



Efficacy and safety of paracetamol for spinal pain and osteoarthritis: systematic review and meta-analysis of randomised placebo controlled trials

Gustavo C Machado,¹ Chris G Maher,¹ Paulo H Ferreira,² Marina B Pinheiro,² Chung-Wei Christine Lin,¹ Richard O Day,^{3,4} Andrew J McLachlan,^{5,6} Manuela L Ferreira^{1,7}

¹The George Institute for Global Health, Sydney Medical School, University of Sydney, Sydney, NSW 2000, Australia

²Faculty of Health Sciences, University of Sydney, Sydney, NSW 2141, Australia

³Department of Clinical Pharmacology, St Vincent's Hospital and University of New South Wales, Sydney, NSW 2010, Australia

⁴School of Medical Sciences, Department of Medicine, University of New South Wales, Sydney, NSW 2033, Australia

⁵Faculty of Pharmacy, University of Sydney, Sydney, NSW 2050, Australia

⁶Centre for Education and Research on Ageing, Concord Hospital, Sydney, NSW 2139, Australia

⁷Institute of Bone and Joint Research, The Kolling Institute, Sydney Medical School, University of Sydney, Sydney, NSW 2065, Australia

Correspondence to: G C Machado gmachado@georgeinstitute.org.au

Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmj.h1225>)

Cite this as: *BMJ* 2014;350:h1225 doi: 10.1136/bmj.h1225

Accepted: 04 February 2015

ABSTRACT OBJECTIVE

To investigate the efficacy and safety of paracetamol (acetaminophen) in the management of spinal pain and osteoarthritis of the hip or knee.

DESIGN

Systematic review and meta-analysis.

DATA SOURCES

Medline, Embase, AMED, CINAHL, Web of Science, LILACS, International Pharmaceutical Abstracts, and Cochrane Central Register of Controlled Trials from inception to December 2014.

ELIGIBILITY CRITERIA FOR SELECTING STUDIES

Randomised controlled trials comparing the efficacy and safety of paracetamol with placebo for spinal pain (neck or low back pain) and osteoarthritis of the hip or knee.

DATA EXTRACTION

Two independent reviewers extracted data on pain, disability, and quality of life. Secondary outcomes were adverse effects, patient adherence, and use of rescue medication. Pain and disability scores were converted to a scale of 0 (no pain or disability) to 100 (worst possible pain or disability). We calculated weighted mean differences or risk ratios and 95% confidence intervals using a random effects model. The Cochrane Collaboration's tool was used for assessing risk of bias, and the GRADE approach was used to evaluate the quality of evidence and summarise conclusions.

RESULTS

12 reports (13 randomised trials) were included. There was "high quality" evidence that paracetamol is

ineffective for reducing pain intensity (weighted mean difference -0.5 , 95% confidence interval -2.9 to 1.9) and disability (0.4 , -1.7 to 2.5) or improving quality of life (0.4 , -0.9 to 1.7) in the short term in people with low back pain. For hip or knee osteoarthritis there was "high quality" evidence that paracetamol provides a significant, although not clinically important, effect on pain (-3.7 , -5.5 to -1.9) and disability (-2.9 , -4.9 to -0.9) in the short term. The number of patients reporting any adverse event (risk ratio 1.0 , 95% confidence interval 0.9 to 1.1), any serious adverse event (1.2 , 0.7 to 2.1), or withdrawn from the study because of adverse events (1.2 , 0.9 to 1.5) was similar in the paracetamol and placebo groups. Patient adherence to treatment (1.0 , 0.9 to 1.1) and use of rescue medication (0.7 , 0.4 to 1.3) was also similar between groups. "High quality" evidence showed that patients taking paracetamol are nearly four times more likely to have abnormal results on liver function tests (3.8 , 1.9 to 7.4), but the clinical importance of this effect is uncertain.

CONCLUSIONS

Paracetamol is ineffective in the treatment of low back pain and provides minimal short term benefit for people with osteoarthritis. These results support the reconsideration of recommendations to use paracetamol for patients with low back pain and osteoarthritis of the hip or knee in clinical practice guidelines.

SYSTEMATIC REVIEW REGISTRATION

PROSPERO registration number CRD42013006367.

Introduction

Low back and neck pain (spinal pain) are leading causes of disability worldwide, and osteoarthritis of the hip or knee is the 11th highest contributor to global disability, when disability is measured by years lived with disability.¹ The point prevalence of spinal pain is 9.4%, and osteoarthritis affects nearly 4% of the global population.²⁻⁴ The increasing healthcare expenditure for these conditions is mostly attributed to the increasing cost of prescription medicines, accounting for about 20% of the total cost.⁵

Prescription of drugs is the most common approach to treatment used by general practitioners for spinal pain and osteoarthritis,⁶ and guidelines consistently recommend the prescription of paracetamol (acetaminophen) as the first line analgesic for these conditions.⁷⁻¹¹ There has, however, been controversy about keeping paracetamol in the most recent guidance on osteoarthritis from the National Institute for Health and Care

WHAT IS ALREADY KNOWN ON THIS TOPIC

Clinical guidelines recommend paracetamol as first line analgesic drug for both spinal pain (neck and low back pain) and osteoarthritis of the hip and knee

The evidence base supporting these recommendations has recently been called into question

WHAT THIS STUDY ADDS

High quality evidence suggests that paracetamol is ineffective in reducing pain and disability or improving quality of life in patients with low back pain

There is high quality evidence that paracetamol offers a small but not clinically important benefit for pain and disability reduction in patients with hip or knee osteoarthritis

Though high quality evidence shows that patients taking paracetamol are nearly four times more likely to have abnormal results on liver function tests compared with those taking oral placebo, the clinical relevance of this is unclear

Excellence,¹² mainly because of previous studies reporting small effects of paracetamol compared with placebo.^{13–15} Moreover, optimal therapeutic benefits of paracetamol might require regular doses of up to 4000 mg/day.¹⁶ There are some concerns regarding safety of the full recommended dose,^{17 18} although the evidence on safety is still debatable.¹⁹ Potential adverse effects and treatment schedule seem to also have a considerable effect on patient adherence²⁰ as taking analgesics constantly and regularly three or four times a day is inconvenient at least.

New randomised controlled trials^{15 21} have been conducted since the last meta-analyses of paracetamol for spinal pain and osteoarthritis of the hip or knee were published. There is still uncertainty, however, whether consideration of new data changes the conclusions regarding the efficacy and safety of paracetamol for these conditions. In this systematic review we investigated the efficacy and safety of paracetamol in patients with spinal pain or osteoarthritis of the hip or knee by including data from placebo controlled trials only, as these represent the highest standard of evidence to inform the optimal use of drugs.²²

Methods

Data sources and searches

We conducted a systematic review following the PRISMA statement²³ and prospectively registered the review on PROSPERO. We carried out a systematic electronic search in Medline, Embase, AMED, CINAHL, Web of Science, LILACS, International Pharmaceutical Abstracts, and Cochrane Central Register of Controlled Trials from inception to 8 December 2014. We used a combination of relevant keywords to construct the search strategy including paracetamol, acetaminophen, back pain, neck pain, osteoarthritis, osteoarthrosis, placebo, randomised, and controlled trial (see appendix 1). One author (GCM) conducted the first screening of potentially relevant records based on titles and abstract, and two authors (GCM and MBP) independently performed the final selection of included trials based on full text evaluation. Citation tracking was also performed on included studies and relevant systematic reviews, and relevant websites and clinical trials registries were searched for unpublished studies. Consensus between the two reviewers was used to resolve any disagreement.

Study selection

We included only randomised controlled trials comparing the efficacy of paracetamol versus placebo. To be eligible, trials had to include participants with non-specific spinal pain (neck or low back pain) or osteoarthritis of the hip or knee. We did not exclude trials in mixed populations of patients with spinal pain and osteoarthritis. The intensity and duration of symptoms were not restricted. There were also no restrictions for languages or publication date. Studies that included patients with a serious spinal pathology (such as cauda equina syndrome, tumour, or infection) were excluded. Studies with mixed populations of patients with

rheumatoid arthritis and osteoarthritis were also excluded, unless separate data were reported for osteoarthritis. Studies in which participants had previous spinal, hip, or knee surgery remained eligible, but trials evaluating analgesia in the immediate postoperative period were not included. We included only full reports in this systematic review (that is, no abstracts).

Trials were eligible for inclusion when they reported at least one of the following primary outcome measures: pain intensity, disability status, and quality of life. Secondary outcome measures were safety (adverse effects), patient adherence, and use of rescue medication.

