head-to-head trials) and indirect evidence (where two treatments are compared indirectly via a common comparator). We will use a class-based hierarchical model (Dias et al., 2018) to estimate the relative effects of different treatment classes (e.g. NSAIDs, opioids) and of individual treatments within a class (e.g. ibuprofen, aspirin, diclofenac). Pharmacological and non-pharmacological studies may differ in patient and study characteristics and type of biases that may exist. As such, we will conduct separate analyses of these two networks along with an analysis of the whole network (providing head-to-head comparisons of pharmacological and non-pharmacological interventions are available) to see if these two approaches yield similar results.

The relative effectiveness of different treatments will be modelled as a function of their performance relative to a placebo reference treatment. This will be presented as a forest plot for class effects and in table form for class and individual effects. Mean ranks with their 95% credible intervals and SUCRA (a simple transformation of the mean rank) will be used to provide a hierarchy of the best treatments.

3.2.1 *Estimation details*

Model parameters will be estimated in WinBUGS using Markov Chain Monte Carlo simulation. Posterior distributions will be derived from binomial (binary outcomes) and normal (continuous) likelihood functions using vague prior distributions. For within-treatment study variability, we will assume a common heterogeneity standard deviation and use a partially informative uniform prior with an upper bound limit based on the outcome

scale used (e.g. U(0, 10) for pain ratings). For within-class variability (of treatments) we will use a uniform prior distribution estimated separately for each class. However, for classes with only a few elements, decisions will be made on whether the within-class variance estimates can be shared across similar classes (e.g. SNRI and SSRI classes). For other parameters we will use wide non-informative normal priors. We will examine Gelman-Rubin trace plots to check that multiple chains achieve convergence during the burn-in period, and base our estimates on 50,000 or more subsequent iterations to ensure MC estimator error is less than 5% of the standard deviation for the treatment effect and heterogeneity parameters. With respect to multi-arm trials, the correlation between multiple treatment comparisons within these trials are naturally accounted for within the Bayesian framework.

The choice between a random-effects (RE) and fixed-effect (FE) model will be informed by a comparison of Deviance Information Criteria (DIC) model fit statistics. If the DIC for the RE model is at least 3 units lower (with lower values indicating better fit) (Dias et al., 2018) we will use a RE model. If the models are otherwise similar, we will choose the more parsimonious FE model provided there is no excessive study heterogeneity from separate pairwise analysis.

3.2.2 *Assessment of consistency*

We will assess whether there is consistency of direct and indirect evidence globally across the whole network (which is a natural consequence of the similarity assumption) using the unrelated mean effects model (Dias et al., 2013). If evidence of inconsistency is found, we will use a node-splitting approach (Dias et al., 2010) to identify possible areas of local inconsistency and if sufficient data exist, run network meta-regression to examine whether

inconsistency (and study heterogeneity) is resolved by a consideration of differences in clinical variables (section 3.1).

In the event of minor unresolved inconsistency, we will proceed with NMA but advise caution in the interpretation of results for comparisons where there are material differences between direct and indirect estimates. If there is evidence of substantive inconsistency, we will consider excluding network nodes.

3.2.3 *Assessment of within-comparison heterogeneity*

Study heterogeneity within each treatment comparison will be examined with forest plots from pairwise meta-analysis for an initial visual assessment (and these will be used to alert us to potential outliers). We will also compute I^2, which indicates the proportion of overall variance in effect sizes due to genuine heterogeneity. I^2>60% can indicate a moderate or greater variation in study effect sizes (Higgins et al., 2019) and will be explored with meta-regression. We will also compute Cochran's Q with p<.10 used to indicate possible presence of heterogeneity, and tau-squared to provide an estimate of effect size heterogeneity for different comparisons

3.3 Meta-regression and sensitivity analysis

Given sufficient data, we will use network meta-regression to explore whether inconsistency/ heterogeneity and group differences in the two primary outcomes is influenced by potential biases such as industry sponsorship, performance in less (vs. more) developed countries (Desai et al., 2019), risk of bias scores, novel agent effects (Salanti et al.,

2010) and researcher allegiance to the study intervention (Dragioti et al., 2015). Two members of the review team will independently assess researcher allegiance (with any disagreement resolved by consensus) using a checklist developed and piloted for the current study (Additional File 3) based on the modified reprint method (Munder et al., 2013). We will also include effect size derivation method (post vs. change scores) as a dummy-coded covariate to check that effect sizes from both methods are similar.

