Cochrane
Library

Cochrane Database of Systematic Reviews

Celecoxib for osteoarthritis (Review)

Puljak L, Marin A, Vrdoljak D, Markotic F, Utrobicic A, Tugwell P

Puljak L, Marin A, Vrdoljak D, Markotic F, Utrobicic A, Tugwell P.
Celecoxib for osteoarthritis.

Cochrane Database of Systematic Reviews 2017, Issue 5. Art. No.: CD009865.

DOI: [10.1002/14651858.CD009865.pub2](https://doi.org/10.1002/14651858.CD009865.pub2).

www.cochranelibrary.com

Celecoxib for osteoarthritis (Review)

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

WILEY

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	4
BACKGROUND	8
OBJECTIVES	8
METHODS	8
RESULTS	12
Figure 1.	13
Figure 2.	15
Figure 3.	16
Figure 4.	24
Figure 5.	25
DISCUSSION	26
AUTHORS' CONCLUSIONS	29
ACKNOWLEDGEMENTS	29
REFERENCES	30
CHARACTERISTICS OF STUDIES	38
DATA AND ANALYSES	109
Analysis 1.1. Comparison 1 Celecoxib versus placebo, Outcome 1 Pain.	109
Analysis 1.2. Comparison 1 Celecoxib versus placebo, Outcome 2 Physical function.	110
Analysis 1.3. Comparison 1 Celecoxib versus placebo, Outcome 3 Number withdrawn due to adverse events.	110
Analysis 1.4. Comparison 1 Celecoxib versus placebo, Outcome 4 Number experiencing any serious adverse events.	111
Analysis 1.5. Comparison 1 Celecoxib versus placebo, Outcome 5 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).	112
Analysis 1.6. Comparison 1 Celecoxib versus placebo, Outcome 6 Number experiencing cardiovascular events (myocardial infarction, stroke).	112
Analysis 2.1. Comparison 2 Celecoxib versus tNSAIDs, Outcome 1 Pain.	113
Analysis 2.2. Comparison 2 Celecoxib versus tNSAIDs, Outcome 2 Physical function.	113
Analysis 2.3. Comparison 2 Celecoxib versus tNSAIDs, Outcome 3 Number withdrawn due to adverse events.	114
Analysis 2.4. Comparison 2 Celecoxib versus tNSAIDs, Outcome 4 Number experiencing any serious adverse events.	114
Analysis 2.5. Comparison 2 Celecoxib versus tNSAIDs, Outcome 5 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).	114
Analysis 2.6. Comparison 2 Celecoxib versus tNSAIDs, Outcome 6 Number experiencing cardiovascular events (myocardial infarction, stroke).	115
Analysis 3.1. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 1 All pain < 24 weeks.	116
Analysis 3.2. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 2 Pain VAS at 6 weeks.	117
Analysis 3.3. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 3 Pain VAS at 12 weeks.	118
Analysis 3.4. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 4 Pain VAS at 13 weeks.	118
Analysis 3.5. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 5 Pain on walking VAS at 6 weeks.	118
Analysis 3.6. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 6 Pain on walking WOMAC at 12 weeks. ...	119
Analysis 3.7. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 7 Pain WOMAC at 6 weeks.	119
Analysis 3.8. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 8 Pain WOMAC at 12 weeks.	119
Analysis 3.9. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 9 Pain WOMAC at 13 weeks.	120
Analysis 3.10. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 10 Pain WOMAC at 24 weeks.	120
Analysis 3.11. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 11 All physical function < 24 weeks.	121
Analysis 3.12. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 12 Physical function WOMAC at 6 weeks. .	121
Analysis 3.13. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 13 Physical function WOMAC at 12 weeks.	122
Analysis 3.14. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 14 Physical function WOMAC at 13 weeks.	122
Analysis 3.15. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 15 Physical function WOMAC at 24 weeks.	123
Analysis 3.16. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 16 Quality of life: SF-36 physical component scores.	123

Analysis 3.17. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 17 Quality of life: SF-36 mental component scores.	123
Analysis 3.18. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 18 Number of responders with at least 50% improvement in WOMAC pain.	123
Analysis 3.19. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 19 Number withdrawn due to adverse events.	124
Analysis 3.20. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 20 Number experiencing any serious adverse events.	125
Analysis 3.21. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 21 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).	125
Analysis 3.22. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 22 Number experiencing cardiovascular events (myocardial infarction, stroke).	126
Analysis 4.1. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 1 All pain under 24 weeks.	127
Analysis 4.2. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 2 All pain over 24 weeks.	127
Analysis 4.3. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 3 All physical function under 24 weeks. ...	127
Analysis 4.4. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 4 Number withdrawn due to adverse events.	128
Analysis 4.5. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 5 Number experiencing any serious adverse events.	128
Analysis 4.6. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 6 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).	129
Analysis 4.7. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 7 Number experiencing cardiovascular events (myocardial infarction, stroke).	129
Analysis 5.1. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 1 All pain ≤ 24 weeks.	130
Analysis 5.2. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 2 Pain VAS at 6 weeks.	131
Analysis 5.3. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 3 Pain VAS at 12 weeks.	131
Analysis 5.4. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 4 Pain WOMAC at 6 weeks.	131
Analysis 5.5. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 5 Pain WOMAC at 12 weeks.	131
Analysis 5.6. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 6 Pain WOMAC at 6 months.	132
Analysis 5.7. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 7 All physical function ≤ 24 weeks.	132
Analysis 5.8. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 8 Physical function WOMAC at 6 weeks.	132
Analysis 5.9. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 9 Physical function WOMAC at 12 weeks.	133
Analysis 5.10. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 10 Physical function WOMAC at 6 months.	133
Analysis 5.11. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 11 Number withdrawn due to adverse events.	133
Analysis 5.12. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 12 Number experiencing any serious adverse events.	134
Analysis 5.13. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 13 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).	134
Analysis 6.1. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 1 Pain VAS at 1 year.	135
Analysis 6.2. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 2 Number withdrawn due to adverse events.	135
Analysis 6.3. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 3 Number experiencing any serious adverse events. .	135
Analysis 6.4. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 4 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).	135
Analysis 6.5. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 5 Number experiencing cardiovascular events (myocardial infarction, stroke).	136
Analysis 7.1. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 1 Pain VAS at 6 weeks.	137
Analysis 7.2. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 2 Pain WOMAC at 6 weeks.	137
Analysis 7.3. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 3 Pain on walking at 12 weeks.	137
Analysis 7.4. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 4 Physical function WOMAC at 6 weeks.	137
Analysis 7.5. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 5 Number withdrawn due to adverse events.	138
Analysis 7.6. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 6 Number experiencing any serious adverse events. .	138
Analysis 7.7. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 7 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).	138
ADDITIONAL TABLES	138

APPENDICES	140
FEEDBACK	141
WHAT'S NEW	142
HISTORY	142
CONTRIBUTIONS OF AUTHORS	143
DECLARATIONS OF INTEREST	143
SOURCES OF SUPPORT	143
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	143
INDEX TERMS	143

[Intervention Review]

Celecoxib for osteoarthritis

Livia Puljak¹, Ana Marin², Davorka Vrdoljak³, Filipa Markotic⁴, Ana Utrobicic⁵, Peter Tugwell⁶

¹Cochrane Croatia, University of Split School of Medicine, Split, Croatia. ²Cochrane Croatia, Split, Croatia. ³Department of Family Medicine, School of Medicine in Split, Split, Croatia. ⁴Centre for Clinical Pharmacology, University Clinical Hospital Mostar, Mostar, Bosnia and Herzegovina. ⁵Central Medical Library, University of Split, School of Medicine, Split, Croatia. ⁶Department of Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada

Contact: Livia Puljak, Cochrane Croatia, University of Split School of Medicine, Soltanska 2, Split, 21000, Croatia.
livia.puljak@gmail.com, livia.puljak@mefst.hr.

Editorial group: Cochrane Musculoskeletal Group.

Publication status and date: Edited (no change to conclusions), comment added to review, published in Issue 7, 2017.

Citation: Puljak L, Marin A, Vrdoljak D, Markotic F, Utrobicic A, Tugwell P. Celecoxib for osteoarthritis. *Cochrane Database of Systematic Reviews* 2017, Issue 5. Art. No.: CD009865. DOI: [10.1002/14651858.CD009865.pub2](https://doi.org/10.1002/14651858.CD009865.pub2).

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Osteoarthritis (OA) is the most common form of arthritis and is caused by degeneration of the joint cartilage and growth of new bone, cartilage and connective tissue. It is often associated with major disability and impaired quality of life. There is currently no consensus on the best treatment to improve OA symptoms. Celecoxib is a selective non-steroidal anti-inflammatory drug (NSAID).

Objectives

To assess the clinical benefits (pain, function, quality of life) and safety (withdrawals due to adverse effects, serious adverse effects, overall discontinuation rates) of celecoxib in osteoarthritis (OA).

Search methods

We searched the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, Embase and clinical trials registers up to April 11, 2017, as well as reference and citation lists of included studies. Pharmaceutical companies and authors of published articles were contacted.

Selection criteria

We included published studies (full reports in a peer reviewed journal) of prospective randomized controlled trials (RCTs) that compared oral celecoxib versus no intervention, placebo or another traditional NSAID (tNSAID) in participants with clinically- or radiologically-confirmed primary OA of the knee or hip, or both knee and hip.

Data collection and analysis

Two authors independently performed data extraction, quality assessment, and compared results. Main analyses for patient-reported outcomes of pain and physical function were conducted on studies with low risk of bias for sequence generation, allocation concealment and blinding of participants and personnel.

Main results

We included 36 trials that provided data for 17,206 adults: 9402 participants received celecoxib 200 mg/day, and 7804 were assigned to receive either tNSAIDs (N = 1869) or placebo (N = 5935). Celecoxib was compared with placebo (32 trials), naproxen (6 trials) and diclofenac (3 trials). Studies were published between 1999 and 2014. Studies included participants with knee, hip or both knee and hip OA; mean OA duration was 7.9 years. Most studies included predominantly white participants whose mean age was 62 (± 10) years; most participants

were women. There were no concerns about risk of bias for performance and detection bias, but selection bias was poorly reported in most trials. Most trials had high attrition bias, and there was evidence of selective reporting in a third of the studies.

Celecoxib versus placebo

Compared with placebo celecoxib slightly reduced pain on a 500-point Western Ontario and McMaster Universities Arthritis Index (WOMAC) pain scale, accounting for 3% absolute improvement (95% CI 2% to 5% improvement) or 12% relative improvement (95% CI 7% to 18% improvement) (4 studies, 1622 participants). This improvement may not be clinically significant (high quality evidence).

Compared with placebo celecoxib slightly improved physical function on a 1700-point WOMAC scale, accounting for 4% absolute improvement (95% CI 2% to 6% improvement), 12% relative improvement (95% CI 5% to 19% improvement) (4 studies, 1622 participants). This improvement may not be clinically significant (high quality evidence).

There was no evidence of an important difference for withdrawals due to adverse events (Peto OR 0.99, 95% CI 0.85 to 1.15) (moderate quality evidence due to study limitations).

Results were inconclusive for numbers of participants experiencing any serious AEs (SAEs) (Peto OR 0.95, 95% CI 0.66 to 1.36), gastro-intestinal events (Peto OR 1.91, 95% CI 0.24 to 14.90) and cardiovascular events (Peto OR 3.40, 95% CI 0.73 to 15.88) (very low quality evidence due to serious imprecision and study limitations). However, regulatory agencies have warned of increased cardiovascular events for celecoxib.

Celecoxib versus tNSAIDs

There were inconclusive results regarding the effect on pain between celecoxib and tNSAIDs on a 100-point visual analogue scale (VAS), showing 5% absolute improvement (95% CI 11% improvement to 2% worse), 11% relative improvement (95% CI 26% improvement to 4% worse) (2 studies, 1180 participants, moderate quality evidence due to publication bias).

Compared to a tNSAID celecoxib slightly improved physical function on a 100-point WOMAC scale, showing 6% absolute improvement (95% CI 6% to 11% improvement) and 16% relative improvement (95% CI 2% to 30% improvement). This improvement may not be clinically significant (low quality evidence due to missing data and few participants) (1 study, 264 participants).

Based on low or very low quality evidence (downgraded due to missing data, high risk of bias, few events and wide confidence intervals) results were inconclusive for withdrawals due to AEs (Peto OR 0.97, 95% CI 0.74 to 1.27), number of participants experiencing SAEs (Peto OR 0.92, 95% CI 0.66 to 1.28), gastro-intestinal events (Peto OR 0.61, 0.15 to 2.43) and cardiovascular events (Peto OR 0.47, 95% CI 0.17 to 1.25).

In comparisons of celecoxib and placebo there were no differences in pooled analyses between our main analysis with low risk of bias and all eligible studies. In comparisons of celecoxib and tNSAIDs, only one outcome showed a difference between studies at low risk of bias and all eligible studies: physical function (6% absolute improvement in low risk of bias, no difference in all eligible studies).

No studies included in the main comparisons measured quality of life. Of 36 studies, 34 reported funding by drug manufacturers and in 34 studies one or more study authors were employees of the sponsor.

Authors' conclusions

We are highly reserved about results due to pharmaceutical industry involvement and limited data. We were unable to obtain data from three studies, which included 15,539 participants, and classified as awaiting assessment. Current evidence indicates that celecoxib is slightly better than placebo and some tNSAIDs in reducing pain and improving physical function. We are uncertain if harms differ among celecoxib and placebo or tNSAIDs due to risk of bias, low quality evidence for many outcomes, and that some study authors and Pfizer declined to provide data from completed studies with large numbers of participants. To fill the evidence gap, we need to access existing data and new, independent clinical trials to investigate benefits and harms of celecoxib versus tNSAIDs for people with osteoarthritis, with longer follow-up and more direct head-to-head comparisons with other tNSAIDs.

PLAIN LANGUAGE SUMMARY

Benefits and harms of celecoxib to treat people with osteoarthritis

Review question

We evaluated benefits and harms of celecoxib, a drug used to treat people with osteoarthritis, to improve pain, movement, quality of life and drug safety.

Background

Osteoarthritis is a joint disease (mostly knees and hips) caused when cartilage wears away leading to pain, joint stiffness and inflammation, which can cause limited mobility and reduces quality of life. Celecoxib is marketed as being as effective as traditional anti-inflammatory drugs (tNSAIDs such as naproxen and diclofenac) but with fewer gut and other problems.

Celecoxib for osteoarthritis (Review)

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Study characteristics

We searched up to April 2017 and included 36 trials that involved 17,206 adults who received 200 mg celecoxib daily (9402 participants), or either another drug (naproxen - 6 trials, or diclofenac - 3 trials, total of 1869 participants) or a fake treatment (placebo - 32 trials, 5935 participants).

On average, participants were aged 62 years and had osteoarthritis for 7.9 years. Participants had knee, hip, or both, osteoarthritis; most were women and white.

Study funding sources

Most studies (34/36) were funded by the studied drugs manufacturers, and (34/36) included authors who worked for drug companies.

Key results

Celecoxib, compared with placebo, slightly reduced pain, improved physical function, and probably did not increase numbers of people who withdrew from trials because of problems with the treatment. However, these results were unlikely to be clinically significant. Very low quality evidence meant that we are uncertain if celecoxib caused harms, such as gut or heart problems when compared with placebo. However, drug regulatory agencies have warned of increased risk of heart problems with celecoxib.

Celecoxib and tNSAIDs reduced pain by the same amount; celecoxib slightly improved physical function compared with tNSAIDs. Evidence comparing celecoxib and tNSAIDs was low to very low quality, so it is uncertain if harms (such as heart or gut problems) occurred from taking these drugs.

None of the high-quality studies assessed quality of life.

Quality of evidence

Evidence quality was poor and results should be interpreted cautiously. Because it is known that sponsorship by the manufacturing drug company may lead to more favorable results and conclusions than sponsorship from other sources, we are highly reserved about results due to significant drug industry involvement and few data.

We were unable to obtain data from three trials which included 15,539 participants (awaiting assessment); Pfizer declined to provide data for two reports that involved 14,042 participants.

We need access to all trials results to better understand the benefits and harms of celecoxib and other treatments. New trials researching benefits and harms of celecoxib versus tNSAIDs for people with osteoarthritis, that have longer follow-up periods, more direct comparisons with other tNSAIDs, and funded by non-commercial sources, are needed.

SUMMARY OF FINDINGS

Summary of findings for the main comparison. Celecoxib versus placebo for osteoarthritis

Patient or population: osteoarthritis
Setting: international ambulatory care
Intervention: celecoxib 200 mg/day
Comparison: placebo

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	N° of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with placebo	Risk with celecoxib				
Pain assessed with: self-report questionnaire WOMAC pain (scale 0 to 500, where 0 is no pain) follow up: range 6 weeks to 24 weeks	The mean pain was 136	The mean pain in the intervention group was 16 lower (9 lower to 24 lower) ¹	-	1622 (4 RCTs)	⊕⊕⊕⊕ HIGH	3% absolute improvement (95% CI 2% to 5%) ¹ , 12% relative improvement (95% CI 7% to 18% improvement), SMD -0.22 (-0.32 to -0.12), NNTB 11 (7 to 18) ²
Physical function assessed with: self-report questionnaire WOMAC physical pain (scale 0 to 1700, where 0 is no limitation) follow up: range 6 weeks to 24 weeks	The mean physical function was 540	The mean physical function in the intervention group was 64 lower (26 lower to 101 lower) ³	-	1622 (4 RCTs)	⊕⊕⊕⊕ HIGH	4% absolute improvement (95% CI 2% to 6% improvement) ³ , 12% relative improvement (95% CI 5% to 19% improvement), SMD -0.17 (-0.27 to -0.07), NNTB 14 (9 to 34) ²
Quality of life	not estimable	see comment	-	(0 RCTs)		No included studies measured this outcome
Number withdrawn due to adverse events follow up: range 6 weeks to 24 weeks	57 per 1000	55 per 1000 (47 to 64)	Peto OR 0.99 (0.85 to 1.15) ⁶	10996 (24 RCTs)	⊕⊕⊕⊖ MODERATE ⁴	0% absolute change (95% CI 1% less to 1% more), 1% relative change (95% CI 15% less to 15% more) (NNTB = NA)
Number experiencing any serious adverse events follow up: range 6 weeks to 24 weeks	10 per 1000	10 per 1000 (7 to 14)	Peto OR 0.95 (0.66 to 1.36) ⁶	10926 (22 RCTs)	⊕⊕⊖⊖ VERY LOW ^{4 5}	0% absolute change (95% CI 0% to 0%), 5% relative change (95% CI 34% less to 36% more) (NNTB = NA)

Number experiencing gastro-intestinal events (perforation, ulcer, bleeds) follow up: range 6 weeks to 24 weeks	1 per 1000	1 per 1000 (0 to 4)	Peto OR 1.91 (0.24 to 14.90) ⁶	3263 (8 RCTs)	⊕⊕⊕⊕ VERY LOW ^{4 5}	0% absolute change (95% CI 0% less to 1% more), 91% relative change (95% CI 76% less to 1390% more) (NNTH = NA)
Number experiencing cardiovascular events (myocardial infarction, stroke) follow up: range 6 weeks to 24 weeks	1 per 1000	7 per 1000 (1 to 57)	Peto OR 3.40 (0.73 to 15.88) ⁶	2112 (4 RCTs)	⊕⊕⊕⊕ VERY LOW ^{4 5}	0% absolute change (95% CI 0% less to 1% more), 240% relative change (95% CI 27% less to 1488% more) (NNTH = NA)

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **NNTB:** number needed to treat for an additional benefit; **NNTH:** number needed to harm; **SMD:** standardized mean difference; **RCT:** randomized controlled trial; **RR:** Risk ratio; **OR:** Odds ratio; **WOMAC:** Western Ontario and McMaster Universities Arthritis Index

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

¹ Calculations based on control group SD for pain of 74.2 (from Clegg 2006); to obtain absolute change, the following formula was used: SMD x SD/500

² Number needed to treat for an additional beneficial outcome (NNTB) for continuous outcomes calculated using the Wells calculator (from the CMSG Editorial office <http://musculoskeletal.cochrane.org/>)

³ Calculations based on control group SD for pain of 374.1 (from Clegg 2006); to obtain absolute change, the following formula was used: SMD x SD/500

⁴ Downgraded one level due to study limitations (all trials had high or unclear risk of at least one type of bias (harms outcomes included all eligible studies))

⁵ Downgraded two levels for serious imprecision (few events and wide confidence intervals)

⁶ Report Peto OR which can be interpreted as an RR due to the low event rate

Summary of findings 2. Celecoxib versus tNSAIDs for osteoarthritis

Patient or population: osteoarthritis

Setting: international ambulatory care

Intervention: celecoxib 200 mg/day

Comparison: tNSAIDs (combined: diclofenac 100 mg/day, diclofenac 150 mg/day, naproxen 1000 mg/day)

Outcomes	Anticipated absolute effects* (95% CI)	Relative effect (95% CI)	Nº of participants (studies)	Quality of the evidence (GRADE)	Comments
----------	--	--------------------------	------------------------------	---------------------------------	----------

	Risk with tNSAIDs	Risk with celecoxib				
Pain assessed with: self-report questionnaire pain VAS (scale 0 to 100, where 0 is no pain) follow up: range 12 weeks to 52 weeks	The mean pain was 41 points	The mean pain in the intervention group was 4.52 points lower (10.65 lower to 1.61 higher)	-	1180 (2 RCTs)	⊕⊕⊕⊖ MODERATE ¹	5% absolute improvement (95% CI 11% improvement to 2% worse), 11% relative improvement (95% CI 26% improvement to 4% worse), MD -4.52 (-10.65 to 1.61)
Physical function assessed with: self-report questionnaire: WOMAC physical function (scale 0 to 100, where 0 is no limitation) follow up: 12 weeks	The mean physical function was 37 points	The mean physical function in the intervention group was 6 points lower (0.6 lower to 11 lower)	-	264 (1 RCT)	⊕⊕⊕⊖ MODERATE ¹	6% absolute improvement (95% CI 6% to 11% improvement), 16% relative improvement (95% CI 2% to 30% improvement), MD: -6 (-11.4 to -0.6), NNTB 9 (5 to 121) ²
Quality of life	not estimable	see comment	-	(0 study)		no included studies measured this outcome
Number withdrawn due to adverse events follow up: range 6 weeks to 52 weeks	74 per 1000	72 per 1000 (56 to 93)	Peto OR 0.97 (0.74 to 1.27) ⁵	3150 (8 RCTs)	⊕⊕⊕⊖ LOW ³	0% absolute change (95% CI 3% less to 2% more), 3% relative change (95% CI 26% less to 27% more) (NNTH = NA)
Number experiencing any serious adverse events follow up: range 6 weeks to 52 weeks	68 per 1000	63 per 1000 (46 to 86)	Peto OR 0.92 (0.66 to 1.28) ⁵	2404 (5 RCTs)	⊕⊕⊕⊖ VERY LOW ^{3 4}	0% absolute change (95% CI 2% less to 1% more), 8% relative change (95% CI 34% less to 28% more) (NNTH = NA)
Number experiencing gastro-intestinal events (perforation, ulcer, bleeds) follow up: range 6 weeks to 52 weeks	6 per 1000	4 per 1000 (3 to 5)	Peto OR 0.61 (0.15 to 2.43) ⁵	1755 (4 RCTs)	⊕⊕⊕⊖ VERY LOW ^{3 4}	0% absolute change (95% CI 1% less to 0% more), 39% relative change (95% CI 85% less to 143% more) (NNTH = NA)
Number experiencing cardiovascular events (myocardial infarction, stroke) follow up: 52 weeks	24 per 1000	11 per 1000 (4 to 31)	Peto OR 0.47 (0.17 to 1.25) ⁵	916 (1 RCT)	⊕⊕⊕⊖ VERY LOW ^{3 4}	1% absolute change (95% CI 3% less to 0% more), 53% relative change (95% CI 83% less to 25% more) (NNTH = NA)

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **NNTB:** number needed to treat for an additional benefit; **NNTH:** number needed to harm; **SMD:** standardized mean difference; **RCT:** randomized controlled trial; **RR:** Risk ratio; **OR:** Odds ratio; **VAS:** visual analogue scale; **WOMAC:** Western Ontario and McMaster Universities Arthritis Index

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

¹ Downgraded one level due to publication bias (missing data from large studies)

² Number needed to treat for an additional beneficial outcome (NNTB) for continuous outcomes calculated using the Wells calculator (from the [CMSG editorial office](#))

³ Downgraded one level for study limitations (all trials had high or unclear risk of at least one type of bias (harms outcomes included all eligible studies)) and one level due to imprecision

⁴ Downgraded two levels for serious imprecision (few events and wide confidence intervals)

⁵ Report Peto OR which can be interpreted as an RR due to the low event rate

BACKGROUND

Description of the condition

Osteoarthritis (OA) is the most common form of arthritis and is caused by degeneration of the joint cartilage and growth of new bone, cartilage and connective tissue. It is often associated with major disability and impaired quality of life, particularly when the knee and hip bones are affected. The onset of OA is age-related and occurs usually between the ages of 50 years and 60 years (Garner 2008).

In 2014 a report from the Global Burden of Disease 2010 study indicated that of the 291 conditions listed, hip and knee OA was ranked globally as the 11th highest contributor to global disability and 38th highest in disability-adjusted life year (DALYs) (Cross 2014). The global age-standardized prevalence of knee OA was 3.8% and hip OA was 0.85%, with no discernible change between 1990 and 2010. Prevalence was higher in women than men and increases with age. The conclusion of the report is that hip and knee OA is one of the leading causes of global disability. It was also emphasized that the real burden of OA may have been underestimated because of methodological limitations. Considering the aging and increasing obesity of the global population, health professionals need to prepare for increased demand for health services to treat hip and knee OA (Cross 2014). The impact of OA on pain, disability, and quality of life results in considerable burdens to individuals, health services and society (Deeks 2002).

Description of the intervention

Non-steroidal anti-inflammatory drugs (NSAIDs) are among the most widely prescribed drugs worldwide in the treatment of rheumatic disorders and other degenerative inflammatory joint diseases (Warner 1999). NSAIDs are used for the treatment of OA because they provide effective relief from symptoms (Deeks 2002). Most clinical guidelines for OA recommend the use of NSAIDs after the failure of acetaminophen as first-line treatment of mild-to-moderate symptoms of OA and as initial treatment for severe symptoms. However, NSAIDs cause numerous side effects, particularly cardiovascular and gastro-intestinal toxicity. A nationwide cohort study conducted in Denmark showed that NSAIDs were significantly associated with increased relative risks of cardiovascular events and death, even in the low-risk population, which is a major public health concern because of the widespread use of such drugs (Fosbøl 2009). Cardiovascular risks of NSAIDs are well established although uncertainty remains about differences among individual NSAIDs (Kearney 2006; Trelle 2011). The gastro-intestinal toxicity of NSAIDs can be broadly categorized into three groups: (i) symptoms such as dyspepsia, heartburn, nausea, vomiting and abdominal pain; (ii) mucosal lesions seen on endoscopy; and (iii) serious gastro-intestinal complications, such as perforated ulcers or bleeding requiring hospitalization (Singh 1998).

The analgesic effect of NSAIDs is explained by inhibition of cyclooxygenase (COX) enzyme, which has two isoforms (COX-1 and COX-2). While COX-1 is a constitutive member of normal cells, COX-2 has a role in mediation of pain, inflammation and fever (Conaghan 2012). Selective COX-2 inhibitors (coxibs) were developed to avoid gastro-intestinal toxicity associated with traditional NSAIDs (tNSAIDs) (Conaghan 2012).

The COX-2 selective inhibitors are marketed as drugs that are as effective as traditional non-selective NSAIDs (tNSAIDs) but with a safer gastro-intestinal profile (Micklewright 2003). As a result of concerns over the long-term use of a coxib drug, rofecoxib, and increased risk of serious cardiovascular events, the manufacturer voluntarily withdrew rofecoxib from the market in September 2004 (Sibbald 2004). Subsequently, the USA Food and Drug Administration's (FDA) Arthritis and Drug Safety and Risk Management Advisory Committees reviewed all available data on selective cyclooxygenase (COX)-2 inhibitors. This led to a request by the FDA to the manufacturer for the voluntary withdrawal of valdecoxib from the market in April 2005 and re-labelling of celecoxib to include a more specific warning of the risks of serious cardiovascular adverse events associated with its use (Chou 2006).

How the intervention might work

NSAIDs primarily act on the COX enzyme that converts arachidonic acid into prostaglandins, which mediate inflammation and pain among other functions (Garner 2002). The constitutive COX-1 is present in, for example in the endothelium, stomach and kidney, whereas COX-2 is induced by pro-inflammatory cytokines and endotoxin in cells in vitro and at inflammatory sites in vivo (Vane 1998). Studies that use in vitro methods show inhibitory activity of putative anti-inflammatory drugs on either COX-1 or COX-2 (Warner 1999). On the basis of these laboratory tests, many of the original NSAIDs were found to have inhibitory activity against both isoforms, they were dual inhibitors of COX-1 and COX-2. The more selective COX-2 inhibitors had fewer gastro-intestinal side effects, and thus, it was predicted that a new generation of anti-inflammatory drugs, selective COX-2 inhibitors, would be free of gastro-intestinal side effects (Willoughby 2000). This group of NSAIDs known as the coxibs includes celecoxib, etoricoxib, lumiracoxib, parecoxib, rofecoxib and valdecoxib (Garner 2008). Celecoxib is the only coxib still on the market in most countries, where it represents a substantial market share. The recommended dose of celecoxib for OA is 200 mg daily, given as a single dose or 100 mg twice a day.