Data extraction and quality assessment

Using a standardised data extraction form, two reviewers (GCM and MBP) independently extracted study characteristics (details of participants, interventions, and outcomes) from the included trials, and a third author (MLF) resolved any disagreement. We extracted means, standard deviations, and sample sizes for our primary outcome measures. Mean estimates were extracted in the following hierarchical order: mean differences, change scores, and final values. For our secondary outcomes, we extracted the number of cases and the total sample size. The safety outcomes extracted from included trials were the number of patients reporting any adverse event, the number of patients reporting any serious adverse event (as defined by each study), the number of patients withdrawn from study because of adverse events, and the number of patients with abnormal results on liver function tests (hepatic enzyme activity ≥ 1.5 times the upper limit of the reference range). We contacted authors to provide further information when there were insufficient data reported in the paper. When authors were unavailable we estimated data using the recommendations in the Cochrane Handbook for Systematic Reviews of Interventions.²⁴

Two reviewers (GCM and MBP) independently assessed the risk of bias of the included studies using the Cochrane Collaboration's tool.^{24 25} Consensus was used to resolve any disagreement. RevMan version 5.3.5 was used to generate figures and summaries. The quality of evidence was rated for each pooled analysis with the GRADE (grading of recommendations assessment, development and evaluation) system,²⁶ with outcomes of interest being ranked according to their relevance for clinical decision making as of limited importance, important, or critical.²⁷ The quality of evidence was downgraded by one level according to the following criteria: limitation of study design, inconsistency of results, imprecision, and publication bias. We did not consider the indirectness criterion in this review because we included a specific population with relevant outcomes and direct comparisons.²⁸ Briefly, GRADE was downgraded by one level for limitation of study design when more than a quarter of the studies included in an analysis were considered at high risk of bias (that is, one or more bias domains were judged as high risk) according to the Cochrane Handbook and thus plausible to affect the interpretation of our results.^{24 29} Results were considered inconsistent if there

was a wide variance of point estimates across studies or if the heterogeneity between trials was large ($I^2 > 50\%$).³⁰ Imprecision was identified when the upper or lower limits of the 95% confidence interval crossed the minimal clinically important difference of 9 points (range 0–100).^{31,32} We visually judged a funnel plot (scatterplot of the effect of estimates from individual studies against its standard error) and used Egger's test to investigate publication bias (small study effects).³³ We included a total of 11 trials in the assessment of small study effects (nine trials including patients with osteoarthritis and two trials including patients with back pain, reporting data on immediate or short term pain intensity). If the Egger's test result was significant (two tailed $P < 0.1$) we would downgrade the quality of evidence (GRADE) by one level for all meta-analyses.³⁴ The quality of evidence was defined as "high quality," "moderate quality," "low quality," and "very low quality."²⁶

Data synthesis and analysis

We grouped the outcomes into four time points of assessment: immediate term (≤ 2 weeks), short term (> 2 weeks but ≤ 3 months), intermediate term (> 3 months but ≤ 12 months), and long term (> 12 months). If studies reported multiple time points within each category, we used the time point closest to one week for immediate term, eight weeks for short term, six months for intermediate term, and 12 months for long term. When studies reported more than one scale to measure pain we extracted the more severe estimate reported at baseline. Scores for pain and disability were converted to a common 0 (no pain or disability) to 100 (worse pain or disability) scale. Pain intensity measures to calculate treatment effects were numerical rating scale scores (range 0–10) or visual analogue scale scores (range 0–100). These two pain measures are highly correlated and can be used interchangeably when transformed.³⁵ Other measures of pain were also obtained from visual analogue scale scores included in the Western Ontario and McMaster Universities osteoarthritis index (WOMAC) pain subscale (VA3 series range 0–100)³⁶ and from the multi-dimensional health assessment questionnaire (MDHAQ) pain subscale (range 0–100).³⁷ Disability measures in the meta-analyses were WOMAC function subscale or WOMAC total scores.³⁸ One study reported pain and disability measures from the WOMAC Likert version (LK series), and the scores were also normalised to a scale of 0 to 100.

To facilitate the interpretation of our pooled estimates, we defined the effects of paracetamol compared with placebo as ineffective when the 95% confidence interval crossed the no effect line, showing no significant difference between groups. We considered the minimal clinically important difference as a difference of 9 mm in a 0–100 mm visual analogue scale. This estimate has been used in past systematic reviews³² to investigate the efficacy of medicines compared with placebo for osteoarthritis and corresponds to the median minimal clinically important difference found in trials investigating patients with osteoarthritis. When our treatment effects were smaller than 9 mm, although

significant, we considered the effect as small and not clinically important.

We used the I^2 statistic to assess heterogeneity between trials, and values higher than 50% were defined to identify high heterogeneity.³⁹ We calculated weighted mean differences or risk ratios and 95% confidence intervals and used the random effects model to pool estimates for each analysis obtained with Comprehensive Meta-Analysis version 2.2.064 (Englewood, NJ, 011).

Secondary exploratory analysis

We performed sensitivity analyses to explore the influence of each risk of bias domain on pooled treatment effects. These stratified analyses were accompanied by meta-regression to generate a P value for interaction between the bias domain and estimate of treatment effect. For these analyses we used data from all osteoarthritis trials included in the meta-analysis on short term pain (seven trials). As a previous study reported that small trials in osteoarthritis tend to report more beneficial treatment effects than large trials,⁴⁰ we also conducted a sensitivity analysis between large trials (sample size ≥ 100 per group) versus small trials (sample size < 100 per group) for all trials investigating patients with osteoarthritis at immediate or short term follow-up. Negative differences in treatment effects indicate that small trials have more beneficial effects than large trials.

Post hoc analysis

We carried out a post hoc analysis to assess the potential impact of a new trial on the current evidence and thus to determine if a further new trial is justified. We used extended funnel plots (graphical augmentations of the funnel plots commonly used to investigate publication bias in meta-analyses)⁴¹ to assess the impact of a new trial in our meta-analysis. The extended funnel plots provide shaded contours that represent the contribution of a new trial to existing evidence based on statistical simulations.⁴² Addition of data from a new trial of a certain sample size and treatment effect could

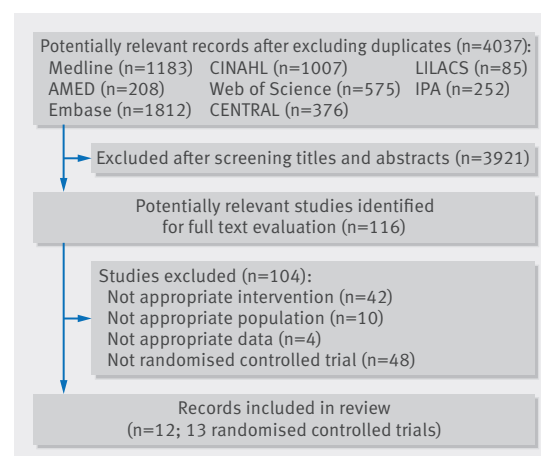


Fig 1 | Flow chart of trials investigating efficacy of paracetamol in spinal pain and osteoarthritis. Numbers of records from each database include duplicates. IPA=International Pharmaceuticals Abstracts, CENTRAL=Cochrane Register of Controlled Trials

result in the new conclusion that the effect of intervention is clearly worthwhile or clearly not worthwhile, for instance. We conducted extended funnel plots to assess the impact a further trial of paracetamol for spinal pain and hip or knee osteoarthritis would have on the current evidence presented in this meta-analysis. Stata 13 (StataCorp, College Station, TX) was used for this analysis.

Results

Our search results yielded 5498 records, and after excluding duplicates we screened 4037 titles and abstracts. Two independent reviewers evaluated 116 potentially relevant studies, and 12 records (13 randomised controlled trials) met the criteria to be included in this review, with one article reporting results of two trials (fig 1).^{43–54} Ten trials reporting data from 3541 patients evaluated the efficacy of paracetamol in patients with osteoarthritis of the hip or knee,^{43 44 46–51 54} and three trials (1825 patients) inves-

tigated the efficacy of paracetamol in people with low back pain.^{45 52 53} We did not identify any trials in patients with neck pain. Overall, the included trials assessed 5366 patients. We identified two randomised trials published as abstracts and excluded them from this review.^{55 56} One randomised trial investigating the efficacy of paracetamol for low back pain did not report results for the placebo group, and attempts to access these data from the authors and the company that funded the study were unsuccessful.⁴⁵ This trial was included in the review but not in the meta-analysis.

