Comparative effectiveness and safety of analgesic medicines for adults with acute non-specific low back pain: systematic review and network meta-analysis

Michael A Wewege,[1,2] Matthew K Bagg,[2,3,4] Matthew D Jones,[1,2] Michael C Ferraro,[1,2] Aidan G Cashin,[1,2] Rodrigo RN Rizzo,[1,2] Hayley B Leake,[2,5] Amanda D Hagstrom,[1] Saurab Sharma,[1,2] Andrew J McLachlan,[6] Christopher G Maher,[7,8] Richard Day,[9,10] Benedict M Wand,[11] Neil E O'Connell,[12] Adriani Nikolakopoulou,[13] Siobhan Schabrun,[2,14,15] Sylvia M Gustin,[2,16] James H McAuley[1,2]

ABSTRACT

OBJECTIVE
To evaluate the comparative effectiveness and safety of analgesic medicines for acute non-specific low back pain.

DESIGN
Systematic review and network meta-analysis.

DATA SOURCES
Medline, PubMed, Embase, CINAHL, CENTRAL, ClinicalTrials.gov, clinicialtrialsregister.eu, and World Health Organization's International Clinical Trials Registry Platform from database inception to 20 February 2022.

ELIGIBILITY CRITERIA FOR STUDY SELECTION
Randomised controlled trials of analgesic medicines (eg, non-steroidal anti-inflammatory drugs, paracetamol, opioids, anti-convulsant drugs, skeletal muscle relaxants, or corticosteroids) compared with another analgesic medicine, placebo, or no treatment. Adults (≥18 years) who reported acute non-specific low back pain (for less than six weeks).

DATA EXTRACTION AND SYNTHESIS
Primary outcomes were low back pain intensity (0-100 scale) at end of treatment and safety (number of participants who reported any adverse event during treatment). Secondary outcomes were low back specific function, serious adverse events, and discontinuation from treatment. Two reviewers independently identified studies, extracted data, and assessed risk of bias. A random effects network meta-analysis was done and confidence was evaluated by the Confidence in Network Meta-Analysis method.

RESULTS
98 randomised controlled trials (15 134 participants, 49% women) included 69 different medicines or combinations. Low or very low confidence was noted in evidence for reduced pain intensity after treatment with tolperisone (mean difference −26.1 (95% confidence intervals −34.0 to −18.2)), aceclofenac plus tizanidine (−26.1 (−38.5 to −13.6)), pregabalin (−24.7 (−34.6 to −14.7)), and 14 other medicines compared with placebo. Low or very low confidence was noted for no difference between the effects of several of these medicines. Increased adverse events had moderate to very low confidence with tramadol (risk ratio 2.6 (95% confidence interval 1.5 to 4.5)), paracetamol plus sustained release tramadol (2.4 (1.5 to 3.8)), baclofen (2.3 (1.5 to 3.4)), and paracetamol plus tramadol (2.1 (1.3 to 3.4)) compared with placebo. These medicines could increase the risk of adverse events compared with other medicines with moderate to low confidence. Moderate to low confidence was also noted for secondary outcomes and secondary analysis of medicine classes.

CONCLUSIONS
The comparative effectiveness and safety of analgesic medicines for acute non-specific low back pain are uncertain. Until higher quality randomised controlled trials of head-to-head comparisons are published, clinicians and patients are recommended to take a cautious approach to manage acute non-specific low back pain with analgesic medicines.

SYSTEMATIC REVIEW REGISTRATION
PROSPERO CRD42019145257

WHAT IS ALREADY KNOWN ON THIS TOPIC?

Analgesic medicines are a common treatment for acute non-specific low back pain

Previous reviews have evaluated analgesic medicines compared with placebo, but the evidence for the comparative effectiveness of these medicines is limited

WHAT THIS STUDY ADDS

Low or very low confidence evidence suggests that some analgesic medicines might be superior for reducing pain intensity, limited by trial risk of bias and imprecision in effect estimates

Evidence of moderate to very low confidence suggests that some analgesic medicines might increase the risk of adverse events during treatment

Clinicians and patients are recommended to take a cautious approach to managing acute non-specific low back pain with analgesic medicines until higher quality trials of head-to-head comparisons are published

Introduction

Acute low back pain (for less than six weeks' duration) is a common presentation in primary care.[1] Acute non-specific low back pain, in which a pathoanatomical cause of pain cannot be reliably determined, represents more than 90% of these presentations.[2] Clinical practice guidelines recommend advice, reassurance, encouragement of physical activity, and self-management of symptoms as first line care.[3] Second line care includes non-pharmacological interventions (eg, manual therapy) and analgesic medicines.[3-6] Surveys about primary care indicate many adults receive an analgesic medicine (48% in the UK and 61% in Australia).[7 8]

Clinicians who prescribe medicines for low back pain must choose between medicines with different analgesic properties and safety profiles. Systematic reviews that compared medicines with placebo only partially inform this decision.[9-17] A network meta-analysis combines direct and indirect information across a network of randomised clinical trials to estimate the comparative effectiveness of multiple treatments.[18] This study type incorporates evidence from placebo controlled trials and trials of comparative effectiveness.[19] A previous network meta-analysis compared the effectiveness of classes of analgesic medicines as part of a broader evaluation of pharmacological and non-pharmacological interventions.[20] However, no comprehensive evaluation of individual medicines is available to inform clinical decision making for the best medicine for acute non-specific low back pain.[21 22]

Our study used a network meta-analysis to evaluate the comparative effectiveness of analgesic medicines for adults with acute non-specific low back pain.

Methods
We followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses-network meta-analysis (PRISMA-NMA) statement for this article.[23] This report is part of a larger project (PROSPERO CRD42019145257) evaluating analgesic medicines for low back pain. The published protocol appears in supplement 1,[24] and protocol updates are in supplement 2a and 2b.

Eligibility criteria
We included randomised controlled trials of adults (≥18 years) with acute non-specific low back pain.[1] We included randomised controlled trials that compared an analgesic medicine with another analgesic medicine, placebo medicine, or no treatment (including continuation of usual care or being placed on a waitlist). We did not restrict our criteria by language or publication status. We excluded randomised controlled trials with enriched enrolment because this method violates the transitivity assumption.[24-26]

We included non-steroidal anti-inflammatory drugs, paracetamol, opioids, anticonvulsants, antidepressants, skeletal muscle relaxants, or corticosteroids from the World Health Organization Anatomical Therapeutic Chemical system (supplement 2c).[27] Medicines must have had a license for use in humans in 2021 by the US Food and Drug Administration,[28] UK Medicine and Healthcare Products Regulatory Agency,[29] European Medicines Agency,[30] or Australian Therapeutic Goods Administration.[31] We included additional licensed medicines in these classes that were identified during the review process. Medicines must have been administered systemically (eg, oral, intravenous, and intramuscular) as a single drug or combination formulations, at any dose. We excluded non-systemic administrations (eg, topical and epidural). Trials that used non-pharmacological co-interventions were included and were considered in the assessment of transitivity.[24]

We only included trials that assessed the effects of medicines that had been administered for a minimum of 24 h or, where single administration was used, outcomes at the end of treatment had to have been measured a minimum of 24 h later. This threshold excluded trials that tested the analgesic effect of medicines on immediate term outcomes only, which typically examined acute emergency care or experimental settings and is different to primary care.[32 33]

Data sources
We searched five electronic databases and three clinical trial registers (Medline, PubMed, Embase, CINAHL, the Cochrane Central Register of Controlled Trials, ClinicalTrials.gov, EU Clinical Trials Register, and the World Health Organization's International Clinical Trial Registry Platform) from database inception until 20 February 2022. Full search strategies appear in supplement 2d. We also searched previous reviews and reference lists of included trials, which returned no additional records.

Study identification
Two authors (MAW and one of MDJ, MCF, AGC, RRNR, HBL, ADH, or SSh) independently screened records by title and abstract and full text in Covidence.[34] Authors were experienced with similar eligibility criteria[17 35-37] and were trained for this review. Discrepancies were resolved through discussion and arbitration from a third author (JHM). If required, the corresponding author of the trial was contacted up to three times to determine record eligibility. All included records underwent linkage to establish unique trials.[38]

Outcomes and data extraction
Two authors (MAW and one of MDJ, MCF, AGC, RRNR, HBL, ADH, or SSh) independently extracted data from included trials into standardised spreadsheets, with discrepancies resolved through discussion. Authors were experienced with these extraction sheets.[17 35-37]

We extracted information on trial characteristics (country, setting, number of trial sites, sample size, duration), participants (diagnosis, duration of low back pain, numbers of men and women, pain intensity at baseline, comorbidities), interventions (medicine, route of administration, duration of intervention, dosage, usage of rescue medication, provision of usual care, co-interventions prescribed by trial investigators), and outcomes.

The primary outcomes were low back pain intensity (0-100 scale, values as integers) at the end of treatment, and safety (number of participants who had any adverse event during the treatment period).[39] The end of treatment endpoint accounts for the different treatment durations of medicines. Secondary outcomes were low back specific function (0-100 scale, values as integers), harm (number of participants who had a serious adverse event during the treatment period),[39 40] and acceptability (number of participants who stopped participation in the trial for any reason before the end of treatment).[41]

32 597
Records identified
31 897 Databases   **700** Registers

6595
Records removed before screening
6595 Duplicate records
0 Records marked as ineligible by automation tools
0 Records removed for other reasons

26 002
Records screened

24 548
Records excluded

1454
Reports sought for retrieval

0
Reports not retrieved

1454
Reports assessed for eligibility

1300
Records excluded
319 Not randomised controlled trial
31 No back pain
155 Chronic back pain
30 Sciatica
7 Operative setting
4 Low back pain caused by specific pathology
222 Does not contain two interventions of interest
261 Duplicate
95 Secondary analysis
53 Enriched design
2 Retracted article
9 Intervention duration
112 No separate data