Why it is important to do this review

Clinical studies, rather than the laboratory assay studies, are the best way to determine whether participants actually benefit from using more selective NSAIDs (Chou 2006). This review examined the clinical benefit and harms of celecoxib for people with OA.

OBJECTIVES

To assess the clinical benefits (pain, function, quality of life) and harms (withdrawals due to adverse effects, serious adverse effects, overall discontinuation rates) of celecoxib in osteoarthritis (OA).

METHODS

Criteria for considering studies for this review

Types of studies

Published (full reports in a peer reviewed journal) prospective randomized controlled trials (RCTs) were included. Studies published in any language were considered. Studies were excluded if they had fewer than 50 participants in each arm (Nüesch 2010) or involved treatment of less than four weeks' duration.

Types of participants

At least 75% of participants with clinically- or radiologically-confirmed primary osteoarthritis (OA) of the knee or hip or both hip and knee OA, of any age and either sex. The diagnosis of OA was based on valid clinical and radiographic findings in accordance with the American College of Rheumatology (ACR) criteria. Trials including only participants with inflammatory arthritis, such as rheumatoid arthritis, were not considered.

Types of interventions

Oral celecoxib 200 mg daily (either as 200 mg once daily or 100 mg twice daily) versus no intervention, placebo or another marketed tNSAID. Dosage of 200 mg daily was used because it is the recommended dosage.

Types of outcome measures

Studies were included if any accepted method to assess outcomes of benefit was used because there is already an extensive body of literature available on adverse events and celecoxib.

For benefit, the major outcomes were: (1) pain, (2) function and (3) quality of life, as currently recommended for OA trials ([Altman 1996](#); [Pham 2004](#)). For harms, the major outcomes were the number of participants: (4) who withdrew because of adverse events; (5) experiencing any serious adverse events; (6) presenting clinically with perforations, ulcer or bleed; and (7) experiencing cardiovascular outcomes. The timing was the end of the RCT unless specified otherwise.

Benefit

Pain

If data on more than one pain scale were provided for a trial, data were extracted on the pain scale that was highest on the following list, according to a previously described hierarchy of pain-related outcomes ([Jüni 2006](#); [Reichenbach 2010](#)).

1. Global pain.
2. Pain on walking.
3. The Western Ontario and McMaster Universities osteoarthritis index (WOMAC) osteoarthritis index pain sub score.
4. Composite pain scores other than WOMAC.
5. Pain on activities other than walking.
6. Rest pain or pain during the night.
7. WOMAC global algofunctional score.
8. Lequesne osteoarthritis index global score.
9. Other algofunctional scale.

Physical function

If data on more than one physical function scale were provided for a trial, data were extracted according to hierarchy presented below:

1. Global disability score.
2. Walking disability.
3. WOMAC disability sub score.
4. Composite disability scores other than WOMAC.
5. Disability other than walking.
6. WOMAC global scale.
7. Lequesne osteoarthritis index global score.

8. Other algofunctional scale.

Quality of life

Quality of life data were extracted from the following instruments, if available: Short Form-12 (SF-12) and 36 (SF-36), EuroQoL, Sickness Impact Profile (SIP) or Nottingham Health Profile (NHP).

At least 50% reduction in pain score

Number of responders with at least 50% reduction in pain scores was recorded.

Harms

- Withdrawals due to adverse events.
- Serious adverse effects were extracted from trials, defined as events resulting in inpatient hospitalization, prolongation of hospitalization, persistent or significant disability, congenital abnormality or birth defect of offspring, life-threatening events or death.
- Gastro-intestinal events, number of participants presenting clinically with perforations, ulcer or bleeds.
- Cardiovascular events, number of participants with cardiovascular events.

Search methods for identification of studies

Electronic searches

The following electronic databases were searched:

- Cochrane Central Register of Controlled Clinical Trials (CENTRAL) via Ovid (*The Cochrane Library*, current issue) ([Appendix 1](#));
- MEDLINE via Ovid (1948 to present) ([Appendix 2](#)); and
- Embase via Ovid (1980 to present) ([Appendix 3](#)).

The MEDLINE search strategy combined the subject search with the Cochrane highly sensitive search strategy for identifying reports of RCTs (as published in Box 6.4.c in the *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0, updated March 2011) and was modified for other databases ([Higgins 2011](#)). Searches were performed up to April 11, 2017.

For assessments on adverse effects, we searched the websites of the regulatory agencies: [USA Food and Drug Administration-Medwatch](#), [European Medicines Evaluation Agency](#), [Australian Adverse Drug Reactions Bulletin](#), and UK Medicines and Healthcare Products Regulatory Agency (MHRA) pharmacovigilance and drug safety updates.

Searching other resources

References of all included studies, other reviews, guidelines and related articles were searched, using both forward (through citation databases such as Web of Science) and backward (examining reference lists) citation searching up to April 11, 2017.

Ongoing studies were searched in the trial registry [ClinicalTrials.gov](#) and World Health Organization International Clinical Trials Registry Platform ([WHO ICTRP](#)).

Data collection and analysis

Selection of studies

The selection of papers and decisions about eligibility was carried out independently by at least two review authors. Full-text copies of all possibly or definitely relevant studies were obtained for further assessment. All disagreements were resolved by discussion and consensus. A statistician was consulted in cases of doubt about inclusion or data extraction, as well as with regard to data analysis.

Study authors were contacted for clarifications and to obtain additional data to perform a systematic review whenever necessary. Studies considered for inclusion were translated into English when necessary.

Data extraction and management

Predefined data extraction forms were pilot tested and used to collate data. Relevant data from selected studies were independently extracted according to inclusion criteria by two review authors. Review authors independently determined study eligibility and discrepancies were resolved by discussion. Standard tables for characteristics of included studies and risk of bias were constructed to minimize errors. Possible doubts and disagreements were discussed and if unresolved, study authors were contacted for clarification. Abstracts were considered together with the full publication. The review authors were not blinded to the study authors, interventions or results obtained in the included studies. For each study, the following information was extracted: participant and study characteristics, intervention and outcomes data. The raw data (means and standard deviations for continuous outcomes and numbers of events and participants for dichotomous outcomes) were extracted for outcomes of interest. Where least squares means (LSM) were available, we assumed this for the mean to enter into RevMan.

In nine studies only LSM were available for the continuous outcomes of interest ([Asmus 2014 Study 1](#); [Asmus 2014 Study 2](#); [Birbara 2006 Study 1](#); [Birbara 2006 Study 2](#); [DeLemos 2011](#); [Essex 2012a](#); [Hochberg 2011 Study 307](#); [Hochberg 2011 Study 309](#); [McKenna 2001a](#)). Five studies provided explanations for adjustments made for calculating LSM; four studies ([Asmus 2014 Study 1](#); [Asmus 2014 Study 2](#); [DeLemos 2011](#); [Essex 2012a](#)) did not provide explanations for adjustments. Two studies made adjustments for treatment group factors: previous OA treatment strata and baseline WOMAC score at randomization ([Birbara 2006 Study 1](#); [Birbara 2006 Study 2](#)). Two studies used baseline values as covariates to estimate LSM changes from baseline for each treatment group ([Hochberg 2011 Study 307](#); [Hochberg 2011 Study 309](#)). One study used baseline score, investigational site and treatment as factors for adjustment ([McKenna 2001a](#)).

Since most continuous outcomes for pain and physical function were reported as changes from baseline, for the remainder of the studies, change from baseline was calculated using baseline and final data using procedures recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* ([Higgins 2011](#)). Two studies ([Williams 2000](#); [Williams 2001](#)) had three arms, comparing continuous and intermittent celecoxib 200 mg per day therapy with placebo. We combined data for the celecoxib groups using procedures recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* ([Higgins 2011](#)).

Two review authors extracted trial information independently using a standardized, piloted extraction form accompanied by a codebook. Disagreements were resolved by discussion. The following data were extracted: both generic and trade names of the experimental intervention, the type of control used, dosage, frequency, route of administration, duration of treatment, participant characteristics (gender, average age and duration of symptoms, types of joints affected), types of measures used for the outcomes, trial design, trial size, duration of follow-up, type and source of financial support, and publication status. Whenever possible, results from intention-to-treat analyses were used. When necessary, means and measures of dispersion were extracted from figures in the reports using open-access [Plot Digitizer](#) software ([Jelicic Kadic 2016](#); [Vucic 2015](#)).

Requests for data

Incomplete reporting of study methods and results may impact negatively on systematic review quality, introduce bias and overestimate treatment effects ([Mullan 2009](#)). Therefore, study authors were contacted to reduce risk of bias and improve the strength of the inferences ([Mullan 2009](#)). For the purpose of this review, corresponding authors of 24 studies were contacted by email, at least twice if the first email did not elicit response. Contact authors of eight studies responded (33% response rate), but only one ([Essex 2012a](#)) provided requested information (5% success rate in obtaining necessary information from authors). All requests were made to obtain missing information or clarify imprecise reporting. Corresponding authors of 16 studies did not respond to requests for additional information or clarification ([Asmus 2014 Study 1](#); [Asmus 2014 Study 2](#); [Birbara 2006 Study 1](#); [Birbara 2006 Study 2](#); [Clegg 2006](#); [Emery 2008](#); [Hochberg 2011 Study 307](#); [Hochberg 2011 Study 309](#); [Kivitz 2001](#); [Leeb 2004](#); [Lehmann 2005](#); [Ozgocmen 2005](#); [Pincus 2004 PACES-a](#); [Pincus 2004 PACES-b](#); [Smugar 2006 Study 1](#); [Smugar 2006 Study 2](#)).

Authors of four studies responded immediately, or after a reminder, indicating that they would look into providing the requested data ([Boswell 2008 Study A](#); [Boswell 2008 Study B](#), funded by GlaxoSmithKline) or that they needed to contact the study sponsor ([Bingham 2007 Study 1](#); [Bingham 2007 Study 2](#), funded by Merck) before responding. However, repeated follow-up queries did not result in further response. One study author responded that he did not have requested data ([Fleischmann 2005](#), funded by Novartis Pharma AG); a second email about where data might be obtained elicited the response "don't know". The corresponding author of a study that reported previously unpublished data responded to our query that: "A formal protocol outlining your research approach will be necessary for our group to participate". After sharing the published protocol of the review, no further response was received ([Iyengar 2013](#), funded by Pfizer). Extensive email exchanges took place with Pfizer employees regarding data from two studies that did not report outcomes separately for all randomized groups ([Lisse 2001](#); [Singh 2006](#)). Both requests were denied by Pfizer, first for data relating to [Singh 2006](#) (communication presented in detail in [Table 1](#)) and then for [Lisse 2001](#) after waiting two years for response.

Assessment of risk of bias in included studies

Two review authors independently assessed risk of bias for included studies using Cochrane's tool for assessing risk of bias ([Higgins 2011](#)). The tool addresses seven specific domains: sequence generation, allocation concealment, blinding

of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting and 'other sources of bias'. To determine each study's risk of bias of a study, the presence of sufficient information and the likelihood of potential bias was evaluated for each criterion. Each criterion was rated as low risk of bias, high risk of bias or unclear risk of bias (either lack of information or uncertainty about the potential for bias). In a consensus meeting, disagreements among review authors were discussed and resolved.

Attrition above 30% in either arm and imbalance in attrition rates between study arms over 10% was considered as high risk of bias, regardless of the data imputation method. Attrition between 10% and 30% in either arm, and imbalance between groups between 5% and 10% was considered as unclear risk of bias.

Since none of the included studies were assessed at low risk of bias for all seven domains, we considered studies that had low risk of bias for the first three domains - related to sequence generation, allocation concealment and blinding of participants and personnel - as overall low risk of bias.

Measures of treatment effect

Meta-analysis was performed for benefit outcomes if study data were sufficiently clinically and statistically homogeneous. If not sufficiently homogenous, heterogeneity was explored in stratified analyses.

For dichotomous data, the results of each RCT were expressed as risk ratio (RR) with corresponding 95% confidence interval (95% CI). RRs were considered clinically relevant if less than 0.7 or larger than 1.5, in favor of the intervention or control group, respectively. For dichotomous data with rare events (< 10%) Peto's odds ratio (Peto OR) was used.

For continuous data, results were analyzed as mean differences (MD) between the intervention and comparator group with 95% CIs. There were many different pain and function outcome measures to pool, so to obtain a clinically useful synthesis, standardized mean differences (SMD) were calculated. The SMD expresses the size of the intervention effect in each study relative to the variability observed in that study. SMDs were calculated by dividing the MD by the standard deviation of outcome among participants. A 95% CI was computed for the SMD. The SMD was interpreted as described by Cohen 1988, that is, SMD of 0.2 is considered to indicate a small beneficial effect, 0.5 a medium effect, and 0.8 a large effect of celecoxib.

Unit of analysis issues

For cross-over trials (Pincus 2004 PACES-a; Pincus 2004 PACES-b), data were extracted from the first period only. If the meta-analyses resulted in statistically significant overall estimates, these results were transformed (pooled estimate of RR, MD or SMD) back into measures which are clinically useful in daily practice, such as the number needed to treat (NNT) and the absolute or relative improvement on the original units, to express the final results of the review. The results were back translated by multiplying SMD by the standard deviation from a representative study (Akl 2011).

Dealing with missing data

For missing summary data, standard deviations were imputed from other statistics such as standard errors, CIs or P values, according to

methods recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011).

In two studies (Kivitz 2001; McKenna 2001a) there were no available relevant statistics for calculating standard deviations for change from baseline of global pain visual analog scale (VAS) outcome. We followed recommendations from the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011, chapter 16.1.3.2. *Imputing standard deviations for changes from baseline*), and included standard deviations from two similar studies (Bensen 1999; McKenna 2001b). In one study there were no relevant statistics for calculating standard deviation for change from baseline for pain on walking VAS (McKenna 2001a) and standard deviation was included for this outcome from a similar study (Smugar 2006 Study 2). Two studies did not have relevant statistics for calculating standard deviation for WOMAC pain (Boswell 2008 Study A; Boswell 2008 Study B) and standard deviation was imputed from two similar studies, Williams 2001 and Bingham 2007 Study 1, respectively.

Whenever possible, the original investigators were contacted to request missing data.

The possible impact of missing data is presented in the Discussion.

Assessment of heterogeneity

Before meta-analysis, studies were first assessed for clinical homogeneity with respect to type of therapy, control group and outcomes. For studies judged as clinically homogeneous, statistical heterogeneity was tested by χ^2 (Q) and I^2 statistics (Higgins 2003). A P value of less than 0.10 or an I^2 value greater than 50% indicates substantial heterogeneity. In cases of substantial heterogeneity, data were explored further, including subgroup analyses, to try to explain the heterogeneity.

Assessment of reporting biases

To determine whether reporting bias was present, the review authors searched for protocols of RCTs that were published before starting recruitment of study participants. Trial protocols were compared with subsequent publication(s) to identify any discrepancies in reported outcomes.

All trials that began enrolment of participants after September 2005 should be registered in a public trials registry at or before the onset of enrolment to be considered for publication in major medical journals that follow International Committee of Medical Journal Editors (ICMJE) recommendations. Therefore, for studies published after 1 July 2005, the Clinical Trial Register at the International Clinical Trials Registry Platform of the World Health Organization was checked. If the protocols were not found, outcomes listed in the methods section of a publication were compared with the reported results. For each study, the review authors evaluated whether selective reporting of outcomes was present. In the case of suspected reporting bias, study authors were contacted for clarification.

A funnel plot was to be constructed to assess the possibility of publication bias for outcomes that were reported in 10 or more studies. A test for funnel plot asymmetry (small study effects) formally examines whether the association between estimated intervention effects and a measure of study size is greater than might be expected to occur by chance.

Data synthesis

The aim was to statistically pool results from individual studies for each intervention to inform an estimate of an overall effect. Meta-analyses were conducted using the random-effects model. In the case of statistically significant results, the number needed to treat for an additional benefit (NNTB) and number needed to harm (NNTH) were also calculated for the main analyses. Analysis was performed using Review Manager 5 (RevMan 2014) and forest plots were produced for all analyses.

Subgroup analysis and investigation of heterogeneity

The following subgroup analyses were planned: participants' age (< 65 years versus ≥ 65 years), gender, duration of OA (≤ 2 years versus > 2 years), location of the study joint (knee or hip), previous OA medication (tNSAID, paracetamol), doses of celecoxib and doses of tNSAIDs in comparison groups. However, there were insufficient data in included manuscripts to justify subgroup analyses in this study.

Sensitivity analysis

Primary meta-analyses were planned to be restricted to studies at low risk of bias. However, because none of the included studies were assessed at low risk of bias for all domains, we considered studies with low risk of bias on the first three domains (randomization, allocation concealment, blinding of participants and personnel) as low risk of bias, and sensitivity analysis was conducted accordingly. We also performed sensitivity analysis for studies with imputed standard deviations. A third set of sensitivity analyses was conducted in an attempt to assess how data from Singh 2006 (a study with 13,274 participants that is waiting assessment) could change the effect estimates.

Summary of findings tables

The main results of the study were presented in a 'Summary of findings' table, which was produced using GRADEpro software (GRADEpro GDT 2014). This table provides key information concerning the quality of the evidence, the magnitude of effect of the interventions examined, and the sum of available data on

the main outcomes. The table included an overall grading of the evidence related to each of the main outcomes using the GRADE approach, as indicated in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011).

The important outcomes that were included in the 'Summary of findings' tables are:

1. pain;
2. function;
3. quality of life;
4. number of participants experiencing any serious adverse events;
5. number of participants who withdrew because of adverse events;
6. number of participants presenting clinically with perforations, or perforations, ulcers, bleeds (PUBs), or both; and
7. number of participants experiencing cardiovascular outcomes.

We used the five GRADE (Grades of Recommendation, Assessment, Development and Evaluation) considerations (study limitations, consistency of effect, imprecision, indirectness and publication bias) to assess the quality of a body of evidence for stated outcomes. Outcomes pooled using SMDs were re-expressed as changes by multiplying by a representative control group baseline SD. We presented the absolute change, the relative change, as well as NNTB and NNTH that were estimated from the OR in the 'Comments' column of the table.

RESULTS

Description of studies

Results of the search

We identified 7627 records from database searches and 2338 additional records from other sources. Following removal of duplicates, we assessed 6237 records, and excluded 6142 based on titles and abstracts (Figure 1). We assessed 93 possibly eligible papers in full text. Of these, 36 studies met the inclusion criteria for this review.

Figure 1. Study flow diagram

The 36 included trials were published in 29 full-text manuscripts; seven reported two separate randomized controlled trials (RCTs) (Asmus 2014 Study 1; Asmus 2014 Study 2; Bingham 2007 Study 1; Bingham 2007 Study 2; Birbara 2006 Study 1; Birbara 2006 Study 2; Boswell 2008 Study A; Boswell 2008 Study B; Hochberg 2011 Study 307; Hochberg 2011 Study 309; Pincus 2004 PACES-a; Pincus 2004 PACES-b; Smugar 2006 Study 1; Smugar 2006 Study 2). Data from one study were published in two separate manuscripts (Bensen 1999; Zhao 1999). All included studies were published in English.

We excluded 52 studies for the reasons given in the [Characteristics of excluded studies](#) table; eight studies were classified as [Studies awaiting classification](#); four ongoing studies were identified and described in [Characteristics of ongoing studies](#).

Included studies

We included 36 studies which are described in [Characteristics of included studies](#). Publication dates ranged from 1999 to 2014. All studies were double-blinded, randomized, parallel-group trials and included a total of 17,206 adults whose mean age was 62 (± 10) years. There were 9402 participants assigned to treatment with

celecoxib 200 mg per day, and 7804 assigned to a comparator group – tNSAIDs (N = 1869) or placebo (N = 5935).

Of the 36 studies, 34 reported receiving funding from pharmaceutical companies that produced study drugs. One study had independent funding (Clegg 2006). One study did not provide information about the source of funding, but according to the author affiliations, three of six authors were employees of Novartis Corporation, and a study drug by Novartis was used in the study (Sheldon 2005). Only two studies (Clegg 2006; Dahlberg 2009) did not include authors who were employed by pharmaceutical companies; all others (34 studies) included one or more (or all) authors who were employed by the sponsor.

Only eight studies reported registration on the ClinicalTrials.gov register (Bingham 2007 Study 1; Bingham 2007 Study 2; Clegg 2006; Conaghan 2013; Hochberg 2011 Study 307; Hochberg 2011 Study 309; Rother 2007; Schnitzer 2011), although 23 studies were published after 2005. Of the eight registered studies, only two deposited results on ClinicalTrials.gov (Hochberg 2011 Study 307; Hochberg 2011 Study 309).

Among the included studies, 32 compared celecoxib 200 mg with placebo (Asmus 2014 Study 1; Asmus 2014 Study 2; Bensen 1999; Bingham 2007 Study 1; Bingham 2007 Study 2; Birbara 2006 Study 1; Birbara 2006 Study 2; Boswell 2008 Study A; Boswell 2008 Study B; Clegg 2006; Conaghan 2013; DeLemos 2011; Essex 2012b; Essex 2014; Fleischmann 2005; Gibofsky 2003; Hochberg 2011 Study 307; Hochberg 2011 Study 309; Kivitz 2001; Lehmann 2005; McKenna 2001a; McKenna 2001b; Pincus 2004 PACES-a; Pincus 2004 PACES-b; Rother 2007; Schnitzer 2011; Sheldon 2005; Smugar 2006 Study 1; Smugar 2006 Study 2; Tannenbaum 2004; Williams 2000; Williams 2001).

Six studies compared celecoxib 200 mg with naproxen 1000 mg/day (Bensen 1999; Essex 2012a; Essex 2012b; Essex 2014; Kivitz 2001; Sowers 2005). Three studies compared celecoxib 200 mg with diclofenac 100 mg or 150 mg/day (Dahlberg 2009; Emery 2008; McKenna 2001b).

Nineteen studies were performed in one country: USA (Asmus 2014 Study 1; Asmus 2014 Study 2; Birbara 2006 Study 1; Birbara 2006 Study 2; Clegg 2006; Essex 2012a; Essex 2012b; Essex 2014; Hochberg 2011 Study 307; Hochberg 2011 Study 309; McKenna 2001a; McKenna 2001b; Smugar 2006 Study 1; Smugar 2006 Study 2; Williams 2000; Williams 2001), UK (Emery 2008) and Germany (Rother 2007). Three studies were conducted in two countries, including USA and Canada (Bensen 1999; Gibofsky 2003; Kivitz 2001; Sheldon 2005), USA and Finland (Pincus 2004 PACES-a; Pincus 2004 PACES-b), and Sweden and Norway (Dahlberg 2009). Other studies were conducted in more than two countries, including “Europe, North America and internationally” (Boswell 2008 Study A; Boswell 2008 Study B), Czech Republic, Germany, Poland and the UK (Conaghan 2013), Canada, USA, Germany, Italy and UK (DeLemos 2011; Schnitzer 2011), Argentina, Chile, Colombia, Peru, Uruguay, Venezuela, Switzerland, Canada, USA (Fleischmann 2005), Canada, France, Belgium, Germany, UK, Hungary, Denmark, Switzerland, Italy, Spain (Tannenbaum 2004), seven countries including Austria, Canada, Chile, Germany, Spain, United Kingdom (Sowers 2005) and “15 countries” (Lehmann 2005).

Only knee, hip or both knee and hip osteoarthritis (OA) was evaluated in the included studies. None of the studies evaluated OA

in other sites such as hands or spine. In three studies only the hip joint was evaluated (Emery 2008; Kivitz 2001; Schnitzer 2011). The knee was evaluated in 24 RCTs (Asmus 2014 Study 1; Asmus 2014 Study 2; Bensen 1999; Birbara 2006 Study 1; Birbara 2006 Study 2; Boswell 2008 Study A; Boswell 2008 Study B; Clegg 2006; Conaghan 2013; Essex 2012a; Essex 2012b; Essex 2014; Fleischmann 2005; Gibofsky 2003; Hochberg 2011 Study 307; Hochberg 2011 Study 309; Lehmann 2005; McKenna 2001a; McKenna 2001b; Rother 2007; Sheldon 2005; Tannenbaum 2004; Williams 2000; Williams 2001). Seven RCTs assessed both knee and hip joints (Dahlberg 2009; DeLemos 2011; Pincus 2004 PACES-a; Pincus 2004 PACES-b; Smugar 2006 Study 1; Smugar 2006 Study 2; Sowers 2005).

Participants with primary OA were evaluated in all studies. In most of the included studies, the diagnosis of OA was based on valid clinical and radiographic findings in accordance with the American College of Rheumatology (ACR) criteria. The Kellgren and Lawrence radiographic gradation of OA (Kellgren 1957) was used to evaluate the radiographic diagnosis of OA in five studies (Boswell 2008 Study A; Boswell 2008 Study B; Clegg 2006; Pincus 2004 PACES-a; Pincus 2004 PACES-b). One study described clinical OA criteria for inclusion, without mentioning referent criteria (Rother 2007). Mean OA duration was 7.9 years.

Duration of studies ranged from 6 weeks to 12 months.

Birbara 2006 reported two separate randomized studies, but reported combined adverse events for both studies (Birbara 2006 Study 1; Birbara 2006 Study 2). Number of events and participants was therefore proportionally divided between the studies and included as such in our analyses. Bingham 2007 also reported two separate randomized studies, where both studies had two placebo groups, but data for both placebo groups were pooled together and reported as such. Pooled data for both placebo groups were used as such in our analyses (Bingham 2007 Study 1; Bingham 2007 Study 2).

Excluded studies

See [Characteristics of excluded studies](#).

We excluded 53 studies for the following reasons: duration shorter than four weeks (this time frame may be too short to assess harms and benefits of the intervention for osteoarthritis) (Bianchi 2007; Detrembleur 2005; Hawel 2003; Simon 1998; Wittenberg 2006), fewer than 50 participants in each arm (Bianchi 2003; Gallelli 2013; Leeb 2004; NCT02192190; Ozgocmen 2005; Sampalis 2012; Taşcioğlu 2004; Trudeau 2015), reviews summarizing primary studies (Andrews 1999; Angiolillo 2014; Ashcroft 2001; Bensen 2000; Bensen 2000b; Bingham 2008; Bingham 2009; Bingham 2011; Dougados 2007; Johnson 2003; Moore 2009; Sakamoto 2011; Stam 2012; Tive 2015), conference abstracts with irrelevant comparators (Castgillo 2015; Castillo 2014; Chan 2015; Henrotin 2015; Hochberg 2014a; Hochberg 2014b; Pelletier 2016), primary study, but not an RCT (Brereton 2014; Hirayama 2014; Ogunlade 2005; Rozenberg 2008; Tsvetkova 2001; Wolfe 2004), RCT comparing celecoxib with comparators that were irrelevant for this systematic review (Battisti 2004; Essex 2014b; Luyten 2007; Sands 2013; Schnitzer 2015; Stengaard-Pedersen 2004; Strand 2011; Tran 2004; Yoo 2014a; Yoo 2014b), RCT including participants with post-traumatic osteoarthritis (Adegbingbe 2008), RCT with irrelevant dosage of celecoxib and no benefit outcomes of interest for this systematic review (Silverstein 2000) and RCT that included patients with OA

and rheumatoid arthritis patients, but patients with OA were not analyzed separately (Nissen 2016).

Studies awaiting classification

See [Characteristics of studies awaiting classification](#).

We identified eight studies that await classification. We contacted the study sponsor (Pfizer Inc.) to request additional data for the largest trial (Singh 2006) and another study (Lisse 2001); however, Pfizer declined to provide further information. We await provision of data after contacting the authors of Iyengar 2013. The published manuscripts of these studies presented data in way that combined outcome data from the same groups in multiple RCTs (Iyengar 2013; Lisse 2001) or the authors presented data from one study, but pooled data for two tNSAID groups and two doses of celecoxib, and presented pain outcome only per country/region, without providing participant numbers (Singh 2006). One registered clinical trial had results posted online, but not in a peer-reviewed journal (EUCTR2005-002772-14-GB). Information published online was incomplete; our email to Novartis regarding studies on celecoxib did not yield response. One conference abstract appeared eligible,

but we were unable to locate the full text (Mastbergen 2010). Two trials were published during preparation of this manuscript and we will include them in the next update (Essex 2016; Gordo 2017), as well as the third trial for which we initially found only conference abstract and its lead investigator informed us in April 2017 that the study was just accepted in the peer-reviewed journal (Reginster 2016).

Ongoing studies

See [Characteristics of ongoing studies](#).

Four ongoing studies that fit the inclusion criteria were found on ClinicalTrials.gov (NCT01765296; NCT02079727; NCT01768520) and WHO ICTRP (EUCTR2011-005398-22-ES), but no findings were available.

Risk of bias in included studies

All studies had two or more domains that were judged as having unclear or high risk of bias. See the risk of bias graphs (Figure 2; Figure 3).