In the included studies paracetamol was primarily administered orally (as tablets/capsules). One trial, however, reported the use of intravenous paracetamol in participants with chronic low back pain.⁵² The total oral dose and dose regimens for paracetamol varied across trials, with 10 trials using a total dose of 3900–4000 mg/day and three trials using 3000 mg/day (table 1). Two trials used a three arm design, one

Table 1 | Characteristics of randomised placebo controlled trials included in review of efficacy and safety of paracetamol for spinal pain and osteoarthritis

| Study | Details of participants | Interventions (Dose regimens) | Outcomes and time points |
|--|--|--|--|
| Spinal pain | | | |
| Nadler, 2002 ⁴⁵ | 133 patients (group 1=113, group 2=20); mean age (SD) group 1=34.9 (11.3), group 2=38.0 (9.1); duration: acute (NR) | Group 1: paracetamol 500 mg, 2 tablets, 4 times daily, 4000 mg total. Group 2: oral placebo | Pain (VRS, 0–5); Roland Morris questionnaire (0–24); adverse events; on days 2 and 4 |
| Wetzel, 2014 ⁵² (cross over) | 40 patients on chronic opioid therapy, mean age (SD)=57.2 (12.8); duration: chronic (>6 months) | Group 1: single intravenous paracetamol 1000 mg dose. Group 2: intravenous placebo | Pain (VAS, 0–10); Roland Morris questionnaire (0–24); immediately after infusion |
| Williams, 2014 ⁵³ | 1652 patients (group 1=550, group 2=549, group 3=553); mean age (SD) group 1=44.1 (14.8), group 2=45.4 (16.7), group 3=45.4 (16.7); duration: acute (<6 weeks) | Group 1: paracetamol 665 mg, 2 tablets, 3 times daily, 3990 mg total. Group 2: paracetamol 500 mg, 1–2 tablets as required, 4–6 hours apart, maximum 8 tablets per day. Group 3: oral placebo. Rescue medication allowed | Pain (NRS, 0–10); Roland Morris questionnaire (0–24); SF-12 physical score (0–100); patient adherence; rescue medication; adverse events; at 1, 2, 4, and 12 weeks |
| Osteoarthritis | | | |
| Amadio, 1983 ⁴³ (cross over) | 25 patients; median age (range)=64 (43–80); duration: NR | Group 1: paracetamol 500 mg, 2 tablets, 4 times daily, 4000 mg total. Group 2: oral placebo | 50 ft (15 m) walking test; adverse events; at 4 weeks |
| Zoppi, 1995 ⁴⁴ | 60 patients (group 1=30, group 2=30); mean age (SD) group 1=57.6 (11.2), group 2=55.3 (11.9); duration: group 1=75.0 (98.2) months, group 2=45.8 (58.6) months | Group 1: effervescent paracetamol 500 mg, 2 tablets, 3 times daily, 3000 mg total. Group 2: effervescent placebo | Pain (VAS, 0–100); adverse events; at 1 week |
| Case, 2003 ⁴⁶ | 57 patients (group 1=29, group 2=28); mean age (SD) group 1=62.1 (11.4), group 2=61.7 (9.0); duration: NR | Group 1: paracetamol 500 mg, 2 tablets, 4 times daily, 4000 mg total. Group 2: Oral placebo | WOMAC pain (VAS, 0–500); WOMAC function (0–1700); adverse events; at 2 and 12 weeks |
| Golden, 2004 ⁴⁷ | 303 patients (group 1=148, group 2=155); mean age (SD) group 1=61.1 (13.1), group 2=60.3 (13.0); duration: NR | Group 1: paracetamol 1000 mg, 1 tablet, 4 times daily, 4000 mg total. Group 2: Oral placebo | Pain intensity on weight bearing (0–4); 50 ft (15 m) walking test; adverse events; at 1 week |
| Miceli-Richard, 2004 ⁴⁸ | 779 patients (group 1=405, group 2=374); mean age (SD)=70 (11); duration: 46 (47) months | Group 1: paracetamol 1000 mg, 1 tablet, 4 times daily, 4000 mg total. Group 2: oral placebo. Rescue medication not allowed | Pain (VAS, 0–100); WOMAC function (0–100); adverse events; patient adherence; at 1 and 6 weeks |
| Pincus, 2004a (PACES-A cross over trial) ⁴⁹ | 524 patients; mean age (SE) group 1=63.7 (1.2), group 2=62.8 (1.3); duration group 1: 8.5 (1.0) years, group 2: 8.1 (1.1) years | Group 1: Paracetamol 1000 mg, 1 tablet, 4 times daily, 4000 mg total. Group 2: Oral placebo. Rescue medication allowed | MDHAQ pain (VAS, 0–100); WOMAC (0–100); adverse events; at 6 weeks |
| Pincus, 2004b (PACES-B cross over trial) ⁴⁹ | 556 patients; mean age (SE) group 1=64.8 (1.3), group 2=63.4 (1.3); duration group 1: 10.4 (1.3) years, group 2: 9.5 (1.1) years | Group 1: paracetamol 1000 mg, 1 tablet, 4 times daily, 4000 mg total. Group 2: oral placebo. Rescue medication allowed | MDHAQ pain (VAS, 0–100); WOMAC (0–100); adverse events; at 6 weeks |
| Herrero-Beaumont, 2007 ⁵¹ | 212 patients (group 1=108, group 2=104); mean age (SD) group 1=63.8 (7.2), group 2=64.5 (6.9); duration: group 1: 6.5 (5.3) years, group 2: 7.2 (5.8) years | Group 1: paracetamol 1000 mg, 1 tablet, 3 times daily, 3000 mg total. Group 2: oral placebo. Rescue medication allowed | WOMAC pain (Likert, 0–20); WOMAC function (0–68); rescue medication; adverse events; at 6 months |
| Altman, 2007 ⁵⁰ | 483 patients (group 1=160, group 2=158, group 3=165); mean age (range)=62.2 (40–90); duration: NR | Group 1: paracetamol ER 1300 mg, 3 times daily, 3900 mg total. Group 2: paracetamol 650 mg, 3 times daily, 1950 mg total. Group 3: oral placebo. Rescue medication allowed | WOMAC pain (VAS, 0–100); WOMAC function (0–100); adverse events; at 12 weeks |
| Prior, 2014 ⁵⁴ | 542 patients (group 1=267, group 2=275); mean age (SD) group 1=61.7 (10.2), group 2=61.7 (10.1); duration: NR | Group 1: paracetamol ER 650 mg, 2 tablets, 3 times daily, 3900 mg total. Group 2: oral placebo. Rescue medication allowed, but limited | WOMAC pain (VAS, 0–100); WOMAC function (0–100); adverse events; at 2 and 12 weeks |

VRS=verbal rating scale, VAS=visual analogue scale, NRS=numeric rating scale, NR=not reported, WOMAC=Western Ontario and McMaster Universities arthritis index, MDHAQ=multi-dimensional health assessment questionnaire, SF-12=12-item short form health survey, Duration=duration of condition, ER=extended release.

included a third group that received paracetamol as required,⁵³ and another included a third group that received a lower dose of paracetamol (650 mg, one tablet, three times/day, 1950 mg total).⁵⁰ All three treatment groups were included in the meta-analyses following the recommendation in the Cochrane Handbook for Systematic Reviews of Interventions.²⁴ The washout period before treatment started varied across trials, ranging from one day to six months. The washout periods were 12 weeks for corticosteroids,⁵¹ six weeks for intra-articular steroids,⁴³ and ranged from three days to two weeks for non-steroidal anti-inflammatories.^{43 46–48} Patients stopped taking simple analgesics from one to 10 days.^{43 46 48 52} One trial reported that the washout for glucosamine drugs was six months,⁵¹ and two trials used “five half lives” to define this period.^{50 54}

We included six trials that reported data from people with chronic pain,^{44 48 49 51 52} and two studies that included people with acute pain only.^{45 53} The remaining studies did not report the duration of pain or disability. Nine trials used the diagnosis of osteoarthritis based on image evidence and clinical assessment,^{43 46–51 54} whereas one trial based the diagnosis solely on image evidence.⁴⁴ Two trials used a clear definition of low back pain,^{52 53} and one trial used a simple question to define patients (“do the muscles of your low back hurt?”).⁴⁵ Table 1 includes more detailed information on included trials.

Figure 2 summarises the assessment of risk of bias for individual trials. Twelve trials had at least one domain judged as unclear risk of bias. Four trials had at least one domain considered as high risk of bias, and only one trial had all bias domains judged as low risk of bias. Most trials (nine) failed to report the method used to generate the sequence allocation, though all reported being randomised studies. Three trials adopted an appropriate method of concealment of allocation, and only one trial failed to report blinding of patients, personnel, and outcome assessors. Eight trials were funded by companies that produce paracetamol and were considered as having unclear risk of bias for the other sources of bias domain. As only one study reported data for intermediate term follow-up, its results were pooled with trials reporting data for short term follow-up. None of the included trials reported data for long term follow-up. The inspection of the funnel plot and the lack of significance of the Egger’s test ($P=0.21$) suggested no serious small study effects (see appendix 2, fig A). We therefore considered that no meta-analysis presented serious publication bias according to the GRADE approach. Figure 3 summarises pooled effect sizes for pain and disability at immediate and short term follow-up. Tables 2 and 3 present individual trial results and calculations of effect sizes.