We will produce treatment-control comparison adjusted funnel plots to explore possible publication bias, and if bias is suspected explore this by including sample size as a covariate. We will also perform a test of excess significance (Ioannidis and Trikalinos, 2007) which is applied to data aggregated across the whole network of interventions (thus offering higher statistical power than pairwise tests) to assess whether there is an excess of statistically significant findings.

We will also assess the robustness of the findings to various decisions by performing sensitivity analyses including removing studies (a) with high risk of bias, (b) where imputations have been performed, (c) where we assumed LBP was non-specific when this could not be definitively determined (section 2.1), and (d) where very high/low dosages were used for off label medications. In addition, we will rerun the analysis after reclassifying McKenzie therapy into mind-body awareness exercises based on feedback from the Lancet LBP working group.

3.4 Unit of analysis issues

For trials that use cluster randomisation without adjusting standard errors for the study's design effect (Hox et al., 2017), we will apply this adjustment ourselves. As intra-class correlations needed to make this correction are seldom reported, we will use values obtained from external literature for the outcome examined (or if these are not available use a single plausible value and examine the impact of varying this value in sensitivity analysis).

4 Discussion

The results from this NMA will provide an important evidence base for clinicians to inform treatment decisions by providing a comparative assessment of a wide range of interventions (Tomlinson et al., 2019). This will help efforts to develop a precision medicine approach to the treatment for non-specific chronic low back pain, which can be used in everyday clinical settings. While there are numerous factors that must be considered in treatment decisions, such as cost effectiveness, individual patient suitability and patient preferences (Kernot et al., 2019), reliable information on the pain-relieving effects and acceptability of a treatment as well as an assessment of how bias-free these results might be are fundamental points in guiding these decisions.

Given the sheer scale of the burden of chronic low back pain we expect the results of the NMA to be of considerable interest to clinicians, academics, guideline developers and policy makers (Leucht et al., 2016) and we will disseminate the findings widely through academic publications, conference presentations and communication with healthcare providers.

559

560 **Abbreviations**

561 **CINeMA:** Confidence In Network Meta-Analysis

562 **COS:** Core Outcome Set

563 **CLBP:** Chronic Low Back Pain

564 **FDA:** Food and Drug Administration

565 **GRADE:** Grading of Recommendations, Assessment, Development and Evaluation

566 **IMMPACT:** Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials

567 **LBP:** Low Back Pain

568 **NMA:** Network Meta-Analysis

569 **NSAIDs:** Non-Steroidal Anti-Inflammatory Drugs

570 **PF:** Physical Functioning

571 **PICOS:** Population, Intervention, Comparator, Outcomes, Study design

572 **PRISMA-P:** Preferred Reporting Items for Systematic review and Meta-Analysis Protocols

573 **RoB:** Risk of Bias

574 **SNRI:** Serotonin–Norepinephrine Reuptake Inhibitor

575 **SSRI:** Selective Serotonin Reuptake Inhibitor

576 **SUCRA:** Surface Under the Cumulative RAnking curve

577 **WHO:** World Health Organisation

578

579 **Declarations**

580 *Ethics approval and consent to participate*

581 Not applicable

582

583 *Consent for publication*

584 Not applicable

585

586 *Availability of data and material*

587 Not applicable

588

589 *Competing interests*

590 The authors declare that they have no competing interests

591 *Funding*

592 None

593

594 *Authors' contributions*

595 TT was responsible for the conception and design of the study and for writing the initial
596 protocol draft. JPAI contributed to the design of the study. TT and SD developed the
597 statistical analysis plan and JPAI advised. SD, DP, SW, LM, CR, JS, JF, NV, ED, BS, MS, CM, AC
598 and JPAI provided critical revisions of all aspects of the review. The final protocol has been
599 approved by all authors.