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies

Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias): Patient-reported outcomes	Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Asmus 2014 Study 1	?	?	+	+	?	-	+	+
Asmus 2014 Study 2	?	?	+	+	?	-	+	+
Bensen 1999	?	?	+	+	?	-	+	?
Bingham 2007 Study 1	?	?	+	+	?	-	+	?
Bingham 2007 Study 2	?	?	+	+	?	-	+	?
Birbara 2006 Study 1	+	?	+	+	?	-	+	?
Birbara 2006 Study 2	+	?	+	+	?	-	+	?
Boswell 2008 Study A	?	?	+	+	?	?	-	?
Boswell 2008 Study B	?	?	+	+	?	?	-	?
Clegg 2006	+	+	+	+	?	?	-	?
Conaghan 2013	+	?	+	+	?	+	+	?
Dahlberg 2009	+	+	+	+	?	-	?	?
DeLemos 2011	?	?	+	+	?	-	+	?
Emery 2008	+	+	+	+	?	-	-	?
Essex 2012a	+	?	+	+	?	-	?	?
Essex 2012b	+	?	+	+	?	-	-	?

Figure 3. (Continued)

Essex 2012a	+	+	+	+	+	+	+	+
Essex 2012b	+	?	+	+	?	-	-	?
Essex 2014	+	?	+	+	+	?	-	+
Fleischmann 2005	?	?	+	+	?	?	-	?
Gibofsky 2003	+	?	+	+	?	-	+	?
Hochberg 2011 Study 307	+	+	+	+	+	?	?	?
Hochberg 2011 Study 309	+	+	+	+	+	?	+	?
Kivitz 2001	?	?	+	+	?	-	?	?
Lehmann 2005	+	?	+	+	?	+	+	?
McKenna 2001a	+	?	+	+	?	?	?	?
McKenna 2001b	?	?	+	+	?	-	+	?
Pincus 2004 PACES-a	?	?	+	+	?	?	-	?
Pincus 2004 PACES-b	?	?	+	+	?	?	-	?
Rother 2007	+	+	+	+	?	+	+	?
Schnitzer 2011	+	?	+	+	+	-	+	?
Sheldon 2005	?	?	+	+	?	-	?	?
Smugar 2006 Study 1	?	?	+	+	?	-	-	?
Smugar 2006 Study 2	?	?	+	+	?	-	-	?
Sowers 2005	+	+	+	+	?	?	-	?
Tannenbaum 2004	?	?	+	+	?	+	-	?
Williams 2000	?	?	+	+	?	-	-	?
Williams 2001	?	?	+	+	?	-	?	?

Allocation

Random sequence generation was adequately described in 17 studies which were assessed as low risk of bias for this domain (Birbara 2006 Study 1; Birbara 2006 Study 2; Clegg 2006; Conaghan 2013; Dahlberg 2009; Emery 2008; Essex 2012a; Essex 2012b; Essex 2014; Gibofsky 2003; Hochberg 2011 Study 307; Hochberg 2011 Study 309; Lehmann 2005; McKenna 2001a; Rother 2007; Schnitzer 2011; Sowers 2005). Thirteen studies used a computer-generated number list as their randomization method (Birbara 2006 Study 1; Birbara 2006 Study 2; Emery 2008; Essex 2012a; Essex 2012b; Essex 2014; Gibofsky 2003; Hochberg 2011 Study 307; Hochberg 2011 Study 309; McKenna 2001a; Rother 2007; Schnitzer 2011; Sowers 2005), three studies used block randomization (Clegg 2006; Conaghan 2013; Dahlberg 2009) and one study used "validated automated system" (Lehmann 2005). We classified 19 studies as having unclear risk of bias for this domain because the method of random sequence generation was not described (Asmus 2014 Study 1; Asmus 2014 Study 2; Bensen 1999; Bingham 2007 Study 1;

Bingham 2007 Study 2; Boswell 2008 Study A; Boswell 2008 Study B; DeLemos 2011; Fleischmann 2005; Kivitz 2001; McKenna 2001b; Pincus 2004 PACES-a; Pincus 2004 PACES-b; Sheldon 2005; Smugar 2006 Study 1; Smugar 2006 Study 2; Tannenbaum 2004; Williams 2000; Williams 2001).

Allocation sequence was not described (unclear risk of bias) in 29 studies. In seven studies participants were allocated to randomized groups centrally (Clegg 2006; Dahlberg 2009; Emery 2008; Hochberg 2011 Study 307; Hochberg 2011 Study 309; Rother 2007; Sowers 2005). No studies were categorized at high risk of bias for selection bias.

Blinding

All studies were assessed at low risk of bias for blinding of participants and personnel, with adequate descriptions for methods. For patient-reported outcomes (pain, physical function) studies were judged as having low risk of detection bias because

participants were blinded. Patients were blinded by using identical pills in all groups.

For non-patient reported outcomes (all safety outcomes) 32 studies presented no information about blinding of outcome assessors and were therefore assessed at unclear risk of bias. Four studies specifically indicated blind assessment of non-patient-reported outcomes ([Essex 2014](#); [Hochberg 2011 Study 307](#); [Hochberg 2011 Study 309](#); [Schnitzer 2011](#)). No studies were assessed at high risk of bias for performance bias or detection bias.

Incomplete outcome data

Most studies were assessed at either high or unclear risk of bias for incomplete outcome data because of high attrition rates or imbalances in attrition between study arms. The last observation carried forward (LOCF) method was used in most studies to impute missing values. One study used baseline observation carried forward (BOCF) for participants with missing values ([Conaghan 2013](#)), two studies used modified intention-to-treat (ITT) analysis ([Bingham 2007 Study 1](#); [Bingham 2007 Study 2](#)); five studies did not indicate methods for handling missing values ([Essex 2014](#); [McKenna 2001a](#); [McKenna 2001b](#); [Rother 2007](#); [Sowers 2005](#)).

Selective reporting

All outcomes mentioned in trials' methods were reported in most studies. However, 11 studies did not provide all outcomes that were indicated in the trials' methods ([Boswell 2008 Study A](#); [Boswell 2008 Study B](#); [Clegg 2006](#); [Essex 2012b](#); [Essex 2014](#); [Fleischmann 2005](#); [Pincus 2004 PACES-a](#); [Pincus 2004 PACES-b](#); [Smugar 2006 Study 1](#); [Smugar 2006 Study 2](#); [Sowers 2005](#)). In one study, the protocol was amended such that change of visual analog scale (VAS) at week 6 rather than week 12 became the primary end point ([Emery 2008](#)). Most studies reported adverse events only if they occurred in more than a certain percentage of participants.

Other potential sources of bias

Use of co-interventions was identified as a potential source of bias in most studies. Three studies did not report use of any co-interventions ([Asmus 2014 Study 1](#); [Asmus 2014 Study 2](#); [Essex 2014](#)); 33 studies reported that participants were permitted to take various co-interventions (some studies indicated allowing more than one co-intervention). Co-interventions included acetaminophen, steroids, aspirin, low-dose aspirin, glucosamine or chondroitin sulphate, antacids, prednisone, anti-platelet agents, corticosteroids, H2-receptor antagonists, physiotherapy, propoxyphene, codeine, tramadol, proton pump inhibitors, cryoprotective agents, topical pain relief, estrogen, progesterone, testosterone. Some studies indicated that participants were not permitted to take co-intervention therapies within 12, 24 or 48 hours before the study visit, mostly for acetaminophen as a rescue medication; there were no limits when participants needed to stop taking co-intervention treatments before a scheduled study visit for other permitted co-interventions. Since none of the 33 studies reported amounts of co-interventions consumed by participants during the study, and it is unclear whether this could have affected benefit and harms outcomes, these studies were assessed at unclear risk of bias.

Effects of interventions

See: [Summary of findings for the main comparison Celecoxib versus placebo for osteoarthritis](#); [Summary of findings 2 Celecoxib versus tNSAIDs for osteoarthritis](#)

Comparison 1: Celecoxib compared with placebo

Studies in this comparison were assessed at low risk of bias for sequence generation, allocation concealment and blinding of participants and personnel.

Benefits

Pain

This was a patient-reported outcome and included data from four studies (N = 1622) with follow-up ranging from 6 weeks to 24 weeks. Celecoxib was statistically significantly superior to placebo, accounting for 3% absolute improvement (95% CI 2% to 5% improvement) or 12% relative improvement (95% CI 7% to 18% improvement) (SMD -0.22, 95% CI -0.32 to -0.12, $P < 0.001$). The number needed to treat to benefit (NNTB) was 11 (95% CI 7 to 18). Heterogeneity was low ($\text{Chi}^2 = 2.11$, $\text{df} = 3$, $P = 0.55$; $I^2 = 0\%$, [Analysis 1.1](#)). Absolute improvement of 3% may not be clinically significant and there is uncertainty about this result (high quality evidence; [Summary of findings for the main comparison](#)).

Physical function

This was a patient-reported outcome and included data from four studies (N = 1622) with follow-up ranging from 6 weeks to 24 weeks. Celecoxib was statistically significantly superior to placebo, accounting for 4% absolute improvement (95% CI 2% to 6% improvement) or 12% relative improvement (95% CI 5% to 19% improvement) (SMD -0.17, 95% CI -0.27 to -0.07, $P < 0.001$; [Analysis 1.2](#)). NNTB was 14 (95% CI 9 to 34). Heterogeneity was low ($\text{Chi}^2 = 1.62$, $\text{df} = 3$, $P = 0.66$, $I^2 = 0\%$). Absolute improvement of 4% may not be clinically significant and there is uncertainty about this result (high quality evidence; [Summary of findings for the main comparison](#)).

Quality of life

No studies reported this outcome.

Numbers of responders with at least 50% reduction in pain score

No studies reported this outcome.

Harms

None of the harms considered for this outcome were patient-reported. All eligible studies were included in analyses. Follow-up ranged from 6 weeks to 24 weeks.

Numbers of participants who withdrew due to adverse events

Based on direct, moderate-quality evidence (downgraded due to risk of bias) we found no evidence of an important difference for withdrawals due to adverse events (Peto OR (can be interpreted as RR due to low event rate) 0.99, 95% CI 0.85 to 1.15; $P = 0.88$; 28 studies, 12,785 participants; [Analysis 1.3](#)). There was no statistically significant difference between celecoxib and placebo. Heterogeneity was low ($\text{Chi}^2 = 34.60$, $\text{df} = 27$, $P = 0.15$, $I^2 = 22\%$).

Numbers of participants who experienced any serious adverse event

This analysis included 28 studies (N = 13,393). Based on direct, very low quality evidence (downgraded due to risk of bias and imprecision), results were inconclusive (Peto OR 0.95, 95% CI 0.66 to 1.36; P = 0.78; 28 studies, 13,393 participants; [Analysis 1.4](#)). There was no statistically significant difference between celecoxib and placebo. Heterogeneity was low ($\chi^2 = 28.39$, df = 25, P = 0.29, $I^2 = 12\%$).

Numbers of participants who experienced gastro-intestinal events (perforation, ulcer, bleeds)

Based on direct, very low quality evidence (downgraded due to risk of bias and imprecision), results were inconclusive (Peto OR 1.91, 95% CI 0.24 to 14.90; P = 0.54; 8 studies, 3263 participants; [Analysis 1.5](#)). There was no statistically significant difference between celecoxib and placebo. Heterogeneity was low ($\chi^2 = 3.93$, df = 3, P = 0.27, $I^2 = 24\%$). Four studies indicated that no gastro-intestinal events of the types specified occurred during the study. In the remaining four studies, three gastro-intestinal events were reported in participants taking celecoxib and one event in a participant taking placebo.

Numbers of participants who experienced cardiovascular events (myocardial infarction, stroke)

Results were inconclusive (very low quality evidence, downgraded due to risk of bias and imprecision). There was no statistically significant difference between celecoxib and placebo (Peto OR 3.40, 95% CI 0.73 to 15.88; P = 0.12; 5 studies, 2947 participants; [Analysis 1.6](#)). Heterogeneity was low ($\chi^2 = 0.54$, df = 4, P = 0.97, $I^2 = 0\%$). There were seven cardiovascular events reported in four studies including one stroke in the celecoxib group ([Clegg 2006](#)), three myocardial infarctions in the celecoxib group ([Rother 2007](#); [Smugar 2006 Study 1](#); [Smugar 2006 Study 2](#)). There were two deaths resulting from cerebrovascular adjudication and an atherosclerotic cardiovascular event in the celecoxib group and one ischemic stroke in the placebo group ([Schnitzer 2011](#)).

Regulatory agencies in Australia, Europe and the UK refer to the USA FDA assessment of increased risk of cardiovascular events ([FDA 2005](#)). An increased risk of cardiovascular events is included in most product labels.

Comparison 2: Celecoxib compared with tNSAIDs

Studies in this comparison were assessed at low risk of bias for sequence generation, allocation concealment and blinding of participants and personnel.

Benefits

Pain

We analysed data from two studies (N = 1180) that compared celecoxib 200 mg/day with naproxen 1000 mg/day with follow-up of 12 weeks ([Sowers 2005](#)) and celecoxib 200 mg/day with diclofenac 100 mg/day with follow-up of 52 weeks ([Dahlberg 2009](#)). Moderate quality evidence (downgraded due to missing data; [Summary of findings 2](#)) provided inconclusive results between celecoxib and tNSAIDs (MD -4.52, 95% CI -10.65 to 1.61; P = 0.15). The pooled analysis indicated 5% absolute improvement (95% CI 11% improvement to 2% worse) and 11% relative improvement (95% CI 26% improvement to 4% worse). Heterogeneity was substantial,

but not statistically significant ($\chi^2 = 2.85$, df = 1, P = 0.09, $I^2 = 65\%$, [Analysis 2.1](#)).

Physical function

Results from one study (N = 264) which compared celecoxib 200 mg/day with naproxen 1000 mg/day with follow-up of 12 weeks ([Sowers 2005](#)) indicated that celecoxib was superior to tNSAID, accounting for 6% absolute improvement (95% CI 6% to 11% improvement) and 16% relative improvement (95% CI 2% to 30% improvement) (MD -6.00, 95% CI -11.40 to -0.60, P = 0.03) in improving physical function. NNTB was 9 (95% CI 5 to 121, [Analysis 2.2](#)). Heterogeneity was not applicable. The improvement may not be of clinical significance (moderate quality evidence due to missing data; [Summary of findings 2](#)).

Quality of life

No studies reported this outcome.

Numbers of responders with at least 50% reduction in pain score

No studies reported this outcome.

Harms

Numbers of participants who withdrew due to adverse events

This analysis included eight studies (N = 3150) that compared celecoxib 200 mg/day with diclofenac 100 mg/day with 52-weeks follow-up ([Dahlberg 2009](#)), diclofenac 150 mg/day with 6-weeks follow-up ([McKenna 2001b](#)), diclofenac 150 mg/day with 12-weeks follow-up ([Emery 2008](#)), naproxen 1000 mg/day with 6-weeks follow-up ([Essex 2012b](#); [Essex 2014](#)) and naproxen 1000 mg/day with 12-weeks follow-up ([Bensen 1999](#); [Kivitz 2001](#); [Sowers 2005](#)). Low quality evidence (downgraded due to missing data, risk of bias, few events and wide confidence intervals; [Summary of findings 2](#)) provided inconclusive results for numbers of participants who withdrew due to adverse events (Peto OR 0.97, 95% CI 0.74 to 1.27, P = 0.82). Heterogeneity was moderate, but not significant ($\chi^2 = 10.63$, df = 7, P = 0.16, $I^2 = 34\%$, [Analysis 2.3](#)).

Numbers of participants who experienced any serious adverse event (SAE)

This analysis included five studies (N = 2404) that compared celecoxib 200 mg/day with diclofenac 100 mg/day with 52-weeks follow-up ([Dahlberg 2009](#)), diclofenac 150 mg/day with 6-weeks follow-up ([McKenna 2001b](#)), diclofenac 150 mg/day with 12-weeks follow-up ([Emery 2008](#)), naproxen 1000 mg/day with 6-weeks follow-up ([Essex 2012b](#)) and naproxen 1000 mg/day with 6-months follow-up ([Essex 2012a](#)). Based on very low quality evidence (downgraded due to missing data, risk of bias, few events and wide confidence intervals; [Summary of findings 2](#)), results were inconclusive for numbers of participants experiencing any SAEs between celecoxib and tNSAIDs (Peto OR 0.92, 95% CI 0.66 to 1.28, P = 0.60). Heterogeneity was moderate, but not significant ($\chi^2 = 5.90$, df = 4, P = 0.21, $I^2 = 32\%$; [Analysis 2.4](#)).

Numbers of participants who experienced gastro-intestinal events (perforation, ulcer, bleeds)

This analysis included four studies (N = 1755) that compared celecoxib 200 mg/day with diclofenac 100 mg/day with 52-weeks follow-up ([Dahlberg 2009](#)), diclofenac 150 mg/day with 12-weeks follow-up ([Emery 2008](#)), naproxen 1000 mg/day with 6-weeks follow-up ([Essex 2014](#)) and naproxen 1000 mg/day with 12-weeks

follow-up (Bensen 1999). The rate of gastro-intestinal events was higher in participants who received tNSAIDs (1.2%, 5/878, 6 per 1000) compared with celecoxib (0.8%, 3/877, 4 per 1000). However, this result was not statistically significant (Peto OR 0.61, 0.15 to 2.43, $P = 0.48$). Heterogeneity was moderate ($\text{Chi}^2 = 4.84$, $df = 3$, $P = 0.18$, $I^2 = 38\%$, Analysis 2.5).

Numbers of participants who experienced cardiovascular events (myocardial infarction, stroke)

This analysis included one study that compared celecoxib 200 mg/day versus diclofenac 100 mg/day with 52-weeks follow-up (Dahlberg 2009). Based on very low quality evidence (downgraded due to missing data, risk of bias, few events and wide confidence intervals; Summary of findings 2), results were inconclusive for numbers of participants who experienced cardiovascular events between celecoxib and tNSAIDs (Peto OR 0.47, 95% CI 0.17 to 1.25, $P = 0.13$; 1 study, 916 participants; Analysis 2.6).

Comparison 3: Celecoxib compared with placebo (all eligible studies)

Differences between comparison 1 (celecoxib vs placebo in studies with low risk of bias) and comparison 3 (celecoxib vs placebo in all eligible studies): in comparisons of celecoxib and placebo there were no significant differences between the analyses (analyses which included studies with low risk of bias on sequence generation, allocation concealment and blinding of participants and personnel, and analyses with all eligible studies).

Benefits

Pain

All reported pain under 24 weeks

This analysis included 31 studies; follow-up ranged from 6 weeks to 24 weeks. Celecoxib was more effective in reducing pain compared with placebo (SMD -0.41, 95% CI -0.49 to -0.32, $P < 0.001$; 31 studies, 13,069 participants; Analysis 3.1). Heterogeneity was high ($\text{Chi}^2 = 163.27$, $df = 30$, $P < 0.001$, $I^2 = 82\%$). Two studies (Bingham 2007 Study 1; Bingham 2007 Study 2) did not report WOMAC pain data for end of study (12 weeks), but averaged data for 2, 4, 8 and 12-week follow-up were provided.

Visual analogue scale (VAS) for pain (0 mm to 100 mm)

Celecoxib was statistically significantly superior in reducing pain compared with placebo at 6 weeks (SMD -0.49, 95% CI -0.64 to -0.34, $P < 0.001$; 11 studies, 3722 participants; $\text{Chi}^2 = 48.62$, $df = 10$, $P < 0.001$, $I^2 = 79\%$, Analysis 3.2); 12 weeks (MD -11.09, 95% CI -12.68 to -9.50, $P < 0.001$; 3 studies, 1226 participants; $\text{Chi}^2 = 1.82$, $df = 2.0$, $P = 0.40$, $I^2 = 0\%$, Analysis 3.3); and 13 weeks (MD -6.35, 95% CI -8.00 to -4.70, $P < 0.001$; 5 studies, 3853 participants; $\text{Chi}^2 = 3.01$, $df = 4$, $P = 0.56$, $I^2 = 0\%$, Analysis 3.4).

Pain on walking VAS (0 to 100 mm) at 6 weeks

Data from four studies informed analysis of this outcome ($N = 1572$). We found evidence of statistically significant differences between celecoxib and placebo in favor of celecoxib (MD -12.42, 95% CI -14.83 to -10.01, $P < 0.001$; $\text{Chi}^2 = 0.15$, $df = 3$, $P = 0.99$, $I^2 = 0\%$, Analysis 3.5).

WOMAC pain

Celecoxib was statistically significantly superior in reducing pain at 6 weeks (MD -3.28, 95% CI -4.51 to -2.06, $P < 0.001$; 12 studies, 4178

participants; $\text{Chi}^2 = 91.72$, $df = 11$, $P < 0.001$, $I^2 = 88\%$, Analysis 3.7); 12 weeks (SMD -0.32; 95% CI -0.40 to -0.23, $P < 0.001$; 9 studies, 3328 participants; $\text{Chi}^2 = 10.64$, $df = 8.0$, $P = 0.22$, $I^2 = 25\%$, Analysis 3.8) (however, two studies (Bingham 2007 Study 1; Bingham 2007 Study 2) did not report WOMAC pain data for end of study (12 weeks), but provided averaged data for 2, 4, 8 and 12-week follow-up); 13 weeks (MD -1.06, 95% CI -1.31 to -0.80, $P < 0.001$; 5 studies, 3853 participants; $\text{Chi}^2 = 3.44$, $df = 4$, $P = 0.49$, $I^2 = 0\%$, Analysis 3.9); and 24 weeks (MD -13.1, 95% CI -24.69 to -1.51, $P = 0.03$; 1 study, 631 participants; Analysis 3.10).

Physical function

All physical function under 24 weeks

Celecoxib was more effective than placebo (SMD -0.35, 95% CI -0.43 to -0.27, $P < 0.001$; 27 studies, 11,940 participants; $\text{Chi}^2 = 113.87$, $df = 26$, $P < 0.001$, $I^2 = 77\%$, Analysis 3.11).

WOMAC physical function

Celecoxib was more effective than placebo at 6 weeks (SMD -0.31, 95% CI -0.41 to -0.22, $P < 0.001$; 12 studies, 4069 participants; moderate, but statistically significant heterogeneity, $\text{Chi}^2 = 22.14$, $df = 11$, $P = 0.02$, $I^2 = 50\%$, Analysis 3.12); 12 weeks (SMD -0.46, 95% CI -0.65 to -0.26, $P < 0.001$; 9 studies, 3185 participants; very high heterogeneity, $\text{Chi}^2 = 54.71$, $df = 8$, $P < 0.001$, $I^2 = 85\%$, Analysis 3.13); 13 weeks (MD -3.70, 95% CI -4.67 to -2.74, $P < 0.001$; 5 studies, 3853 participants; moderate heterogeneity, $\text{Chi}^2 = 5.89$, $df = 4$, $P = 0.21$, $I^2 = 32\%$, Analysis 3.14); at 24 weeks, there was no statistically significant difference between celecoxib 200 mg and placebo (MD -32.6, 95% CI -81.07 to 15.87, $P = 0.19$; 1 study, 631 participants; Analysis 3.15).

Quality of life

One study assessed quality of life (DeLemos 2011, $N = 402$). Data were reported as SF-36 physical component scores (PCS) and mental component scores (MCS). SF-36 PCS was higher in participants taking celecoxib compared to placebo (MD 2.20, 95% CI 0.54 to 3.86, $P = 0.01$, Analysis 3.16). There was no difference in the SF-36 MCS between celecoxib and placebo (MD 0.20, 95% CI -1.46 to 1.86, $P = 0.81$, Analysis 3.17).

Numbers of responders with at least 50% reduction in pain score

Based on data from four studies ($N = 1816$) with 12-week (Bingham 2007 Study 1; Bingham 2007 Study 2; Conaghan 2013) and 24-week follow-up (Clegg 2006), that reported number of responders with at least 50% improvement in WOMAC pain, celecoxib was more effective than placebo (RR 1.47, 95% CI 1.16 to 1.87, $P = 0.001$). Heterogeneity was high ($\text{Chi}^2 = 8.92$, $df = 3$, $P = 0.03$, $I^2 = 66\%$, Analysis 3.18). Data about this outcome for two studies (Bingham 2007 Study 1; Bingham 2007 Study 2) were not published in the primary publication about trials; they were provided by Merck Research Laboratories for the purpose of a meta-analysis published by Moore et al in 2010. Prof. Andrew R. Moore approved usage of these additional data for the purpose of this review (personal communication).

Harms

Analysis 3.19; Analysis 3.20; Analysis 3.21; Analysis 3.22 report the same data as the corresponding analyses in Comparison 1: numbers of participants withdrawn due to adverse events (Analysis 1.3); experiencing any serious adverse event (Analysis 1.4);

experiencing gastro-intestinal events (perforation, ulcer, bleeds) ([Analysis 1.5](#)) and experiencing cardiovascular events (myocardial infarction, stroke) ([Analysis 1.6](#)).

Comparison 4: Celecoxib compared with tNSAIDs (all eligible studies)

Differences between Comparison 2 (celecoxib vs tNSAIDs in studies with low risk of bias) and Comparison 4 (celecoxib vs tNSAIDs in all eligible studies) were analyzed. For most outcomes there were no significant differences between outcomes sets (analyses including studies with low risk of bias for sequence generation, allocation concealment and blinding of participants and personnel, and analyses with all eligible studies). The exception was physical function; absolute improvement of 6% (relative improvement = 16%) in physical function was observed in one eligible study with low risk of bias ([Analysis 2.2](#)), but not in analyses of all eligible studies ([Analysis 4.3](#)).

Benefits

Pain

All pain under 24 weeks

Eight studies (N = 2277) compared celecoxib 200 mg/day with diclofenac 150 mg/day with 6-weeks follow-up ([McKenna 2001b](#)), diclofenac 150 mg/day with 12-weeks follow-up ([Emery 2008](#)), naproxen 1000 mg/day with 6-weeks follow-up ([Essex 2012b](#); [Essex 2014](#)), naproxen 1000 mg/day with 12-weeks follow-up ([Bensen 1999](#); [Kivitz 2001](#); [Sowers 2005](#)) and naproxen 1000 mg/day with 6-months follow-up ([Essex 2012a](#)). Pooled analysis (SMD 0.00, 95% CI -0.10 to 0.10; [Analysis 4.1](#)) showed that there was no statistically significant difference between celecoxib and tNSAIDs (P = 0.93) in pain outcomes in studies that followed participants for fewer than 24 weeks. Heterogeneity was low ($\text{Chi}^2 = 9.90$, df = 7, P = 0.19, $I^2 = 29\%$).

All pain over 24 weeks

There was no difference between celecoxib 200 g/day and diclofenac 100 mg/day (P = 0.24) in pain at 1 year (MD -2.00, 95% CI -5.32 to 1.32, P = 0.24, 1 study, 916 participants; [Analysis 4.2](#)).

Physical function

All physical function under 24 weeks

Seven studies (N = 2176) provided data for this outcome which compared celecoxib 200 mg/day with diclofenac 150 mg/day with 6-weeks follow-up ([McKenna 2001b](#)), naproxen 1000 mg/day with 6-weeks follow-up ([Essex 2012b](#); [Essex 2014](#)), naproxen 1000 mg/day with 12-weeks follow-up ([Bensen 1999](#); [Kivitz 2001](#); [Sowers 2005](#)) and naproxen 1000 mg/day with 6-months follow-up ([Essex 2012a](#)). Pooled analysis (SMD 0.06, 95% CI -0.09 to 0.21, [Analysis 4.3](#)) showed no statistically significant difference between celecoxib and tNSAIDs (P = 0.46) in physical function outcomes in studies that followed participants for fewer than 24 weeks. Heterogeneity was high ($\text{Chi}^2 = 18.58$, df = 6, P = 0.005, $I^2 = 68\%$).

Harms

Numbers of participants who withdrew due to adverse events

Studies compared celecoxib 200 mg/day with diclofenac 100 mg/day with 52-weeks follow-up ([Dahlberg 2009](#)), diclofenac 150 mg/day with 6-weeks follow-up ([McKenna 2001b](#)), diclofenac 150 mg/day with 12-weeks follow-up ([Emery 2008](#)), naproxen 1000 mg/day

day with 6-weeks follow-up ([Essex 2012b](#); [Essex 2014](#)), naproxen 1000 mg/day with 12-weeks follow-up ([Bensen 1999](#); [Kivitz 2001](#); [Sowers 2005](#)) and naproxen 1000 mg/day with 6-months follow-up ([Essex 2012a](#)). There was no statistically significant difference among groups (Peto OR 0.85, 95% CI 0.68 to 1.07, P = 0.42; 9 studies, 3739 participants; $\text{Chi}^2 = 13.56$, df = 8, P = 0.09, $I^2 = 41\%$; [Analysis 4.4](#)).

Numbers of participants who experienced any serious adverse event

Studies compared celecoxib 200 mg/day with diclofenac 100 mg/day with 52-weeks follow-up ([Dahlberg 2009](#)), diclofenac 150 mg/day with 6-weeks follow-up ([McKenna 2001b](#)), diclofenac 150 mg/day with 12-weeks follow-up ([Emery 2008](#)), naproxen 1000 mg/day with 6-weeks follow-up ([Essex 2012b](#)) and naproxen 1000 mg/day with 6-months follow-up ([Essex 2012a](#)). There was no statistically significant difference among groups (Peto OR 0.92, 95% CI 0.66 to 1.28, P = 0.60; 4 studies, 2404 participants; $\text{Chi}^2 = 5.90$, df = 4, P = 0.21, $I^2 = 32\%$; [Analysis 4.5](#)).