Spinal pain

Immediate term follow-up

Two trials including 1692 patients with low back pain tested the effect of paracetamol compared with placebo

| | Random sequence generation (selection bias) | Allocation concealment (selection bias) | Blinding of participants and personnel (performance bias) | Blinding of outcome assessment (detection bias) | Incomplete outcome data (attrition bias) | Selective reporting (reporting bias) | Other bias |
|------------------------|---|---|---|---|--|--------------------------------------|------------|
| Altman 2007 | ? | ? | + | + | + | + | + |
| Amadio & Cummings 1983 | ? | ? | + | + | + | + | + |
| Case 2003 | ? | ? | + | + | + | + | + |
| Golden 2004 | ? | ? | + | + | + | + | ? |
| Herrero-Beaumont 2007 | + | + | + | + | ? | + | ? |
| Miceli-Richard 2004 | ? | ? | + | + | + | + | + |
| Nadler 2002 | ? | ? | ? | ? | + | + | ? |
| Pincus 2004a | ? | ? | + | + | ? | + | ? |
| Pincus 2004b | ? | ? | + | + | ? | + | ? |
| Prior 2014 | + | + | + | + | ? | + | ? |
| Wetzel 2014 | + | ? | + | + | + | + | ? |
| Williams 2014 | + | + | + | + | + | + | + |
| Zoppi 1995 | ? | ? | + | + | + | + | + |

Fig 2 | Risk of bias summary showing review authors’ judgments about each risk of bias domain in placebo controlled trials on efficacy of paracetamol for spinal pain and osteoarthritis. Randomised clinical trials are listed alphabetically by author name

in pain reduction.^{52 53} Pooling showed no effect of paracetamol on pain (weighted mean difference 1.4, 95% confidence interval –1.3 to 4.1; “moderate quality” evidence, downgraded for limitation of study design). For disability, one trial evaluating 1652 patients found no difference between paracetamol and placebo (–1.9, –4.8 to 1.0).⁵³ The quality of evidence for disability in the immediate term was rated “high quality” according to the GRADE approach.

Short term follow-up

Only one trial investigated the short term efficacy of paracetamol in 1652 patients with low back pain.⁵³ This trial showed no effect of paracetamol on pain intensity (weighted mean difference –0.5, 95% confidence interval –2.9 to 1.9), disability (0.4, –1.7 to 2.5), or quality of life measured by the 12-item short form health survey (SF-12 version 2) (0.4, –0.9 to 1.7) at short term follow-up. The quality of evidence (GRADE) for

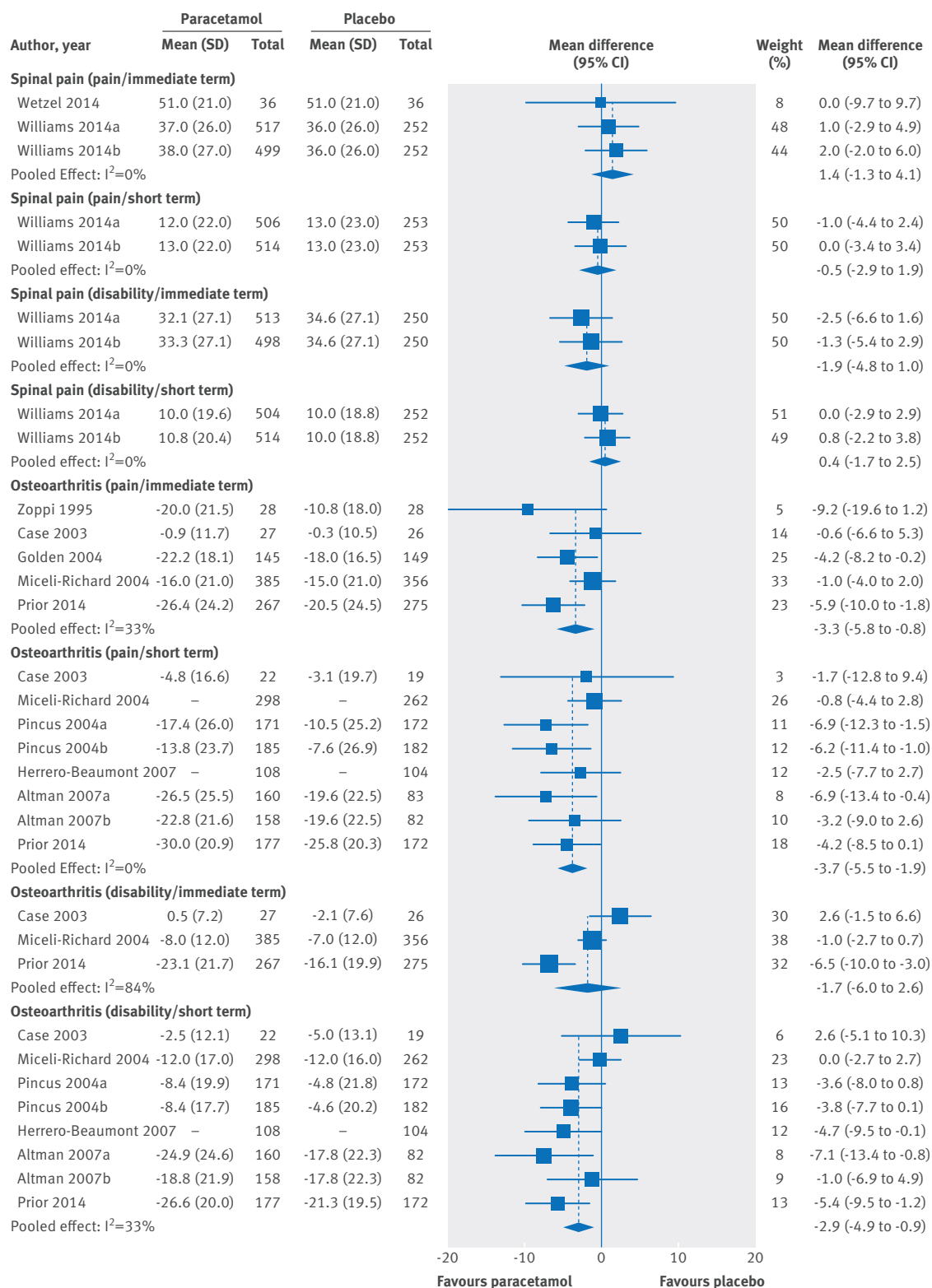


Fig 3 | Weighted mean differences for pain and disability in placebo controlled trials on efficacy of paracetamol for spinal pain and hip or knee osteoarthritis. Pain and disability are expressed on scale of 0–100. Immediate term=follow-up ≤ 2 weeks; short term=follow-up evaluations > 2 weeks but ≤ 3 months. Studies ordered chronologically within subgroups

all three outcomes was rated as “high quality.” Tables 4 and 5 summarise the findings and quality assessment (GRADE) for outcomes ranked as critical for decision making.

Osteoarthritis

Immediate term follow-up

Five trials reported data from 1741 patients with hip or knee osteoarthritis and were included in a meta-analysis

Table 2 | Calculation of effect sizes for immediate and short term pain and disability outcome measures in people with spinal pain randomised to paracetamol or placebo

| | Outcome scale | Range | Mean (SD or SE), extracted | | Mean (SD), converted* | | No of patients | | Mean difference (95% CI) | Analytic method* |
|--------------------------------|---------------|-------|----------------------------|-----------|-----------------------|-------------|----------------|---------|--------------------------|------------------|
| | | | Paracetamol | Placebo | Paracetamol | Placebo | Paracetamol | Placebo | | |
| Pain/immediate term | | | | | | | | | | |
| Wetzel, 2014 ⁵² | VAS | 0–10 | 5.1 (2.1) | 5.1 (2.1) | 51.0 (21.0) | 51.0 (21.0) | 36 | 36 | 0.0 (–9.7 to 9.7) | FV |
| Williams, 2014a ^{53†} | NRS | 0–10 | 3.7 (2.6) | 3.6 (2.6) | 37.0 (26.0) | 36.0 (26.0) | 517 | 252 | 1.0 (–2.9 to 4.9) | FV |
| Williams, 2014b ^{53‡} | NRS | 0–10 | 3.8 (2.7) | 3.6 (2.6) | 38.0 (27.0) | 36.0 (26.0) | 499 | 252 | 2.0 (–2.0 to 6.0) | FV |
| Pain/short term | | | | | | | | | | |
| Williams, 2014a ^{53†} | NRS | 0–10 | 1.2 (2.2) | 1.3 (2.3) | 12.0 (22.0) | 13.0 (23.0) | 506 | 253 | –1.0 (–4.4 to 2.4) | FV |
| Williams, 2014b ^{53‡} | NRS | 0–10 | 1.3 (2.2) | 1.3 (2.3) | 13.0 (22.0) | 13.0 (23.0) | 514 | 253 | 0.0 (–3.4 to 3.4) | FV |
| Disability/immediate term | | | | | | | | | | |
| Williams, 2014a ^{53†} | RMQ | 0–24 | 7.7 (6.5) | 8.3 (6.5) | 32.1 (27.1) | 34.6 (27.1) | 513 | 250 | –2.5 (–6.6 to 1.6) | FV |
| Williams, 2014b ^{53‡} | RMQ | 0–24 | 8.0 (6.5) | 8.3 (6.5) | 33.3 (27.1) | 34.6 (27.1) | 498 | 250 | –1.3 (–5.4 to 2.9) | FV |
| Disability/short term | | | | | | | | | | |
| Williams, 2014a ^{53†} | RMQ | 0–24 | 2.4 (4.7) | 2.4 (4.5) | 10.0 (19.6) | 10.0 (18.8) | 504 | 252 | 0.0 (–2.9 to 2.9) | FV |
| Williams, 2014b ^{53‡} | RMQ | 0–24 | 2.6 (4.9) | 2.4 (4.5) | 10.8 (20.4) | 10.0 (18.8) | 514 | 252 | 0.8 (–2.2 to 3.8) | FV |

NRS=numerical rating scale, VAS=visual analogue scale, RMQ=Roland-Morris questionnaire, FV=final value.