600

601 *Acknowledgements*

602 We are extremely grateful to several members of the Lancet Low Back Pain Series Working
603 Group who provided vital commentaries on the most appropriate ways to classify the
604 different CLBP interventions. Specifically, we would like to thank Prof Judith Turner
605 (University of Washington, US), Prof Peter Croft (Keele University, UK), Prof Douglas Gross

(University of Alberta, Canada), Prof Julie Fritz (University of Utah, US) and Professor Martin Underwood (University of Warwick, UK) for their invaluable comments. Many thanks also to Cathryn Peppard, a health librarian at the University of Greenwich, for assistance with the construction of search strings. Andrea Cipriani is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility, by an NIHR Research Professorship (grant RP-2017-08-ST2-006), by the NIHR Oxford and Thames Valley Applied Research Collaboration and by the NIHR Oxford Health Biomedical Research Centre (grant BRC-1215-20005). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health.

Authors' information

School of Human Sciences, University of Greenwich, London, UK

Trevor Thompson, Damian Poulter, Lucy Marsh, Claire Rossato

Centre for Reviews and Dissemination, University of York, York, UK

Sofia Dias

School of Health Sciences, University of Greenwich, London, UK

Sharon Weldon

Barts Health NHS Trust, The Royal London Hospital, London, UK

Sharon Weldon

Department of Pediatrics, Yonsei University College of Medicine, Seoul, Republic of Korea

Jae Il Shin

NICM Health Research Institute, Western Sydney University, Sydney, Australia

Joseph Firth

Division of Psychology and Mental Health, University of Manchester, UK

Joseph Firth

National Research Council, Neuroscience Institute, Aging Branch, Padova, Italy

Nicola Veronese

Pain and Rehabilitation Centre and Departments of Health, Medicine and Caring Sciences, Linköping University, Sweden

Elena Dragioti

King's College London and South London and Maudsley NHS Foundation Trust, UK

Brendon Stubbs

Neurosciences Department, University of Padua, Italy

Marco Solmi

Sydney School of Public Health, The University of Sydney, Australia

Christopher G Maher

Department of Psychiatry, University of Oxford, UK

Andrea Cipriani

Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK

Andrea Cipriani

Meta-Research Innovation at Stanford (METRICS) and Departments of Medicine, Health Research and Policy, Biomedical Science and Statistics, Stanford University, USA

John P.A. Ioannidis

References

Abenhaim, L., Rossignol, M., Valat, J. P., Nordin, M., Avouac, B., Blotman, F., . . . Vautravers, P. (2000). The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. *Spine (Phila Pa 1976)*, *25*(4 Suppl), 1S-33S. doi:10.1097/00007632-200002151-00001

Bardin, L. D., King, P., & Maher, C. G. (2017). Diagnostic triage for low back pain: a practical approach for primary care. *Med J Aust*, *206*(6), 268-273. doi:10.5694/mja16.00828

Beneciuk, J. M., Hill, J. C., Campbell, P., Afolabi, E., George, S. Z., Dunn, K. M., & Foster, N. E. (2017). Identifying Treatment Effect Modifiers in the STarT Back Trial: A Secondary Analysis. *Journal of Pain*, *18*(1), 54-65. doi:10.1016/j.jpain.2016.10.002

Boonstra, A. M., Stewart, R. E., Köke, A. J., Oosterwijk, R. F., Swaan, J. L., Schreurs, K. M., & Schiphorst Preuper, H. R. (2016). Cut-Off Points for Mild, Moderate, and Severe Pain on the Numeric Rating Scale for Pain in Patients with Chronic Musculoskeletal Pain:

677 Variability and Influence of Sex and Catastrophizing. *Frontiers in Psychology*, *7*, 1466.
678 doi:10.3389/fpsyg.2016.01466
679 Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. R. (2009). *Introduction to Meta-*
680 *Analysis*. West Sussex: Wiley.
681 Buchbinder, R., & Underwood, M. (2012). Prognosis in people with back pain. *CMAJ*,
682 *184*(11), 1229-1230. doi:10.1503/cmaj.120627
683 Caldwell, D. M., Ades, A. E., Dias, S., Watkins, S., Li, T., Taske, N., . . . Welton, N. J. (2016). A
684 threshold analysis assessed the credibility of conclusions from network meta-analysis.
685 *Journal of Clinical Epidemiology*, *80*, 68-76. doi:10.1016/j.jclinepi.2016.07.003
686 Chiarotto, A., Boers, M., Deyo, R. A., Buchbinder, R., Corbin, T. P., Costa, L. O. P., . . . Ostelo,
687 R. W. (2018). Core outcome measurement instruments for clinical trials in nonspecific
688 low back pain. *Pain*, *159*(3), 481-495. doi:10.1097/j.pain.0000000000001117
689 Chou, R., Deyo, R., Friedly, J., Skelly, A., Weimer, M., Fu, R., . . . Grusing, S. (2017). Systemic
690 Pharmacologic Therapies for Low Back Pain: A Systematic Review for an American
691 College of Physicians Clinical Practice Guideline. *Annals of Internal Medicine*, *166*(7),
692 480-492. doi:10.7326/M16-2458
693 Cipriani, A., Higgins, J. P., Geddes, J. R., & Salanti, G. (2013). Conceptual and technical
694 challenges in network meta-analysis. *Annals of Internal Medicine*, *159*(2), 130-137.
695 doi:10.7326/0003-4819-159-2-201307160-00008
696 Cooper, H., Hedges, L., & Valentine, J. (2009). *Handbook of Research Synthesis and Meta-*
697 *analysis.* NY: Russell Sage Foundation.
698 Costa, L., Maher, C. G., Hancock, M. J., McAuley, J. H., Herbert, R. D., & Costa, L. O. (2012).
699 The prognosis of acute and persistent low-back pain: a meta-analysis. *CMAJ*, *184*(11),
700 E613-24. doi:10.1503/cmaj.111271

da Costa, B. R., Nüesch, E., Rutjes, A. W., Johnston, B. C., Reichenbach, S., Trelle, S., . . . Jüni, P. (2013). Combining follow-up and change data is valid in meta-analyses of continuous outcomes: a meta-epidemiological study. *Journal of Clinical Epidemiology*, *66*(8), 847-855. doi:10.1016/j.jclinepi.2013.03.009

Desai, K., Carroll, I., Asch, S., Hernandez-Boussard, T., & Ioannidis, J. P. A. (2019). Extremely large outlier treatment effects may be a footprint of bias in trials from less developed countries: randomized trials of gabapentinoids. *Journal of Clinical Epidemiology*, *106*, 80-87. doi:10.1016/j.jclinepi.2018.10.012

Dias, S., & Caldwell, D. M. (2019). Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed*, *104*(1), F8-F12. doi:10.1136/archdischild-2018-315224

Dias, S., Welton, N. J., Caldwell, D. M., & Ades, A. E. (2010). Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*, *29*(7-8), 932-944. doi:10.1002/sim.3767

Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G., & Ades, A. E. (2013). Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making*, *33*(5), 641-656. doi:10.1177/0272989X12455847

Dias, S., Ades, A. E., Welton, N. J., Jansen, J. P., & Sutton, A. J. (2018). *Network Meta-Analysis for Decision-Making (Statistics in Practice)*. Wiley.

Dominici, F., Parmigiani, G., Wolpert, R. L., & Hasselblad, V. (1999). Meta-Analysis of Migraine Headache Treatments: Combining Information from Heterogeneous Designs. *Journal of the American Statistical Association*, *94*(445), 16. doi:10.2307/2669674

Dragioti, E., Dimoliatis, I., Fountoulakis, K. N., & Evangelou, E. (2015). A systematic appraisal of allegiance effect in randomized controlled trials of psychotherapy. *Ann Gen Psychiatry*, *14*, 25. doi:10.1186/s12991-015-0063-1

Dwan, K., Gamble, C., Williamson, P. R., Kirkham, J. J., & Reporting, B. G. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS ONE*, *8*(7), e66844. doi:10.1371/journal.pone.0066844

Dworkin, R. H., Turk, D. C., McDermott, M. P., Peirce-Sandner, S., Burke, L. B., Cowan, P., . . . Sampaio, C. (2009). Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. *Pain*, *146*(3), 238-244. doi:10.1016/j.pain.2009.08.019

Eady, A. M., Wilczynski, N. L., & Haynes, R. B. (2008). PsycINFO search strategies identified methodologically sound therapy studies and review articles for use by clinicians and researchers. *Journal of Clinical Epidemiology*, *61*(1), 34-40. doi:10.1016/j.jclinepi.2006.09.016

Finley, C. R., Chan, D. S., Garrison, S., Korownyk, C., Kolber, M. R., Campbell, S., . . . Allan, G. M. (2018). What are the most common conditions in primary care? Systematic review. *Can Fam Physician*, *64*(11), 832-840.