Numbers of participants who experienced gastro-intestinal events (perforation, ulcer, bleeds)

Studies compared celecoxib 200 mg/day with diclofenac 100 mg/day with 52-weeks follow-up ([Dahlberg 2009](#)), diclofenac 150 mg/day with 12-weeks follow-up ([Emery 2008](#)), naproxen 1000 mg/day with 6-weeks follow-up ([Essex 2014](#)) and naproxen 1000 mg/day with 12-weeks follow-up ([Bensen 1999](#)). There was no statistically significant difference between celecoxib and tNSAIDs for major gastro-intestinal adverse events (Peto OR 0.61, 95% CI 0.15 to 2.43, P = 0.48; 4 studies, 1755 participants; $\text{Chi}^2 = 4.84$, df = 3, P = 0.18, $I^2 = 38\%$; [Analysis 4.6](#)).

Numbers of participants who experienced cardiovascular events (myocardial infarction, stroke)

There was no statistically significant difference between celecoxib 200 mg/day and diclofenac 100 mg/day with 52-weeks follow-up for cardiovascular events (Peto OR 0.47, 95% CI 0.17 to 1.25, P = 0.13; 1 study, 916 participants; [Analysis 4.7](#)).

Comparison 5: Celecoxib compared with naproxen 1000 mg

Differences between Comparison 5 (celecoxib vs naproxen 1000 mg/day) and Comparison 4 (celecoxib vs tNSAIDs in all eligible studies) were analyzed. There were no differences in results.

Benefits

Pain

All pain up to 24 weeks

There was no statistically significant difference between celecoxib and naproxen (SMD -0.04, 95% CI -0.14 to 0.05; 6 studies, 1781 participants; P = 0.35; $\text{Chi}^2 = 2.23$, df = 5, P = 0.82, $I^2 = 0\%$; [Analysis 5.1](#)).

VAS for pain (0 mm to 100 mm)

There was no statistical difference in pain measured with VAS between celecoxib 200 mg and naproxen 1000 mg/day at 6 weeks (MD 0.55, 95% CI 3.97 to 5.07, P = 0.81; 2 studies, 398 participants; $\text{Chi}^2 = 0.34$, df = 1, P = 0.56, $I^2 = 0\%$; [Analysis 5.2](#)).

There was no statistical difference in pain measured with VAS between celecoxib 200 mg and naproxen 1000 mg/day at 12 weeks (MD -0.97, 95% CI -2.69 to 0.76, P = 0.27; 3 studies, 975 participants; $\text{Chi}^2 = 0.68$, df = 2, P = 0.71, $I^2 = 0\%$; [Analysis 5.3](#)).

Celecoxib for osteoarthritis (Review)

WOMAC pain

We did not find evidence of a clear difference between celecoxib 200 mg and naproxen 1000 mg at 6 weeks (MD 0.35, 95% CI -0.54 to 1.23, $P = 0.44$; 2 studies, 503 participants; $\text{Chi}^2 = 1.26$, $df = 1$, $P = 0.26$, $I^2 = 20\%$, [Analysis 5.4](#)) or at 12 weeks (MD -0.103, 95% CI -0.65 to 0.60, $P = 0.93$; 3 studies, 1073 participants; $\text{Chi}^2 = 2.84$, $df = 2$, $P = 0.24$, $I^2 = 30\%$, [Analysis 5.5](#)). There was no statistically significant difference at 6 months (MD 0.10, 95% CI -0.46 to 0.66, $P = 0.72$; 1 study, 310 participants; [Analysis 5.6](#)).

Physical function

All physical function up to 24 weeks

There was no statistically significant difference between celecoxib and naproxen at up to 24 weeks (SMD -0.01, 95% CI -0.18 to 0.16; $P = 0.93$; 6 studies, 1817 participants). Heterogeneity was substantial ($\text{Chi}^2 = 16.32$, $df = 5$, $P = 0.006$, $I^2 = 69\%$, [Analysis 5.7](#)).

WOMAC physical function

There was no statistically significant difference between celecoxib 200 mg and naproxen 1000 mg in physical function at 6 weeks (MD -2.99, 95% CI -6.43 to 0.44, $P = 0.08$; 2 studies, 313 participants; $\text{Chi}^2 = 1.07$, $df = 1$, $P = 0.30$, $I^2 = 7\%$, [Analysis 5.8](#)); 12 weeks (MD -1.27, 95% CI -4.32 to 1.78, $P = 0.41$; 3 studies, 1073 participants) but heterogeneity was substantial ($\text{Chi}^2 = 19.53$, $df = 2$, $P < 0.001$, $I^2 = 89\%$, [Analysis 5.9](#)); or at 6 months (MD 0.30, 95% CI -1.94 to 2.54, $P = 0.79$; 1 study, 301 participants; [Analysis 5.10](#)).

Quality of life

No studies reported this outcome.

Numbers of responders with at least 50% reduction in pain score

No studies reported this outcome. However, [Essex 2012a](#) reported number of responders, defined as at least 20% decrease in WOMAC pain score; there was no statistically significant difference between celecoxib 200 mg/day and naproxen 1000 mg/day (RR 1.1, 95% CI 0.94 to 1.28; $P = 0.22$).

Harms

Numbers of participants who withdrew due to adverse events

There was no statistically significant difference between celecoxib 200 mg and naproxen 1000 mg (pooled OR 0.81, 95% CI 0.54 to 1.23; 6 studies, 2173 participants; $P = 0.33$). Heterogeneity was moderate, but not statistically significant ($\text{Chi}^2 = 8.60$, $df = 5$, $P = 0.13$, $I^2 = 42\%$, [Analysis 5.11](#)).

Numbers of participants who experienced any serious adverse event

We did not find evidence of a clear difference between treatments (Peto OR 1.11, 95% CI 0.45 to 2.75, $P = 0.83$; 2 studies, 841 participants; $\text{Chi}^2 = 0.00$, $df = 1$, $P = 0.94$, $I^2 = 0\%$, [Analysis 5.12](#)). There were no subgroups in this analysis.

Numbers of participants who experienced gastro-intestinal events (perforation, ulcer, bleeds)

We did not find evidence of a clear difference between treatments (Peto OR 0.37, 95% CI 0.05 to 2.62, $P = 0.32$; 2 studies, 587 participants; $\text{Chi}^2 = 0.33$, $df = 1$, $P = 0.56$, $I^2 = 0\%$, [Analysis 5.13](#)). One event was observed among participants taking celecoxib and three among those taking naproxen 1000 mg.

Numbers of participants who experienced cardiovascular events (myocardial infarction, stroke)

No studies reported this outcome.

Comparison 6: Celecoxib compared with diclofenac 100 mg

We compared results of Comparison 6 (celecoxib vs. diclofenac 100 mg/day) with analyses from Comparison 4 (celecoxib vs. all tNSAIDs). There were no differences.

Benefits

Pain

There was no statistically significant difference between celecoxib 200 mg and diclofenac 100 mg in VAS pain after 1 year of therapy (MD -2.0, 95% CI -5.32 to 1.32; $P = 0.24$; 1 study, 916 participants; [Analysis 6.1](#)).

Physical function

No studies reported this outcome.

Quality of life

No studies reported this outcome.

Numbers of responders with at least 50% reduction in pain score

No studies reported this outcome.

Harms

Numbers of participants who withdrew due to adverse events

There was no clear difference between celecoxib 200 mg and diclofenac 100 mg (Peto OR 1.44, 95% CI 0.80 to 2.61, $P = 0.23$; 1 study, 916 participants; [Analysis 6.2](#)).

Numbers of participants who experienced any serious adverse event

There was no clear difference between celecoxib and diclofenac 100 mg (OR 0.90, 95% CI 0.62 to 1.30, $P = 0.57$; 1 study, 916 participants; [Analysis 6.3](#)).

Numbers of participants who experienced gastro-intestinal events (perforation, ulcer, bleeds)

There was no statistically significant difference between celecoxib 200 mg and diclofenac 100 mg (Peto OR 0.14, 95% CI 0.01 to 2.16, $P = 0.16$; 1 study, 916 participants; [Analysis 6.4](#)).

Numbers of participants who experienced cardiovascular events (myocardial infarction, stroke)

There was no statistically significant difference between celecoxib 200 mg and diclofenac 100 mg (Peto OR 0.47, 95% CI 0.17 to 1.25, $P = 0.13$; 1 study, 916 participants; [Analysis 6.5](#)).

Comparison 7: Celecoxib compared with diclofenac 150 mg

We compared results of Comparison 7 (celecoxib vs. diclofenac 150 mg/day) with analyses from Comparison 4 (celecoxib vs. all tNSAIDs). There were no differences.

Benefits

Pain

VAS for pain (0 mm to 100 mm) at 6 weeks

There was no statistically significant difference between celecoxib 200 mg and diclofenac 150 mg (MD 1.90 (95% CI -3.68 to 7.48, $P = 0.58$; 1 study, 398 participants; [Analysis 7.1](#)).

Pain WOMAC at 6 weeks

There was no clear difference between celecoxib and diclofenac 150 mg (MD 0.3, 95% CI -0.52 to 1.12; 1 study, 398 participants; [Analysis 7.2](#)).

Pain on walking at 12 weeks

Diclofenac 150 mg was statistically significantly better in reducing pain on walking compared with celecoxib 200 mg (MD 13.00 (95% CI 3.11 to 22.89; 1 (12 week) study, 98 participants; [Analysis 7.3](#)).

Physical function

There was no statistically significant difference between celecoxib 200 mg and diclofenac 150 mg on physical function measured on the WOMAC scale after 6 weeks follow up (MD 1.90, 95% CI -1.26 to 5.06, $P = 0.24$; 1 study, 398 participants; [Analysis 7.4](#)).

Quality of life

No studies reported this outcome.

Numbers of responders with at least 50% reduction in pain score

No studies reported this outcome.

Harms

Numbers of participants who withdrew due to adverse events

There was no statistically significant difference between celecoxib 200 mg and diclofenac 150 mg in withdrawals due to adverse events

(pooled OR (random-effects model) 0.78, 95% CI 0.46 to 1.32, $P = 0.35$; 2 studies, 650 participants; $\text{Chi}^2 = 1.11$, $df = 1$, $P = 0.29$, $I^2 = 10\%$, [Analysis 7.5](#)).

Numbers of participants who experienced any serious adverse event

There was no statistically significant difference between celecoxib 200 mg and diclofenac 150 mg in withdrawals due to adverse events (Peto OR 0.79, 95% CI 0.21 to 2.93, $P = 0.72$; 2 studies, 649 participants; substantial heterogeneity $\text{Chi}^2 = 5.66$, $df = 1$, $P = 0.02$, $I^2 = 82\%$; [Analysis 7.6](#)).

Numbers of participants who experienced gastro-intestinal events (perforation, ulcer, bleeds)

There was no statistically significant difference between celecoxib 200 mg and diclofenac 150 mg in gastro-intestinal events including perforation, ulcer and bleeds (Peto OR 7.45, 95% CI 0.46 to 119.74, $P = 0.16$; 1 study, 252 participants; [Analysis 7.7](#)).

Numbers of participants who experienced cardiovascular events (myocardial infarction, stroke)

No studies reported this outcome.

Publication bias

The potential for small-study effects was explored using funnel plots in outcomes with more than 10 studies; this included pain ([Analysis 3.1](#); [Analysis 3.2](#); [Analysis 3.7](#)), physical function ([Analysis 3.11](#); [Analysis 3.12](#)), numbers of withdrawals due to adverse events ([Analysis 1.3](#); [Analysis 3.19](#)), numbers experiencing any serious side effects ([Analysis 1.4](#); [Analysis 3.20](#)), and numbers of participants experiencing gastro-intestinal events ([Analysis 1.5](#)). Based on the appearance of these funnel plots, we judged that the pooled results for these outcomes were not biased by small-study effects. Examples of these funnel plots are shown in [Figure 4](#) and [Figure 5](#).

Figure 4. Funnel plot of comparison: 1 Celecoxib versus placebo, outcome: 1.3 Number withdrawn due to adverse events.

Figure 5. Funnel plot of comparison: 1 Celecoxib versus placebo, outcome: 1.4 Number experiencing any serious adverse events.

However, because there are missing data for thousands of people who participated in studies, this presents serious publication bias. We therefore downgraded the quality of evidence for outcomes in the main comparisons of celecoxib and tNSAIDs (comparisons of studies that were judged as having low risk of bias).

Sensitivity analyses

We conducted sensitivity analyses to assess risk of bias. Because analyses of benefit outcomes for all included studies showed high statistical heterogeneity, we run sensitivity analysis on both comparisons of the patient-reported outcomes for studies assessed at low risk of bias for three domains (randomization, allocation concealment and blinding of participants and personnel). In comparisons of celecoxib and placebo there were no statistically significant differences in these analyses. Sensitivity analyses of celecoxib and tNSAIDs showed no differences in all benefit and harms outcomes. However, analysis of low risk of bias studies only showed that celecoxib was significantly better than tNSAIDs in improving physical function. One low risk of bias study showed that celecoxib caused significantly fewer gastro-intestinal adverse events compared with tNSAIDs.

Sensitivity analyses were also conducted to test if exclusion of studies with imputed standard deviations (Kivitz 2001; McKenna 2001a; Boswell 2008 Study A; Boswell 2008 Study B) influenced results. Exclusion of these studies did not statistically significant impact effect sizes.

Sensitivity analyses to estimate the effects of including Singh 2006 could have on review results were also conducted. Singh 2006 was by far the largest study we found (N = 13,274). However, the report available does not present summary data for each study arm. The study sponsor, Pfizer, declined to provide data for inclusion in this systematic review. We conducted three analyses with the data that could be extracted. Data for numbers of gastro-intestinal and cardiovascular events were provided for celecoxib 200 mg daily and 400 mg daily combined, and for diclofenac and naproxen combined. This is significant because we included only analyses of celecoxib 200 mg daily because this is now the clinically recommended dose of celecoxib for people with osteoarthritis. Data for these combined groups were added to relevant analyses in the comparison of celecoxib vs tNSAIDs. For numbers of participants who experienced gastro-intestinal events (Analysis 2.5) statistical significance changed after adding data from Singh 2006 study. After inclusion of data from Singh 2006, the number of participants totaled 14,949; total numbers of events in the celecoxib group was 86/9677 (0.9%) and 66/5272 (1.3%) in the tNSAIDs group. Statistically significant difference was then observed in favor of celecoxib (OR 0.68, 95% CI 0.49 to 0.93, P = 0.02, no heterogeneity). Without data from Singh 2006, there was no difference between groups (Analysis 2.5). Data on cardiovascular events from Singh 2006 were added to Analysis 2.6 (N = 14,110). Numbers of events in the celecoxib group was 29/9258 (0.31%) and 18/4852 (0.37%) in the tNSAIDs group. There was no statistically significant difference between groups when data from Singh 2006 were added (OR 1.03, 95% CI 0.56 to 1.91, P = 0.92); heterogeneity

was statistically significant (I^2 73%, $P = 0.05$). Analysis including generic inverse variance for VAS pain (100 mm scale) comparing celecoxib 200 mg daily and tNSAIDs was also conducted. However, these data were not presented as a summary effect; [Singh 2006](#) presented mean differences and 95% CIs for participants from 15 regions separately, but numbers of participants per region was not reported. When data from all 15 regions were combined using generic inverse variance, results were statistically significant in favor of tNSAIDs (pooled MD 1.21, 95% CI 0.06 to 2.35, $P = 0.03$). Heterogeneity was low and not statistically significant ($\chi^2 = 17.35$, $df = 14$, $P = 0.24$, $I^2 = 19\%$). We did not attempt to add these data to other analyses because numbers of participants were not available. However, considering the large number of participants in [Singh 2006](#), it is possible that addition of these data would change results in favor of tNSAIDs.

Exploring potential sources of heterogeneity

Because analyses of all included studies showed statistically significant heterogeneity, we analyzed baseline participants' variables in detail to identify any systematic differences among studies. However, we did not find any consistent differences that could explain heterogeneity. Furthermore, for assessment of pain in celecoxib versus placebo we analyzed summary effects separately at three follow-up times (6 weeks, 12/13 weeks and 24 weeks) and for two different pain instruments (pain VAS and pain WOMAC) ([Table 2](#)). We did not find reduced heterogeneity using either of the pain instruments or follow-up times.

DISCUSSION

Summary of main results

We included 36 randomized controlled trials (RCTs) that assessed the benefits and harms of celecoxib for people with osteoarthritis (OA). We found that compared with placebo or traditional non-steroidal anti-inflammatory drugs (tNSAIDs), celecoxib provided slight symptomatic benefit for people with OA of the knee, hip, or both knee and hip; however, it is unlikely that the small benefit has clinical significance. Benefits ranged from 3% to 6% improvement, whereas 15% improvement is considered to be a minimal clinically important difference (MCID) ([Dworkin 2008](#)).

Harms outcomes were inconclusive for celecoxib versus placebo or celecoxib versus tNSAIDs, with the exception of withdrawals due to adverse events in the placebo comparison in which there was no evidence of an important difference. These results were assessed as providing low to very low quality evidence. Quality of evidence for most outcomes was downgraded due to risk of bias, imprecision and publication bias. Considering that most studies had high or unclear risk of bias on at least one domain, harms outcomes suffered from imprecision due to wide confidence intervals and few events that were reported. Furthermore, we were unable to obtain results from studies that included 15,539 participants, and hence, our confidence in results presented is not high.

Our decision to provide separate analyses for low risk of bias studies that assessed patient-reported outcomes was made because of high heterogeneity observed in analyses for all applicable studies. We investigated baseline variables of participants, but we were unable to identify variables that could explain statistical heterogeneity. We also conducted separate analyses for pain outcomes at different follow-up times and a range of pain

measurement instruments. Except for showing declining effect estimates over time, this analysis did not explain sources of statistical heterogeneity. Because high risk of bias studies can also be relevant to inform readers about the robustness of findings, we also presented separate analyses for patient-reported outcomes. There were no major differences between these analyses.

Overall completeness and applicability of evidence

This review had several limitations. Most studies were short (6 weeks to 13 weeks duration). Only three of 36 studies were longer than 13 weeks, including two with 6-month durations ([Clegg 2006](#); [Essex 2012a](#)) and one 12-month study ([Dahlberg 2009](#)). Most studies permitted additional use of analgesics and other co-interventions such as antacids, topical therapies and physical therapy during the trial, which could have influenced results related to pain and evaluation of adverse effects. Furthermore, OA can affect joints other than hips and knees, but these were not evaluated in this review.

Most studies were conducted in high-income countries in North America and Europe. Two studies specified that the study was conducted in low- and middle-income South American countries ([Fleischmann 2005](#); [Sowers 2005](#)). Three studies did not specify all study locations, indicating that the study was conducted "internationally" ([Boswell 2008 Study A](#); [Boswell 2008 Study B](#)), or that the study was conducted in "15 countries" ([Lehmann 2005](#)). Two studies specifically targeted minority participants and included only African-Americans ([Essex 2012b](#)) or Hispanic participants ([Essex 2014](#)). All other studies included predominantly white participants, ranging from 71% to 99% Caucasians in the study. It is known that there are complex pain disparities related to race and ethnicity ([Green 2003](#)). There are also racial and ethnic differences in the experience of chronic pain, which may be clinically meaningful ([Riley 2002](#)). Therefore, findings of included studies may not be applicable to people with OA of all races and ethnicities.

Two studies were conducted with participants who had previously failed naproxen (at least 750 mg/day for 2 weeks) and ibuprofen (at least 1200 mg/day for 2 weeks) within the past five years due to either lack of benefit, tolerability or both ([Asmus 2014 Study 1](#); [Asmus 2014 Study 2](#)). These participants are not representative of the entire OA population.

All studies listed multiple morbidities as exclusion criteria. Therefore, review findings may not be applicable to people with OA who have multiple morbidities. It is known that RCTs regularly exclude participants who have multiple morbidities, which is not representative of the general population that a physician would encounter in clinical practice. Research results intended for application in medical practice should take the complex reality of effective treatment of these participants into consideration ([Fortin 2006](#)).

Previous meta-analyses on coxibs and tNSAIDs have shown that these drugs are associated with an increased risk of cardiovascular disease and upper gastro-intestinal complications, but the nature and magnitude of these risks was unclear, especially in participants at increased risk of coronary heart disease ([Kearney 2006](#); [McGettigan 2011](#); [Ofman 2002](#); [Rostom 2007](#); [Solomon 2005](#); [Trelle 2011](#)). It has been suggested that the vascular risk might be highest for celecoxib in people at greatest risk of coronary heart disease

(Solomon 2008). The included studies did not permit analysis of specified outcomes in this subset of the population.

In the included studies, women of childbearing age were regularly excluded if they were pregnant, lactating or not using adequate contraception. The findings of this review may not be applicable to these women.

Only one included study provided data for quality of life (DeLemos 2011). Four studies reported analysis of quality of life in their respective methods sections (Clegg 2006; Fleischmann 2005; Pincus 2004 PACES-a; Pincus 2004 PACES-b), but none reported outcomes.

Systematic reviews can identify gaps that help in developing new evidence-based research and finding high-priority research needs. This gap can be characterized by analysing participants, interventions, comparisons and outcomes (PICOs) of included studies, and if the gap exists because of insufficient or imprecise information, biased information, inconsistency or unknown consistency, or not the right information (Robinson 2011; Saldanha 2013). Based on the analyzed evidence, further research about the benefits and harms of celecoxib for people with osteoarthritis should include participants from different ethnicities, those with multiple morbidities and those not taking co-interventions. Other desirable characteristics are: more head-to-head comparisons with active comparators, relevant outcomes recommended for osteoarthritis, and study designs to accomplish low levels of risk of bias, and that are independently funded.

During preparation of this review, the number of responders with at least 50% pain relief was included as a new outcome to keep in line with the latest guidance on reporting pain outcomes (Busse 2015). Only four of the included trials reported this outcome; three with 12-week follow-up and one with 24-week follow-up. The one with the longest follow-up (Clegg 2006) reported the smallest effect and as much as 42% placebo response rate. Future trials should include 50% pain relief as an important outcome.

Quality of the evidence

Quality of the overall evidence showed serious limitations. None of the studies were assessed at low risk of bias on all seven domains. Poor reporting was a major weakness of the included studies, because many did not report relevant details about methods, leading to judgments of unclear risk of bias. Furthermore, attrition was very high. Most included studies had attrition rates over 30% in at least one arm, and a third of studies had attrition rates between 20% and 30% (Characteristics of included studies). Regardless of the imputation method used for missing data, the high number of participants lost from follow-up is a cause for concern, especially because these were short-term studies. Only one study had a duration over six months. Furthermore, most studies permitted participants to take multiple co-interventions which may have influenced results.

The GRADE approach was used to rate overall quality of evidence for the key outcomes presented in the Summary of findings tables. For patient-reported outcomes we included only studies assessed at low risk of bias on the first three domains. In the comparison of celecoxib versus placebo (Summary of findings for the main comparison) pain and physical function outcomes were rated as high quality evidence; in the comparison of celecoxib versus

tNSAIDs (Summary of findings 2) quality of pain and physical function outcomes were downgraded one level because a large amount of data were missing from studies that were published, but not adequately reported.

Evidence quality of harms outcomes was assessed as low to very low. These studies were assessed at unclear or high risk of bias on at least one domain, missing data, few reported events and imprecision due to wide confidence intervals that included statistically significant effects for both intervention and control. Therefore, downgrading evidence levels related to problems with risk of bias, publication bias and imprecision.

Analyses of patient-reported outcomes revealed statistically significant heterogeneity that we could not explain following systematic analysis of participants' baseline characteristics, different pain instruments, or follow-up times.

We were unable to obtain relevant data for three manuscripts, which reported multiple studies that involved a total of 15,539 participants. Attempts to contact corresponding authors (of studies funded by GlaxoSmithKline, Merck and Novartis) and sponsor (Pfizer) with requests to provide missing data were unsuccessful. This meant that a significant amount of data; especially considering that the 36 RCTs included in this review represent data from a total of 17,206 participants. Our attempts to conduct sensitivity analyses indicated those data could have a significant impact on overall conclusions of this review. Release of missing data and further research is very likely to have an important impact on our confidence in the estimate of effect, and is likely to change the estimate. Furthermore, it is likely that review conclusions about benefits and harms may change following inclusion of relevant information from those studies.

Potential biases in the review process

Strengths of this systematic review include detailed electronic searching; all important databases were included in the search. At least two independent authors were involved in all steps of the review, including screening and data extraction. A third author was consulted when necessary to achieve consensus.

We conducted two sets of analyses for patient-reported outcomes in studies that compared celecoxib versus placebo and celecoxib versus all tNSAIDs. Analysis was conducted including studies assessed at low risk of bias for three domains (sequence generation, allocation concealment, and blinding of participants and personnel) and repeated to include all eligible studies. Our definition for low risk of bias may be considered arbitrary. However, there is no universal definition for a low risk of bias study. Nevertheless, these two approaches did not result in major differences in results. In comparisons of celecoxib and placebo there were no differences in pooled analyses between studies with low risk of bias for patient-reported outcomes and pooled analyses of all eligible studies. In comparisons of celecoxib and tNSAIDs, the only difference in pooled analyses between studies with low risk of bias for patient-reported outcomes and pooled analyses of all eligible studies was observed for one outcome - physical function. We found 6% absolute improvement with celecoxib compared to tNSAIDs in low risk of bias studies and no difference in all eligible studies. Improvement of 6% cannot be considered clinically significant, and therefore, it could be argued

there was no difference between analyses for celecoxib versus tNSAIDs.

Limitations of this systematic review include lack of available data due to unsuccessful requests for missing data and clarifications made to study authors and sponsors. Attempts to contact corresponding authors either did not yield response, or replies that requested data were no longer available. Several study authors indicated that data need to be requested from the study sponsors. Data from two studies sponsored by Pfizer were requested using a complex Investigator-Initiated Research (IIR) request because data were not presented for each randomized group separately; data from several groups were combined in the published manuscripts (Lisse 2001; Singh 2006). Pfizer refused to provide data for both studies. The two studies included a total of 14,042 participants and comparisons of celecoxib 200 mg with placebo, naproxen and diclofenac. Considering lack of head-to-head comparisons, inclusion of data from those studies in this systematic review would be very valuable.

Current spotlight on transparency of data from clinical trials, through the AllTrials campaign, stipulates that "All trials should be registered, with their full methods and results reported, and routine audit on the extent of information withheld" (Goldacre 2014). However, attempts to obtain even aggregate data for each randomized group in studies considered for this systematic review were unsuccessful.

Only one included study was not sponsored by a pharmaceutical company (Clegg 2006); one did not report information about funding, but because its authors were industry employees, it was also likely to have been industry funded (Sheldon 2005). Of the 36 studies in this review, 34 included one or more study authors who were company employees. It is known that sponsorship by the manufacturing company leads to more favorable results and conclusions than sponsorship from other sources (Lundh 2012). Therefore, conclusions of this systematic review should be interpreted cautiously.

Although we did not find evidence of publication bias in our analyses, it is possible that a number of negative studies was not published because almost all studies were linked to the pharmaceutical industry.

Agreements and disagreements with other studies or reviews

In 2002, a systematic review investigating the benefits, tolerability and upper gastro-intestinal safety of celecoxib for both people with OA and those with rheumatoid arthritis (RA) was published (Deeks 2002). The review included published and unpublished studies of at least 12 week duration. Efficacy outcomes at 12 weeks showed that celecoxib resulted in statistically significant improvement in all components of the WOMAC scale, including pain and physical function sub scales, compared to placebo. There were no statistically significant differences between celecoxib and naproxen 500 mg twice daily, with both drugs equally effective for all WOMAC sub scales. Those results cannot be easily compared with our results for two reasons. In their efficacy analyses for comparison of celecoxib and placebo in OA Deeks 2002 included three studies (Bensen 1999, Zhao 1999 and unpublished Study 054, referred to in the literature as *Study 054. Integrated clinical and statistical report for a double-blind placebo controlled, randomised*

comparison study of the efficacy and safety of SC-58635 50 mg, 100 mg and 200 mg BID and naproxen 500 mg BID in treating the signs and symptoms of osteoarthritis of the hip. Pharmacia:Data on file. 1997). However, despite slight differences in data presentation, Bensen 1999 and Zhao 1999 are two reports of the same study; hence, Deeks 2002 unintentionally double-counted participants from the same study. This was noted in a letter from Jüni 2003 published in the British Medical Journal (BMJ). In their response in the same BMJ issue, Deeks et al wrote "We can confirm that the errors Jüni et al point out are not duplicate publications but simple labelling errors that do not affect the results". Furthermore, Study 054 had arms receiving celecoxib 200 mg/day and 400 mg/day and both groups were analyzed together for efficacy outcomes, as judged by the numbers presented in Figure 1. Therefore, this pooled analysis cannot be compared to our analysis of celecoxib 200 mg/day alone. Harms data cannot be compared because harms outcomes in Deeks 2002 were shown for both OA and RA participants combined. It is unclear if Deeks 2002 exhibited sponsorship-related risk of bias because it was funded by Pfizer and Searle pharmaceutical companies.