*Used to calculate treatment effect.

†Paracetamol (as recommended; paracetamol 665 mg, 2 tablets, 3 times daily, 3990 mg total) v placebo. Placebo group sample size was divided by 2.

‡Paracetamol (as required; paracetamol 500 mg, 1–2 tablets as required, 4–6 hours apart, maximum 8 tablets per day) v placebo. Placebo group sample size was divided by 2.

to evaluate the immediate effect of paracetamol in pain reduction.^{44 46–48 54} Pooling showed that paracetamol has a small benefit when compared with placebo in reducing pain (weighted mean difference –3.3, 95% confidence interval –5.8 to –0.8; “high quality” evidence). For disability, pooling of three trials with 1378 patients showed no immediate effect of paracetamol (–1.7, –6.0 to 2.6; “moderate quality” evidence, downgraded for inconsistency).^{46 48 54}

Short term follow-up

At short term follow-up, seven trials including 3153 patients with hip or knee osteoarthritis were pooled to estimate the efficacy of paracetamol in reducing pain and disability.^{46 48–51 54} Pooling showed a significant small effect favouring paracetamol for pain (weighted mean difference –3.7, 95% confidence interval –5.5 to –1.9). Similarly, a significant but small benefit of paracetamol was found for short term reduction in disability (–2.9, –4.9 to –0.9). The quality of evidence (GRADE) for both pooling was rated as “high quality.”

Secondary outcomes

Our secondary outcomes included adverse effects, patient adherence, and use of rescue medication. Fig 4 summarises the results.

Adverse effects

The type of adverse events reported by patients varied substantially between trials. Nine trials investigated the number of participants reporting any adverse event.^{43 44 47–50 53 54} There was no difference in the number of patients reporting adverse events between the paracetamol and placebo groups (risk ratio 1.0, 95% confidence interval 0.9 to 1.1; “moderate quality” evidence). The number of patients reporting any serious adverse event (as defined by each study) was also similar in both paracetamol and placebo groups (1.2, 0.7 to 2.1; “moderate quality” evidence).^{48–51 53 54} Ten trials reported data on the number of patients withdrawn from the study

because of adverse events, with three of these trials reporting no drop outs from adverse events. We found no significant difference between groups for this outcome (1.2, 0.9 to 1.5; “high quality” evidence).^{44 46 48–51} Three trials evaluated the results of liver function tests to detect adverse effects of paracetamol (activities of alanine aminotransferase, and/or aspartate aminotransferase) in participants with osteoarthritis,^{50 51 54} where an abnormal test was defined as hepatic enzyme activity 1.5 times the upper limit of the reference range or over. Pooling showed that participants taking paracetamol are nearly four times more likely to have abnormal results on liver function tests than participants taking placebo (3.8, 1.9 to 7.4; “high quality” evidence).

Patient adherence

Two trials in patients with low back pain and osteoarthritis investigated adherence to study treatments, defined as the number of patients reporting consumption of more than 70%⁵³ or 85%⁴⁸ of the recommended dose. We found no difference in the number of participants adhering to study treatments between paracetamol and placebo groups from the pooling of two trials (risk ratio 1.0, 95% confidence interval 0.9 to 1.1; “moderate quality” evidence, downgraded for inconsistency).

Use of rescue medication

This was measured as the number of patients using a rescue medication (naproxen 250 mg, two tablets initially then one tablet every six to eight hours as needed,⁵³ or ibuprofen 400 mg, one tablet every eight hours for a maximum of three days⁵¹) during the trial. Pooled analysis of two trials in low back pain and osteoarthritis showed no difference between the paracetamol and placebo groups (risk ratio 0.7, 95% confidence interval 0.4 to 1.3; “high quality” evidence).

Secondary exploratory analysis

The results from our secondary analyses on the potential impact of individual risk of bias domains on our

Table 3 | Calculation of effect sizes for immediate and short term pain and disability outcome measures in people with osteoarthritis randomised to paracetamol or placebo

| Outcome scale | Range | Mean (SD or SE), extracted | | Mean (SD), converted* | | No of patients | | MD (95% CI) | Analytical method* |
|---------------------------|--------|----------------------------|---------------|-----------------------|---------------|----------------|---------|----------------------|--------------------|
| | | Paracetamol | Placebo | Paracetamol | Placebo | Paracetamol | Placebo | | |
| Pain/immediate term | | | | | | | | | |
| VAS | 0–100 | –20.0† (21.5‡) | –10.8† (18‡) | –20.0 (21.5) | –10.8 (18.0) | 28 | 28 | –9.2 (–19.6 to 1.2) | CS |
| WOMAC pain (VAS) | 0–500 | –4.7 (58.4) | –1.5 (52.3) | –0.9 (11.7) | –0.3 (10.5) | 27 | 26 | –0.6 (–6.6 to 5.3) | CS |
| Pain on WB | 0–4 | –0.9† (18.1§) | –0.7† (16.5§) | –22.2 (18.1) | –18.0 (16.5) | 145 | 149 | –4.2 (–8.2 to –0.2) | CS |
| VAS | 0–100 | –16.0 (21.0) | –15.0 (21) | –16.0 (21.0) | –15.0 (21.0) | 385 | 356 | –1.0 (–4.0 to 2.0) | CS |
| WOMAC pain (VAS) | 0–100 | –26.4 (1.5) | –20.5 (1.5) | –26.4 (24.2¶) | –20.5 (24.5¶) | 267 | 275 | –5.9 (–10.0 to –1.8) | CS |
| Pain/short term | | | | | | | | | |
| WOMAC pain (VAS) | 0–500 | –23.8 (83.2) | –15.3 (98.7) | –4.8 (16.6) | –3.1 (19.7) | 22 | 19 | –1.7 (–12.8 to 9.4) | CS |
| VAS | 0–100 | NA** | NA** | NA** | NA** | 298 | 262 | –0.8 (–4.4 to 2.8) | ANCOVA |
| MDHAQ pain (VAS) | 0–100 | –17.4 (2.0) | –10.5 (1.9) | –17.4 (26.0¶) | –10.5 (25.2¶) | 171 | 172 | –6.9 (–12.3 to –1.5) | CS |
| MDHAQ pain (VAS) | 0–100 | –13.8 (1.7) | –7.6 (2.0) | –13.8 (23.7¶) | –7.6 (26.9¶) | 185 | 182 | –6.2 (–11.4 to –1.0) | CS |
| WOMAC pain (LK3) | 0–20 | NA** | NA** | NA** | NA** | 108 | 104 | –2.5 (–7.7 to 2.7) | ANCOVA |
| WOMAC pain (VAS) | 0–100 | –26.5 (25.5) | –19.6 (22.5) | –26.5 (25.5) | –19.6 (22.5) | 160 | 83 | –6.9 (–13.4 to –0.4) | CS |
| WOMAC pain (VAS) | 0–100 | –22.8 (21.6) | –19.6 (22.5) | –22.8 (21.6) | –19.6 (22.5) | 158 | 82 | –3.2 (–9.0 to 2.6) | CS |
| WOMAC pain (VAS) | 0–100 | –30.0 (1.6) | –25.8 (1.5) | –30.0 (20.9¶) | –25.8 (20.3¶) | 177 | 172 | –4.2 (–8.5 to 0.1) | CS |
| Disability/immediate term | | | | | | | | | |
| WOMAC function | 0–1700 | 7.8 (123.1) | –35.6 (129.9) | 0.5 (7.2) | –2.1 (7.6) | 27 | 26 | 2.6 (–1.5 to 6.6) | CS |
| WOMAC function | 0–100 | –8.0 (12.0) | –7.0 (12.0) | –8.0 (12.0) | –7.0 (12.0) | 385 | 356 | –1.0 (–2.7 to 0.7) | CS |
| WOMAC function | 0–100 | –23.1 (1.3) | –16.6 (1.2) | –23.1 (21.7¶) | –16.1 (19.9¶) | 267 | 275 | –6.5 (–10.0 to –3.0) | CS |
| Disability/short term | | | | | | | | | |
| WOMAC function | 0–1700 | –41.8 (205.6) | –85.6 (223.2) | –2.5 (12.1) | –5.0 (13.1) | 22 | 19 | 2.6 (–5.1 to 10.3) | CS |
| WOMAC function | 0–100 | –12.0 (17.0) | –12.0 (16.0) | –12.0 (17.0) | –12.0 (16.0) | 298 | 262 | 0.0 (–2.7 to 2.7) | CS |
| WOMAC total | 0–100 | –8.4 (1.5) | –4.8 (1.7) | –8.4 (19.9¶) | –4.8 (21.8¶) | 171 | 172 | –3.6 (–8.0 to 0.8) | CS |
| WOMAC total | 0–100 | –8.4 (1.3) | –4.6 (1.5) | –8.4 (17.7¶) | –4.6 (20.2¶) | 185 | 182 | –3.8 (–7.7 to 0.1) | CS |
| WOMAC function | 0–68 | NA** | NA** | NA** | NA** | 108 | 104 | –4.7 (–9.5 to –0.1) | ANCOVA |
| WOMAC function | 0–100 | –24.9 (24.6) | –17.8 (22.3) | –24.9 (24.6) | –17.8 (22.3) | 160 | 82 | –7.1 (–13.4 to –0.8) | CS |
| WOMAC function | 0–100 | –18.8 (21.9) | –17.8 (22.3) | –18.8 (21.9) | –17.8 (22.3) | 158 | 82 | –1.0 (–6.9 to 4.9) | CS |
| WOMAC function | 0–100 | –26.6 (1.5) | –21.3 (1.5) | –26.6 (20.0¶) | –21.3 (19.5¶) | 177 | 172 | –5.4 (–9.5 to –1.2) | CS |