Foster, N. E., Anema, J. R., Cherkin, D., Chou, R., Cohen, S. P., Gross, D. P., . . . Lancet, L. B. P. S. W. G. (2018). Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet*, *391*(10137), 2368-2383. doi:10.1016/S0140-6736(18)30489-6

Furlan, A. D., Malmivaara, A., Chou, R., Maher, C. G., Deyo, R. A., Schoene, M., . . . Editorial Board of the Cochrane Back, N. G. (2015). 2015 Updated Method Guideline for

Systematic Reviews in the Cochrane Back and Neck Group. *Spine (Phila Pa 1976)*, *40*(21), 1660-1673. doi:10.1097/BRS.0000000000001061

Furukawa, T. A., Barbui, C., Cipriani, A., Brambilla, P., & Watanabe, N. (2006). Imputing missing standard deviations in meta-analyses can provide accurate results. *Journal of Clinical Epidemiology*, *59*(1), 7-10. doi:10.1016/j.jclinepi.2005.06.006

Furukawa, T. A., Cipriani, A., Barbui, C., Brambilla, P., & Watanabe, N. (2005). Imputing response rates from means and standard deviations in meta-analyses. *Int Clin Psychopharmacol*, *20*(1), 49-52. doi:10.1097/00004850-200501000-00010

Gagnier, J. J., Oltean, H., van Tulder, M. W., Berman, B. M., Bombardier, C., & Robbins, C. B. (2016). Herbal Medicine for Low Back Pain: A Cochrane Review. *Spine (Phila Pa 1976)*, *41*(2), 116-133. doi:10.1097/BRS.0000000000001310

GBD. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, *390*(10100), 1211-1259. doi:10.1016/S0140-6736(17)32154-2

Glanville, J., Dooley, G., Wisniewski, S., Foxlee, R., & Noel-Storr, A. (2019). Development of a search filter to identify reports of controlled clinical trials within CINAHL Plus. *Health Info Libr J*, *36*(1), 73-90. doi:10.1111/hir.12251

Gurung, T., Ellard, D. R., Mistry, D., Patel, S., & Underwood, M. (2015). Identifying potential moderators for response to treatment in low back pain: A systematic review. *Physiotherapy*, *101*(3), 243-251. doi:10.1016/j.physio.2015.01.006

Hayden, J. A., van Tulder, M. W., & Tomlinson, G. (2005). Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. *Annals of Internal Medicine*, *142*(9), 776-785. doi:10.7326/0003-4819-142-9-200505030-00014

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane Handbook for Systematic Reviews of Interventions* (2nd ed.). Chichester, UK: John Wiley & Sons.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. London: Routledge.

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clin Trials*, *4*(3), 245-253. doi:10.1177/1740774507079441

Kernot, C., Tomlinson, A., Chevance, A., & Cipriani, A. (2019). One step closer to personalised prescribing of antidepressants: using real-world data together with patients and clinicians' preferences. *Evid Based Ment Health*, *22*(3), 91-92. doi:10.1136/ebmental-2019-300105

Leucht, S., Chaimani, A., Cipriani, A. S., Davis, J. M., Furukawa, T. A., & Salanti, G. (2016). Network meta-analyses should be the highest level of evidence in treatment guidelines. *Eur Arch Psychiatry Clin Neurosci*, *266*(6), 477-480. doi:10.1007/s00406-016-0715-4

Maher, C., Underwood, M., & Buchbinder, R. (2017). Non-specific low back pain. *Lancet*, *389*(10070), 736-747. doi:10.1016/S0140-6736(16)30970-9

Mallen, C. D., Peat, G., Thomas, E., Dunn, K. M., & Croft, P. R. (2007). Prognostic factors for musculoskeletal pain in primary care: a systematic review. *Br J Gen Pract*, *57*(541), 655-661.