A 2005 study assessed tolerability and adverse events in clinical trials of celecoxib in OA and RA in a systematic review and meta-analysis of data from company clinical trial reports (Moore 2005). Studies lasting two or more weeks were included. Doses of celecoxib ranged from 50 mg to 800 mg daily. However, data were not presented for OA and RA participants separately. The study concluded, among other findings, that celecoxib was associated with higher incidence of myocardial infarction, but this was not statistically significant. Moore 2005 was sponsored by Pfizer and one author was a Pfizer employee.

A 2006 systematic review and meta-analysis on the risk of cardiovascular events associated with celecoxib included studies of at least six weeks duration (Caldwell 2006). However, celecoxib treatment in the included studies was administered for a wide range of medical conditions including RA, OA, Alzheimer's disease, and the prevention of colorectal adenoma in people at high risk. Six studies were included in the meta-analysis. Caldwell 2006 concluded that the use of celecoxib was associated with a 2.26-fold increased risk of myocardial infarction when compared with placebo, and a 1.88-fold increased risk of myocardial infarction when compared with all comparator treatment groups. The published review indicated that no pharmaceutical company had any role in this systematic review (Caldwell 2006).

A 2008 systematic review of RCTs analyzing cyclooxygenase-2 selective non-steroidal anti-inflammatory drugs (NSAIDs) for OA and RA included 40 RCTs (Chen 2008). Data were not presented separately for OA and RA. The summary conclusion for both diseases was that celecoxib in doses ranging from 200 mg to 800 mg per day, compared with non-selective NSAIDs, has equal benefit and superior gastro-intestinal tolerability, but statistically significant higher risk of myocardial infarction (RR 1.77, 95% CI 1.00 to 3.11). This study was conducted by NIHR Health Technology Assessment program (Chen 2008).

A 2012 systematic review of RCTs analyzing the benefit of etoricoxib, celecoxib, lumiracoxib, non-selective NSAIDs, and acetaminophen in OA showed that celecoxib 200 mg was more efficacious than placebo for pain and physical function. Comparisons of celecoxib with active comparators were less clear. This systematic review was

focused mainly on etoricoxib and was sponsored by Merck (Stam 2012).

In 2013, a meta-analysis of vascular and upper gastro-intestinal effects of NSAIDs, based on individual participant data (IPD) from randomized trials, was published (CNT 2013). CNT 2013 included data from trials that analyzed daily celecoxib doses of 100 mg, 200 mg, 400 mg and 800 mg daily. The IPD analysis found that celecoxib statistically significantly increased the risk of major cardiovascular events (RR 1.36, 95% CI 1.00 to 1.84), with smaller proportional additional risk of major vascular events with lower celecoxib doses in placebo-controlled trials. The study also found that the higher doses of celecoxib yield larger proportional excesses in risk of ulcers. The conclusion was that although there was a trend towards less risk with lower celecoxib doses, the vascular effects of celecoxib 200 mg daily (the most widely used coxib regimen) were statistically uncertain. This analysis was funded by the UK Medical Research Council and British Heart Foundation (CNT 2013).

A 2016 network meta-analysis assessed the effectiveness of NSAIDs for the treatment of knee and hip pain in people with osteoarthritis (Da Costa 2016). Da Costa 2016 included 74 RCTs analyzing NSAIDs and paracetamol. Celecoxib 200 mg/day was the most frequently investigated intervention, with the most randomized participants. The network meta-analysis concluded that diclofenac at the maximum daily dose of 150 mg was the most effective intervention for the treatment of pain and physical disability in osteoarthritis, and superior to the maximum doses of frequently used NSAIDs, including ibuprofen, naproxen, and celecoxib (Da Costa 2016). In this systematic review, we found only two studies that directly compared diclofenac 150 mg/day with celecoxib 200 mg/day (Emery 2008; McKenna 2001b). Both studies included data for only some of the outcomes of interest for this review, and indicated that there were no difference between celecoxib and diclofenac in pain at six weeks, but diclofenac 150 mg/day was better in reducing pain on walking at 12 weeks. There was no difference in physical function or harms outcomes. Furthermore, Da Costa 2016 reported that although celecoxib 200 mg/day improved osteoarthritic pain and physical function outcomes when compared with placebo, there was no statistical evidence to support minimum clinically important effect for both outcomes (Da Costa 2016). We reached the same conclusion in this systematic review; although celecoxib 200 mg/day was statistically superior than placebo, absolute improvement of 3% for pain and 4% for physical function is unlikely to be clinically significant.

In conclusion, benefit data on celecoxib versus placebo and active comparators are in line with previous systematic reviews. Harms analyses in this review indicate higher numbers of serious adverse events, major gastro-intestinal and cardiovascular events in celecoxib-treated participants, but without statistically significant differences compared with placebo or active comparators – probably due to few participants with these events.

AUTHORS' CONCLUSIONS

Implications for practice

We found that celecoxib was slightly better than placebo and select traditional non-steroidal anti-inflammatory drugs (tNSAIDs) in alleviating pain and improving physical function, but it is unlikely that these small improvements have clinical significance. These findings may not be applicable to people of all races and ethnicities, those with comorbidities, risk of coronary heart disease and pregnant or lactating women. Results were inconclusive for harms between celecoxib and placebo or celecoxib and tNSAIDs. There are known risks of harm (e.g. cardiovascular events) with NSAID use overall in the general population, but we are uncertain about differences between individual NSAIDs, and uncertain about the risk of celecoxib in persons with osteoarthritis. However, we cannot be confident about these results due to observed risk of bias, potential favorable results related to industry funding, and that attempted contact with corresponding authors (of studies funded by GlaxoSmithKline, Merck and Novartis) to elicit data and clarifications were unsuccessful. Pfizer refused to release data from completed studies which included large numbers of participants. These studies (classified as awaiting assessment) included data from 15,539 participants.

Implications for research

This review provides the most recent evidence on benefits and harms of celecoxib for people participants with osteoarthritis (OA). However, the evidence is based on studies with a number of shortcomings. Very few studies over 13 weeks duration were available; most compared celecoxib with placebo, and only two of 36 studies did not report funding from commercial drug companies or included authors who were employees of a sponsor. To fill the evidence gap, access to the existing data denied by study sponsors, is required. New studies investigating benefits and harms of celecoxib versus tNSAIDs for osteoarthritis are required. Future studies should be designed to include longer-term follow-up periods, robust study design exhibiting high methodological quality, independent funding sources, and more head-to-head comparisons with other tNSAIDs. Future studies should also include larger samples of participants to detect rare events. Quality of life should be included as a study outcome, analyzed and reported in future studies, as well as 50% pain relief. Benefits and harms of celecoxib in participants with comorbidities and those with higher risks of coronary heart disease should be further explored.

ACKNOWLEDGEMENTS

We would like to thank the Cochrane Musculoskeletal Group Editorial Team, including Lara Maxwell, Jordi Pardo Pardo, Vivian Welch and Elizabeth Ghogomu for their tremendous help, mentoring and kindness. We are also grateful to Drs Tea Andabaka, Tina Poklepovic Pericic, Vanja Giljaca, Goran Poropat and Prof Zarko Alfirevic for providing valuable methodological advice. Many thanks to Ms. Tamara Rader for helping us with electronic searches and obtaining full text copies of studies. Special gratitude goes to Camillia and Branko Mamic for their generous support throughout the years.

REFERENCES

References to studies included in this review

Asmus 2014 Study 1 {published data only}

* Asmus MJ, Essex MN, Brown PB, Mallen SR. Efficacy and tolerability of celecoxib in osteoarthritis patients who previously failed naproxen and ibuprofen: results from two trials. *International Journal of Clinical Rheumatology* 2014;**9**(6):551-8.

Asmus MJ, Essex MN, Brown PB, Mallen SR. Efficacy and tolerability of celecoxib in patients with osteoarthritis who previously did not respond to or did not tolerate naproxen and ibuprofen: results from 2 identically designed randomized trials. *Osteoarthritis and Cartilage*. 2013; Vol. 21:S253.

Asmus 2014 Study 2 {published data only}

* Asmus MJ, Essex MN, Brown PB, Mallen SR. Efficacy and tolerability of celecoxib in osteoarthritis patients who previously failed naproxen and ibuprofen: results from two trials. *International Journal of Clinical Rheumatology* 2014;**9**(6):551-8.

Asmus MJ, Essex MN, Brown PB, Mallen SR. Efficacy and tolerability of celecoxib in patients with osteoarthritis who previously did not respond to or did not tolerate naproxen and ibuprofen: results from 2 identically designed randomized trials. *Osteoarthritis and Cartilage*. 2013; Vol. 21:S253.

Bensen 1999 {published data only}

* Bensen WG, Fiechtner JJ, McMillen JI, Zhao WW, Yu SS, Woods EM, et al. Treatment of osteoarthritis with celecoxib, a cyclooxygenase-2 inhibitor: a randomized controlled trial. *Mayo Clinic Proceedings* 1999;**74**(11):1095-105.

Zhao SZ, McMillen JI, Markenson JA, Dedhiya SD, Zhao WW, Osterhaus JT, et al. Evaluation of the functional status aspects of health-related quality of life of patients with osteoarthritis treated with celecoxib. *Pharmacotherapy* 1999;**19**(11):1269-78.

Bingham 2007 Study 1 {published data only}

* Bingham CO, Sebba AI, Rubin BR, Ruoff GE, Kremer J, Bird S, et al. Efficacy and safety of etoricoxib 30 mg and celecoxib 200 mg in the treatment of osteoarthritis in two identically designed, randomized, placebo-controlled, non-inferiority studies. *Rheumatology* 2007;**46**(3):496-507.

Moore RA, Moore OA, Derry S, Peloso PM, Gammaitoni AR, Wang H. Responder analysis for pain relief and numbers needed to treat in a meta-analysis of etoricoxib osteoarthritis trials: bridging a gap between clinical trials and clinical practice. *Annals of the Rheumatic Diseases* 2010;**69**:374-79.

Bingham 2007 Study 2 {published data only}

* Bingham CO, Sebba AI, Rubin BR, Ruoff GE, Kremer J, Bird S, et al. Efficacy and safety of etoricoxib 30 mg and celecoxib 200 mg in the treatment of osteoarthritis in two identically designed, randomized, placebo-controlled, non-inferiority studies. *Rheumatology* 2007;**46**(3):496-507.

Moore RA, Moore OA, Derry S, Peloso PM, Gammaitoni AR, Wang H. Responder analysis for pain relief and numbers needed

to treat in a meta-analysis of etoricoxib osteoarthritis trials: bridging a gap between clinical trials and clinical practice. *Annals of the Rheumatic Diseases* 2010;**69**:374-79.

Birbara 2006 Study 1 {published data only}

Birbara C, Ruoff G, Sheldon E, Valenzuela C, Rodgers A, Petruschke RA, et al. Efficacy and safety of rofecoxib 12.5 mg and celecoxib 200 mg in two similarly designed osteoarthritis studies. *Current Medical Research and Opinion* 2006;**22**(1):199-210.

Birbara 2006 Study 2 {published data only}

Birbara C, Ruoff G, Sheldon E, Valenzuela C, Rodgers A, Petruschke RA, et al. Efficacy and safety of rofecoxib 12.5 mg and celecoxib 200 mg in two similarly designed osteoarthritis studies. *Current Medical Research and Opinion* 2006;**22**(1):199-210.

Boswell 2008 Study A {published data only}

Boswell DJ, Ostergaard K, Philipson RS, Hodge RA, Blum D, Brown JC, et al. Evaluation of GW406381 for treatment of osteoarthritis of the knee: two randomized, controlled studies. *Medscape Journal of Medicine* 2008;**10**(11):259-78.

Boswell 2008 Study B {published data only}

Boswell DJ, Ostergaard K, Philipson RS, Hodge RA, Blum D, Brown JC, et al. Evaluation of GW406381 for treatment of osteoarthritis of the knee: two randomized, controlled studies. *Medscape Journal of Medicine* 2008;**10**(11):259-78.

Clegg 2006 {published data only}

* Clegg DO, Reda DJ, Harris CL, Klein MA, O'Dell JR, Hooper MM, et al. Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. *New England Journal of Medicine* 2006;**354**(8):795-808.

Sawitzke AD, Shi H, Finco MF, Dunlop DD, Harris CL, Singer NG, et al. Clinical efficacy and safety of glucosamine, chondroitin sulphate, their combination, celecoxib or placebo taken to treat osteoarthritis of the knee: 2-year results from GAIT. *Annals of the Rheumatic Diseases* 2010;**69**(8):1459-64.

Conaghan 2013 {published data only}

Conaghan PG, Dickson J, Bolten W, Cevc G, Rother M. A multicentre, randomized, placebo- and active-controlled trial comparing the efficacy and safety of topical ketoprofen in Transfersome gel (IDEA-033) with ketoprofen-free vehicle (TDT 064) and oral celecoxib for knee pain associated with osteoarthritis. *Rheumatology* 2013;**52**(7):1303-12.

Dahlberg 2009 {published data only}

Dahlberg LE, Holme I, Høye K, Ringertz B. A randomized, multicentre, double-blind, parallel-group study to assess the adverse event-related discontinuation rate with celecoxib and diclofenac in elderly patients with osteoarthritis. *Scandinavian Journal of Rheumatology* 2009;**38**(2):133-43.

DeLemos 2011 {published data only}

DeLemos BP, Xiang J, Benson C, Gana TJ, Pascual ML, Rosanna R, et al. Tramadol hydrochloride extended-release once-daily in the treatment of osteoarthritis of the knee and/or hip: a double-blind, randomized, dose-ranging trial. *American Journal of Therapeutics* 2011;**18**(3):216-26.

Emery 2008 {published data only}

Emery P, Koncz T, Pan S, Lowry S. Analgesic effectiveness of celecoxib and diclofenac in patients with osteoarthritis of the hip requiring joint replacement surgery: a 12-week, multicenter, randomized, double-blind, parallel-group, double-dummy, noninferiority study. *Clinical Therapeutics* 2008;**30**(1):70-83.

Essex 2012a {published data only}

Essex MN, Bhadra P, Sands GH. Efficacy and tolerability of celecoxib versus naproxen in patients with osteoarthritis of the knee: a randomized, double-blind, double-dummy trial. *Journal of International Medical Research* 2012;**40**(4):1357-70.

Essex 2012b {published data only}

Essex MN, O'Connell M, Bhadra Brown P. Response to nonsteroidal anti-inflammatory drugs in African Americans with osteoarthritis of the knee. *Journal of International Medical Research* 2012;**40**(4):2251-66.

Essex 2014 {published data only}

Essex MN, Behar R, O'Connell MA, Bhadra Brown P. Efficacy and tolerability of celecoxib and naproxen versus placebo in Hispanic patients with knee osteoarthritis. *International Journal of General Medicine* 2014;**7**:227-35. [DOI: [10.2147/IJGM.S61297](https://doi.org/10.2147/IJGM.S61297)]

Fleischmann 2005 {published data only}

Fleischmann R, Sheldon E, Maldonado-Cocco J, Dutta D, Yu S, Sloan VS. Lumiracoxib is effective in the treatment of osteoarthritis of the knee: a prospective randomized 13-week study versus placebo and celecoxib. *Clinical Rheumatology* 2005;**25**(1):42-53.

Gibofsky 2003 {published data only}

Gibofsky A, Williams GW, McKenna F, Fort JG. Comparing the efficacy of cyclooxygenase 2-specific inhibitors in treating osteoarthritis. *Arthritis and Rheumatism* 2003;**48**(11):3102-11.

Hochberg 2011 Study 307 {published data only}

Hochberg MC, Fort JG, Svensson O, Hwang C, Sostek M. Fixed-dose combination of enteric-coated naproxen and immediate-release esomeprazole has comparable efficacy to celecoxib for knee osteoarthritis: two randomized trials. *Current Medical Research and Opinion* 2011;**27**(6):1243-53.

Hochberg 2011 Study 309 {published data only}

Hochberg MC, Fort JG, Svensson O, Hwang C, Sostek M. Fixed-dose combination of enteric-coated naproxen and immediate-release esomeprazole has comparable efficacy to celecoxib for knee osteoarthritis: two randomized trials. *Current Medical Research and Opinion* 2011;**27**(6):1243-53.

Kivitz 2001 {published data only}

Kivitz AJ, Moskowitz RW, Woods E, Hubbard RC, Verburg KM, Lefkowitz JB, et al. Comparative efficacy and safety of celecoxib

and naproxen in the treatment of osteoarthritis of the hip.

Journal of International Medical Research 2001;**29**(6):467-79.

Lehmann 2005 {published data only}

Lehmann R, Brzosko M, Kopsa P, Nischik R, Kreiss A, Thurston H, et al. Efficacy and tolerability of lumiracoxib 100 mg once daily in knee osteoarthritis: a 13-week, randomized, double-blind study vs. placebo and celecoxib. *Current Medical Research and Opinion* 2005;**21**(4):517-26.

McKenna 2001a {published data only}

McKenna F, Weaver A, Fiechtner JJ, Bello AE, Fort JG. COX-2 specific inhibitors in the management of osteoarthritis of the knee: a placebo-controlled, randomized, double-blind study. *Journal of Clinical Rheumatology* 2001;**7**(3):151-9.

McKenna 2001b {published data only}

McKenna F, Borenstein D, Wendt H, Wallemark C, Lefkowitz JB, Geis GS. Celecoxib versus diclofenac in the management of osteoarthritis of the knee: a placebo-controlled, randomised, double-blind comparison. *Scandinavian Journal of Rheumatology* 2001;**30**(1):11-8.

Pincus 2004 PACES-a {published data only}

Pincus T, Koch G, Lei H, Mangal B, Sokka T, Moskowitz R, 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. *Annals of the Rheumatic Diseases* 2004;**63**(8):931-9.

Pincus 2004 PACES-b {published data only}

Pincus T, Koch G, Lei H, Mangal B, Sokka T, Moskowitz R, 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. *Annals of the Rheumatic Diseases* 2004;**63**(8):931-9.

Rother 2007 {published data only}

Rother M, Lavins BJ, Kneer W, Lehnhardt K, Seidel EJ, Mazgareanu S. Efficacy and safety of epicutaneous ketoprofen in Transfersome (IDEA-033) versus oral celecoxib and placebo in osteoarthritis of the knee: multicentre randomised controlled trial. *Annals of the Rheumatic Diseases* 2007;**66**(9):1178-83.

Schnitzer 2011 {published data only}

Schnitzer TJ, Dattani ID, Serio B, Schneider H, Moore A, Tseng L, et al. A 13-week, multicenter, randomized, double-blind study of lumiracoxib in hip osteoarthritis. *Clinical Rheumatology* 2011;**30**(11):1433-46.

Sheldon 2005 {published data only}

* Sheldon E, Beaulieu A, Paster Z, Dutta D, Yu S, Sloan VS. Efficacy and tolerability of lumiracoxib in the treatment of osteoarthritis of the knee: a 13-week, randomized, double-blind comparison with celecoxib and placebo. *Clinical Therapeutics* 2005;**27**(1):64-77.

Sheldon EA, Beaulieu A, Paster Z, Yu S, Rebuli R. Long-term efficacy and safety of lumiracoxib 100 mg: an open-

label extension of a 13-week randomized controlled trial in patients with primary osteoarthritis of the knee. *Clinical and Experimental Rheumatology* 2008;**26**(4):611-9.

Smugar 2006 Study 1 {published data only}

Smugar SS, Schnitzer TJ, Weaver AL, Rubin BR, Polis AB, Tershakovec AM. Rofecoxib 12.5 mg, rofecoxib 25 mg and celecoxib 200 mg in the treatment of symptomatic osteoarthritis: results of two similarly designed studies. *Current Medical Research and Opinion* 2006;**22**(7):1353-67.

Smugar 2006 Study 2 {published data only}

Smugar SS, Schnitzer TJ, Weaver AL, Rubin BR, Polis AB, Tershakovec AM. Rofecoxib 12.5 mg, rofecoxib 25 mg and celecoxib 200 mg in the treatment of symptomatic osteoarthritis: results of two similarly designed studies. *Current Medical Research and Opinion* 2006;**22**(7):1353-67.

Sowers 2005 {published data only}

Sowers JR, White WB, Pitt B, Whelton A, Simon LS, Winer N, et al. The effects of cyclooxygenase-2 inhibitors and nonsteroidal anti-inflammatory therapy on 24-hour blood pressure in patients with hypertension, osteoarthritis, and type 2 diabetes mellitus. *Archives of Internal Medicine* 2005;**165**(2):161-8.

Tannenbaum 2004 {published data only}

Tannenbaum H, Berenbaum F, Reginster JY, Zacher J, Robinson J, Poor G, et al. Lumiracoxib is effective in the treatment of osteoarthritis of the knee: a 13 week, randomised, double blind study versus placebo and celecoxib. *Annals of the Rheumatic Diseases* 2004;**63**(11):1419-26.

Williams 2000 {published data only}

Williams GW, Ettlinger RE, Ruderman EM, Hubbard RC, Lonien ME, Yu SS, et al. Treatment of osteoarthritis with a once-daily dosing regimen of celecoxib: a randomized, controlled trial. *Journal of Clinical Rheumatology* 2000;**6**(2):65-74.

Williams 2001 {published data only}

Williams GW, Hubbard RC, Yu SS, Zhao W, Geis GS. Comparison of once-daily and twice-daily administration of celecoxib for the treatment of osteoarthritis of the knee. *Clinical Therapeutics* 2001;**23**(2):213-27.

References to studies excluded from this review

Adegbehingbe 2008 {published data only}

Adegbehingbe OO, Adesanya SA, Idowu TO, Okimi OC, Oyelami OA, Iwalewa EO. Clinical effects of Garcinia kola in knee osteoarthritis. *Journal of Orthopaedic Surgery and Research* 2008;**3**(34):1-10.

Andrews 1999 {published data only}

Andrews SA, Wallace CK, Davis RL. Celecoxib: a COX-2 inhibitor. *American Journal of Managed Care* 1999;**5**(4):511-8.

Angiolillo 2014 {published data only}

Angiolillo DJ, Datto C, Raines S, Yeomans ND. Impact of concomitant low-dose aspirin on the safety and tolerability of naproxen and esomeprazole magnesium delayed-release tablets in patients requiring chronic nonsteroidal anti-

inflammatory drug therapy: an analysis from 5 Phase III studies. *Journal of Thrombosis and Thrombolysis* 2014;**38**(1):11-23.

Ashcroft 2001 {published and unpublished data}

Ashcroft DM, Chapman SR, Clark WK, Millson DS. Upper gastroduodenal ulceration in arthritis patients treated with celecoxib. *Annals of Pharmacotherapy* 2001;**35**(7-8):829-34.

Battisti 2004 {published data only}

Battisti WP, Katz NP, Weaver AL, Matsumoto AK, Kivitz AJ, Polis AB, et al. Pain management in osteoarthritis: A focus on onset of efficacy—a comparison of rofecoxib, celecoxib, acetaminophen, and nabumetone across four clinical trials. *Journal of Pain* 2004;**5**(9):511-20.

Bensen 2000 {published data only}

Bensen WG. Antiinflammatory and analgesic efficacy of COX-2 specific inhibition: from investigational trials to clinical experience. *Journal of Rheumatology* 2000;**60**(Supp):17-24.

Bensen 2000b {published data only}

Bensen WG, Zhao SZ, Burke TA, Zabinski RA, Makuch RW, Maurath CJ, et al. Upper gastrointestinal tolerability of celecoxib, a COX-2 specific inhibitor, compared to naproxen and placebo. *Journal of Rheumatology* 2000;**27**(8):1876-83.

Bianchi 2003 {published data only}

Bianchi M, Broggin M. A randomised, double-blind, clinical trial comparing the efficacy of nimesulide, celecoxib and rofecoxib in osteoarthritis of the knee. *Drugs* 2003;**63**(Suppl 1):37-46.

Bianchi 2007 {published data only}

Bianchi M, Broggin M, Balzarini P, Franchi S, Sacerdote P. Effects of nimesulide on pain and on synovial fluid concentrations of substance P, interleukin-6 and interleukin-8 in patients with knee osteoarthritis: comparison with celecoxib. *International Journal of Clinical Practice* 2007;**61**(8):1270-7.

Bingham 2008 {published data only}

Bingham CO 3rd, Bird SR, Smugar SS, Xu X, Tershakovec AM. Responder analysis and correlation of outcome measures: pooled results from two identical studies comparing etoricoxib, celecoxib, and placebo in osteoarthritis. *Osteoarthritis and Cartilage* 2008;**16**(11):1289-93.

Bingham 2009 {published data only}

Bingham CO 3rd, Smugar SS, Wang H, Tershakovec AM. Early response to COX-2 inhibitors as a predictor of overall response in osteoarthritis: pooled results from two identical trials comparing etoricoxib, celecoxib and placebo. *Rheumatology* 2009;**48**(9):1122-7.

Bingham 2011 {published data only}

Bingham CO 3rd, Smugar SS, Wang H, Peloso PM, Gammaitoni A. Predictors of response to cyclo-oxygenase-2 inhibitors in osteoarthritis: pooled results from two identical trials comparing etoricoxib, celecoxib, and placebo. *Pain Medicine* 2011;**12**(3):352-61.

Brereton 2014 {published data only}

Brereton N, Pennington B, Ekelund M, Akehurst R. A cost-effectiveness analysis of celecoxib compared with diclofenac in the treatment of pain in osteoarthritis (OA) within the Swedish health system using an adaptation of the NICE OA model. *Journal of Medical Economics* 2014;**17**(9):677-84.

Castgillo 2015 {published data only}

Castgillo JR, Hochberg HC, Martel-Pelletier J, Monfort J, Moller I, Arden N, et al. Combined chondroitin sulfate and glucosamine versus celecoxib for painful knee osteoarthritis: Post-hoc analyses by Kellgren and Lawrence grade and C-reactive protein level from a randomized, double-blind, multicentre clinical trial. *Clinical Therapeutics* 2015;**37**(8):e122. [DOI: [10.1016/j.clinthera.2015.05.349](https://doi.org/10.1016/j.clinthera.2015.05.349)]

Castillo 2014 {published data only}

Castillo JR, Hochberg MC, Martel-Pelletier J, Monfort J, Mooller I, Arden N, et al. Multicentric osteoarthritis intervention study with sysadua (moves): Effects of combined glucosamine hydrochloride and chondroitin sulfate vs. celecoxib for painful knee osteoarthritis. *Basic and Clinical Pharmacology and Toxicology* 2014;**115**:12.

Chan 2015 {published data only}

Chan FKL, Ching J, Cheung C, Lam LYK, Au KWL, Ng SC, et al. Prevention of recurrent ulcer bleeding in arthritis patients with high cardiovascular and high gastrointestinal risks: A 18-month, double-blind, randomized trial. *Gastroenterology* 2015;**148**(4 Suppl 1):S157-8.

Detrembleur 2005 {published data only}

Detrembleur C, De Nayer J, van den Hecke A. Celecoxib improves the efficiency of the locomotor mechanism in patients with knee osteoarthritis. A randomised, placebo, double-blind and cross-over trial. *Osteoarthritis and Cartilage* 2005;**13**(3):206-10.

Dougados 2007 {published data only}

Dougados M, Moore A, Yu S, Gitton X. Evaluation of the patient acceptable symptom state in a pooled analysis of two multicentre, randomised, double-blind, placebo-controlled studies evaluating lumiracoxib and celecoxib in patients with osteoarthritis. *Arthritis Research and Therapy* 2007;**9**(1):R11.

Essex 2014b {published data only}

Essex MN, Brown PB, Sands GH. The efficacy of continuous versus intermittent celecoxib treatment in osteoarthritis patients aged < 60 and > 60 years. *International Journal of Clinical Rheumatology* 2014;**9**(1):13-20.

Gallelli 2013 {published data only}

Gallelli L, Galasso O, Falcone D, Southworth S, Greco M, Ventura V, et al. The effects of nonsteroidal anti-inflammatory drugs on clinical outcomes, synovial fluid cytokine concentration and signal transduction pathways in knee osteoarthritis. A randomized open label trial. *Osteoarthritis and Cartilage* 2013;**21**(9):1400-8.

Hawel 2003 {published data only}

Hawel R, Klein G, Singer F, Mayrhofer F, Kahler ST. Dexibuprofen in a special crystal form versus celecoxib in the management of osteoarthritis of the hip. *Wiener Medizinische Wochenschrift* 2002;**152**(Suppl 112):13.

* Hawel R, Klein G, Singer F, Mayrhofer F, Kähler ST. Comparison of the efficacy and tolerability of dexibuprofen and celecoxib in the treatment of osteoarthritis of the hip. *International Journal of Clinical Pharmacology and Therapeutics* 2003;**41**(4):153-64.

Henrotin 2015 {published data only}

Henrotin Y, Arden N, Berenbaum F, Blanco FJ, Castillo JR, Conaghan PG, et al. Combined chondroitin sulfate and glucosamine is more efficient than celebrex in reducing serum levels of Coll2-1, a cartilage degradation biomarker, in patients with severe OA: Results from a randomized, doubleblind, multicentric clinical trial. *Osteoarthritis and Cartilage* 2015;**23**(Suppl 2):A86.