questionnaire, WOMAC=Western Ontario McMaster osteoarthritis index.

*Used to calculate treatment effect.
†Mean calculated from graphs.
‡SD from baseline.
§Average SD adopted from similar studies, Zoppi, Case, and Miceli-Richard.
¶SD calculated using SE and sample size.
**Weighted mean difference and 95% CI provided.
††Paracetamol (paracetamol extended release 1300 mg, 3 times daily, 3900 mg total) v placebo. Placebo sample size was divided by 2.
‡‡Paracetamol (paracetamol 650 mg, 3 times daily, 1950 mg total) v placebo. Placebo sample size was divided by 2.

Table 4 | Summary of findings and quality of evidence assessment for outcomes classified as critical for clinical decision making in patients with spinal pain randomised to paracetamol or placebo

| Time point | Summary of findings | | | Quality of evidence assessment (GRADE) | | | |
|-------------------|---------------------|-----------------------|------------------|--|-------------|----------|------------|
| | No of patients | Effect size* (95% CI) | Study limitation | Inconsistency | Imprecision | Quality | Importance |
| Pain | | | | | | | |
| Immediate term | 1592 (2 trials) | 1.4 (–1.3 to 4.1) | –1 | None | None | Moderate | Critical |
| Short term | 1526 (1 trial) | –0.5 (–2.9 to 1.9) | None | None | None | High | Critical |
| Disability | | | | | | | |
| Immediate term | 1511 (1 trial) | –1.9 (–4.8 to 1.0) | None | None | None | High | Critical |
| Short term | 1522 (1 trial) | 0.4 (–1.7 to 2.5) | None | None | None | High | Critical |

*Weighted mean difference. Negative value favours paracetamol.

†>25% of studies included in analysis had at least one bias domain judged as high risk of bias according to Cochrane Collaboration's tool.

Table 5 | Summary of findings and quality of evidence assessment for outcomes classified as critical for clinical decision making in patients with osteoarthritis randomised to paracetamol or placebo

| Time point | Summary of findings | | | Quality of evidence assessment (GRADE) | | | |
|---|-------------------------|-----------------------|------------------|--|-------------|----------|------------|
| | No of patients (trials) | Effect size* (95% CI) | Study limitation | Inconsistency | Imprecision | Quality | Importance |
| Pain | | | | | | | |
| Immediate term | 1686 (5) | –3.3 (–5.8 to –0.8) | None | None | None | High | Critical |
| Short term | 2355 (7) | –3.7 (–5.5 to –1.9) | None | None | None | High | Critical |
| Disability | | | | | | | |
| Immediate term | 1336 (3) | –1.7 (–6.0 to 2.6) | None | –1† | None | Moderate | Critical |
| Short term | 2354 (7) | –2.9 (–4.9 to –0.9) | None | None | None | High | Critical |
| Adverse events (all short term)‡ | | | | | | | |
| Any | 4846 (9) | 1.0 (0.9 to 1.1) | None | –1† | None | Moderate | Critical |
| Serious§ | 4852 (7) | 1.2 (0.7 to 2.1) | None | –1† | None | Moderate | Critical |
| Drop out¶ | 3023 (7) | 1.2 (0.9 to 1.5) | None | None | None | High | Critical |
| Liver** | 1237 (3) | 3.8 (1.9 to 7.4) | None | None | None | High | Critical |

*Weighted mean difference (negative value favours paracetamol) for pain and disability; risk ratio for adverse events.

†Wide variance of point estimates across studies or large heterogeneity between trials ($I^2 > 50\%$).

‡Includes patients with hip/knee osteoarthritis and low back pain.

§As defined by each study.

¶Patients withdrawn from study because of adverse events.

**No of patients with abnormal results on liver function test (AST/ALN > 1.5 ULN).

treatment effects are presented in fig B in appendix 2. None of the individual domains had a significant influence on the estimated treatment effect. Our stratified analysis between small and large trials showed a difference of effects of 1.4 (95% confidence interval –2.8 to 5.6), indicating that smaller trials tend to report less beneficial effects, though this difference was not significant ($P=0.51$).

Extended funnel plot assessment

After consideration of the results we carried out a post hoc analysis to assess the effect of a new trial in our meta-analysis using extended funnel plots. Our results confirm that the results of a new trial added to current evidence would not change the conclusion that paracetamol does not deliver a clinically important benefit (at least 9 points out of a 0–100 range) for spinal pain and osteoarthritis (see fig C in appendix 2).

Discussion

There is “high quality” evidence that paracetamol has a significant but small effect in patients with hip or knee osteoarthritis compared with placebo in the short term. The small effects, < 4 points on a 0–100 point scale, are not likely to be meaningful for clinicians or patients. “High quality” evidence shows that parac-

etamol is ineffective for low back pain, but we found no trials investigating neck pain. We also found “high quality” evidence that paracetamol increases the risk of having an abnormal result on liver function tests by nearly fourfold, although the impact of this on clinically relevant patient outcomes is unclear. Adherence to the treatment protocol was similar in both paracetamol and placebo groups, and there was also no difference in the use of rescue medication. Overall, our results are based on “high quality” evidence (GRADE), and therefore further research is unlikely to change this evidence. This systematic review should inform clinical practice and policy with regard to first line care of these patients.

Strengths and weaknesses of the study

This systematic review was prospectively registered, and we followed the protocol thoroughly. We included only placebo controlled trials in the review as they provide the best evidence on the efficacy of pharmacological treatment.²² We included 13 randomised trials, 10 in people with hip or knee osteoarthritis, and three investigating people with low back pain. We included two more trials than the last meta-analysis investigating people with osteoarthritis,¹⁵ and three more than the last review on people with spinal pain.²¹ To facilitate the