154
Eligible records
124
Eligible trials

26
Trial registrations terminated, ongoing, or unknown
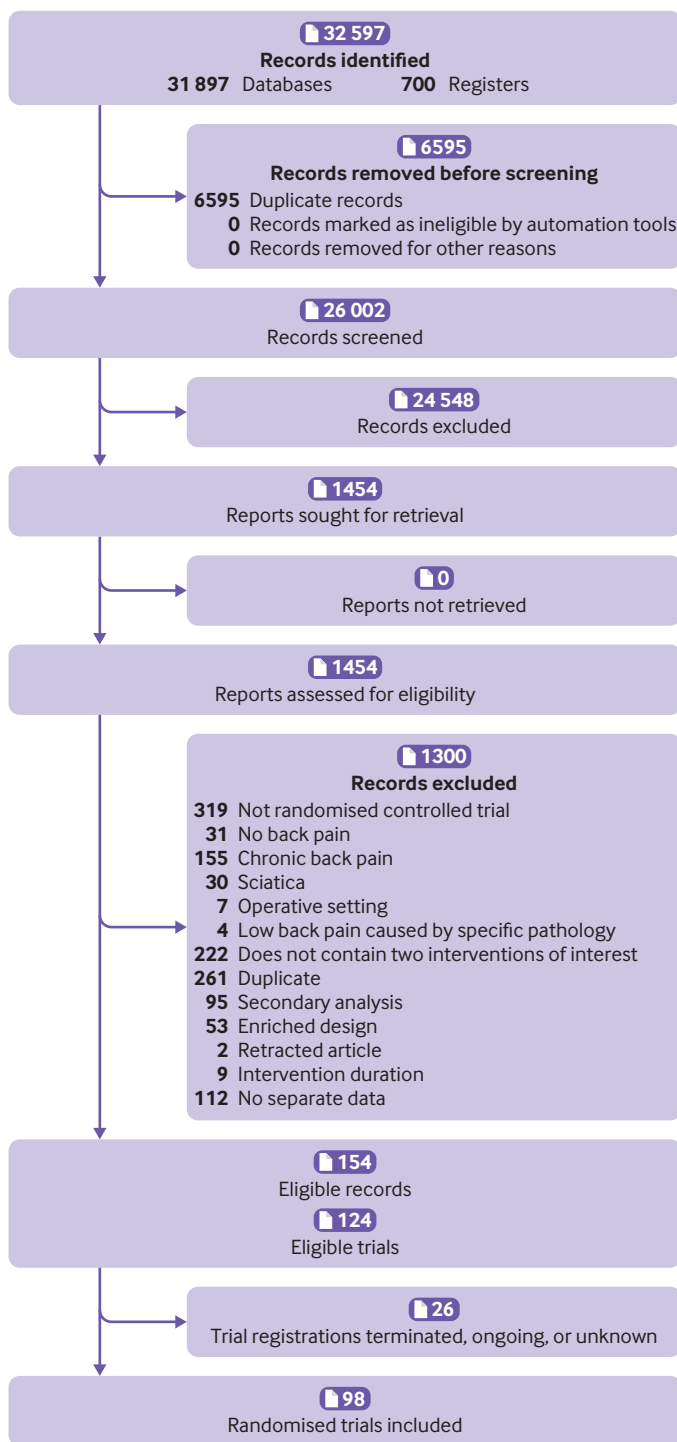
98
Randomised trials included

Fig 1 | Flow diagram of study identification, screening, and inclusion

For pain intensity and function, data from continuous self-reported scales were extracted at the time point closest to end of treatment. The hierarchy for extraction of data formats was (1) group mean and standard deviation at end of treatment, (2) group mean change from baseline and standard deviation, and (3) between group differences. Data from studies reporting multiple measures for pain intensity were prioritised as follows: 100 mm visual analogue scale, 10 cm visual analogue scale, 11 point numerical rating scale,

rating scale from a composite measure, and ordinal scale.[17][36] Data from studies that reported multiple measures for function were prioritised similarly: Oswestry Disability Index,[42] Roland Morris Disability Questionnaire,[43] rating scale from a composite measure, ordinal scale.[17][36] Data for pain intensity and function were normalised to 0-100 scales before analysis to improve clinical interpretability.[9][10][44] Data presented in other forms (eg, median or standard error) were transformed.[45][46] If measures of variance were not reported and unobtainable, the median standard deviation value from included studies with low risk of bias was imputed (30/100 for pain intensity and 35/100 for function). The number of participants per group who had one or more events was extracted for safety, harm, and acceptability.

The corresponding author of a trial was contacted up to three times via email to request missing outcomes (eg, mean and standard deviation for pain intensity or function and number of participants who had adverse events) and demographic data (eg, age, sex, baseline pain intensity).

Risk of bias

Two authors (MAW and one of MDJ, MCF, AGC, RRNR, HBL, ADH, or SSh) independently appraised outcome level risk of bias using the Cochrane tool for assessing risk of bias in randomised trials (RoB 2).[47] For each outcome, we assessed risk of bias across five domains: randomisation process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. We visualised risk of bias ratings using the robvis tool.[48]

Data synthesis and analysis
Evaluation of transitivity
Transitivity, the key assumption for valid estimation of indirect comparisons, was assessed before conducting analyses.[18][49][50] The distributions of prespecified effect modifiers were examined across network comparisons: baseline pain intensity (continuous), presence of co-interventions (binary), sample size (continuous),[51] whether participants were required to be previously untreated to the test medicine (binary), and medicine dose (binary).[24] Dose was classified as within or above the standard dosing range, sourced from the Prescriber's Digital Reference,[52] Monthly Index of Medical Specialties,[53] or Australian Medicines Handbook.[54] If unavailable, the licensed dosing range was used.

Measures of effect
We analysed comparisons of between group level mean and standard deviation values for pain intensity and function at end of treatment using mean difference with 95% confidence intervals on a 0-100 scale (values as integers). We also analysed comparisons of between group level event rates for safety, harm, and acceptability by risk ratio with 95% confidence intervals. Effects were considered statistically significant when the 95% confidence interval did not

Table 1 | Summary of studies included in the review

Characteristics	No. (%)
Patients:	
Women (reported in 71 trials)	5700 (49)
Men (reported in 71 trials)	5927 (51)
Baseline pain intensity (0-100) (reported in 60 trials)*	65 (57-72)
Treatments:	
Individual medicines	42/69 (61)
Medicine combinations	27/69 (39)
Treatment duration, days†	1-42
Within standard/licensed dosing range*	168/172 (98)
Trial design:	
Placebo controlled	38/98 (39)
Double blind (participant and clinician)	66/98 (67)
Industry sponsorship/control	40/98 (41)
Single centre (reported in 76 trials)	28/76 (37)
Multisite (reported in 76 trials)	48/76 (63)
No. of trials analysed for each outcome for medicines‡:	
Pain intensity	66/98 (67)
Safety	68/92 (74)
Function	33/44 (75)
Acceptability	74/79 (94)
No. of trials analysed for each outcome for classes‡:	
Pain intensity	45/65 (69)
Safety	46/59 (78)
Function	23/29 (79)
Acceptability	53/53 (100)

Data are n (%), n/N (%), unless indicated otherwise.
*Data are median (interquartile range). Data from other scales were normalised to a 0-100 scale before analysis.
†Data are range.
‡No. of trials included in the network meta-analysis (after evaluation of network diagnostics) from No. of trials that measured the outcome. Values below 100% indicate data was missing from one or more trials for the outcome.

cross the null. For pain intensity and function, between group differences were considered small if 5-10 points, moderate if more than 10-20 points, and large if more than 20 points.[55 56] Confidence in the effect estimates was judged using Confidence in Network Meta-Analysis (CINeMA),[57 58] which considered six domains: trial level risk of bias, reporting bias, indirectness, imprecision, heterogeneity, and incoherence. Descriptions of how we considered each domain are available in our protocol.[24]

Analytical approach
We performed a random effects network meta-analysis using the netmeta package in R, which implements a frequentist method based on a graph theoretical approach, according to electrical network theory.[59 60] The method follows a two stage approach, in which study effect estimates and their variances are synthesised and weighted by the inverse of their variance. We assumed a common heterogeneity variance across the network for each outcome, which was added to each comparison of the network and estimated via the generalised DerSimonian-Laird method of moments estimator.[61 62] Dependent observations from trials with more than three groups were accounted for with a back calculation of variances.[59] Results from the network meta-analysis were presented as summary relative effect sizes (mean difference or risk ratio) along with 95% confidence intervals, derived assuming a normal distribution of the effects, for each possible pair of treatments. We calculated P scores (the frequentist equivalent of the surface under the cumulative ranking curve (SUCRA)) to measure the extent of certainty that a treatment is better than any other treatment.[63] Estimates of heterogeneity and the proportion of variability that was not due to sampling error were calculated for each comparison. Statistics were also calculated for heterogeneity across the network, within designs, and between designs. We evaluated coherence (statistical agreement between direct and indirect treatment effects in closed loops)[19] by use of these heterogeneity statistics, and complemented with the design by treatment interaction model,[64 65] the net heat plot,[66] and the Separating Indirect from Direct Evidence (node splitting) approach.[67] Small trial effects were evaluated using comparison adjusted funnel plots, with reference to placebo (supplement 3).[68]

Node definitions
The nodes for the primary analysis of each outcome were defined at the level of the medicines. Each single drug or combination formulation was a separate node. We considered licensed sustained release formulations as separate nodes to conventional formulations of the same medicine. Different routes of administration for the same medicine (or combination) were merged into the same node. Where trials reported more than one intervention group within the same dosing range, we combined the outcome data.[46]

The secondary analysis considered classes of medicines as separate nodes in the network for each outcome. Medicines were combined into classes based on expertise of the author team, clinical guidelines, and previous reviews[9-17] (supplement 2a).

Additional analyses
Prespecified sensitivity analyses of the primary outcomes (pain intensity and safety) assessed the effect of removing trials with overall high risk of bias, removing medicines with dosages above the standard or licensed dosing range, removing groups with baseline pain intensity above 70/100, removing trials with total sample sizes of fewer than 50 participants, and removing trials where data were imputed. These analyses were done where the network structure remained the same as the primary analysis. We also conducted a post hoc sensitivity analysis in which we removed two trials that were published in predatory journals, with concerns for research integrity (supplement 2b). We were asked during peer review to perform a post hoc sensitivity analysis on industry sponsorship.

Patient and public involvement
This study did not involve any patient representatives or members of the public in a formal capacity. As a result of limited funding, we were not able to engage with consumer groups and the review protocol was drafted before the involvement of patients and the public in reviews became standard practice. The review team provided the results of the review to their clinical

Fig 2 | Network plot for pain intensity for medicines for placebo network. Within each network, the node size is proportional to the sample size of each intervention and the line thickness is proportional to the number of trials in the comparison (also indicated by the numbers). Light purple shading indicates trials with more than two arms. The two trials that did not connect to the network and that were not included were hydrocodone plus ibuprofen versus oxycodone plus paracetamol (Palangio 2002; for full details of references see supplement 2); and etodolac plus thiocolchicoside versus etodolac plus tolperisone (Garg 2019)

colleagues and individuals from the general public with whom they had personal relationships. The team sought informal feedback from these individuals based on their experiences with low back pain as either patients or clinicians.