Hirayama 2014 {published data only}

Hirayama A, Tanahashi N, Daida H, Ishiguro N, Chcacchin M, Sugioka T, et al. Accept study investigators in Japan. Assessing the cardiovascular risk between celecoxib and nonselective nonsteroidal antiinflammatory drugs in patients with rheumatoid arthritis and osteoarthritis. *Circulation Journal* 2014;**78**(1):194-205.

Hochberg 2014a {published data only}

Hochberg M, Martel-Pelletier J, Monfort J, Moller I, Castillo JR, Arden N, et al. Multicentric osteoarthritis intervention study with sysadua (MOVES): Effects of combined glucosamine hydrochloride and chondroitin sulfate vs celecoxib for painful knee osteoarthritis. *Annals of the Rheumatic Diseases*. 2014; Vol. 73, issue 11.

Hochberg 2014b {published data only}

Hochberg MC, Martel-Pelletier J, Monfort J, Moller I, Du Souich P, Pelletier JP. Randomized, double-blind, multicenter, non inferiority clinical trial with combined glucosamine and chondroitin sulfate vs celecoxib for painful knee osteoarthritis. *Osteoarthritis and cartilage*. 2014; Vol. 22, issue 24.

Johnson 2003 {published data only}

Johnson M, Seaton T. Comparing celecoxib with traditional nonsteroidal anti-inflammatory drugs. *Journal of Family Practice* 2003;**52**(2):96-8.

Leeb 2004 {published data only}

Leeb BF, Bucsi L, Keszthelyi B, Böhmova J, Valesova M, Hawel R, et al. Treatment of osteoarthritis of the knee joint. Efficacy and tolerance to acetaminol slow release in comparison to celecoxib [Behandlung der Gonarthrose. Wirksamkeit und Verträglichkeit von retardiertem Acemetacin im Vergleich zu Celecoxib]. *Der Orthopäde* 2004;**33**(9):1032-41.

Luyten 2007 {published data only}

Luyten FP, Geusens P, Malaise M, De Clerck L, Westhovens R, Raeman F, et al. A prospective randomised multicentre study comparing continuous and intermittent treatment with

celecoxib in patients with osteoarthritis of the knee or hip. *Annals of the Rheumatic Diseases* 2007;**66**(1):99-106.

Moore 2009 {published data only}

Moore RA, Peloso PM, Mehta A, Gammaitoni A. Baseline predictors of improvement in pain using a pooled analysis of 7 randomized controlled studies in patients with osteoarthritis (OA). *European Journal of Pain* 2009;**13**(S1):S82.

NCT02192190 {published data only}

* Jin Y, Smith C, Monteith D, Brown R, Camporeale A, McNearney T, Deeg M, Raddad E, de la Pena A, Kivitz A, Schnitzer T. LY2951742, a monoclonal antibody against CGRP, failed to reduce signs and symptoms of knee osteoarthritis. *Osteoarthritis and Cartilage* 2016;**24**(S1):S50.

NCT02192190. A study of LY2951742 in participants with mild to moderate osteoarthritis knee pain. clinicaltrials.gov/show/NCT02192190 (first received 14 July 2014).

Smith C, Jin Y, Raddad E, McNearney T, Ni X, Monteith D, Brown R, Deeg M, Schnitzer T. Applications of Bayesian statistical methodology to success criteria: A case study of a phase 2 design with interim futility assessment in patients with knee osteoarthritis. *Osteoarthritis and Cartilage* 2016;**24**(S1):S185.

Nissen 2016 {published data only}

Nissen SE, Yeomans ND, Solomon DH, Lüscher TF, Libby P, Husni ME, Graham DY, Borer JS, Wisniewski LM, Wolski KE, Wang Q, Menon V, Ruschitzka F, Gaffney M, Beckerman B, Berger MF, Bao W, Lincoff AM, PRECISION Trial Investigators. Cardiovascular Safety of Celecoxib, Naproxen, or Ibuprofen for Arthritis. *N Engl J Med* 2016;**375**(26):2519-29.

Ogunlade 2005 {published data only}

Ogunlade SO, Oginni LM, Nwadiaro HC, Popoola O, Enweani UN, Onabowale BO, et al. The efficacy and toleration of celecoxib (Celebrex) in the treatment of osteoarthritis in Nigeria: a multicentre study. *West African Journal of Medicine* 2005;**24**(3):263-7.

Ozgocmen 2005 {published data only}

Ozgocmen S, Ardicoglu O, Erdogan H, Fadillioglu E, Gudul H. In vivo effect of celecoxib and tenoxicam on oxidant/ anti-oxidant status of patients with knee osteoarthritis. *Annals of Clinical and Laboratory Science* 2005;**35**(2):137-43.

Pelletier 2016 {published data only}

Pelletier JP, Raynauld JP, Beaulieu A, Bessette L, Morin F, Fernandes AJ, et al. In a two-year double-blind randomized controlled multicenter study, chondroitin sulfate was significantly superior to celecoxib at reducing cartilage loss with similar efficacy at reducing disease symptoms in knee osteoarthritis patients. *Osteoarthritis and Cartilage* 2016;**24**(Suppl. 1):S189-S190.

Rozenberg 2008 {published data only}

Rozenberg S, Méric G, Jeanpetit Y. Changes in quality of life in patients with osteoarthritis treated with celecoxib: the Qualice study. *Presse Médicale* 2008;**37**(4 Pt 1):571-8.

Sakamoto 2011 {published data only}

Sakamoto C, Soen S. Efficacy and safety of the selective cyclooxygenase-2 inhibitor celecoxib in the treatment of rheumatoid arthritis and osteoarthritis in Japan. *Digestion* 2011;**83**(1-2):108-23.

Sampalis 2012 {published data only}

Sampalis JS, Brownell LA. A randomized, double blind, placebo and active comparator controlled pilot study of UP446, a novel dual pathway inhibitor anti-inflammatory agent of botanical origin. *Nutrition Journal* 2012;**11**:21.

Sands 2013 {published data only}

Sands GH, Brown PB, Essex MN. The efficacy of continuous versus intermittent celecoxib treatment in osteoarthritis patients with body mass index ≥ 30 and < 30 kg/m². *Open Rheumatology Journal* 2013;**7**:32-7.

Schnitzer 2015 {published data only}

Schnitzer TJ, Ekman EF, Spierings ELH, Greenberg HS, Smith MD, Brown MT, et al. Efficacy and safety of tanezumab monotherapy or combined with non-steroidal anti-inflammatory drugs in the treatment of knee or hip osteoarthritis pain. *Annals of the Rheumatic Diseases* 2015;**74**(6):1202-11.

Silverstein 2000 {published data only}

Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A, et al. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: A randomized controlled trial. Celecoxib Long-term Arthritis Safety Study. *JAMA* 2000;**284**(10):1247-55.

Simon 1998 {published data only}

Simon LS, Lanza FL, Lipsky PE, Hubbard RC, Talwalker S, Schwartz BD, et al. Preliminary study of the safety and efficacy of SC-58635, a novel cyclooxygenase 2 inhibitor: efficacy and safety in two placebo-controlled trials in osteoarthritis and rheumatoid arthritis, and studies of gastrointestinal and platelet effects. *Arthritis and Rheumatism* 1998;**41**(9):1591-602.

Stam 2012 {published data only}

Stam W, Jansen J, Taylor S. Efficacy of etoricoxib, celecoxib, lumiracoxib, non-selective NSAIDs, and acetaminophen in osteoarthritis: A mixed treatment comparison. *Open Rheumatology Journal* 2012;**6**:6-20.

Stengaard-Pedersen 2004 {published data only}

Stengaard-Pedersen K, Ekesbo R, Karvonen AL, Lyster M. Celecoxib 200 mg q.d. is efficacious in the management of osteoarthritis of the knee or hip regardless of the time of dosing. *Rheumatology* 2004;**43**(5):592-5.

Strand 2011 {published data only}

Strand V, Simon LS, Dougados M, Sands GH, Bhadra P, Breazna A, et al. Treatment of osteoarthritis with continuous versus intermittent celecoxib. *Journal of Rheumatology* 2011;**38**(12):2625-34.

Taşcıoğlu 2004 {published data only}

Taşcıoğlu F, Öner C, Aydemir A. Comparison of the efficacy of celecoxib and diclofenac sodium in the treatment of knee osteoarthritis [Diz osteoartritinin tedavisinde selekoksib ve diklofenak sodyum etkinliğinin karşılaştırılması]. *Türkiye Fiziksel Tıp ve Rehabilitasyon Derneği* 2004;**50**(4):7-12.

Tive 2015 {published data only}

Tive L, Radin D, Bello A, Nguyen H, Brown MT, West CR, et al. Pooled efficacy and safety from phase 3 controlled studies of tanezumab in patients with osteoarthritis. *Arthritis & Rheumatology* 2015;**67**(Suppl 10).

Tran 2004 {published data only}

Tran F, Boggie DT, Delattre ML, Schaefer MG, Morreale AP, Plowman BK. Therapeutic interchange involving replacement of rofecoxib or celecoxib with valdecoxib. *American Journal of Health-System Pharmacy* 2004;**61**(13):1391-4.

Trudeau 2015 {published data only}

Trudeau J, Van Inwegen R, Eaton T, Bhat G, Paillard F, Ng D, et al. Assessment of pain and activity using an electronic pain diary and actigraphy device in a randomized, placebo-controlled crossover trial of celecoxib in osteoarthritis of the knee. *Pain Practice* 2015;**15**(3):247-55.

Tsvetkova 2001 {published data only}

Tsvetkova ES, Alekseeva LI, Balabanova RM, Chichasova NV, Shostak NA, Shmidt EI. Effectiveness and tolerance of celebrex in osteoarthritis (data of a Russian study). *Terapevticheskii Arkhiv* 2001;**73**(5):61-3.

Wittenberg 2006 {published data only}

Wittenberg RH, Schell E, Krehan G, Maeumbaed R, Runge H, Schluter P, et al. First-dose analgesic effect of the cyclooxygenase-2 selective inhibitor lumiracoxib in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled comparison with celecoxib [NCT00267215]. *Arthritis Research and Therapy* 2006;**8**(2):R35.

Wolfe 2004 {published data only}

Wolfe F, Michaud K, Burke TA, Zhao SZ. Longer use of COX-2-specific inhibitors compared to nonspecific nonsteroidal antiinflammatory drugs: longitudinal study of 3639 patients in community practice. *Journal of Rheumatology* 2004;**31**(2):355-8.

Yoo 2014a {published data only}

Yoo MC, Yoo WH, Kang SB, Park YW, Kim SS, Moon KH, et al. Etoricoxib in the treatment of Korean patients with osteoarthritis. *International Journal of Rheumatic Diseases* 2014;**17**(Suppl 1):128.

* Yoo MC, Yoo WH, Kang SB, Park YW, Kim SS, Moon KH, et al. Etoricoxib in the treatment of Korean patients with osteoarthritis in a double-blind, randomized controlled trial. *Current Medical Research and Opinion* 2014;**30**(12):2399-2408.

Yoo 2014b {published data only}

Yoo WH, Yoo HG, Park S-H, Baek HL, Lee YJ, Shim SC, et al. Efficacy and safety of PG201 (Layla) and celecoxib in the treatment of symptomatic knee osteoarthritis: a double-

blinded, randomized, multi-center, active drug comparative, parallel-group, non-inferiority, phase III study. *Rheumatology International* 2014;**34**(10):1369-78.

References to studies awaiting assessment
Essex 2016 {published data only}

Essex MN, O'Connell MA, Behar R, Bao W. Efficacy and safety of nonsteroidal anti-inflammatory drugs in Asian patients with knee osteoarthritis: summary of a randomized, placebo-controlled study. *Int J Rheum Dis* 2016;**19**(3):262-70.

EUCTR2005-002772-14-GB {unpublished data only}

EUCTR2005-002772-14-GB. A 13-week, multicenter, randomized, double-blind, double-dummy, placebo-controlled, parallel group trial of lumiracoxib (COX189) 100 mg o.d. in patients with primary hip osteoarthritis using celecoxib (200 mg o.d.) as a positive control. <http://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2005-002772-14-GB> (first received 14 October 2005).

Gordo 2017 {published data only}

Gordo AC, Walker C, Armada B, Zhou D. Efficacy of celecoxib versus ibuprofen for the treatment of patients with osteoarthritis of the knee: A randomized double-blind, non-inferiority trial. *J Int Med Res* 2017;**45**(1):59-74.

Iyengar 2013 {published data only}

Iyengar RL, Gandhi S, Aneja A, Thorpe K, Razzouk L, Greenberg J, et al. NSAIDs are associated with lower depression scores in patients with osteoarthritis. *American Journal of Medicine* 2013;**126**(11):1017.

Lisse 2001 {published data only}

Lisse J, Espinoza L, Zhao SZ, Dedhiya SD, Osterhaus JT. Functional status and health-related quality of life of elderly osteoarthritic patients treated with celecoxib. *Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 2001;**56**(3):M167-75.

Mastbergen 2010 {published data only}

Mastbergen SC, De Boer TN, Huisman AM, Polak AA, Bijlsma JWJ, Lafeber FPJG. 9th World Congress of the International Cartilage Repair Society; 2010, Nov 26-29, Barcelona, Spain.. 2010; Vol. 1, issue 2(S1):46S.

Reginster 2016 {published data only}

Reginster JY. Pharmaceutical Grade Chondroitin Sulfate Improves Knee Osteoarthritis Symptoms More Than Placebo and As Much As Celecoxib: Results of the Chondroitin Vs Celecoxib Vs Placebo Trial (CONCEPT). *Arthritis & Rheumatology* 2016;**68**(S10).

Singh 2006 {published data only}

Singh G, Fort JG, Goldstein JL, Levy RA, Hanrahan PS, Bello AE, et al. Celecoxib versus naproxen and diclofenac in osteoarthritis patients: SUCCESS-I study. *American Journal of Medicine* 2006;**119**(3):255-66.

References to ongoing studies

EUCTR2011-005398-22-ES {unpublished data only}

EUCTR2011-005398-22-ES. Clinical study comparing a new single dose chewable tablet as treatment knee osteoarthritis with moderate to severe pain, compared to placebo and celecoxib. apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2011-005398-22-ES (first received 11 January 2012).

NCT01765296 {published data only}

NCT01765296. Phase III study of CG100649 in osteoarthritis patients. clinicaltrials.gov/show/NCT01765296 (first received 8 January 2013).

NCT01768520 {unpublished data only}

NCT01768520. Evaluation of the efficacy and safety of entelon tab. 150 mg in patients with osteoarthritis of knee. clinicaltrials.gov/ct2/show/NCT01768520 (first received 14 January 2013).

NCT02079727 {published data only}

NCT02079727. Condrosulf vs celebrex vs placebo in the treatment of knee OA. clinicaltrials.gov/show/NCT02079727 (first received 28 February 2014).

Additional references

Akl 2011

Akl EA, Oxman AD, Herrin J, Vist GE, Terrenato I, Sperati F, et al. Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews* 2011, Issue 3. [DOI: [10.1002/14651858.CD006776.pub2](https://doi.org/10.1002/14651858.CD006776.pub2)]

Altman 1996

Altman R, Brandt K, Hochberg M, Moskowitz R, Bellamy N, Bloch DA, et al. Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Results from a workshop. *Osteoarthritis Cartilage* 1996;**4**(4):217-43.

Busse 2015

Busse JW, Bartlett SJ, Dougados M, Johnston BC, Guyatt GH, Kirwan JR, et al. Optimal strategies for reporting pain in clinical trials and systematic reviews: recommendations from an OMERACT 12 workshop. *Journal of Rheumatology* 2015;**42**(10):1962-70.

Caldwell 2006

Caldwell B, Aldington S, Weatherall M, Shirtcliffe P, Beasley R. Risk of cardiovascular events and celecoxib: a systematic review and meta-analysis. *Journal of the Royal Society of Medicine* 2006;**99**(3):132-40.

Chen 2008

Chen YF, Jobanputra P, Barton P, Bryan S, Fry-Smith A, Harris G, et al. Cyclooxygenase-2 selective non-steroidal anti-inflammatory drugs (etodolac, meloxicam, celecoxib, rofecoxib, etoricoxib, valdecoxib and lumiracoxib) for osteoarthritis and rheumatoid arthritis: a systematic review and economic evaluation. *Health Technology Assessment* 2008;**12**(11):1-278.

Chou 2006

Chou R, Helfand M, Peterson K, Dana T, Roberts C. Drug Class Review: Cyclo-oxygenase (COX)-2 Inhibitors and Non-steroidal Anti-inflammatory Drugs (NSAIDs): Final Report Update 3. Portland, OR: Oregon Health & Science University, 2006.

CNT 2013

Coxib and traditional NSAID Trialists' (CNT) Collaboration. Bhalra N, Emberson J, Merhi A, Abramson S, Arber N, Baron JA, et al. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. *Lancet* 2013;**382**(9894):769-79.

Cohen 1988

Cohen J. Statistical power analysis for the behavioral sciences. 2nd Edition. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

Conaghan 2012

Conaghan PG. A turbulent decade for NSAIDs: update on current concepts of classification, epidemiology, comparative efficacy, and toxicity. *Rheumatology International* 2012;**32**(6):1491-502.

Cross 2014

Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. *Annals of the Rheumatic Diseases* 2014;**73**(7):1323-30.

Da Costa 2016

Da Costa BR, Reichenbach S, Keller N, Nartey L, Wandel S, Jüni P, et al. Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. *Lancet* 2016;**387**(10033):2093-105.

Deeks 2002

Deeks JJ, Smith LA, Bradley MD. Efficacy, tolerability, and upper gastrointestinal safety of celecoxib for treatment of osteoarthritis and rheumatoid arthritis: systematic review of randomised controlled trials. *British Medical Journal* 2002;**325**(7365):619.

Dworkin 2008

Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *Journal of Pain* 2008;**9**(2):105-21.

FDA 2005

Food, Drug Administration. Information for healthcare professionals: Celecoxib (marketed as Celebrex). www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm124655.htm (accessed February 20, 2017).

Fortin 2006

Fortin M, Dionne J, Pinho G, Gignac J, Almirall J, Lapointe L. Randomized controlled trials: do they have external validity for patients with multiple comorbidities?. *Annals of Family Medicine* 2006;**4**(2):104-8.

Fosbøl 2009

Fosbøl EL, Gislason GH, Jacobsen S, Folke F, Hansen ML, Schramm TK, et al. Risk of myocardial infarction and death associated with the use of nonsteroidal anti-inflammatory drugs (NSAIDs) among healthy individuals: a nationwide cohort study. *Clinical Pharmacology and Therapeutics* 2009;**85**(2):190-7.

Garner 2002

Garner S, Fidan D, Frankish R, Judd M, Shea B, Towheed T, et al. Celecoxib for rheumatoid arthritis. *Cochrane Database of Systematic Reviews* 2002, Issue 1. [DOI: [10.1002/14651858.CD003831](https://doi.org/10.1002/14651858.CD003831)]

Garner 2008

Garner SE, Fidan DD, Frankish R, Maxwell L. Rofecoxib for osteoarthritis. *Cochrane Database of Systematic Reviews* 2008, Issue 4. [DOI: [10.1002/14651858.CD005115](https://doi.org/10.1002/14651858.CD005115)]

Goldacre 2014

Goldacre B, Heneghan C. Improving, and auditing, access to clinical trial results. *British Medical Journal* 2014;**348**:g213.

GRADEpro GDT 2014 [Computer program]

GRADE Working Group, McMaster University. GRADEpro GDT. Version (accessed prior to March 21, 2017). Hamilton (ON): GRADE Working Group, McMaster University, 2014.

Green 2003

Green CR, Anderson KO, Baker TA, Campbell CL, Decker S, Fillingim RB, et al. The unequal burden of pain: confronting racial and ethnic disparities in pain. *Pain Medicine* 2003;**4**(3):277-94.

Higgins 2003

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistencies in meta-analyses. *British Medical Journal* 2003;**327**:557-60.

Higgins 2011

Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Jelicic Kadic 2016

Jelicic Kadic A, Vucic K, Dosenovic S, Sapunar D, Puljak L. Extracting data from figures with software was faster, with higher inter-rater reliability than manual extraction. *Journal of Clinical Epidemiology* 2016;**74**:119-23.

Jüni 2003

Jüni P, Sterchi R, Dieppe P. Systematic review of celecoxib for osteoarthritis and rheumatoid arthritis. Problems compromise review's validity. *British Medical Journal* 2003;**326**(7384):334.

Jüni 2006

Jüni P, Reichenbach S, Dieppe P. Osteoarthritis: rational approach to treating the individual. *Best Practice & Research. Clinical Rheumatology* 2006;**20**(4):721-40.

Kearney 2006

Kearney PM, Baigent C, Godwin J, Halls H, Emberson JR, Patrono C. Do selective cyclo-oxygenase-2 inhibitors and traditional non-steroidal anti-inflammatory drugs increase the risk of atherothrombosis? Meta-analysis of randomised trials. *British Medical Journal* 2006;**332**(7553):1302-8.

Kellgren 1957

Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. *Ann Rheum Dis* 1957;**16**:494-501.

Lundh 2012

Lundh A, Sismondo S, Lexchin J, Busuioac OA, Bero L. Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews* 2012;**12**:MR000033. DOI: [10.1002/14651858.MR000033.pub2](https://doi.org/10.1002/14651858.MR000033.pub2).

McGettigan 2011

McGettigan P, Henry D. Cardiovascular risk with non-steroidal anti-inflammatory drugs: systematic review of population-based controlled observational studies. *PLoS Medicine* 2011;**8**:e1001098.

Micklewright 2003

Micklewright R, Lane S, Linley W, McQuade C, Thompson F, Maskrey N. Review article: NSAIDs, gastroprotection and cyclo-oxygenase-II-selective inhibitors. *Alimentary Pharmacology and Therapeutics* 2003;**17**(3):321-32.

Moore 2005

Moore RA, Derry S, Makinson GT, McQuay HJ. Tolerability and adverse events in clinical trials of celecoxib in osteoarthritis and rheumatoid arthritis: systematic review and meta-analysis of information from company clinical trial reports. *Arthritis Research and Therapy* 2005;**7**(3):R644-65.

Mullan 2009

Mullan RJ, Flynn DN, Carlberg B, Tleyjeh IM, Kamath CC, LaBella ML, et al. Systematic reviewers commonly contact study authors but do so with limited rigor. *Journal of Clinical Epidemiology* 2009;**62**(2):138-42.

Nüesch 2010

Nüesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *British Medical Journal* 2010;**341**:c3515.

Ofman 2002

Ofman JJ, MacLean CH, Straus WL, Morton SC, Berger ML, Roth EA, Shekelle P. A metaanalysis of severe upper gastrointestinal complications of nonsteroidal antiinflammatory drugs. *Journal of Rheumatology* 2002;**29**(4):804-12.

Pham 2004

Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. *Osteoarthritis and Cartilage* 2004;**12**(5):389-99.

Reichenbach 2010

Reichenbach S, Rutjes AW, Nüesch E, Trelle S, Jüni P. Joint lavage for osteoarthritis of the knee. *Cochrane Database of Systematic Reviews* 2010, Issue 5. [DOI: [10.1002/14651858.CD007320.pub2](https://doi.org/10.1002/14651858.CD007320.pub2)]

RevMan 2014 [Computer program]

Nordic Cochrane Centre, The Cochrane Collaboration. Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

Riley 2002

Riley JL, Wade JB, Myers CD, Sheffield D, Papas RK, Price DD. Racial/ethnic differences in the experience of chronic pain. *Pain* 2002;**100**(3):291–8.

Robinson 2011

Robinson KA, Saldanha IJ, McKoy NA. Development of a framework to identify research gaps from systematic reviews. *Journal of Clinical Epidemiology* 2011;**64**(12):1325–30.

Rostom 2007

Rostom A, Muir K, Dubé C, Jolicoeur E, Boucher M, Joyce J, et al. Gastrointestinal safety of cyclooxygenase-2 inhibitors: a Cochrane Collaboration systematic review. *Clinical Gastroenterology and Hepatology* 2007;**5**(7):818–28.

Saldanha 2013

Saldanha IJ, Wilson LM, Bennett WL, Nicholson WK, Robinson KA. Development and pilot test of a process to identify research needs from a systematic review. *Journal of Clinical Epidemiology* 2013;**66**(5):538–45.

Sibbald 2004

Rofecoxib (Vioxx) voluntarily withdrawn from market. *CMAJ* 2004; Vol. 171, issue 9:1027–8.

Singh 1998

Singh G. Recent considerations in nonsteroidal anti-inflammatory drug gastropathy. *American Journal of Medicine* 1998;**105**(1B):31S–8S.

Solomon 2005

Solomon SD, McMurray JJ, Pfeffer MA, Wittes J, Fowler R, Finn P, et al. Cardiovascular risk associated with celecoxib in a clinical

trial for colorectal adenoma prevention. *New England Journal of Medicine* 2005;**352**(11):1092–102.

Solomon 2008

Solomon SD, Wittes J, Finn PV, Fowler R, Viner J, Bertagnolli MM, et al. Cardiovascular risk of celecoxib in 6 randomized placebo-controlled trials: the cross trial safety analysis. *Circulation* 2008;**117**(16):2104–13.

Trelle 2011

Trelle S, Reichenbach S, Wandel S, Hildebrand P, Tschannen B, Villiger PM, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. *British Medical Journal* 2011;**342**:c7086.

Vane 1998

Vane JR, Bakhle YS, Botting RM. Cyclooxygenases 1 and 2. *Annual Review of Pharmacology and Toxicology* 1998;**38**:97–120.

Vucic 2015

Vucic, Jelicic Kadic A, Puljak L. Survey of Cochrane protocols found methods for data extraction from figures not mentioned or unclear. *Journal of Clinical Epidemiology* 2015;**68**(10):1161–4.

Warner 1999

Warner TD, Giuliano F, Vojnovic I, Bukasa A, Mitchell JA, Vane JR. Nonsteroid drug selectivities for cyclo-oxygenase-1 rather than cyclo-oxygenase-2 are associated with human gastrointestinal toxicity: a full in vitro analysis. *Proceedings of the National Academy of Sciences of the United States of America* 1999;**96**(17):7563–8.

Willoughby 2000

Willoughby DA, Moore AR, Colville-Nash PR. COX-1, COX-2, and COX-3 and the future treatment of chronic inflammatory disease. *Lancet* 2000;**355**(9204):646–8.