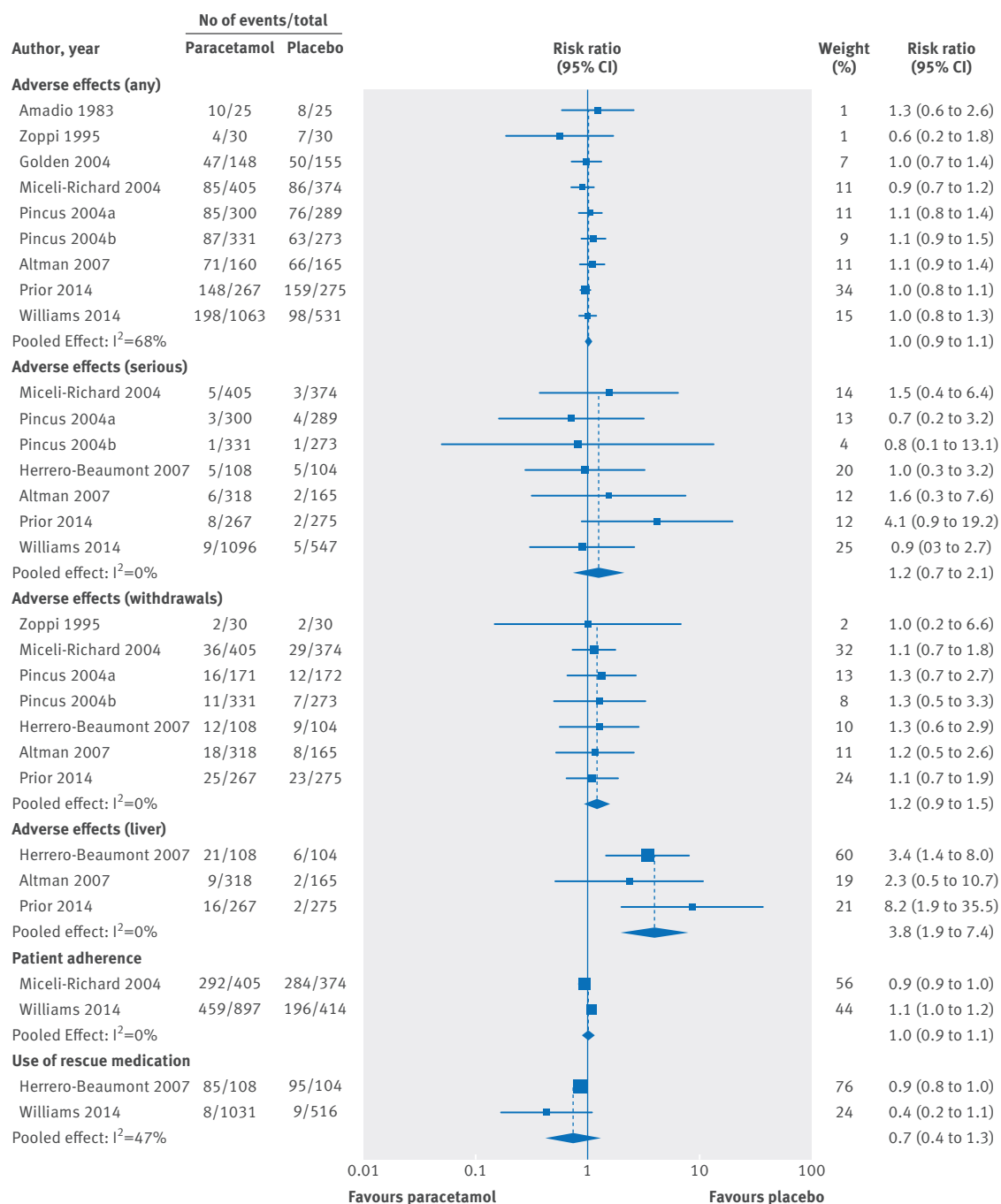


Fig 4 | Risk ratio for safety outcome measures, patient adherence, and use of rescue medication in placebo controlled trials on efficacy of paracetamol compared with placebo. Any=No of patients reporting any adverse event; serious=No of patients reporting any serious adverse event (as defined by each study); withdrawals=No of patients withdrawn from study because of adverse events; liver=No of patients with abnormal results on liver function tests. Studies are ordered chronologically within subgroups

interpretation of our results, we provide precise estimates and clinically interpretable scores on 0–100 point scales of pain and disability. Overall, the quality of evidence for our outcomes considered critical for clinical decision making was ranked “high” according to the GRADE system. Moreover, this is the first review to report evidence of changes in hepatic enzyme activity associated with paracetamol, patient adherence, and use of rescue medication in patients with osteoarthritis and spinal pain. Other strengths of our review included

lack of restrictions to publication language or date and use of hand search of clinical trial registries (for example, ClinicalTrials.gov) and relevant websites for unpublished trials.

The number of studies in each meta-analysis was relatively small because of small number of trials available on this topic (paracetamol versus placebo for spinal pain and osteoarthritis). For instance, in the meta-analyses investigating the efficacy of paracetamol on pain reduction for back pain we have

included a maximum number of two trials, and for osteoarthritis we included a maximum number of seven trials in a meta-analysis. Moreover, none of the trials reported data for long term follow-up, and our results are limited to the immediate and short term efficacy of paracetamol. Although we included three trials investigating spinal pain, none of these trials included patients with neck pain. In addition, one of the included trials did not report results for the placebo group,⁴⁵ and attempts to gain access to these data were unsuccessful. Most of the included trials used the maximum dose of 4000 mg/day recommended by the US Food and Drug Administration: seven trials used 4000 mg/day as the maximum dose, two trials used 3990 mg as the maximum dose, and two trials used 3900 mg as the maximum dose. Only two trials used 3000 mg/day as the maximum dose.

Strengths and weaknesses in relation to other studies

Previous meta-analyses have concluded that paracetamol significantly reduces pain in people with hip or knee osteoarthritis.^{13–15} One of these reviews reported no difference in toxicity, defined by the number of patients reporting any adverse event.¹⁴ All endorsed the use of paracetamol for pain reduction in such patients. Our review included two trials not previously identified in the most recent previous meta-analysis, and our results show only a small clinically irrelevant benefit of paracetamol for pain and disability at short term follow-up.

Suprathreshold doses of paracetamol can overwhelm the normal metabolic pathways and protective mechanisms in the liver and produce dangerous amounts of a toxic metabolite, *N*-acetyl-*p*-benzoquinoneimine.⁵⁷ Most commonly this is seen in intentional overdoses, and the consequence can be liver failure. However, the drug has been used extensively for decades for chronic musculoskeletal conditions, and there is scant evidence for clinically significant toxicity with regular doses of up to 4000 mg/day in otherwise healthy adults, although some researchers contest this.¹⁷ The significant effect on hepatic enzymes that we show is well known,⁵⁸ but a link with clinically important toxicity is still uncertain.

Implications for clinicians and policymakers

Interventions such as drugs that aim to provide symptomatic relief have been associated with improvement of physical function in people with osteoarthritis.^{59 60} Similarly, there is a high correlation of changes in pain scores and function scores in people with low back pain.^{61 62} This evidence supports the use of drugs for pain relief to improve function in these conditions, and, overall, we have shown consistent results across pain and disability outcome measures. We found that paracetamol is ineffective on both pain and disability outcomes for low back pain in the immediate and short term and is not clinically superior to placebo on both pain and disability outcomes for osteoarthritis.

Although thresholds for clinically important differences between groups are unknown for osteoarthritis, a recent study has used a minimal clinically important

difference of 0.9 on a 0–10 scale (or 9 on a 0 to 100 scale) based on the median difference found in previous large trials including patients with osteoarthritis.³² Our largest observed effect size of –3.7 points on a 0–100 pain scale, favouring paracetamol, is unlikely to be considered clinically important by patients or clinicians. Moreover, the lower boundary of the 95% confidence interval of this effect size was –5.5 and still did not reach the minimal clinically important difference of –9 defined in this review. Our results therefore provide an argument to reconsider the endorsement of paracetamol in clinical practice guidelines for low back pain and hip or knee osteoarthritis.

Recent evidence on lower limb osteoarthritis shows that exercises (such as strengthening exercise) compared with no exercise control result in large treatment effects for pain reduction (mean difference –2.3, 95% confidence interval –2.8 to –1.26; on a 10 cm visual analogue scale).⁶³ This effect size is much larger than the largest effect size from our pooled analyses on short term effects of paracetamol for hip or knee osteoarthritis. Paracetamol alone therefore might not be sufficient to treat hip or knee osteoarthritis and might need to be accompanied by other management strategies, such as exercises and advice/education. Future trials, however, are needed to assess the combined effect of these interventions in patients with osteoarthritis.

Unanswered questions and future research

This systematic review shows precise and clinically interpretable estimates of the size of the effect of paracetamol compared with placebo in the management of spinal pain and osteoarthritis of the hip or knee. Although our results provide “high quality” evidence that paracetamol does not provide a clinically important effect in the short term, the long term effect of this drug in the treatment of spinal pain and osteoarthritis remains unknown. Moreover, we found higher risk of abnormal results on liver function tests in patients taking paracetamol, though the clinical implications of this are uncertain. The effects of paracetamol for neck pain are unknown as we found no trials including participants with this condition.

Contributors: GCM, MLF, CGM, PHF, C-WCL, ROD, and AJMcL were involved in the conception and design of the review. GCM, MLF, and MBP developed the search strategy and performed study selection. GCM and MBP extracted data from included studies. GCM and MLF were involved in the data analysis. GCM, MLF, CGM, PHF, C-WCL, ROD, and AJMcL were involved in the interpretation and discussion of results. GCM drafted the manuscript, and MLF, CGM, and PHF contributed to the drafting of the review. CCL, ROD, MBP, and AJMcL revised it critically for important intellectual content. All authors approved the final version of the article. All authors had access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. MLF is guarantor.

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. GCM and MBP hold an international postgraduate research scholarship/postgraduate award from the Australian Government. CGM is supported by a research fellowship from the National Health and Medical Research Council. MLF holds a fellowship from Sydney Medical Foundation/Sydney Medical School. CCL holds a career development fellowship from the National Health and Medical Research Council, Australia. RD is a chief investigator on NH&MRC Programme Grant No 1054146. AJM is an investigator on the NHMRC Centre for Research Excellence in Medicines and Ageing.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: AJM received support from GlaxoSmithKline for a PhD scholarship, and AJM, ROD, CGM, and C-WCL received support from GlaxoSmithKline for the PACE trial.

Ethical approval: Not required.

Data sharing: No additional data available.