Results

We identified 154 eligible records corresponding to 124 eligible trials. Twenty six trial registrations were noted as terminated, ongoing, or unknown. Therefore, we included 98 randomised controlled trials published
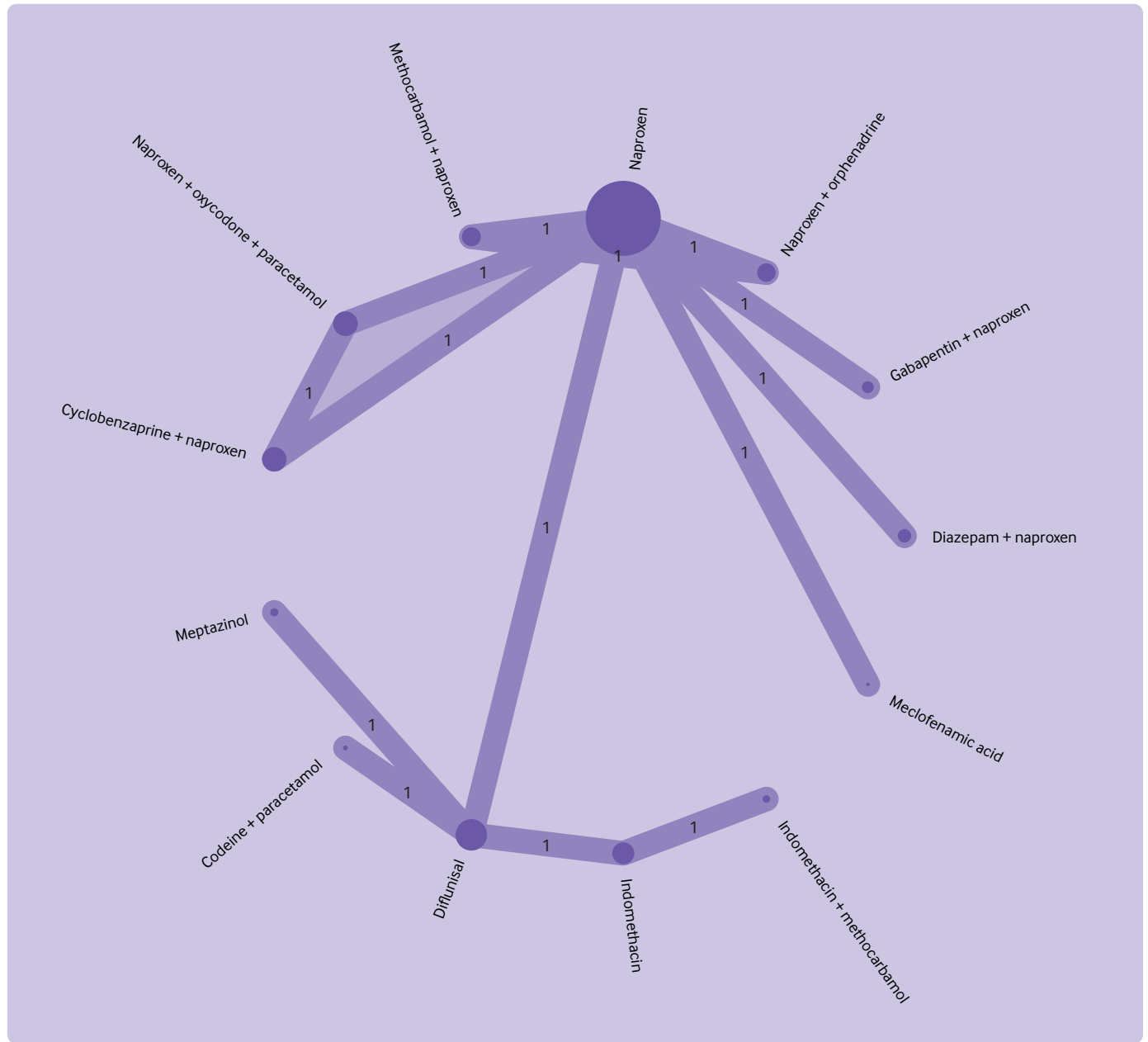
5

Fig 3 | Naproxen network plot for pain intensity for medicines. Within each network, the node size is proportional to the sample size of each intervention and the line thickness is proportional to the number of trials in the comparison (also indicated by the numbers). Light purple shading indicates trials with more than two arms

between 1964 and 2021 (fig 1). The 1300 records excluded during full text screening are provided in supplement 2e. The 98 included trials (n=15 134 participants) evaluated 70 unique interventions (69 medicines or combinations, and placebo; supplement 2f). No trials included a no treatment group.

Participant characteristics (table 1) reflected typical acute non-specific low back pain populations: 49% women, mean age mostly between 30 and 60 years, low back pain duration ranged from 24 h to 21 days, and median pain intensity at baseline of 65/100 (interquartile range 57-72) across included trials. Thirty eight (39%) of 98 trials were placebo controlled, 66 trials (67%) masked both participants

and clinicians, and 40 trials (41%) reported industry sponsorship. Analyses on industry sponsorship are reported in supplement 2. Characteristics about participants, interventions, and outcomes are available in supplement 2g and 2h. Characteristics about trial registrations noted as terminated, ongoing, or unknown are available in supplement 2i.

Forty two medicines were administered as a monotherapy and 27 as combinations (supplement 2f). Treatment duration ranged from one day (single administration) to 42 days. Eighty (82%) of 98 trials administered medicines orally, and 168 (98%) of 172 medicines were administered within a standard or licensed dosing range (table 1). Two trials[69 70] reported

Treatment	Direct comparisons	P score	Mean difference (95% CI)	Mean difference (95% CI)	Confidence
Tolperisone	0	0.94		-26.1 (-34.0 to -18.2)	Low
Aceclofenac + tizanidine	0	0.93		-26.1 (-38.5 to -13.6)	Very low
Pregabalin	0	0.92		-24.7 (-34.6 to -14.7)	Low
Diclofenac + thiocolchicoside	1	0.86		-21.6 (-32.7 to -10.5)	Very low
Ketoprofen + thiocolchicoside	0	0.85		-21.0 (-30.9 to -11.0)	Low
Thiocolchicoside	3	0.85		-20.4 (-26.3 to -14.5)	Very low
Ketoprofen	0	0.81		-20.1 (-37.1 to -3.2)	Very low
Cyclobenzaprine	1	0.75		-16.7 (-29.4 to -3.9)	Very low
Aceclofenac + thiocolchicoside	0	0.70		-14.7 (-27.8 to -1.5)	Very low
Etoricoxib	0	0.67		-13.2 (-23.7 to -2.7)	Very low
Ketorolac	0	0.65		-12.8 (-23.6 to -2.0)	Very low
Carisoprodol	2	0.65		-12.4 (-20.3 to -4.6)	Very low
Dexketoprofen	0	0.62		-13.0 (-32.5 to 6.4)	Very low
Nimesulide	0	0.61		-11.2 (-19.9 to -2.3)	Very low
Aceclofenac	0	0.59		-10.8 (-19.4 to -2.2)	Very low
Chlorzoxazone	0	0.56		-11.3 (-36.0 to 13.4)	Very low
Diclofenac + tolperisone	0	0.54		-9.5 (-18.7 to -0.2)	Very low
Meloxicam	0	0.51		-8.6 (-16.1 to -1.0)	Very low
Piroxicam	1	0.49		-8.0 (-14.6 to -1.3)	Very low
Tramadol	0	0.46		-7.6 (-23.3 to 8.0)	Very low
Tizanidine	3	0.42		-6.3 (-13.6 to 0.9)	Very low
Paracetamol + tramadol SR	1	0.41		-6.2 (-14.2 to 1.7)	Very low
Ibuprofen + paracetamol	0	0.40		-5.9 (-17.6 to 5.8)	Very low
Aceclofenac + chlorzoxazone + paracetamol	0	0.39		-5.8 (-21.0 to 9.5)	Very low
Baclofen + ibuprofen	0	0.37		-5.1 (-17.6 to 7.5)	Very low
Etodolac + paracetamol	0	0.35		-4.5 (-18.2 to 9.2)	Very low
Paracetamol + tramadol	0	0.34		-4.5 (-15.7 to 6.6)	Very low
Diclofenac	2	0.33		-4.6 (-9.6 to 0.5)	Very low
Etodolac	0	0.32		-4.1 (-13.3 to 5.1)	Very low
Ibuprofen	2	0.31		-4.0 (-9.5 to 1.4)	Very low
Celecoxib	0	0.38		-2.6 (-13.9 to 8.6)	Very low
Ibuprofen + tizanidine	0	0.27		-2.5 (-13.6 to 8.6)	Very low
Paracetamol	1	0.26		-2.8 (-8.9 to 3.3)	Very low
Tenoxicam	1	0.26		-2.3 (-10.8 to 6.2)	Very low
Lornoxicam	0	0.23		-1.6 (-11.6 to 8.5)	Very low
Ibuprofen + metaxalone	0	0.19		0.3 (-12.5 to 13.1)	Very low
Methylprednisolone	1	0.18		2.0 (-14.0 to 18.0)	Very low
Prednisone	1	0.07		7.7 (-7.9 to 23.2)	Very low