References to other published versions of this review

Vrdoljak 2012

Vrdoljak D, Selimovic M, Marin A, Utrobicic A, Tugwell P, Puljak L. Celecoxib for osteoarthritis. *Cochrane Database of Systematic Reviews* 2012, Issue 5. [DOI: [10.1002/14651858.CD009865](https://doi.org/10.1002/14651858.CD009865)]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Asmus 2014 Study 1

Methods	Study design: randomized, double-blind, placebo-controlled trial, 2 parallel groups, multicenter study
	Duration: 6 weeks
	Study dates: Not indicated
	Locations: 29 centers in the USA

Celecoxib for osteoarthritis (Review)

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Asmus 2014 Study 1 (Continued)

Participants	Randomized: N = 380 Completed: N = 278 Mean age: 60.3 ± 10 years Female: 60.5% Inclusion: age ≥ 40 years, with diagnosed, active symptomatic osteoarthritis (OA) of the knee in a flare state diagnosed according to the American College of Rheumatology (ACR) criteria, according to the following inclusion criteria: had failed prior treatment with both prescription strength naproxen (at least 750 mg/day for 2 weeks) and ibuprofen (at least 1200 mg/day for 2 weeks) within the past 5 years due to either lack of efficacy and/or tolerability; women of childbearing potential had a negative urine pregnancy test and had to be using an adequate method of contraception; if taking chronic non-steroidal anti-inflammatory drug (NSAID) therapy, participants were to complete a wash out period for a minimum of 2 days; participants were to have a functional capacity class of I–III; a willingness to participate for 6 weeks and ability to provide informed consent
Interventions	Celecoxib 200 mg (N = 190) Placebo (N = 190)
Outcomes	Primary efficacy outcomes: change from baseline to week 6 in the Patient's Assessment of Arthritis Pain using visual analog scale (VAS) measured on a 0 (no pain) to 100 (very severe pain) scale Secondary outcomes: change in Western Ontario and McMaster Universities Osteoarthritis Index (WOM-AC) from baseline to week 6, change in Patient's and Physician's Global Assessment of Pain from baseline to week 6 Harms: number of participants reporting at least one adverse event, number of participants who discontinued due to adverse events, treatment-related adverse events that led to discontinuation, number of deaths, treatment-emerging serious adverse events
Funding	Pfizer, Inc
Declaration of interest	All authors were employees of Pfizer, Inc. The authors indicated they had no other conflicts of interest
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomization not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind study
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded

Asmus 2014 Study 1 (Continued)

Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition was 19.5% (37/190) in celecoxib group and 34.2% (65/190) in the placebo group. In the celecoxib group 4.2% of attrition was related to study drug (adverse events, lack of efficacy), while this percent was 13.7% in the placebo group (Table 1, p. 553). Missing values were imputed using last observation carried forward (LOCF)
Selective reporting (reporting bias)	Low risk	All primary and secondary outcomes reported in the methods are included in the report of results
Other bias	Low risk	From the data provided, no indication of other important risks of bias

Asmus 2014 Study 2

Methods	Study design: randomized, double-blind, placebo-controlled trial, 2 parallel groups, multicentre study Duration: 6 weeks Study dates: Not indicated Locations: 30 centers in the USA
Participants	Randomized: N = 388 Completed: N = 294 Mean age: 58.6 ± 10 years Female: 68% Inclusion: age ≥ 40 years, with diagnosed, active symptomatic osteoarthritis (OA) of the knee in a flare state diagnosed according to the ACR criteria, according to the following inclusion criteria: had failed prior treatment with both prescription strength naproxen (at least 750 mg/day for 2 weeks) and ibuprofen (at least 1200 mg/day for 2 weeks) within the past 5 years due to either lack of efficacy and/or tolerability; women of childbearing potential had a negative urine pregnancy test and had to be using an adequate method of contraception; if taking chronic non-steroidal anti-inflammatory drug (NSAID) therapy, participants were to complete a wash-out period for a minimum of 2 days; participants were to have a functional capacity class of I–III; a willingness to participate for 6 weeks and ability to provide informed consent
Interventions	Celecoxib 200 mg (N = 195) Placebo (N = 193)
Outcomes	Primary efficacy outcomes: change from baseline to week 6 in the Patient's Assessment of Arthritis Pain (VAS measured on a 0 (no pain) to 100 (very severe pain) scale. Secondary outcomes: change in Western Ontario and McMaster Universities Osteoarthritis Index (WOM-AC) from baseline to week 6, change in Patient's and Physician's Global Assessment of Pain from baseline to week 6 Harms: number of participants reporting at least one adverse event, number of participants who discontinued due to adverse events, treatment-related adverse events that led to discontinuation, number of deaths, treatment-emerging serious adverse events

Celecoxib for osteoarthritis (Review)

Asmus 2014 Study 2 (Continued)

Funding	Pfizer, Inc	
Declaration of interest	All authors were employees of Pfizer, Inc. The authors indicated that they had no other conflicts of interest	
Notes		
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomization not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind study
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition was 19% (38/195) in celecoxib group and 29% (56/193) in the placebo group. In the celecoxib group 5.6% of attrition was related to study drug (adverse events, lack of efficacy), while this percent was 18.1% in the placebo group (Table 1, p. 553). Missing values were imputed using last observation carried forward (LOCF)
Selective reporting (reporting bias)	Low risk	All primary and secondary outcomes reported in the methods are included in the report of results
Other bias	Low risk	From the data provided, no indication of other important risks of bias

Bensen 1999

Methods	Study design: randomized, double-blind, placebo and active controls, 5 parallel groups, multicentre study Duration: 12 weeks Study dates: August 1996 to July 1997 Locations: 71 clinical sites in the USA and Canada	
Participants	Randomized N = 1003 Completed N = 569	

Celecoxib for osteoarthritis (Review)

Bensen 1999 (Continued)

Mean age: 62 years (range 21 to 89 years)

Women: 72%

Inclusion: osteoarthritis (OA) of the knee, stage I-III according to the American College of Rheumatology (ACR) criteria

Interventions	<p>Celecoxib 100 mg/day (50 mg twice daily) (N = 197)</p> <p>Celecoxib 200 mg/day (100 mg twice daily) (N = 201)</p> <p>Celecoxib 400 mg/day (200 mg twice daily) (N = 202)</p> <p>Naproxen 1000 mg/day (500 mg twice daily) (N = 198)</p> <p>Placebo (N = 203)</p> <p>Stable doses of aspirin, 325 mg/d or less, and acetaminophen, up to 2 g/d, taken for no longer than 3 consecutive days – except during the 48-hour period prior to arthritis assessments – were allowed</p>
Outcomes	<p>Efficacy outcomes: patient's and physician's global assessment of arthritis (1-5 scale); the patient's assessment of arthritis pain on a visual analog scale (VAS 0-100 mm); the OA Severity Index (0-24 scale); Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)</p> <p>Harms outcomes: incidence and type of adverse reactions; the rate of withdrawal because of adverse reactions; laboratory abnormalities</p>
Funding	Supported in part by GD Searle & Co
Declaration of interest	<p>Not reported</p> <p>7/10 authors were employees of Searle Research and Development</p>

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	All treatment regimens were fully masked with the use of a double-dummy technique (p. 1096)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described

Celecoxib for osteoarthritis (Review)

Bensen 1999 (Continued)

Incomplete outcome data (attrition bias) All outcomes	High risk	High attrition was observed, as 43% of participants did not complete the study (p. 1097). Attrition per group was reported for withdrawals due to treatment failure: 19.2% in placebo group, 10% in the celecoxib 200 mg/day group and 14% in the naproxen group (Table 2) and for withdrawals due to adverse events as 4% in the placebo group, 8% in the celecoxib group and 4% in the naproxen group (Table 4). Other reasons for attrition were not provided. All efficacy analyses were performed on the intention-to-treat (ITT) cohort (all randomized participants who took one dose or more of study medication) and missing values were imputed by carrying forward the last observation for any participant who discontinued the study for any reason (including treatment failure) prior to the full 12 weeks of therapy (p. 1097)
Selective reporting (reporting bias)	Low risk	All outcomes reported in the methods are also reported in the results section
Other bias	Unclear risk	The results may have been influenced by a selection bias in favor of participants tolerant of naproxen, and thus the tolerability comparison in this study may underestimate any actual difference between celecoxib and naproxen in the arthritis population as a whole (p. 1103). Participants were allowed to take stable doses of aspirin (325 mg/d or less) and acetaminophen (up to 2 g/d taken for no longer than 3 consecutive days except during the 48 h period prior to arthritis assessment) (p. 1096). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Bingham 2007 Study 1

Methods	<p>Study design: two randomized studies, double-blind, placebo and active controls, 4 parallel groups, multicenter studies</p> <p>Duration: 26 weeks; each study consisting of part one (12 weeks) and part two (14 weeks)</p> <p>Dates of study: March 2004 to February 2005</p> <p>Locations: 74 centers in the USA</p>
Participants	<p>Randomized N = 599</p> <p>Completed N = 468</p> <p>Mean age: 62.3 ± 9.5 years</p> <p>Women: 67.1%</p> <p>Inclusion: Osteoarthritis (OA) of the knee > 6 months, aged ≥ 40 years, American Rheumatology Association (ARA) functional class I, II or III</p>
Interventions	<p>Celecoxib 200 mg (N = 241)</p> <p>Etoricoxib 30 mg (N = 231)</p> <p>First placebo group (N = 64)</p> <p>Second placebo group (N = 63)</p> <p>Participants who successfully completed part one were enrolled directly into part two, an active comparator 14-week follow-up. Participants on active treatment in part one remained on the same treatment in part two; participants receiving placebo in part one received either etoricoxib 30 mg or cele-</p>

Bingham 2007 Study 1 (Continued)

coxib 200 mg in part two, based on their initial randomization schedule at enrolment. Only the first part was considered for this systematic review

Outcomes	<p>Efficacy outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain sub scale using average of the first five questions of the WOMAC measured by visual analog scale (VAS) 0-100 mm; WOMAC physical function sub scale (average of questions 8 through 24 of the WOMAC measured by VAS 0-100 mm); Patient Global Assessment of Disease status (PGADS) measured by VAS 0-100 mm</p> <p>Harms outcomes: clinical and laboratory assessments and patient reported adverse events (AEs); pre-defined AEs: discontinuation due to any AE, discontinuation due to oedema, hypertension or gastrointestinal (GI) event or congestive heart failure (CHF)</p>
Funding	The study was funded by Merck & Co., Inc
Declaration of interest	<p>All authors reported potential conflict of interest related to funding from Merck & Co</p> <p>5/10 authors were employees of Merck & Co</p>
Notes	<p>Two placebo groups in both studies were shown as one group. Corresponding author was contacted, and he responded on May 8, 2013 that he is trying to get the information requested from Merck. The corresponding author did not respond to subsequent multiple queries about the missing information. Therefore, combined data for both placebo group were presented, as in the manuscript</p> <p>Trial registration: ClinicalTrials.gov, http://clinicaltrials.gov/, NCT00092768</p>

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind, double-dummy study (p. 497)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	<p>Attrition was 20% (49/241) in the celecoxib group (33/49 due to lack of efficacy or adverse events), 36% (23/64) in the first placebo group (20/23 due to lack of efficacy or adverse events) and 30% (19/63) in the second placebo group (17/19 due to lack of efficacy or adverse events)</p> <p>The primary efficacy analysis was a modified intention-to-treat (mITT) approach on time-weighted average (TWA) response. All patients with a baseline value and at</p>

Bingham 2007 Study 1 (Continued)

		least one post-baseline observation were included in the primary efficacy analysis. Only observed data were included in each patient's TWA response; no data were carried forward or imputed for this computation. A secondary per-protocol analysis removing pre-specified protocol violators was also carried out
Selective reporting (reporting bias)	Low risk	All primary and secondary outcomes and harms/adverse events data were included in report of results. Adverse events were categorised and described for each group
Other bias	Unclear risk	Acetaminophen 325 mg tablets were available as rescue analgesia at a maximum daily dose of 2600 mg. Patients were requested to use as little acetaminophen as possible, and discontinued acetaminophen use at least 12 h before visits 2 to 8. A tablet count of study medication was performed at each visit. Patients who missed > 20% of scheduled doses were considered non-compliant (p. 497). However, number of rescue medication that the patients in each group used, or number of non-compliant patients were not reported. It is unclear whether this could have influenced the study results

Bingham 2007 Study 2

Methods	<p>Study design: two randomized studies, double-blind, placebo and active controls, 4 parallel groups, multicenter studies</p> <p>Duration: 26 weeks; each study consisting of part one (12 weeks) and part two (14 weeks)</p> <p>Dates of study: March 2004 to February 2005</p> <p>Locations: 74 centers in the USA</p>
Participants	<p>Randomized N = 608</p> <p>Completed N = 474</p> <p>Mean age: 61.7 ± 9.2 years</p> <p>Women: 65.5%</p> <p>Inclusion: osteoarthritis (OA) of the knee > 6 months, age ≥ 40 years, American Rheumatology Association (ARA) functional class I, II or III</p>
Interventions	<p>Celecoxib 200 mg (N = 247)</p> <p>Etoricoxib 30 mg (N = 244)</p> <p>First placebo group (N = 58)</p> <p>Second placebo group (N = 59)</p> <p>Participants who successfully completed part one were enrolled directly into part two, an active comparator 14-week follow-up. Participants on active treatment in part one remained on the same treatment in part two; participants receiving placebo in part one received either etoricoxib 30 mg or celecoxib 200 mg in part two, based on their initial randomization schedule at enrolment. Only the first part was considered for this systematic review</p>
Outcomes	<p>Efficacy outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain sub scale using average of the first five questions of the WOMAC measured by visual analog scale (VAS) 0-100 mm; WOMAC physical function sub scale (average of questions 8 through 24 of the WOMAC mea-</p>

Bingham 2007 Study 2 (Continued)

sured by VAS 0-100 mm); Patient Global Assessment of Disease status (PGADS) measured by VAS 0-100 mm

Harms outcomes: clinical and laboratory assessments and patient reported adverse events (AEs); pre-defined AEs: discontinuation due to any AE, discontinuation due to oedema, hypertension or gastrointestinal (GI) event or chronic heart failure (CHF)

Funding	The study was funded by Merck & Co., Inc
Declaration of interest	All authors reported potential conflict of interest related to funding from Merck & Co 5/10 authors were employees of Merck & Co
Notes	Two placebo groups in both studies were shown as one group. Corresponding author was contacted, and he responded on May 8, 2013 that he is trying to get the information requested from Merck. The corresponding author did not respond to subsequent multiple queries about the missing information. Therefore, combined data for both placebo group were presented, as in the manuscript Trial registration: ClinicalTrials.gov, http://clinicaltrials.gov/ , NCT00092791

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind, double-dummy study (p. 497)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition was 18% (45/247) in the celecoxib group (32/45 due to lack of efficacy or adverse events), 48% (28/58) in the first placebo group (26/28 due to lack of efficacy or adverse events) and 49% (29/59) in the second placebo group (25/29 due to lack of efficacy or adverse events) The primary efficacy analysis was a modified intention-to-treat (mITT) approach on time-weighted average (TWA) response. All patients with a baseline value and at least one post-baseline observation were included in the primary efficacy analysis. Only observed data were included in each patient's TWA response; no data were carried forward or imputed for this computation. A secondary per-protocol analysis removing pre-specified protocol violators was also carried out

Bingham 2007 Study 2 (Continued)

Selective reporting (reporting bias)	Low risk	All primary and secondary outcomes and harms/adverse events data were included in report of results. Adverse events were categorised and described for each group
Other bias	Unclear risk	Acetaminophen 325 mg tablets were available as rescue analgesia at a maximum daily dose of 2600 mg. Patients were requested to use as little acetaminophen as possible, and discontinued acetaminophen use at least 12 h before visits 2 to 8. A tablet count of study medication was performed at each visit. Patients who missed > 20% of scheduled doses were considered non-compliant (p. 497). However, number of rescue medication that the patients in each group used, or number of non-compliant patients were not reported. It is unclear whether this could have influenced the study results

Birbara 2006 Study 1

Methods	<p>Study design: randomized, double-blind, placebo and active controls, 3 parallel groups, multicenter study</p> <p>Duration: 6 weeks</p> <p>Study dates: February to November 2003</p> <p>Locations: sites in the USA</p>
Participants	<p>Randomized N = 395</p> <p>Completed N = 345</p> <p>Mean age: 60.4 ± 10.3 years</p> <p>Female: 72%</p> <p>Inclusion: osteoarthritis (OA) of the knee ≥ 6 months, class I-III according to the American Rheumatology Association (ARA) criteria, visual analog scale (VAS) 0-100 pain walking on flat surface at least 40 mm, which had increased at least 15 mm since discontinuing their original OA medication. Also worsening in Investigator Global Assessment of Disease Status (IGADS) of at least 1 unit on a 5-point scale.</p> <p>Participants previously taking acetaminophen for OA symptoms were permitted to continue with their medication until 12 hours before the second visit. Acetaminophen participants were required to have a measure of at least 40 mm for Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain walking on a flat surface and a rating of 'very poor', 'poor' or 'fair' on IGADS at both the screening visit and randomization visit while on acetaminophen; no flare of disease was required. Participants who used low-dose aspirin (81 mg or less daily) for cardioprotective effects were permitted to continue low-dose aspirin use during the studies. Glucosamine and chondroitin sulfate, if taken for longer than 6 months, were also permitted if taken at the same stable dose for the duration of the studies</p>
Interventions	<p>Celecoxib 200 mg (N = 157)</p> <p>Rofecoxib 12.5 mg (N = 160)</p> <p>Placebo (N = 78)</p>
Outcomes	<p>Efficacy outcomes: Patient global assessment of response to therapy (PGART, 0-4 scale); WOMAC pain (VAS 0-100 mm); WOMAC physical function (VAS 0-100 mm)</p> <p>Harms outcomes: spontaneously reported adverse events (AEs); incidence of serious adverse events (SAEs), drug-related AEs, AEs which cause discontinuation, edema-related AEs causing discontinua-</p>

Celecoxib for osteoarthritis (Review)

Birbara 2006 Study 1 (Continued)

tions, hypertension-related AEs causing discontinuations, congestive heart failure, and gastro-intestinal AEs

Funding	Merck & Co. Inc
Declaration of interest	Not reported. 5/8 authors were employees of the Merck & Co. Inc. according to their affiliation data. In September 2004 the drug manufacturer announced the withdrawal of rofecoxib from the market; rofecoxib is no longer available for use
Notes	The manuscript where this study was published reported data for two separate randomized controlled trials (Study 1 and Study 2). Data for adverse events were shown for both studies pooled together. These data were divided proportionally between the two studies and included in the relevant analyses as such

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Participants were randomly assigned by computer-generated allocation schedule in a 2:2:1 ratio (p. 201)
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	All treatment regimens were fully masked with the use of a double-dummy technique, and both participants and investigators were blinded (p. 201)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	This manuscript reported two separate studies (Study 1 and Study 2) and adverse events were shown as combined for both studies (Table 3) High attrition was observed, with 27% (21/78) in the placebo group (24.4% due to adverse experience or lack of efficacy) and 8.9% (14/157) in the celecoxib group (5.1% due to adverse experience or lack of efficacy), and the main reasons for attrition were lack of efficacy and clinical adverse experience (Table 1). All randomized participants who took at least one dose of study medication, and had at least one on-treatment assessment, were included in a modified ITT analysis. The last on-treatment observation was carried forward in the event of missing data for all efficacy analyses (p. 202)
Selective reporting (reporting bias)	Low risk	All primary and secondary outcomes and harms/adverse events data were included in report of results. All serious adverse events were adjudicated. Three independent, blinded committees evaluated cardiovascular and gastro-intestinal AEs. Each of them had three separate subspecialty committees of three members (p. 202)

Birbara 2006 Study 1 (Continued)

Other bias	Unclear risk	Participants were allowed to take low-dose aspirin (81 mg or less daily) for cardioprotective effect; glucosamine and chondroitin sulfate (if taken for longer than 6 months, at the same stable dose); acetaminophen (max. dose 2 600 mg/d, as rescue therapy for OA pain if the study medication did not provide adequate pain control, and were instructed to discontinue use 12 hours before study visits) (p. 201). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results
------------	--------------	---

Birbara 2006 Study 2

Methods	<p>Study design: randomized, double-blind, placebo and active controls, 3 parallel groups, multicentre study</p> <p>Duration: 6 weeks</p> <p>Study dates: February to November 2003</p> <p>Locations: sites in the USA</p>
Participants	<p>Randomized N = 413</p> <p>Completed N = 344</p> <p>Mean age: 60.8 ± 10.9 years</p> <p>Female: 66%</p> <p>Inclusion: osteoarthritis (OA) of the knee ≥ 6 months, class I-III according to the American Rheumatology Association (ARA), VAS 0-100 pain walking on flat surface at least 40 mm, which had increased at least 15 mm since discontinuing their original OA medication. Also worsening in Investigator Global Assessment of Disease Status (IGADS) of at least 1 unit on a 5-point scale.</p> <p>Participants previously taking acetaminophen for OA symptoms were permitted to continue with their medication until 12 hours before the second visit. Acetaminophen participants were required to have a measure of at least 40 mm for Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain walking on a flat surface and a rating of 'very poor', 'poor' or 'fair' on IGADS at both the screening visit and randomization visit while on acetaminophen; no flare of disease was required. Participants who used low-dose aspirin (81 mg or less daily) for cardioprotective effects were permitted to continue low-dose aspirin use during the studies. Glucosamine and chondroitin sulfate, if taken for longer than 6 months, were also permitted if taken at the same stable dose for the duration of the studies.</p>
Interventions	<p>Celecoxib 200 mg (N = 169)</p> <p>Rofecoxib 12.5 mg (N = 159)</p> <p>Placebo (N = 85)</p>
Outcomes	<p>Efficacy outcomes: patient global assessment of response to therapy (PGART, 0-4 scale); WOMAC pain (VAS 0-100 mm); WOMAC physical function (VAS 0-100 mm)</p> <p>Harms outcomes: spontaneously reported adverse events (AEs); incidence of serious adverse events (SAEs), drug-related AEs, AEs which cause discontinuation, edema-related AEs causing discontinuations, hypertension-related AEs causing discontinuations, congestive heart failure, and gastro-intestinal AEs</p>
Funding	Merck & Co. Inc

Celecoxib for osteoarthritis (Review)

Birbara 2006 Study 2 (Continued)

Declaration of interest Not reported. 5/8 authors were employees of the Merck & Co. Inc. according to their affiliation data.

In September 2004 the drug manufacturer announced the withdrawal of rofecoxib from the market; rofecoxib is no longer available for use

Notes The manuscript where this study was published reported data for two separate randomized controlled trials (Study 1 and Study 2). Data for adverse events were shown for both studies pooled together. These data were divided proportionally between the two studies and included in the relevant analyses as such

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Participants were randomly assigned by computer-generated allocation schedule in a 2:2:1 ratio (p. 201)
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	All treatment regimens were fully masked with the use of a double-dummy technique, and both participants and investigators were blinded (p. 201)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	This manuscript reported two separate studies (Study 1 and Study 2) and adverse events were shown as combined for both studies (Table 3) High attrition was observed, with 30.4% (26/85) in the placebo group (18.8% due to adverse experience and lack of efficacy), while it was 15.4% (26/169) in the celecoxib group (10.1% due to adverse experience and lack of efficacy), and the main reasons for attrition were lack of efficacy and clinical adverse experience (Table 1). All randomized participants who took at least one dose of study medication, and had at least one on-treatment assessment, were included in a modified intention-to-treat (ITT) analysis. The last on-treatment observation was carried forward in the event of missing data for all efficacy analyses (p. 202)
Selective reporting (reporting bias)	Low risk	All primary and secondary outcomes and harms/adverse events data were included in report of results. All serious adverse events were adjudicated. Three independent, blinded committees evaluated cardiovascular and gastro-intestinal AEs. Each of them had three separate subspecialty committees of three members (p. 202)
Other bias	Unclear risk	Participants were allowed to take low-dose aspirin (81 mg or less daily) for cardioprotective effect; glucosamine and chondroitin sulfate (if taken for longer than 6 months, at the same stable dose); acetaminophen (max. dose 2 600 mg/d, as rescue therapy for OA pain if the study medication did not provide adequate pain control, and were instructed to discontinue use 12 hours before

Birbara 2006 Study 2 (Continued)

study visits) (p. 201). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Boswell 2008 Study A

Methods	<p>Study design: two multicenter, randomized, double-blind, placebo- and active-controlled, parallel-group, dose-ranging studies; study A – a 6-week, phase 2, non-flare design study; study B – a 12-week, phase 3, flare design study</p> <p>Study dates: September 2003 to April 2004</p> <p>Locations: study A – 50 centers in Europe and Australia; study B – 187 centers in Europe, North America and internationally</p>
Participants	<p>Randomized N = 649</p> <p>Completed N = 556</p> <p>Mean age: 63.4 years</p> <p>Female: 68%</p> <p>Inclusion: osteoarthritis (OA) of the knee ≥ 3 months; American College of Rheumatology (ACR) clinical criteria for OA of the knee; recent (≤ 12 months) radiographic evidence of tibiofemoral OA (grade 2 or 3 on the Kellgren & Lawrence scale); and ARA functional class rating I, II or III</p>
Interventions	<p>Celecoxib 200 mg (N = 109)</p> <p>GW406381 10 mg (N = 106)</p> <p>GW406381 20 mg (N = 101)</p> <p>GW406381 35 mg (N = 1108)</p> <p>GW406381 50 mg (N = 109)</p> <p>placebo (N = 107)</p>
Outcomes	<p>Primary efficacy outcomes:</p> <p>Study A – Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) OA index pain sub score change for GW406381 compared with placebo; “pain on walking on a flat surface” (WOMAC pain Q1) for participants receiving GW406381 compared with celecoxib</p> <p>Study B – change from baseline to week 12 in WOMAC pain sub score; WOMAC function sub score; Patient Global Assessment of Change (PGAC) for participants receiving GW406381 compared with placebo</p> <p>Secondary efficacy outcomes (if not specified as primary): the change from baseline in WOMAC pain Q1; WOMAC pain, stiffness, and physical function sub scores at end of treatment; physician’s and patient’s global assessment of arthritis; percentage of participants fulfilling the Outcome Measures in Rheumatology Clinical Trials – Osteoarthritis Research Society International (OMERACT-OARSI) responder criteria; rate of discontinuation for lack of efficacy; pattern of use of supplemental paracetamol</p> <p>Harms and tolerability outcomes: laboratory testing; electrocardiography; vital signs; assessment of pedal edema; reporting of adverse events and serious adverse events</p>
Funding	GlaxoSmithKline

Boswell 2008 Study A (Continued)

Declaration of interest	All authors disclosed potential conflict of interest. Of the eight authors, five were current and one was a former employee of the GlaxoSmithKline. One author owned GlaxoSmithKline stock. The eighth author received payment from GlaxoSmithKline for this study	
Notes	GW 406381 is an investigational, highly selective COX-2 inhibitor	
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	This was a double-blind study
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Blinded investigators assessed the knee with greatest pain at screening (“index knee”) throughout the study. For other outcomes, the process of outcome assessment is not explicitly described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total attrition was 14% (93/649), but attrition and reasons for attrition were not reported for each arm separately. The safety population included all participants who received at least one dose of study medication. The efficacy population was a modified intention-to-treat (mITT) population that included all participants who received at least one dose of drug and had at least one post-baseline assessment. Last observation carried forward was used to impute missing data points, and efficacy data are presented for the ITT population using this imputation
Selective reporting (reporting bias)	High risk	Data are not provided for the secondary outcomes: physician's and patient's global assessment of arthritis; percentage of participants fulfilling the Outcome Measures in Rheumatology Clinical Trials - Osteoarthritis Research Society International (OMERACT-OARSI) responder criteria; rate of discontinuation for lack of efficacy; and pattern of use of supplemental paracetamol. Statistical measures of dispersion not reported for Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) physical function sub scale. Adverse events shown only if their incidence was above 5% in any treatment group
Other bias	Unclear risk	Participants were allowed to take paracetamol (max. 3 g/d, if pain intensity became unacceptable at any stage of the study, with the exception of the 24-hour period before a clinic visit) (p. 261). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Boswell 2008 Study B

Methods	<p>Study design: randomized, double-blind, placebo and active controls, 7 parallel groups, dose-ranging phase 3, flare design multicenter study</p> <p>Duration: 12 weeks</p> <p>Study dates: May to December 2005</p> <p>Locations: 187 centers in Europe, North America and internationally</p>
Participants	<p>Randomized N = 1331</p> <p>Completed N = 1038</p> <p>Mean age: 66.5 years</p> <p>Female: 70%</p> <p>Inclusion: osteoarthritis (OA) of the knee ≥ 3 months; American College of Rheumatology (ACR) clinical criteria for OA of the knee; recent (≤ 12 months) radiographic evidence of tibiofemoral OA (grade 2 or 3 on the Kellgren & Lawrence scale); and American Rheumatology Association (ARA) functional class rating I, II or III</p>
Interventions	<p>Celecoxib 200 mg (N = 185)</p> <p>GW406381 1 mg (N = 188)</p> <p>GW406381 5 mg (N = 187)</p> <p>GW406381 10 mg (N = 188)</p> <p>GW406381 25 mg (N = 189)</p> <p>GW406381 50 mg (N = 186)</p> <p>Placebo (N = 186)</p>
Outcomes	<p>Primary efficacy outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain on walking on a flat surface change from baseline to week 6 (Q1)</p> <p>Secondary efficacy outcomes: the change from baseline in WOMAC pain Q1; WOMAC pain, stiffness, and physical function; physician's and patient's global assessment of arthritis; percentage of s fulfilling the Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trial Response Criteria Initiative (OMERACT-OARSI) responder criteria; rate of discontinuation for lack of efficacy; pattern of use of supplemental paracetamol</p> <p>Harms and tolerability outcomes: laboratory testing; electrocardiography; vital signs; assessment of pedal edema; reporting of adverse events and serious adverse events</p>
Funding	GlaxoSmithKline
Declaration of interest	All authors disclosed potential conflict of interest. Of the eight authors, five were current and one was former employee of the GlaxoSmithKline. One author owned GlaxoSmithKline stock. The eighth author received payment from GlaxoSmithKline for this study
Notes	GW 406381 is an investigational, highly selective COX-2 inhibitor

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described

Celecoxib for osteoarthritis (Review)

Boswell 2008 Study B (Continued)

Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	This was a double-blind study
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Blinded investigators assessed the knee with greatest pain at screening ("index knee") throughout the study. For other outcomes, the process of outcome assessment is not explicitly described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Total attrition was high with the total of 22% (293/1131) participants that did not complete the study; attrition and reasons for attrition were not reported for each arm separately. The safety population included all participants who received at least one dose of study medication. The efficacy population was a modified intention-to-treat (mITT) population that included all patients who received at least one dose of drug and had at least one post-baseline assessment. Last observation carried forward was used to impute missing data points, and efficacy data are presented for the ITT population using this imputation
Selective reporting (reporting bias)	High risk	Data are not provided for the secondary outcomes: physician's and patient's global assessment of arthritis; percentage of patients fulfilling the Outcome Measures in Rheumatology Clinical Trials - Osteoarthritis Research Society International (OMERACT-OARSI) responder criteria; rate of discontinuation for lack of efficacy; and pattern of use of supplemental paracetamol. Statistical measures of dispersion not reported for WOMAC physical function sub scale. Adverse events shown only if their incidence was above 5% in any treatment group
Other bias	Unclear risk	Patients were allowed to take paracetamol (max. 3 g/d, if pain intensity became unacceptable at any stage of the study, with the exception of the 24-hour period before a clinic visit) (p. 261). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Clegg 2006

Methods	<p>Study design: randomized, double-blind, placebo and active controls, 5 parallel groups, multicenter study</p> <p>Study acronym: GAIT (Glucosamine/Chondroitin Arthritis Intervention Trial)</p> <p>Duration: 24 weeks</p> <p>Study dates: November 2000 to July 2004</p> <p>Locations: 16 clinical centers in the USA</p>
Participants	Randomized N = 1583

Celecoxib for osteoarthritis (Review)