Transparency: The lead author (GCM) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; no important aspects of the study have been omitted.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380:2163–96.
- Cross M, Smith E, Hoy D, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014;73:1323–30.
- Hoy D, March L, Brooks P, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014;73:968–74.
- Hoy D, March L, Woolf A, et al. The global burden of neck pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014;73:1309–15.
- United States Bone and Joint Initiative. The burden of musculoskeletal diseases in the United States. 3rd ed. American Academy of Orthopaedic Surgeons, 2014. www.boneandjointburden.org.
- Bertin P, Becquemont L, Corruble E, et al. The therapeutic management of chronic pain in ambulatory care patients aged 65 and over in France: the S.AGES Cohort. Baseline data. *J Nutr Health Aging* 2013;17:681–6.
- Chou R, Qaseem A, Snow V, et al. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147:478–91.
- Zhang W, Doherty M, Arden N, et al. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISt). *Ann Rheum Dis* 2005;64:669–81.
- Jordan KM, Arden NK, Doherty M, et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISt). *Ann Rheum Dis* 2003;62:1145–55.
- Hochberg MC, Altman RD, April KT, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res* 2012;64:465–74.
- McAlindon TE, Bannuru RR, Sullivan MC, et al. OARS guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis Cartilage* 2014;22:363–88.
- Wise J. NICE keeps paracetamol in UK guidelines on osteoarthritis. *BMJ* 2014;348:g1545.
- Zhang W, Jones A, Doherty M. Does paracetamol (acetaminophen) reduce the pain of osteoarthritis? A meta-analysis of randomised controlled trials. *Ann Rheum Dis* 2004;63:901–7.
- Towheed TE, Maxwell L, Judd MG, et al. Acetaminophen for osteoarthritis. *Cochrane Database Syst Rev* 2006;1:CD004257.
- Zhang W, Nuki G, Moskowitz RW, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage* 2010;18:476–99.
- Tylenol (acetaminophen) professional product information. McNeil Consumer Healthcare, 2010.
- Jalan R, Williams R, Bernuau J. Paracetamol: are therapeutic doses entirely safe? *Lancet* 2006;368:2195–6.
- Graham GG, Davies MJ, Day RO, et al. The modern pharmacology of paracetamol: therapeutic actions, mechanism of action, metabolism, toxicity and recent pharmacological findings. *Inflammopharmacology* 2013;21:201–32.
- Graham GG, Day RO, Graudins A, et al. FDA proposals to limit the hepatotoxicity of paracetamol (acetaminophen): are they reasonable? *Inflammopharmacology* 2010;18:47–55.
- Laba TL, Brien JA, Fransen M, et al. Patient preferences for adherence to treatment for osteoarthritis: the MEDication Decisions in Osteoarthritis Study (MEDOS). *BMC Musculoskelet Disord* 2013;14:160.
- Davies RA, Maher CG, Hancock MJ. A systematic review of paracetamol for non-specific low back pain. *Eur Spine J* 2008;17:1423–30.
- Freedman B. Placebo-controlled trials and the logic of clinical purpose. *IRB* 1990;12:1–6.
- Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
- Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions. Version 5.1.0. Cochrane Collaboration, 2009.
- Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395–400.
- Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 2011;64:1303–10.
- Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407–15.
- Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol* 2011;64:1294–302.
- Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 2011;64:1283–93.
- Wandel S, Juni P, Tendal B, et al. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. *BMJ* 2010;341:c4675.
- Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629–34.
- Guyatt GH, Oxman AD, Montori V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64:1277–82.
- Hjermstad MJ, Fayers PM, Haugen DF, et al. Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manage* 2011;41:1073–93.
- Kersten P, White PJ, Tennant A. The visual analogue WOMAC 3.0 scale—internal validity and responsiveness of the VAS version. *BMC Musculoskelet Disord* 2010;11:80.
- Pincus T, Yazici Y, Bergman M. Development of a multi-dimensional health assessment questionnaire (MDHAQ) for the infrastructure of standard clinical care. *Clin Exp Rheumatol* 2005;23:S19–28.
- Bellamy N. WOMAC osteoarthritis index user guide IV. University of Queensland, 2000.
- Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539–58.
- Nuesch E, Trelle S, Reichenbach S, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ* 2010;341:241.
- Langan D, Higgins JP, Gregory W, et al. Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis. *J Clin Epidemiol* 2012;65:511–9.
- Ferreira ML, Herbert RD, Crowther MJ, et al. When is a further clinical trial justified? *BMJ* 2012;345:e5913.
- Amadio PJ, Cummings DM. Evaluation of acetaminophen in the management of osteoarthritis of the knee. *Curr Ther Res Clin Exp* 1983;34:59–66.
- Zoppi M, Peretti G, Boccia E. Placebo-controlled study of the analgesic efficacy of an effervescent formulation of 500 mg paracetamol in arthritis of the knee or the hip. *Eur J Pain* 1995;16:42–48.
- Nadler SF, Steiner DJ, Erasala GN, et al. Continuous low-level heat wrap therapy provides more efficacy than ibuprofen and acetaminophen for acute low back pain. *Spine* 2002;27:1012–7.
- Case JP, Baliunas AJ, Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis: a randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium. *Arch Intern Med* 2003;163:169–78.
- Golden HE, Moskowitz RW, Minic M. Analgesic efficacy and safety of nonprescription doses of naproxen sodium compared with acetaminophen in the treatment of osteoarthritis of the knee. *Am J Ther* 2004;11:85–94.
- Miceli-Richard C, Le Bars M, Schmidely N, et al. Paracetamol in osteoarthritis of the knee. *Ann Rheum Dis* 2004;63:923–30.
- Pincus T, Koch G, Lei H, et al. Patient preference for placebo, acetaminophen (paracetamol) or celecoxib efficacy studies (PACES): two randomised, double blind, placebo controlled, crossover clinical trials in patients with knee or hip osteoarthritis. *Ann Rheum Dis* 2004;63:931–39.

- 50 Altman RD, Zinsenheim JR, Temple AR, et al. Three-month efficacy and safety of acetaminophen extended-release for osteoarthritis pain of the hip or knee: a randomized, double-blind, placebo-controlled study. *Osteoarthritis Cartilage* 2007;15:454–61.
- 51 Herrero-Beaumont G, Ivorra JA, Del Carmen Trabado M, et al. Glucosamine sulfate in the treatment of knee osteoarthritis symptoms: a randomized, double-blind, placebo-controlled study using acetaminophen as a side comparator. *Arthritis Rheum* 2007;56:555–67.
- 52 Wetzel L, Zadrazil M, Paternostro-Sluga T, et al. Intravenous nonopioid analgesic drugs in chronic low back pain patients on chronic opioid treatment: a crossover, randomised, double-blinded, placebo-controlled study. *Eur J Anaesthesiol* 2014;31:35–40.
- 53 Williams C, Maher C, Latimer J, et al. Efficacy of paracetamol for acute low-back pain: a double-blind, randomised controlled trial. *Lancet* 2014;384:1586–96.
- 54 Prior MJ, Harrison DD, Frustaci ME. A randomized, double-blind, placebo-controlled 12 week trial of acetaminophen extended release for the treatment of signs and symptoms of osteoarthritis. *Curr Med Res Opin* 2014;30:2377–87.
- 55 Glimet TJ, Lermusiaux JL, Masse C, et al. Etude controlee de l'action antalgique du paracetamol effervescent dans la gonarthrose. *Rev Rhum* 1989;56:269.
- 56 Altman RD. Ibuprofen, acetaminophen and placebo in osteoarthritis of the knee: a six-day double-blind study. *Arthritis Rheum* 1999;42:S403.
- 57 Amar PJ, Schiff ER. Acetaminophen safety and hepatotoxicity—where do we go from here? *Expert Opin Drug Saf* 2007;6:341–55.
- 58 Watkins PB, Kaplowitz N, Slattery JT, et al. Aminotransferase elevations in healthy adults receiving 4 grams of acetaminophen daily—a randomized controlled trial. *JAMA* 2006;296:87–93.
- 59 Barthel HR, Peniston JH, Clark MB, et al. Correlation of pain relief with physical function in hand osteoarthritis: randomized controlled trial post hoc analysis. *Arthritis Res Ther* 2010;12:R7.
- 60 Schein JR, Kosinski MR, Janagap-Benson C, et al. Functionality and health-status benefits associated with reduction of osteoarthritis pain. *Curr Med Res Opin* 2008;24:1255–65.
- 61 Gronblad M, Huipli M, Wennerstrand P, et al. Intercorrelation and test-retest reliability of the pain disability index (PDI) and the Oswestry disability questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain* 1993;9:189–95.
- 62 Sheldon EA, Bird SR, Smugar SS, et al. Correlation of measures of pain, function, and overall response: results pooled from two identical studies of etoricoxib in chronic low back pain. *Spine* 2008;33:533–8.
- 63 Uthman OA, van der Windt DA, Jordan JL, et al. Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. *BMJ* 2013;347:f5555.

© BMJ Publishing Group Ltd 2015

Appendix 1: Search strategies

Appendix 2: Supplementary figures A-C