-40 -30 -20 -10 0 10 20 30 40

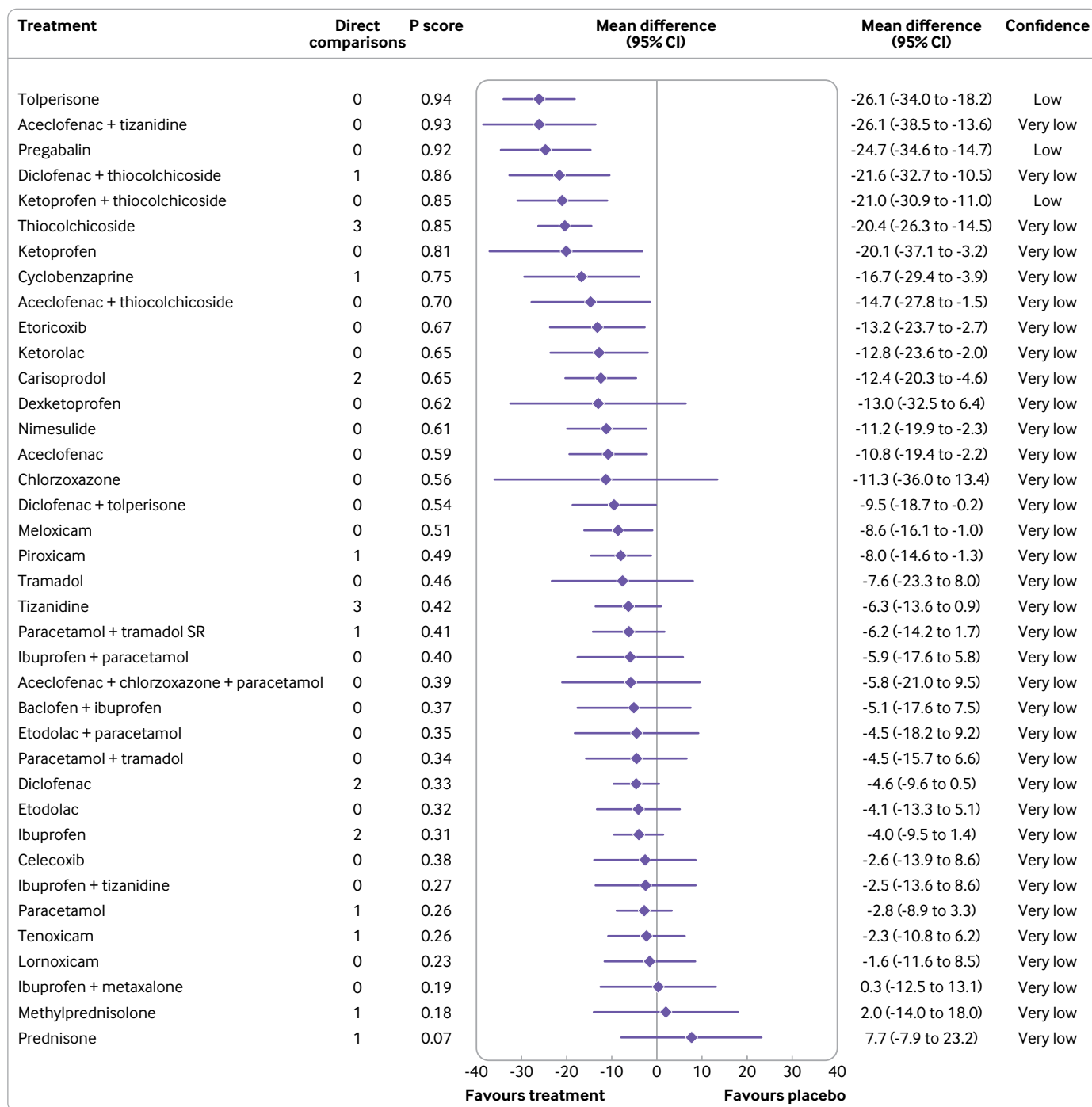Favours treatment   Favours placebo

Fig 4 | Forest plot for analgesic medicines and pain intensity. Medicines are ordered according to their P score ranking and compared with placebo. Point estimates refer to the mean difference. The bars indicate 95% confidence interval. Direct comparisons refer to the number of included studies comparing the intervention to placebo. Random effects model: τ^2=11.51. CI=confidence interval; SR=sustained release

two or more intervention groups within the same dosing range that we combined.

Assessment of transitivity and incoherence

A comprehensive assessment of transitivity was limited by the small number of trials per comparison (supplement 2j). During the evaluation of network diagnostics, four trials were identified that had methodological discrepancies inconsistent with the network (two based on incoherence within the

network and two based on heterogeneity within treatment comparisons) and were removed from all analyses (supplement 2k). We then re-evaluated the diagnostics for the updated models and agreed to proceed to interpreting treatment estimates. However, some comparisons show evidence of unexplained incoherence and, therefore, should be interpreted with caution. Important examples are the network meta-analysis effects for the outcome pain intensity for the comparisons ibuprofen versus placebo and
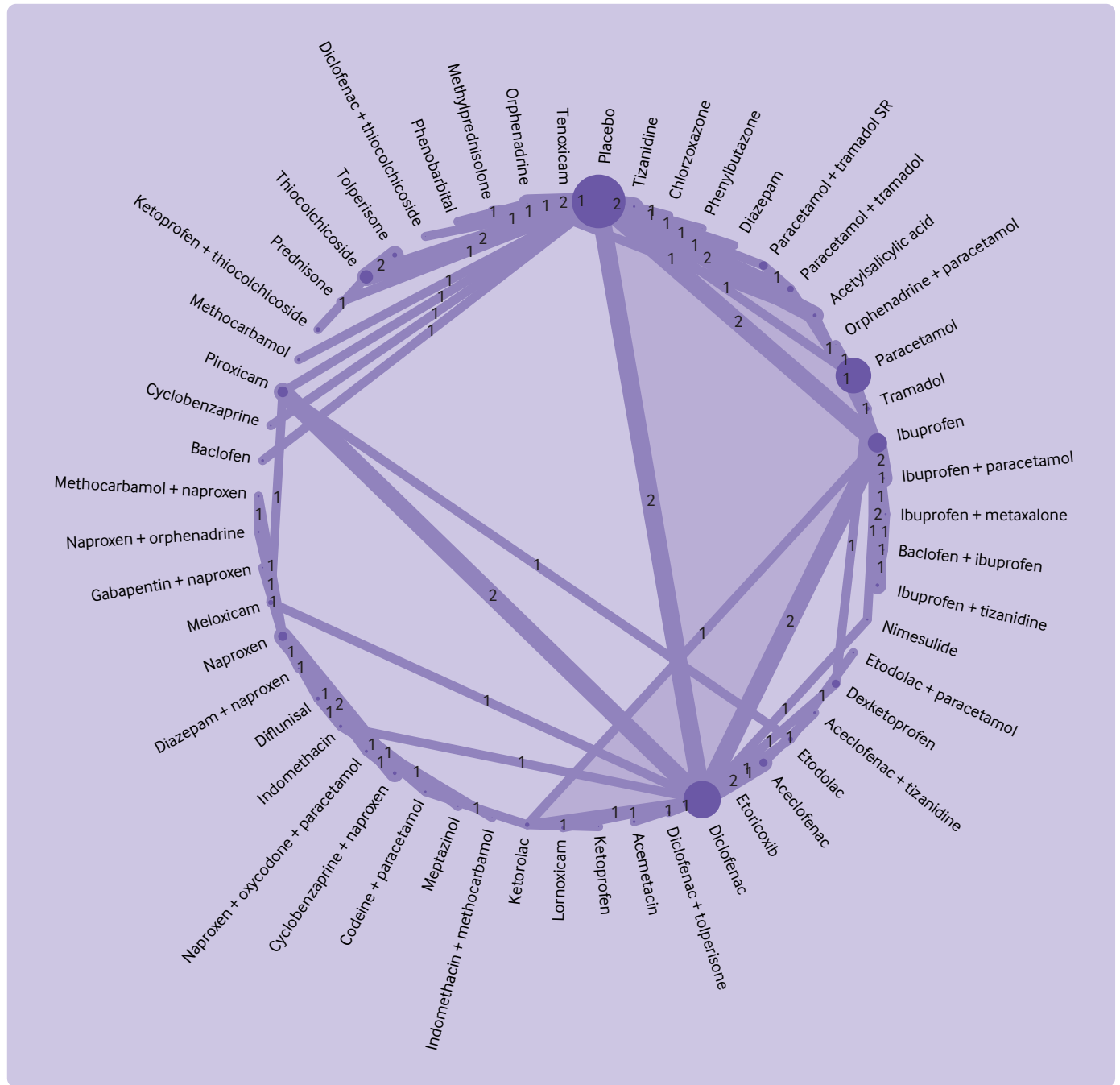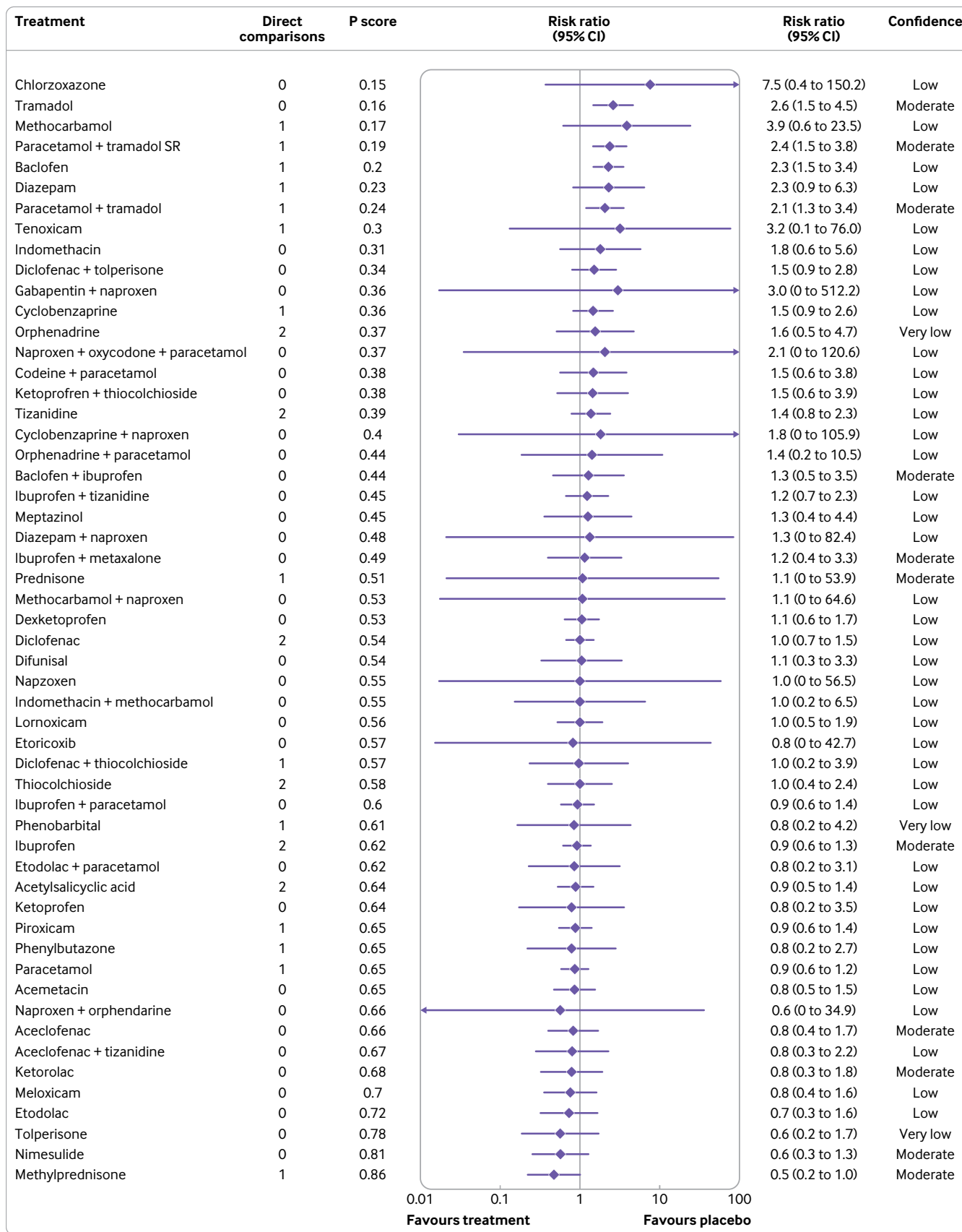
Fig 5 | Network plot for safety for medicines. The node size in proportional the sample size of each intervention. The line thickness is proportional the number of trials in the comparison (also indicated by the numbers). Light blue shading indicates trials with more than two arms. The two trials that did not connect to the network and were not included were hydrocodone plus ibuprofen versus oxycodone plus paracetamol (Palangio 2002; for full details of references see supplement 2); and etodolac plus thiocolchicoside versus etodolac plus tolperisone (Garg 2019)