Manríquez, J. J. (2008). A highly sensitive search strategy for clinical trials in Literatura Latino Americana e do Caribe em Ciências da Saúde (LILACS) was developed. *Journal of Clinical Epidemiology*, *61*(4), 407-411. doi:10.1016/j.jclinepi.2007.06.009

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . PRISMA-P, G. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*, 1. doi:10.1186/2046-4053-4-1

Morlion, B. (2013). Chronic low back pain: pharmacological, interventional and surgical strategies. *Nat Rev Neurol*, *9*(8), 462-473. doi:10.1038/nrneurol.2013.130

Munder, T., Brütsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: an overview of reviews. *Clinical Psychology Review*, *33*(4), 501-511. doi:10.1016/j.cpr.2013.02.002

NatCen Social Research. (2014). Health Survey for England, 2012., *University College London, Department of Epidemiology and Public Health*. doi:10.5255/UKDA-SN-7480-1

National Institute for Health and Care Excellence. (2016). *Low Back Pain and Sciatica in Over 16s: Assessment and Management. NICE Guideline, No. 59.* London: National Guideline Centre (UK)., National Institute for Health and Care Excellence (UK).

Oliveira, C. B., Maher, C. G., Pinto, R. Z., Traeger, A. C., Lin, C. C., Chenot, J. F., . . . Koes, B. W. (2018). Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. *Eur Spine J*, *27*(11), 2791-2803. doi:10.1007/s00586-018-5673-2

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. doi:10.1186/s13643-016-0384-4

Page, M. J., McKenzie, J. E., Kirkham, J., Dwan, K., Kramer, S., Green, S., & Forbes, A. (2014). Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. *The Cochrane database of systematic reviews*, 10), MR000035. doi:10.1002/14651858.MR000035.pub2

817 Phillippo, D. M., Dias, S., Welton, N. J., Caldwell, D. M., Taske, N., & Ades, A. E. (2019).
818 Threshold Analysis as an Alternative to GRADE for Assessing Confidence in Guideline
819 Recommendations Based on Network Meta-analyses. *Annals of Internal Medicine*,
820 *170*(8), 538-548. doi:10.7326/M18-3542
821 Salanti, G., Del Giovane, C., Chaimani, A., Caldwell, D. M., & Higgins, J. P. (2014). Evaluating
822 the quality of evidence from a network meta-analysis. *PLoS ONE*, *9*(7), e99682.
823 doi:10.1371/journal.pone.0099682
824 Salanti, G., Dias, S., Welton, N. J., Ades, A. E., Golfinopoulos, V., Kyrgiou, M., . . . Ioannidis, J.
825 P. (2010). Evaluating novel agent effects in multiple-treatments meta-regression. *Stat*
826 *Med*, *29*(23), 2369-2383. doi:10.1002/sim.4001
827 Samara, M. T., Spineli, L. M., Furukawa, T. A., Engel, R. R., Davis, J. M., Salanti, G., & Leucht,
828 S. (2013). Imputation of response rates from means and standard deviations in
829 schizophrenia. *Schizophr Res*, *151*(1-3), 209-214. doi:10.1016/j.schres.2013.10.029
830 Sterne, J. A. C., Savović, J., Page, MJ, Elbers, RG, Blencowe, NS, Boutron, I., . . . Higgins, J. P. T.
831 (in press). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*
832 *Open*.
833 Tomlinson, A., Furukawa, T. A., Efthimiou, O., Salanti, G., De Crescenzo, F., Singh, I., &
834 Cipriani, A. (2019). Personalise antidepressant treatment for unipolar depression
835 combining individual choices, risks and big data (PETRUSHKA): rationale and protocol.
836 *Evid Based Ment Health*. doi:10.1136/ebmental-2019-300118
837 van Tulder, M., & Koes, B. (2013). Low back pain. In S. McMahon, M. Koltzenburg, I. Tracey,
838 & D. Turk (Eds.), *Textbook of pain* (6th ed.). London: Elsevier.

Whittle, S. L., Richards, B. L., Husni, E., & Buchbinder, R. (2011). Opioid therapy for treating rheumatoid arthritis pain. *The Cochrane database of systematic reviews*, 11), CD003113. doi:10.1002/14651858.CD003113.pub3

Wong, S. S., Wilczynski, N. L., & Haynes, R. B. (2006). Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. *J Med Libr Assoc*, *94*(1), 41-47.

World Health Organization. (2003). The burden of musculoskeletal conditions at the start of the new millennium. *World Health Organ Tech Rep Ser*, *919*, i-x, 1.