paracetamol versus placebo, in which discrepancies between direct and indirect evidence resulted in a P<0.10 for the Separating Indirect from Direct Evidence approach. The full network diagnostics for the updated models are presented in supplement 3. Any remaining concerns about network heterogeneity and incoherence were addressed via downgrading confidence in estimates. A summary of confidence in effect estimates is provided in supplement 2l. Common reasons for downgrading confidence in

estimates were imprecision, heterogeneity, and risk of bias. League tables with estimates and confidence for all comparisons are provided in the supplement and spreadsheets are available on the Open Science Framework.

Primary analysis: nodes as medicines
Pain intensity
Pain intensity was measured in all 98 trials. Three trials measured pain intensity only during movement.

Treatment	Direct comparisons	P score	Risk ratio (95% CI)	Risk ratio (95% CI)	Confidence
Chlorzoxazone	0	0.15		7.5 (0.4 to 150.2)	Low
Tramadol	0	0.16		2.6 (1.5 to 4.5)	Moderate
Methocarbamol	1	0.17		3.9 (0.6 to 23.5)	Low
Paracetamol + tramadol SR	1	0.19		2.4 (1.5 to 3.8)	Moderate
Baclofen	1	0.2		2.3 (1.5 to 3.4)	Low
Diazepam	1	0.23		2.3 (0.9 to 6.3)	Low
Paracetamol + tramadol	1	0.24		2.1 (1.3 to 3.4)	Moderate
Tenoxicam	1	0.3		3.2 (0.1 to 76.0)	Low
Indomethacin	0	0.31		1.8 (0.6 to 5.6)	Low
Diclofenac + tolperisone	0	0.34		1.5 (0.9 to 2.8)	Low
Gabapentin + naproxen	0	0.36		3.0 (0 to 512.2)	Low
Cyclobenzaprine	1	0.36		1.5 (0.9 to 2.6)	Low
Orphenadrine	2	0.37		1.6 (0.5 to 4.7)	Very low
Naproxen + oxycodone + paracetamol	0	0.37		2.1 (0 to 120.6)	Low
Codeine + paracetamol	0	0.38		1.5 (0.6 to 3.8)	Low
Ketoprofen + thiocolchioside	0	0.38		1.5 (0.6 to 3.9)	Low
Tizanidine	2	0.39		1.4 (0.8 to 2.3)	Low
Cyclobenzaprine + naproxen	0	0.4		1.8 (0 to 105.9)	Low
Orphenadrine + paracetamol	0	0.44		1.4 (0.2 to 10.5)	Low
Baclofen + ibuprofen	0	0.44		1.3 (0.5 to 3.5)	Moderate
Ibuprofen + tizanidine	0	0.45		1.2 (0.7 to 2.3)	Low
Meptazinol	0	0.45		1.3 (0.4 to 4.4)	Low
Diazepam + naproxen	0	0.48		1.3 (0 to 82.4)	Low
Ibuprofen + metaxalone	0	0.49		1.2 (0.4 to 3.3)	Moderate
Prednisone	1	0.51		1.1 (0 to 53.9)	Moderate
Methocarbamol + naproxen	0	0.53		1.1 (0 to 64.6)	Low
Dexketoprofen	0	0.53		1.1 (0.6 to 1.7)	Low
Diclofenac	2	0.54		1.0 (0.7 to 1.5)	Low
Difunisal	0	0.54		1.1 (0.3 to 3.3)	Low
Napzoxen	0	0.55		1.0 (0 to 56.5)	Low
Indomethacin + methocarbamol	0	0.55		1.0 (0.2 to 6.5)	Low
Lornoxicam	0	0.56		1.0 (0.5 to 1.9)	Low
Etoricoxib	0	0.57		0.8 (0 to 42.7)	Low
Diclofenac + thiocolchioside	1	0.57		1.0 (0.2 to 3.9)	Low
Thiocolchioside	2	0.58		1.0 (0.4 to 2.4)	Low
Ibuprofen + paracetamol	0	0.6		0.9 (0.6 to 1.4)	Low
Phenobarbital	1	0.61		0.8 (0.2 to 4.2)	Very low
Ibuprofen	2	0.62		0.9 (0.6 to 1.3)	Moderate
Etodolac + paracetamol	0	0.62		0.8 (0.2 to 3.1)	Low
Acetylsalicyclic acid	2	0.64		0.9 (0.5 to 1.4)	Low
Ketoprofen	0	0.64		0.8 (0.2 to 3.5)	Low
Piroxicam	1	0.65		0.9 (0.6 to 1.4)	Low
Phenylbutazone	1	0.65		0.8 (0.2 to 2.7)	Low
Paracetamol	1	0.65		0.9 (0.6 to 1.2)	Low
Acemetacin	0	0.65		0.8 (0.5 to 1.5)	Low
Naproxen + orphendarine	0	0.66		0.6 (0 to 34.9)	Low
Aceclofenac	0	0.66		0.8 (0.4 to 1.7)	Moderate
Aceclofenac + tizanidine	0	0.67		0.8 (0.3 to 2.2)	Low
Ketorolac	0	0.68		0.8 (0.3 to 1.8)	Moderate
Meloxicam	0	0.7		0.8 (0.4 to 1.6)	Low
Etodolac	0	0.72		0.7 (0.3 to 1.6)	Low
Tolperisone	0	0.78		0.6 (0.2 to 1.7)	Very low
Nimesulide	0	0.81		0.6 (0.3 to 1.3)	Moderate
Methylprednisone	1	0.86		0.5 (0.2 to 1.0)	Moderate

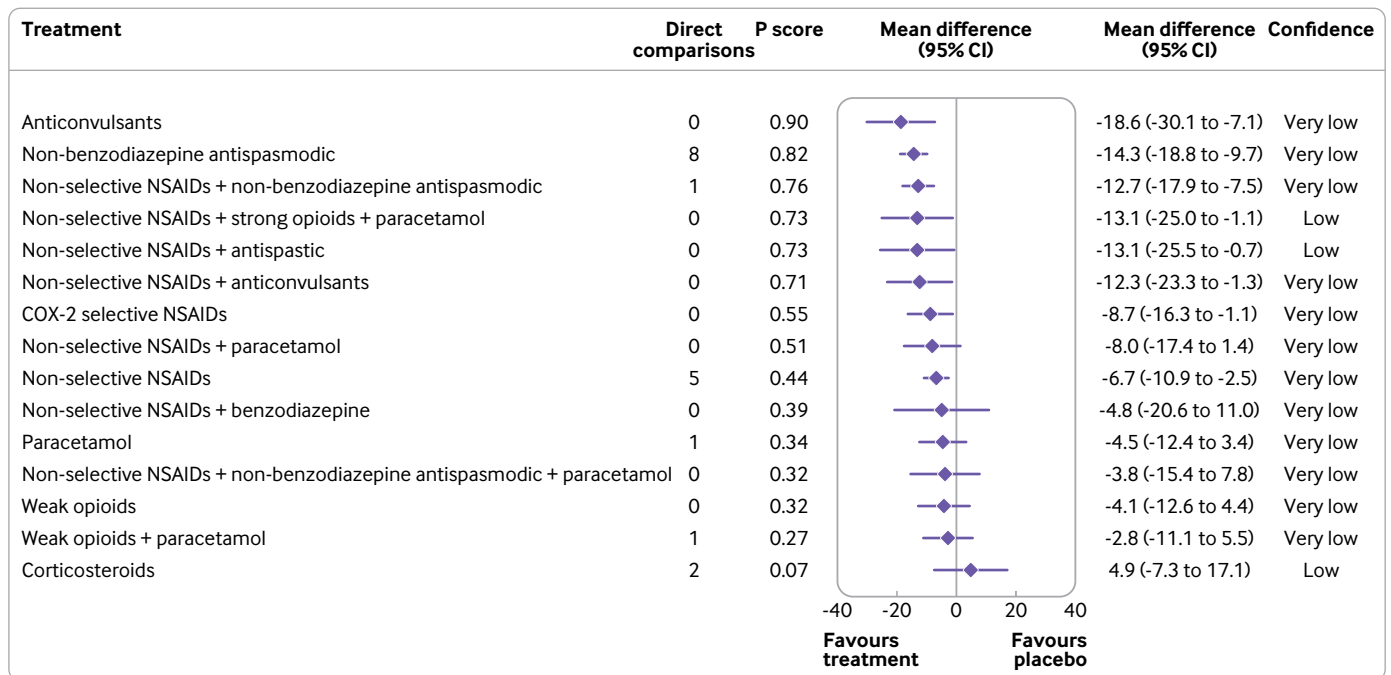0.01 0.1 1 10 100

Favours treatment   Favours placebo

Fig 6 | Forest plot for analgesic medicines with safety (any adverse event). Medicines are ordered according to their P score ranking and compared with placebo. Point estimates refer to the risk ratio. The bars indicate the 95% confidence interval. Direct comparisons refer to the number of included studies comparing the intervention to placebo. Random effects model: τ^2=0.015. CI=confidence interval; SR=sustained release

Pain intensity was measured with a 100 mm visual analogue scale (23 trials), a 10 cm visual analogue scale (16 trials), a 11 point numerical rating scale (seven trials), or another ordinal scale (24 trials). Data for pain intensity were analysed in 66 (67%) of 98 trials. Ten trials were at low risk of bias, 36 trials had some concerns, and 20 trials were at high risk of bias (supplement 2m). Endpoint data were reported in 50 trials and changes from baseline were reported in 16 trials. Fifteen trials (23%) required standard deviation imputation. Pain intensity data were transformed in 16 trials: 12 used count data, two used 95% confidence interval for group mean, one used median, one used range. The 66 trials did not form a connected network (fig 2, fig 3). The placebo network compared 39 interventions (38 medicines and the central node of placebo) in 54 trials (fig 2). Most comparisons consisted of a single trial, ranging from one to three, and had a limited number of closed loops. Direct evidence was available for 52 (7%) of 741 comparisons. The naproxen network compared 13 medicines in 10 trials (the central node was naproxen) with one trial per comparison (fig 3). Direct evidence was available for 14 (18%) of 78 comparisons. Two trials were not included in either network because these trials do not connect to any part of the network).

Data were of very low confidence in 648 (87%) of 741 comparisons and of low confidence in 93 (13%) of 741 of comparisons in the placebo network (supplement 2l). Tolperisone (mean difference −26.1 (95% confidence interval −34.0 to −18.2), low confidence), aceclofenac plus tizanidine (−26.1 (−38.5 to −13.6), very low confidence), and pregabalin (−24.7 (−34.6 to −14.7), low confidence) might be associated with the largest reductions in pain intensity compared with placebo (fig 4). Additionally, for statistically significant reductions, very low confidence was reported for large reductions (mean difference of >20 points) for four medicines, moderate reductions (>10-20 points) for seven medicines; and small reductions (5-10 points) for three medicines (fig 4). The estimates of comparative effectiveness and rankogram are in supplement 2n. No significant differences were noted between all medicines with large reductions in pain intensity compared with placebo, with data low or very low confidence. Similarly, low or very low confidence in evidence was reported for no significant differences between the medicines with large reductions in pain intensity and some medicines with moderate reduction in pain intensity compared with placebo. Some significant differences between medicines were noted; for example, low confidence data suggested that tolperisone is superior to carisoprodol at reducing pain intensity (mean difference −13.7 (−24.9 to −2.5)).

Confidence could not be evaluated for the naproxen network because of the small number of trials. Six medicines might be associated with a statistically significant reduction in pain intensity compared with naproxen (supplement 2o). The estimates of comparative effectiveness and rankogram are in supplement 2p. Sensitivity and post hoc analyses for pain intensity with nodes as medicines are reported in supplement 2q.

Treatment	Direct comparisons	P score	Mean difference (95% CI)	Mean difference (95% CI)	Confidence
Anticonvulsants	0	0.90		-18.6 (-30.1 to -7.1)	Very low
Non-benzodiazepine antispasmodic	8	0.82		-14.3 (-18.8 to -9.7)	Very low
Non-selective NSAIDs + non-benzodiazepine antispasmodic	1	0.76		-12.7 (-17.9 to -7.5)	Very low
Non-selective NSAIDs + strong opioids + paracetamol	0	0.73		-13.1 (-25.0 to -1.1)	Low
Non-selective NSAIDs + antispastic	0	0.73		-13.1 (-25.5 to -0.7)	Low
Non-selective NSAIDs + anticonvulsants	0	0.71		-12.3 (-23.3 to -1.3)	Very low
COX-2 selective NSAIDs	0	0.55		-8.7 (-16.3 to -1.1)	Very low
Non-selective NSAIDs + paracetamol	0	0.51		-8.0 (-17.4 to 1.4)	Very low
Non-selective NSAIDs	5	0.44		-6.7 (-10.9 to -2.5)	Very low
Non-selective NSAIDs + benzodiazepine	0	0.39		-4.8 (-20.6 to 11.0)	Very low
Paracetamol	1	0.34		-4.5 (-12.4 to 3.4)	Very low
Non-selective NSAIDs + non-benzodiazepine antispasmodic + paracetamol	0	0.32		-3.8 (-15.4 to 7.8)	Very low
Weak opioids	0	0.32		-4.1 (-12.6 to 4.4)	Very low
Weak opioids + paracetamol	1	0.27		-2.8 (-11.1 to 5.5)	Very low
Corticosteroids	2	0.07		4.9 (-7.3 to 17.1)	Low

-40 -20 0 20 40

Favours treatment   **Favours placebo**

Fig 7 | Forest plot for analgesic medicine classes with pain intensity. Medicine classes are ordered according to their P score ranking and compared with placebo. Point estimates refer to the mean difference. The bars indicate 95% confidence interval. Direct comparisons refer to the number of included studies comparing the intervention to placebo. Random effects model: τ^2=23.93. CI=confidence interval; COX-2=cyclooxygenase 2; NSAIDs=non-steroidal anti-inflammatory drugs

Safety

Ninety two trials reported measuring safety, but only 68 trials (74%) were analysed for the number of participants who reported an adverse event. The primary reasons for data unavailability were reports of only numbers of adverse events, rather than number of participants, or no data for the subset of participants with acute non-specific low back pain. Nine trials were at low risk of bias, 41 trials had some concerns, and 18 trials were at high risk of bias (supplement 2r). One network compared 55 interventions (54 medicines and placebo) in 66 trials (fig 5), and two trials did not connect to the network. All comparisons in the network consisted of a one or two trials and the number of closed loops was small. Direct evidence was available for 70 (4.7%) of 1485 comparisons. Effect estimates were analysed as risk ratios.

Comparisons were of very low confidence in 34 (2%) of 1485, low confidence in 1274 (86%) of 1485, moderate confidence in 168 (11%) of 1485, and high confidence in nine (1%) of 1485 (supplement 2l). Tramadol (risk ratio 2.6 (95% confidence interval 1.5 to 4.5), moderate confidence), paracetamol plus sustained release tramadol (2.4 (1.5 to 3.8), moderate confidence), baclofen (2.3 (1.5 to 3.4), low confidence), and paracetamol plus tramadol (2.1 (1.3 to 3.4), moderate confidence) might be associated with increased adverse events during treatment compared with placebo (fig 6). The estimates of comparative effectiveness and rankogram are provided in supplement 2s. Data had high to very low confidence that these four medicines were also more likely to increase adverse events compared with other medicines. For example, moderate confidence data suggested that tolperisone was associated with fewer adverse events than tramadol (0.2 (0.1 to 0.7)) and high confidence data suggested that paracetamol was associated with fewer adverse events than paracetamol plus sustained release tramadol (0.4 (0.2 to 0.6)).

The quality of adverse event measurement and reporting varied across trials. Generally, trials did not distinguish between an adverse event (an untoward medical occurrence) and an adverse effect (an untoward medical occurrence judged as related to treatment). Brief descriptions of adverse events reported in each trial are available in supplement 2h. Most commonly reported adverse events were related to the gastrointestinal system (nausea, dyspepsia, vomiting, diarrhoea) and the nervous system (drowsiness, dizziness, headache). Sensitivity and post hoc analyses for safety with nodes as medicines are reported in supplement 2t.

Secondary analysis: medicine classes

The secondary analysis of medicine classes included 65 trials (n=1107 participants; 33 trials that only compared medicines within the same class, primarily non-selective non-steroidal anti-inflammatory drugs, were excluded). A list of the 22 interventions (21 different classes or combinations, and placebo) is available in supplement 2u.

Pain intensity

Pain intensity was analysed in 45 trials. One network compared 16 interventions (15 classes and placebo) in 44 trials, and one trial did not connect to the network (supplement 2v). Direct evidence was available for 22 (18%) of 120 comparisons. Of the 120 comparisons, evidence was of very low confidence in 114 (95%), low confidence in five (4%), and moderate confidence in one (1%) (supplement 2l).

Anticonvulsants (mean difference −18.6 (95% confidence interval −30.1 to −7.1), very low confidence), non-benzodiazepine antispasmodic (−14.3 (−18.8 to −9.7), very low confidence), non-selective non-steroidal anti-inflammatory drugs plus non-benzodiazepine antispasmodic (−12.7 (−17.9 to −7.5), very low confidence), non-selective non-steroidal anti-inflammatory drugs plus strong opioids plus paracetamol (−13.1 (−25.0 to −1.1), low confidence), non-selective non-steroidal anti-inflammatory drugs plus antispastic (−13.1 (−25.5 to −0.7), low confidence), non-selective non-steroidal anti-inflammatory drugs plus anticonvulsants (−12.3 (−23.3 to −1.3), very low confidence) might be associated with the moderate reductions in pain intensity compared with placebo (fig 7). The estimates of comparative effectiveness and rankogram are in supplement 2w. Very low confidence was shown for no statistically significant differences between any of the medicine classes that reduced pain intensity compared with placebo. Some differences between classes were noted; for example, evidence showed very low confidence that anticonvulsants were superior to weak opioids for reducing pain intensity (−14.5 (−28.7 to −0.4)). Sensitivity and post hoc analyses for pain intensity with nodes as medicine classes are reported in supplement 2x.

Safety

Safety was analysed in 46 trials. One network compared 19 interventions (18 classes and placebo) in 45 trials, and one trial did not connect to the network (supplement 2y). Direct evidence was available for 27 (16%) of 171 comparisons. Of 171 comparisons, seven (4%) were of very low confidence, 109 (64%) were of low confidence, 50 (29%) were of moderate confidence, and five (3%) were of high confidence (supplement 2l).

Compared with placebo, increased adverse events during treatment might be associated with antispastic drugs (risk ratio 2.3 (95% confidence interval 1.4 to 3.8), low confidence), weak opioids (1.9 (1.3 to 2.9), moderate confidence), non-selective non-steroidal anti-inflammatory drugs plus strong opioids plus paracetamol (1.9 (1.1 to 3.2), high confidence), weak opioids plus paracetamol (1.9 (1.3 to 2.7), moderate confidence), and non-selective non-steroidal anti-inflammatory drugs plus non-benzodiazepine antispasmodic (1.5 (1.1 to 2.1), moderate confidence) (supplement 2z). The estimates of comparative effectiveness and the rankogram are in supplement 2aa. Findings were of high to very low confidence that

these classes were also more likely to increase adverse events compared with the other classes. For example, non-selective non-steroidal anti-inflammatory drugs plus strong opioids plus paracetamol were of high confidence and was associated with more adverse events than paracetamol (2.2 (1.2 to 4.4)). Sensitivity and post hoc analyses for safety with nodes as medicine classes are reported in supplement 2ab.

Secondary outcomes
We also analysed secondary outcomes with nodes as medicines and nodes as medicine classes. We did not perform sensitivity and post hoc analyses for the secondary outcomes. Results for function are reported in supplement 2ac (nodes as medicines) and supplement 2ad (nodes as medicine classes). Results for acceptability are reported in supplement 2ae (nodes as medicines) and supplement 2af (nodes as medicine classes). Results for harm are reported in supplement 2ag.

Discussion
Our review of analgesic medicines for acute non-specific low back pain found considerable uncertainty around effects for pain intensity and safety. The findings were of low or very low confidence that several medicines might be associated with large reductions in pain intensity compared with placebo, and some medicines might be more effective than other medicines. Several other medicines might be associated with an increased risk of adverse events compared with placebo, as well as compared with other medicines. In the secondary analysis of medicine classes, low or very low confidence evidence showed that seven classes might be associated with small to moderate reductions in pain intensity compared with placebo, with no statistically significant differences between these classes. However, low confidence showed that two of these classes increased the risk of adverse events compared with placebo.

Implications for clinicians and policy makers
Judgements of low or very low confidence in this review warrant caution for the clinical interpretation of these effects, which might change markedly with future research. Most effects were derived solely from indirect evidence and the findings were not robust to sensitivity analyses, with many effects becoming non-significant after the removal of trials based on different methodological considerations (eg, risk of bias). Similar findings of moderate to large effects for pain intensity but low confidence have been reported for several non-pharmacological interventions used for acute non-specific low back pain: superficial heat, massage, manual therapy, and acupuncture.[2] Similar levels of uncertainty were identified in a network meta-analysis published in 2022 of 46 randomised controlled trials that compared pharmacological and non-pharmacological interventions for acute and subacute low back pain.[20]

Clinical practice guidelines recommend non-pharmacological treatments in first line and second line care for acute non-specific low back pain.[3] Given the favourable natural history for most patients,[71] we believe that clinicians and patients should take a cautious approach to the use of analgesic medicines. Similarly, policy makers should recommend a cautious approach when considering analgesic medicines, prioritising the minimisation of harm. Another consideration for clinicians and guideline developers is the legal availability of medicines. We included medicines licensed across the UK, Australia, USA, and Europe, which might not include medicines licensed in other countries and does not imply that the same medicines are available everywhere. Our estimates of comparative effectiveness suggest no differences between several medicines that were superior to placebo, meaning clinicians can incorporate our findings, a medicine's availability, clinical expertise, and patient preferences when choosing an analgesic medicine.

Strengths and limitations
We believe that this review is the most comprehensive in the field. We preregistered and published the protocol and made our updates transparent. Our comprehensive search included published and unpublished literature in any language. Our rigorous method ensured as much data as possible were included, with scrutiny by an expert team. We closely examined network diagnostics to explore network heterogeneity, inconsistency, and incoherence (steps that are not often adequately undertaken)[72] and we attempted to resolve these issues when they arose. However, this study has limitations. Firstly, we aspired to select a sample reflective of acute non-specific low back pain, but patients might differ across clinical settings.[32] Secondly, most included studies had concerns related to risk of bias. Thirdly, data were missing and imputation was required for continuous outcomes, despite attempts to contact authors. Fourthly, no network meta-analysis methods can account for the uncertainty of variance estimates (analogous to the Hartung-Knapp approach for pairwise meta-analysis) and we were unable to thoroughly explore the influence of potential effect modifiers (eg, treatment duration, route of administration) because of the limited data and poor network structure. Finally, adverse event data in some trials were reported in a way that made them unable to be included in this study. In future trials, we encourage investigators to report the number of participants who had any adverse event, as well as type and severity of those adverse events.

Future research
The evidence base includes many different analgesic medicines or combinations, mostly compared to placebo. Relatively few randomised controlled trials evaluate comparative effectiveness. The structure of this information is not yet optimal to inform clinical decision making and the potential for network meta-analysis to contribute improved estimates of effects was under-realised. Most estimates were derived solely from indirect evidence, a key contributor to the low or

very low confidence. Confidence was not substantially improved in the secondary analysis.

Other aspects of trial conduct might be improved in future work. Key limitations were moderate to high risk of bias and missing data, which have established influences on effect estimates.[51] Analgesic medicines with larger effect sizes came from trials with lower methodological quality. Similarly, wide confidence intervals often arose from smaller studies. This uncertainty is propagated when networks make many comparisons via indirect evidence only. Concerns exist about research integrity and the large decrease in pain intensity from pregabalin was no longer apparent in sensitivity analyses. Synthesis of the trials at low risk of bias was not possible with conventional methods for network meta-analysis because they did not form a connected network. Our review, together with established methods for future trial design,[73 74] might be an important guiding contribution to further research. We identified 10 ongoing trials that could contribute additional data to future updates of this study, described briefly in supplement 2i. No further reviews are needed until high quality randomised controlled trials are published.

Conclusion

Despite nearly 60 years of research involving more than 15 000 patients, high quality evidence to guide clinical decisions on analgesic medicines for acute non-specific low back pain remains limited. Similarly, evidence from the secondary analysis of medicine classes had low confidence. Clinicians and patients are advised to take a cautious approach to the use of analgesic medicines. No further reviews are needed until high quality studies are published.

AUTHOR AFFILIATIONS

[1]School of Health Sciences, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia

[2]Centre for Pain IMPACT, Neuroscience Research Australia, Sydney, NSW, Australia

[3]Curtin Health Innovation Research Institute, Faculty of Health Sciences, Curtin University, Perth, WA, Australia

[4]Perron Institute for Neurological and Translational Science, Perth, WA, Australia

[5]IIMPACT in Health, University of South Australia, Adelaide, SA, Australia

[6]Sydney Pharmacy School, Faculty of Medicine and Health, University of Sydney, Gadigal Country, Sydney, NSW, Australia

[7]Sydney Musculoskeletal Health, University of Sydney, Gadigal Country, Sydney, NSW, Australia

[8]Institute for Musculoskeletal Health, Sydney Local Health District, Sydney, NSW, Australia

[9]Clinical Pharmacology and Toxicology, St Vincent's Hospital, Sydney, NSW, Australia

[10]St Vincent's Clinical School, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia

[11]Faculty of Medicine, Nursing and Midwifery and Health Sciences, University of Notre Dame Australia, Fremantle, WA, Australia

[12]Department of Health Sciences, Centre for Health and Wellbeing Across the Lifecourse, Brunel University London, Uxbridge, UK

[13]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Centre, University of Freiburg, Freiburg, Germany

[14]School of Physical Therapy, University of Western Ontario, London, ON, Canada

[15]The Gray Centre for Mobility and Activity, Parkwood Institute, London, ON, Canada

[16]NeuroRecovery Research Hub, School of Psychology, University of New South Wales, Sydney, NSW, Australia

1 Furlan AD, Malmivaara A, Chou R, et al, Editorial Board of the Cochrane Back, Neck Group. 2015 Updated method guideline for systematic reviews in the Cochrane Back and Neck Group. *Spine (Phila Pa 1976)* 2015;40:1660-73. doi:10.1097/BRS.0000000000001061

2 Maher C, Underwood M, Buchbinder R. Non-specific low back pain. *Lancet* 2017;389:736-47. doi:10.1016/S0140-6736(16)30970-9

3 Almeida M, Saragiotto B, Richards B, Maher CG. Primary care management of non-specific low back pain: key messages from recent clinical guidelines. *Med J Aust* 2018;208:272-5. doi:10.5694/mja17.01152

4 Oliveira CB, Maher CG, Pinto RZ, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. *Eur Spine J* 2018;27:2791-803. doi:10.1007/s00586-018-5673-2

5 Qaseem A, Wilt TJ, McLean RM, et al, Clinical Guidelines Committee of the American College of Physicians. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. *Ann Intern Med* 2017;166:514-30. doi:10.7326/M16-2367

6 National Institute for Health and Care Excellence. Guidelines. Low Back Pain and Sciatica in Over 16s: Assessment and Management. London: National Institute for Health and Care Excellence (NICE); 2016.

7 Ahern M, Dean CM, Dear BF, Willcock SM, Hush JM. The experiences and needs of people seeking primary care for low-back pain in Australia. *Pain Rep* 2019;4:e756. doi:10.1097/PR9.0000000000000756

8 Ndlovu M, Bedson J, Jones PW, Jordan KP. Pain medication management of musculoskeletal conditions at first presentation in primary care: analysis of routinely collected medical record data. *BMC Musculoskelet Disord* 2014;15:418. doi:10.1186/1471-2474-15-418

9 Abdel Shaheed C, Maher CG, Williams KA, Day R, McLachlan AJ. Efficacy, Tolerability, and Dose-Dependent Effects of Opioid Analgesics for Low Back Pain: A Systematic Review and Meta-analysis. *JAMA Intern Med* 2016;176:958-68. doi:10.1001/jamainternmed.2016.1251

10 Abdel Shaheed C, Maher CG, Williams KA, McLachlan AJ. Efficacy and tolerability of muscle relaxants for low back pain: Systematic review and meta-analysis. *Eur J Pain* 2017;21:228-37. doi:10.1002/ejp.907

11 Chou R, Pinto RZ, Fu R, et al. Systemic corticosteroids for radicular and non-radicular low back pain. *Cochrane Database Syst Rev* 2022;10:CD012450.

12 Enke O, New HA, New CH, et al. Anticonvulsants in the treatment of low back pain and lumbar radicular pain: a systematic review and meta-analysis. *CMAJ* 2018;190:E786-93. doi:10.1503/cmaj.171333

13 Machado GC, Maher CG, Ferreira PH, Day RO, Pinheiro MB, Ferreira ML. Non-steroidal anti-inflammatory drugs for spinal pain: a systematic review and meta-analysis. *Ann Rheum Dis* 2017;76:1269-78. doi:10.1136/annrheumdis-2016-210597

14 Machado GC, Maher CG, Ferreira PH, et al. Efficacy and safety of paracetamol for spinal pain and osteoarthritis: systematic review and meta-analysis of randomised placebo controlled trials. *BMJ* 2015;350:h1225. doi:10.1136/bmj.h1225

15 Saragiotto BT, Machado GC, Ferreira ML, Pinheiro MB, Abdel Shaheed C, Maher CG. Paracetamol for low back pain. *Cochrane Database Syst Rev* 2016;2016:CD012230.

16 van der Gaag WH, Roelofs PD, Enthoven WTM, van Tulder MW, Koes BW. Non-steroidal anti-inflammatory drugs for acute low back pain. *Cochrane Database Syst Rev* 2020;4:CD013581. doi:10.1002/14651858.CD013581

17 Cashin AG, Folly T, Bagg MK, et al. Efficacy, acceptability, and safety of muscle relaxants for adults with non-specific low back pain: systematic review and meta-analysis. *BMJ* 2021;374:n1446. doi:10.1136/bmj.n1446

18 Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods* 2012;3:80-97. doi:10.1002/jrsm.1037

19 Cipriani A, Higgins JP, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013;159:130-7. doi:10.7326/0003-4819-159-2-201307160-00008

20 Gianola S, Bargeri S, Del Castillo G, et al. Effectiveness of treatments for acute and subacute mechanical non-specific low back pain: a systematic review with network meta-analysis. *Br J Sports Med* 2022;56:41-50. doi:10.1136/bjsports-2020-103596

21 Del Fiol G, Workman TE, Gorman PN. Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA Intern Med* 2014;174:710-8. doi:10.1001/jamainternmed.2014.368

22 Lim YZ, Chou L, Au RTM, et al. People with low back pain want clear, consistent and personalised information on prognosis, treatment options and self-management strategies: a systematic review. *J Physiother* 2019;65:124-35. doi:10.1016/j.jphys.2019.05.010

23 Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med* 2015;162:777-84. doi:10.7326/M14-2385

24 Wewege MA, Bagg MK, Jones MD, McAuley JH, ANiMALIA investigators. Analgesic medicines for adults with low back pain: protocol for a systematic review and network meta-analysis. *Syst Rev* 2020;9:255. doi:10.1186/s13643-020-01506-3

25 Furlan A, Chaparro LE, Irvin E, Mailis-Gagnon A. A comparison between enriched and nonenriched enrollment randomized withdrawal trials of opioids for chronic noncancer pain. *Pain Res Manag* 2011;16:337-51. doi:10.1155/2011/465281

26 Yamato TP, Maher CG, Saragiotto BT, et al. Comparison of effect sizes between enriched and nonenriched trials of analgesics for chronic musculoskeletal pain: a systematic review. *Br J Clin Pharmacol* 2017;83:2347-55. doi:10.1111/bcp.13350

27 World Health Organization Collaborating Centre for Drug Statistics Methodology. ATC/DDD Index 2022 https://www.whocc.no/atc_ddd_index/.

28 US Food and Drug Administration. Drugs@FDA: FDA-Approved Drugs 2022. https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm.

29 Medicines and Healthcare products Regulatory Agency. Find product information about medicines 2022 https://www.gov.uk/guidance/find-product-information-about-medicines.

30 European Medicines Agency. https://www.ema.europa.eu/en/medicines. *Medicines (Basel)* 2022.

31 Therapeutic Goods Administration. Australian Register of Therapeutics Goods 2022 https://www.tga.gov.au/australian-register-therapeutic-goods.

32 Oliveira CB, Hamilton M, Traeger A, et al. Do patients with acute low back pain in emergency departments have more severe symptoms than those in general practice? A systematic review with meta-analysis. *Pain Med* 2021:pnab260.

33 Oliveira CB, Amorim HE, Coombs DM, et al. Emergency department interventions for adult patients with low back pain: a systematic review of randomised controlled trials. *Emerg Med J* 2021;38:59-68. doi:10.1136/emermed-2020-209588

34 Veritas Health Innovation. Covidence systematic review software. Melbourne, Australia: Veritas Health Innovation.

35 Rizzo RRN, Ferraro MC, Wewege MA, et al, Targeting neurotrophic factors for low back pain and sciatica: a systematic review and meta-analysis. *Rheumatology* 2021:keab785.

36 Bagg MK, McLachlan AJ, Maher CG, et al. Paracetamol, NSAIDS and opioid analgesics for chronic low back pain: a network meta analysis. *Cochrane Database Syst Rev* 2018;6:CD013045. doi:10.1002/14651858.CD013045.

37 Ferraro MC, Bagg MK, Wewege MA, et al. Efficacy, acceptability, and safety of antidepressants for low back pain: a systematic review and meta-analysis. *Syst Rev* 2021;10:62. doi:10.1186/s13643-021-01599-4

38 Bagg MK, O'Hagan E, Zahara P, et al. Systematic reviews that include only published data may overestimate the effectiveness of analgesic medicines for low back pain: a systematic review and meta-analysis. *J Clin Epidemiol* 2020;124:149-59. doi:10.1016/j.jclinepi.2019.12.006

39 US Food and Drug Administration. Code of Federal Regulations Title 21. 2018.

40 Tomlinson A, Efthimiou O, Boaden K, et al. Side effect profile and comparative tolerability of 21 antidepressants in the acute treatment of major depression in adults: protocol for a network meta-analysis. *Evid Based Ment Health* 2019;22:61-6. doi:10.1136/ebmental-2019-300087

41 Furukawa TA, Salanti G, Atkinson LZ, et al. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. *BMJ Open* 2016;6:e010919. doi:10.1136/bmjopen-2015-010919

42 Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271-3.

43 Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976)* 1983;8:141-4. doi:10.1097/00007632-198303000-00004

44 Wewege MA, Jones MD, Williams SA, Kamper SJ, McAuley JH. Rescaling pain intensity measures for meta-analyses of analgesic medicines for low back pain appears justified: an empirical examination from randomised trials. *BMC Med Res Methodol* 2022;22:285. doi:10.1186/s12874-022-01763-x

45 Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/ or interquartile range. *BMC Med Res Methodol* 2014;14:135. doi:10.1186/1471-2288-14-135

46 Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3 (updated 2022)*. Cochrane, 2011.

47 Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi:10.1136/bmj.l4898

48 McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. *Res Synth Methods* 2021;12:55-61. doi:10.1002/jrsm.1411

49 Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med* 2013;11:159. doi:10.1186/1741-7015-11-159

50 Jansen JP, Schmid CH, Salanti G. Directed acyclic graphs can help understand bias in indirect and mixed treatment comparisons. *J Clin Epidemiol* 2012;65:798-807. doi:10.1016/j.jclinepi.2012.01.002

51 Dechartres A, Trinquart L, Faber T, Ravaud P. Empirical evaluation of which trial characteristics are associated with treatment effect estimates. *J Clin Epidemiol* 2016;77:24-37. doi:10.1016/j.jclinepi.2016.04.005

52 Prescriber's Digital Reference. 2019. https://www.pdr.net/.

53 MIMS. 2019. https://www.mims.co.uk/2019.

54 Australian Medicines Handbook. 2019. https://amhonline.amh.net.au/.

55 Chou R, Deyo R, Friedly J, et al. Systemic pharmacologic therapies for low back pain: a systematic review for an American College of Physicians clinical practice guideline. *Ann Intern Med* 2017;166:480-92. doi:10.7326/M16-2458

56 Ostelo RW, Deyo RA, Stratford P, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. *Spine (Phila Pa 1976)* 2008;33:90-4. doi:10.1097/BRS.0b013e31815e3a10

57 Nikolakopoulou A, Higgins JPT, Papakonstantinou T, et al. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. *PLoS Med* 2020;17:e1003082. doi:10.1371/journal.pmed.1003082

58 Papakonstantinou T, Nikolakopoulou A, Higgins J, Egger M, Salanti G. CINeMA: Software for semiautomated assessment of the confidence in the results of network meta analysis. *Campbell Syst Rev* 2020;16:e1080. doi:10.1002/cl2.1080.

59 Rücker G, Krahn U, König J, et al. netmeta: Network Meta-Analysis using Frequentist Methods. R package version 2.1-0. https://CRAN.R-project.org/package=netmeta2022.

60 Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* 2012;3:312-24. doi:10.1002/jrsm.1058

61 Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;15:2733-49. doi:10.1002/(SICI)1097-0258(19961230)15:24<2733::AID-SIM562>3.0.CO;2-0

62 Jackson D, White IR, Riley RD. A matrix-based method of moments for fitting the multivariate random effects model for meta-analysis and meta-regression. *Biom J* 2013;55:231-45. doi:10.1002/bimj.201200152

63 Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol* 2015;15:58. doi:10.1186/s12874-015-0060-8

64 Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* 2012;3:98-110. doi:10.1002/jrsm.1044

65 White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Synth Methods* 2012;3:111-25. doi:10.1002/jrsm.1045

66 Krahn U, Binder H, König J. A graphical tool for locating inconsistency in network meta-analyses. *BMC Med Res Methodol* 2013;13:35. doi:10.1186/1471-2288-13-35

67 Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010;29:932-44. doi:10.1002/sim.3767

68 Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Res Synth Methods* 2012;3:161-76. doi:10.1002/jrsm.57

69 Serfer GT, Wheeler WJ, Sacks HJ. Randomized, double-blind trial of carisoprodol 250 mg compared with placebo and carisoprodol 350 mg for the treatment of low back spasm. *Curr Med Res Opin* 2010;26:91-9. doi:10.1185/03007990903382428

70 Williams CM, Maher CG, Latimer J, et al. Efficacy of paracetamol for acute low-back pain: a double-blind, randomised controlled trial. *Lancet* 2014;384:1586-96. doi:10.1016/S0140-6736(14)60805-9

71 da C Menezes Costa L,, Maher CG, Hancock MJ, McAuley JH, Herbert RD,Costa LOP. The prognosis of acute and persistent low-back pain: a meta-analysis. *CMAJ* 2012;184:E613-24. doi:10.1503/cmaj.111271.

72 Veroniki AA, Tsokani S, White IR, et al. Prevalence of evidence of inconsistency and its association with network structural characteristics in 201 published networks of interventions. *BMC Med Res Methodol* 2021;21:224. doi:10.1186/s12874-021-01401-y

73 Nikolakopoulou A, Mavridis D, Salanti G. Using conditional power of network meta-analysis (NMA) to inform the design of future clinical trials. *Biom J* 2014;56:973-90. doi:10.1002/bimj.201300216

74 Salanti G, Nikolakopoulou A, Sutton AJ, et al. Planning a future randomized clinical trial based on a network of relevant past trials. *Trials* 2018;19:365. doi:10.1186/s13063-018-2740-2 Web appendix: Published protocol

Web appendix: Protocol updates, method, results

Web appendix: Network diagnostics