Clegg 2006 (Continued)

Completed N = 1258

Mean age: 58.6 ± 10.3 years

Female: 64%

Inclusion: osteoarthritis (OA) of the knee, age ≥ 40 years, clinical evidence (knee pain for at least six months and on the majority of days during the preceding month) and radiographic evidence (Kellgren and Lawrence grade 2 or 3). Summed pain score of 125 to 400 on the index (more symptomatic) knee according to the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and class I, II, or III American Rheumatology Association (ARA) criteria

Interventions	<p>Celecoxib 200 mg (N = 318)</p> <p>Glucosamine 1500 mg/day (500 mg three times a day) (N = 317)</p> <p>Chondroitin sulfate 1200 mg/day (400 mg three times a day) (N = 318)</p> <p>Glucosamine 1500 mg plus chondroitin sulfate 1200 mg daily (N = 317)</p> <p>Placebo (N = 313)</p>
Outcomes	<p>Primary efficacy outcome: 20% decrease in the summed score for the WOMAC pain sub scale from baseline to week 24</p> <p>Secondary efficacy outcomes: Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trial Response Criteria Initiative (OMER-ACT-OARSI) response rate; scores for the stiffness and function sub scales of WOMAC; the patient's global assessments of disease status and response to therapy using visual analog scale (VAS) 0-100 mm; the investigator's global assessment of disease status (VAS 0-100 mm); the presence or absence of soft-tissue swelling, effusion, or both in the index knee; Short-form Health Survey (SF-36); alternative disability score and pain score on the Health Assessment Questionnaire (HAQ); and acetaminophen use, according to diary entries and tablet counts</p> <p>Harms outcomes: reported adverse events (AEs) and serious adverse events (SAEs); complete blood counts; measurement of serum aspartate aminotransferase, alanine aminotransferase, glucose, creatinine, and partial thromboplastin time; urine analysis</p>
Funding	Supported by a contract from the National Center for Complementary and Alternative Medicine and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NO1AR92236)
Declaration of interest	Authors reported financial ties with companies producing the study drugs, including consulting fees, sitting on company advisory boards, lecture fees, grant support or expert consultants. One author reported receiving royalties from books related to osteoarthritis
Notes	<p>Recruitment began in November 2000, at 13 clinical centers, and 3 centers were added to the study in February 2003 to ensure timely recruitment.</p> <p>Trial registration: ClinicalTrials.gov, http://clinicaltrials.gov/, NCT00032890</p>

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Permuted block randomisation was used with random block sizes stratified according to baseline WOMAC pain stratum and 16 clinical centres. The randomisation code list was developed by the Veterans Affairs Cooperative Studies Program Data Coordinating Centre in Hines, Illinois (p. 796)
Allocation concealment (selection bias)	Low risk	The randomisation code list was developed centrally. During data collection, neither the clinical centres nor the coordinating centre at the University of

Celecoxib for osteoarthritis (Review)

Clegg 2006 (Continued)

Utah had access to the randomisation codes or statistical summaries of follow-up data (p. 796)		
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Study drug assignment and the investigators were blinded. Study drugs containing glucosamine hydrochloride, chondroitin sulfate, their combination and placebo drugs were manufactured to look the same, celecoxib was over-encapsulated (for masking) and matching placebo was prepared (p. 797)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Adverse events and serious events were assessed by the blinded investigator at each study visit. For other non-patient reported outcomes, the process of outcome assessment is not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Attrition of 20.8% (65/313) was reported in the placebo group (10.5% due to adverse event or lack of efficacy) and 16.4% (52/318) in the celecoxib group (5.7% due to adverse event or lack of efficacy) (Figure 1). Analysis of the primary outcome measure was conducted according to the ITT population (p. 797). The last observation carried forward (LOCF) method was used in the analysis of all outcomes among participants who made at least one follow-up visit but who did not complete follow-up (p. 798)
Selective reporting (reporting bias)	High risk	Certain secondary outcomes, including measure for quality of life, were reported only for baseline, not for the subsequent follow-up period. Adverse events per each study group were insufficiently described
Other bias	Unclear risk	Participants were allowed to take acetaminophen (max. 4 g/d, except during the 24 hours before a clinical evaluation for joint pain) (p. 796). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Conaghan 2013

Methods	Study design: randomized, double blind, placebo and active controls, 6 parallel groups, multicenter Duration: 12 weeks Study dates: May 2008 to April 2009 Locations: 71 centers in the Czech Republic, Germany, Poland and the UK
Participants	Randomized N = 1399 Completed N = 1256 Mean age: 61.2 years Female: 66% Inclusion: age > 45 years with a primary diagnosis of functional class I-III knee osteoarthritis (OA), meeting American College of Rheumatology (ACR) clinical classification criteria for knee OA, defined by knee pain and 4 or more of the following: (i) age > 50 years; (ii) morning stiffness of < 30 minute duration; (iii) crepitus on active motion; (iv) bony tenderness; (v) bony enlargement and (vi) no palpable warmth of the synovium

Celecoxib for osteoarthritis (Review)

Conaghan 2013 (Continued)

Interventions	<p>Celecoxib tablets 200 mg (N = 235)</p> <p>IDEA-033/ketoprofen 50 mg (N = 233)</p> <p>IDEA-033/ketoprofen 100 mg (N = 230)</p> <p>2.2 g TDT 064/vehicle (N = 238)</p> <p>4.4 g TDT 064/vehicle (N = 235)</p> <p>Oral placebo (N = 228)</p>
Outcomes	<p>Primary efficacy outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain sub scale</p> <p>Secondary efficacy outcomes: patient global assessment of response to therapy, WOMAC function sub scale, WOMAC stiffness sub scale, proportion of participants defined as responders and time to onset of clinically meaningful, sustainable pain relief</p> <p>Harms: adverse events, the number of participants requiring therapy with omeprazole and the dose received, vital signs, clinical laboratory assessments</p>
Funding	This work was supported by IDEA AG, Germany
Declaration of interest	2/5 authors were employees of the sponsor; authors received consultancy fees, investigator grants and honoraria for expert advice from companies producing study drugs
Notes	Trial registration: ClinicalTrials.gov, http://clinicaltrials.gov/ , NCT00716547

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Participants randomized according to a center-specific assignment list generated by a random permuted block scheme (p. 1304)
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Investigators and participants were blinded to whether they were allocated to active or placebo treatment, but not to whether they were allocated topical or oral medication (p. 1304)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition was 11% (26/235) in celecoxib group (7.7% due to adverse events or lack of efficacy) and 17% (39/228) in control group (15% due to adverse events or lack of efficacy) (Figure 1). The primary efficacy analysis was performed on participants who received at least one dose of study medication, the intention-to-treat (ITT) population. Missing follow-up values were replaced using baseline observation carried forward (BOCF) (p. 1305)

Celecoxib for osteoarthritis (Review)

Conaghan 2013 (Continued)

Selective reporting (reporting bias)	Low risk	All outcomes reported in the methods are also reported in the results section
Other bias	Unclear risk	Participants were allowed to take paracetamol (500 mg up to four times per day, although not within 24 h of the next study visit or between B1 i B2). Participants requiring ≥ 2 g/day of paracetamol or other analgesic medication for longer than 3 consecutive days considered treatment failures and withdrawn from study (p. 1304). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Dahlberg 2009

Methods	Study design: randomized, double-blind, active controls, 2 parallel groups, multicentre study Duration: 52 weeks Study dates: August 1999 to October 2001 Locations: 124 general practitioner centers in Sweden (45 sites) and Norway (79 sites)	
Participants	Randomized N = 925 Completed N = 550 Mean age: 71 ± 7.2 years Female: 69% Inclusion: at least 60 years, clinical diagnosis of osteoarthritis (OA) of the hip or knee, class I-III according to the American College of Rheumatology (ACR) criteria	
Interventions	Celecoxib 200 mg (N = 463) Diclofenac 100 mg/day (50 mg twice daily) (N = 462)	
Outcomes	Primary efficacy outcome: incidence of discontinuation of study drug due to AEs Secondary efficacy outcomes: time to discontinuation of study medication; the Patient's and Physician's Global Assessment of Arthritis (PaGAA, PhGAA) on a 5-point scale; the Patient's Assessment of Arthritis Pain (PAAP) at rest on a visual analog scale (VAS) 0-100 mm scale; patient's and physician's satisfaction assessments on a 10-point scale; Harms outcomes: incidence and severity of adverse events (AEs); changes in vital signs; haematology and blood chemistry laboratory results; incidence and severity of gastrointestinal (GI) symptoms evaluated with the Severity of Dyspepsia Assessment (SODA); healthcare resource utilization	
Funding	Pfizer Inc	
Declaration of interest	Authors reported financial ties with companies producing the study drugs, including consulting fees for their work in a steering committee of a clinical trial and clinical research funding	
Notes		
Risk of bias		
Bias	Authors' judgement	Support for judgement

Celecoxib for osteoarthritis (Review)

Dahlberg 2009 (Continued)

Random sequence generation (selection bias)	Low risk	Randomisation was done according to a computer-generated list prepared by Biostat AS unit in Oslo, Norway. Randomisation was stratified by centre with a block size of six (p. 134)
Allocation concealment (selection bias)	Low risk	Central randomisation (p. 134)
Blinding of participants and personnel (performance bias) All outcomes	Low risk	To conceal allocation, the double-blind supplies were identical in physical appearance and were provided to the investigator in bottles marked with randomisation number and visit number (p. 134)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition was high, with 39% (181/463) of participants lost in the placebo group (153/181 due to adverse event or lack of efficacy) and 40% (185/462) in the celecoxib group (152/462 due to adverse event or lack of efficacy). Part of the patients who discontinued drug remained in health economic study (Figure 1). All randomized participants who had taken at least one dose of study medication were included in the modified intention-to-treat (mITT) and the safety populations. Unless otherwise indicated, adverse events (AE)-related discontinuation, efficacy, and health economic analyses were performed on the mITT population. Safety analyses were performed on the safety population (identical to the mITT population). The last observation carried forward (LOCF) approach was used for missing observations (p. 135)
Selective reporting (reporting bias)	Unclear risk	All outcomes described in this manuscript were pre-specified, and all major pre-specified outcomes are presented in report of results (p. 135). Incidence of adverse events reported only if they occurred in at least 5% of participants
Other bias	Unclear risk	Participants were allowed to take paracetamol (for relief of arthritis pain, up to 3 g daily for short periods); aspirin (≤ 75 mg/d, for cardioprotection); occasional use of antacids was also permitted (p. 134). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

DeLemos 2011

Methods	Study design: randomized, double-blind, placebo and active controls, 5 parallel groups, multicentre study Duration: 12 weeks Study dates: September 2002 to August 2003 Locations: Canada (42 centers), USA (71 centers), Germany (30 centers), Italy (12 centers), and UK (7 centers)
Participants	Randomized N = 1011

Celecoxib for osteoarthritis (Review)

DeLemos 2011 (Continued)

Completed N = 555

Mean age: 60 years (range 20 to 80 years)

Female: 63%

Inclusion: osteoarthritis (OA) of the knee and/or hip, stage I-III according to the American College of Rheumatology (ACR) criteria, baseline pain intensity of ≥ 40 on a 100-mm visual analogue scale (VAS) where 0 = no pain, 100 = extreme pain

Interventions	Celecoxib 200 mg (N = 202) Tramadol ER 100 mg (N = 201) Tramadol ER 200 mg (N = 199) Tramadol ER 300 mg (N = 199) Placebo (N = 200)	
Outcomes	<p>Efficacy outcomes: patients completed the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), and they rated arthritis pain intensity in the index and non-index joints during the past 48 hours using a 100-mm VAS (0 = no pain, 100 = extreme pain); patient and physician global assessments of disease activity were recorded at each study visit on 100-mm VAS (0= very good, 100= very poor in response to the question, “how is the patient doing today” and “considering all the ways your arthritis condition affects you, how are you doing today”, respectively); patients assessed sleep with the Chronic Pain Sleep Inventory, which included a 100-mm visual analogue scale for overall quality of sleep over the past week (0 = very poor, 100 = excellent); patients completed the Short Form-36 (SF-36) Health Survey</p> <p>Harms outcomes: reports of adverse events, either spontaneously or in response to non-directed questioning, and results of physical examinations, vital signs, clinical laboratory tests, and electrocardiograms at study visits; the 49-item Short Form Addiction Research Center Inventory (ARCI) questionnaire at baseline and week 12; the 16-item Physical Dependence Questionnaire (PDQ) at baseline, week 12 (or early discontinuation), and week 13 (or 1 week after early discontinuation).</p>	
Funding	Biovail Corporation and Ortho-McNeil Janssen Scientific Affairs, LLC	
Declaration of interest	Not reported. According to the author affiliation, 3/6 study authors were employees of Ortho-McNeil Janssen Scientific Affairs, LLC., one author was an employee of Biopharmatech Consulting Inc and two authors were employees of Biovail Corporation	
Notes		
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	The double-blind was maintained during the study (p. 218)
Blinding of outcome assessment (detection bias):	Low risk	Participants blinded

Celecoxib for osteoarthritis (Review)

DeLemos 2011 (Continued)

Patient-reported out-comes

Blinding of outcome assessment (detection bias): Non-patient-reported out-comes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition was high. In the placebo group 97/200 (49%) participants discontinued study treatment (Adverse event: 15, Lack of efficacy: 65, Subject choice: 4, Other: 13). In the celecoxib 200 mg group 67/202 (33%) participants discontinued study treatment (Adverse event: 20, Lack of efficacy: 30, Subject choice: 2, Other: 15) (p. 219)
Selective reporting (reporting bias)	Low risk	All primary and secondary outcomes were included in report of results. (p. 222). Adverse events were defined and reported (p. 223).
Other bias	Unclear risk	The authors write that "No concomitant or rescue analgesic therapy was permitted during the study." However, then they indicate: "Aspirin up to 325 mg/d for cardiovascular prophylaxis was allowed, as was acetaminophen up to 2000 mg/d for no more than 3 consecutive days for reasons other than osteoarthritis or chronic pain if absolutely necessary. The use of acetaminophen was to be avoided during the washout period and within 48 hours before each post-screening study visit" (p. 218). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Emery 2008

Methods	Study design: randomized, double-blind, active controls, 2 parallel groups, multicenter study Duration: 12 weeks Study dates: August 2003 to February 2005 Locations: 40 centers in the UK
Participants	Randomized N = 249 Completed N = 149 Mean age: 64 ± 8.9 years Female: 46% Inclusion: age ≥ 45 years, osteoarthritis (OA) of the hip class I-III according to the American College of Rheumatology (ACR) criteria requiring joint replacement surgery in the opinion of the investigator
Interventions	Celecoxib 200 mg (N = 126) Diclofenac 150 mg (50 mg three times a day) (N = 123)
Outcomes	Primary efficacy outcome: change from baseline to week 6 in the patient's assessment of arthritis pain on walking on a flat surface using visual analog scale (VAS) 0-100 mm Secondary efficacy outcomes: change from baseline to week 12 in the patient's assessment of arthritis pain on walking (VAS 0-100 mm); the Pain Satisfaction Scale score at weeks 6 and 12; the patient's and physician's global assessment of arthritis scores (5-point scale)

Celecoxib for osteoarthritis (Review)

Emery 2008 (Continued)

Harms outcomes: adverse events (AEs) and serious adverse events (SAEs); physical examination findings; sitting blood pressure; pulse

Funding	Pfizer Inc
Declaration of interest	3/4 authors were employees of Pfizer. The fourth author declared commercial ties with companies producing the study drugs, including undertaking clinical trials and providing expert advice
Notes	<p>The study protocol was amended such that change in VAS at week 6 rather than week 12 became the primary end point. The authors stated that the amendment did not affect the power of the study, but it allowed participation of patients if they had a period of ≥ 8 weeks before any planned surgery</p> <p>The corresponding author Tamas Koncz was contacted twice to clarify pain outcomes. The author did not respond</p>

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Eligible participants were randomized in the order in which they were admitted to the study according to a computer-generated schedule provided by sponsor (p. 72)
Allocation concealment (selection bias)	Low risk	Central randomisation by sponsor (p. 72)
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Study medication was administered in a double-blind, double-dummy fashion (p. 72)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition rates were high, with 43% (54/126) participants lost in the celecoxib group (30/54 due to lack of effectiveness and adverse events) and 37% (45/123) in the naproxen group (25/45 due to lack of effectiveness and adverse events (Figure 1, p. 74). The primary effectiveness outcome was determined in the evaluable population. VAS scores at week 6 were also analyzed in the modified intention-to-treat (mITT) population, which included all participants who were randomized and received at least one dose of study medication, had a baseline visual analog scale (VAS) score, and had a VAS score at week 6, week 12, or at the withdrawal visit. Although it is indicated that missing data were imputed using last observation carried forward (LOCF) approach (p. 73), this was not done for the outcome measure "arthritis pain on walking" (p. 76)
Selective reporting (reporting bias)	High risk	All outcomes specified in the methods were also reported in the results section. The study protocol was amended such that change in VAS at week 6 rather than week 12 became the primary end point. This amendment was reviewed and approved by the institutional review board and the Multicenter Research Ethics Committee. The amendment did not affect the power of the study, but it allowed participation of participants if they had a period of ≥ 8

Celecoxib for osteoarthritis (Review)

Emery 2008 (Continued)

weeks before any planned surgery (p. 72, 73). Only most frequently reported adverse events were presented, which had incidence above 2% in at least one treatment arm. Data for "withdrawal due to adverse events" from Figure 1 and the same outcome mentioned in text on page 78 do not match. It is indicated that 6 serious adverse events were recorded, but it is unclear in which study group(s)

Other bias	Unclear risk	Participants were allowed to take acetaminophen (at a max. daily dose of 4 g as rescue medication for arthritis flare or non-arthritis pain, but the medication had to be discontinued ≥ 8 hours before arthritis assessments at baseline, week 6, and week 12) and aspirin (for non-arthritis reasons, at a dose of ≤ 75 mg four times daily) (p. 72). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results
------------	--------------	--

Essex 2012a

Methods	Study design: randomized, double-blind, active controls, 2 parallel groups, multicentre study Duration: 6 months Study dates: March 2004 to January 2005 Locations: 47 centers in the USA
Participants	Randomized N = 589 Completed N = 391 Mean age: 60.4 $\pm$ 10.9 years Female: 66% Inclusion: age ≥ 40 years, osteoarthritis (OA) of the knee class I-III according to the American College of Rheumatology (ACR) criteria
Interventions	Celecoxib 200 mg (N = 296) Naproxen 1000 mg/day (500 mg twice daily) (N = 293)
Outcomes	Primary efficacy outcome: 20% improvement from baseline to 6 months in Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) total score Secondary efficacy outcomes: changes from baseline to month 6 in total and sub scale WOMAC scores for participants with 20% improvement; the patient's and physician's global assessments of arthritis (5-point scale); the patient's assessment of arthritis pain using visual analog scale (VAS) 0-100 mm Harms outcomes: AEs; clinically significant changes in vital signs
Funding	Pfizer Inc
Declaration of interest	All three authors were full-time employees of Pfizer, Inc
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
------	--------------------	-----------------------

Essex 2012a (Continued)

Random sequence generation (selection bias)	Low risk	Participants were randomized at study entry using a computer-generated randomisation scheme (p. 1359)
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	This was a double-blind, double-dummy study (p. 1358)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	AEs were subjectively classified by the blinded investigator as mild, moderate or severe (p. 1360). For other non-patient-reported outcomes, the process of outcome assessment is not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition of participants was high with 32% (94/296) of participants lost in the celecoxib group (44/94 due to adverse events) and 35% (104/293) in the naproxen group (64/104 due to adverse events) as shown in Figure 1 (p. 1362). Efficacy analyses were performed on the modified intention-to-treat (mITT) population (all who received at least one dose of study drug and had at least one post-baseline efficacy measurement), and safety analyses were performed on the safety population (all who received at least one dose of study medication) (p. 1360)
Selective reporting (reporting bias)	Unclear risk	All outcomes specified in the methods are also reported in the results section. Adverse events reported if their incidence was above 2% in any group (Table 4). All details about statistical analyses were not reported (information about factors used for calculating least squares mean were missing)
Other bias	Unclear risk	Participants were allowed to take aspirin (at stable dose of ≤ 325 mg/d for cardiovascular prophylaxis); calcium-containing antacids (for osteoporosis prevention); paracetamol (≤ 2 g/day for ≤ 3 consecutive days, for reasons other than arthritis was permitted if absolutely necessary, but not within 24 h of any visit); topical pain relief (to non-index joints was allowed except within 48 h of any visit) (pp. 1358-9). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Essex 2012b

Methods	Study design: randomized, double blind, placebo and active controls, 3 parallel groups, multicenter Duration: 6 weeks Study dates: 31 October 2001 to 20 November 2002 Locations: 28 centers in the USA
Participants	Randomized N = 322 Completed N = 253

Celecoxib for osteoarthritis (Review)

Essex 2012b (Continued)

Mean age: 58 ± 8.6 years

Female: 79.3%

Inclusion: African American participants aged ≥ 45 years, with osteoarthritis (OA) of the knee diagnosed according to American College of Rheumatology (ACR) guidelines, in a flare state, and with a functional capacity classification of I – III

Interventions	Celecoxib 200 mg (N = 127) Naproxen 1000 mg/day (500 mg twice daily) (N = 128) Placebo (N = 67)
Outcomes	Primary efficacy outcomes: Patient's Assessment of Arthritis Pain measured on a 0-100 mm visual analog scale (VAS) Secondary efficacy outcomes: Patient's and Physician's Global Assessments of Arthritis, Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), American Pain Society (APS) pain scores, Pain Satisfaction Scale, Patient Health Questionnaire-9 (PHQ-9) scores Harms: measurement of upper gastro-intestinal (UGI) tolerability
Funding	Pfizer, Inc
Declaration of interest	All three authors full-time employees of Pfizer, Inc. Editorial support was funded by Pfizer
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Computer-generated randomization schedule (p. 2255)
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind study
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	High attrition observed in the study arms, with celecoxib 20% (9/25 due to adverse events of lack of efficacy), naproxen 16% (9/21 due to adverse events or lack of efficacy), placebo 34% (12/23 due to adverse events or lack of efficacy) as shown in Figure 1. Primary efficacy analysis was carried out on evaluable population, while secondary efficacy analyses were performed using modified intention-to-treat populations (p. 2356)

Celecoxib for osteoarthritis (Review)

Essex 2012b (Continued)

Selective reporting (reporting bias)	High risk	Data not shown for all secondary outcomes
Other bias	Unclear risk	Participants were allowed to take acetaminophen (up to 2 g/day, for ≤ 3 consecutive days, and not within 24 h before any arthritis assessments), aspirin (≤ 325 mg/d, if they had been taking a stable dose for ≥ 30 days before the first dose of study medication) (p. 2255). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Essex 2014

Methods	Study design: randomized, double blind, placebo and active controls, 3 parallel groups, multicenter Duration: 6 weeks Study dates: Not indicated Locations: 31 centers in the USA
Participants	Randomized N = 318 Completed N = 236 Mean age: 60.6 ± 10.6 years Female: 65.3% Inclusion: age ≥ 45 years, of self-reported Hispanic descent with osteoarthritis (OA) of the knee diagnosed according to American College of Rheumatology (ACR) criteria, who were determined to be in a flare state and had a functional capacity classification of I to III
Interventions	Celecoxib 200 mg (N = 127) Naproxen 1000 mg (500 mg twice daily) (N = 129) Placebo (N = 62)
Outcomes	Primary efficacy outcomes: Patient’s Assessment of Arthritis Pain on a 0-100 mm visual analog scale (VAS) Secondary efficacy outcomes: Patient’s and Physician’s Global Assessments of Arthritis, Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), American Pain Society (APS) pain scores, Pain Satisfaction Scale and Patient Health Questionnaire-9 (PHQ-9) scores Harms: upper gastrointestinal tolerability
Funding	Pfizer, Inc
Declaration of interest	All four authors were full-time employees of Pfizer, Inc. Editorial support was funded by Pfizer, Inc
Notes	
Risk of bias	
Bias	Authors' judgement Support for judgement

Essex 2014 (Continued)

Random sequence generation (selection bias)	Low risk	Predetermined computer randomization schedule (p. 228)
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind study
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Low risk	All assessments were made by individuals who had been blinded (p. 228)
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Attrition was high with 24% participants lost in the celecoxib group (5/30 due to adverse events or lack of efficacy), 28% in the naproxen group (11/36 due to adverse events or lack of efficacy) and 26% in the placebo group (6/16 due to adverse events or lack of efficacy) as shown in Figure 1. The population evaluable for efficacy was used for the primary efficacy analysis and the modified intent-to-treat population was used for secondary efficacy analyses (pp. 228-9). It is unclear how missing values were handled
Selective reporting (reporting bias)	High risk	Data not shown for all secondary outcomes
Other bias	Low risk	From the data provided, no indication of other important risks of bias

Fleischmann 2005

Methods	Study design: randomized, double-blind, placebo and active controls, 4 parallel groups, multicenter study Duration: 13 weeks Study dates: not reported Locations: Argentina, Chile, Colombia, Peru, Uruguay, Venezuela, Switzerland, Canada, USA
Participants	Randomized N = 1608 Completed N = 1238 Mean age: 61.2 ± 11.3 years Female: 66% Inclusion: age ≥18 years, primary osteoarthritis (OA) of the knee ≥ 3 months according to the American College of Rheumatology (ACR) criteria, visual analog scale (VAS) 0-100 mm pain in the target knee ≥ 40 mm
Interventions	Celecoxib 200 mg (N = 446)

Celecoxib for osteoarthritis (Review)

Fleischmann 2005 (Continued)

Lumiracoxib 200 mg (N = 465)

Lumiracoxib 400 mg (N = 465)

Placebo (N = 232)

Outcomes	<p>Primary efficacy outcomes: pain intensity (VAS mm) in the target knee; patient's global assessment of disease activity (VAS mm); Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain and function</p> <p>Secondary efficacy outcomes: OA pain intensity in the target knee by visit and physician's and patient's global assessments of disease activity by visit</p> <p>Harms outcomes: adverse events (AEs) and serious AEs, standard laboratory tests and urinalysis were regularly collected, electrocardiogram measurements were obtained at screening and at week 13</p>
Funding	Novartis Pharma AG
Declaration of interest	Not reported. According to the author affiliation, 3/6 authors were employees of Novartis
Notes	The corresponding author Roy Fleischmann was contacted twice to clarify pain outcomes. The author did not respond

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding was maintained by means of a double-dummy technique (p. 43)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Attrition was high, with 22% (99/446) participants lost in the celecoxib group (66/99 due to lack of efficacy or adverse events) and 29% (68/232) in the placebo group (52/68 due to lack of efficacy or adverse events) as shown in Figure 1. Efficacy evaluations were performed using the intention-to-treat (ITT) population, with last observation carried forward (LOCF) techniques used for imputing missing results. Some of the efficacy variables were analysed using the ITT modified data set (with no efficacy data being carried forwards) and the per protocol populations at an exploratory level. Tolerability evaluations were performed in all participants randomized to treatment who had been exposed to study medication (p. 44)

Fleischmann 2005 (Continued)

Selective reporting (reporting bias)	High risk	Quality of life data, measured by SF-36, not provided. Data for all adverse events that were specified as safety outcomes not provided, only summary of adverse events (Table 4)
Other bias	Unclear risk	Participants were allowed to take paracetamol (≤ 2 g/d, as rescue medication during the screening and treatment periods); low-dose aspirin (≤ 325 mg/d for a CV indication) (p. 43). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Gibofsky 2003

Methods	<p>Study design: randomized, double-blind, placebo and active controls, 3 parallel groups, multicenter study</p> <p>Duration: 6 weeks</p> <p>Study dates: not reported</p> <p>Locations: 61 centers in the USA and Canada</p>
Participants	<p>Randomized N = 477</p> <p>Completed N = 383</p> <p>Mean age: 62.9 ± 10.2 years</p> <p>Female: 67%</p> <p>Inclusion: age ≥ 40 years, osteoarthritis (OA) of the knee class I-III according to the American College of Rheumatology (ACR) criteria, OA in a flare state at baseline</p>
Interventions	<p>Celecoxib 200 mg (N = 189)</p> <p>Rofecoxib 25 mg (N = 190)</p> <p>Placebo (N = 98)</p>
Outcomes	<p>Primary efficacy outcomes: patient's assessment of arthritis pain on the visual analog scale (VAS) to measure OA pain and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) total domain score at week 6</p> <p>Secondary efficacy outcomes: patient's and physician's global assessments patient's assessment of arthritis pain on the VAS (pain on walking), and WOMAC sub scales for pain, stiffness, and physical function</p> <p>Harms outcomes: adverse events (AEs) and serious AS, laboratory tests</p>
Funding	Pharmacia Corporation
Declaration of interest	One out of four authors was an employee of the Pharmacia Corporation
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
------	--------------------	-----------------------

Gibofsky 2003 (Continued)

Random sequence generation (selection bias)	Low risk	Treatment assignment was based on a computer-generated randomisation schedule prepared prior to the beginning of the study (p. 3103)
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Regimens were masked (double-dummied) so as to be unidentifiable to both participants and study personnel (p. 3103)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	High overall attrition rates were observed, with attrition of 16% (31/189) in the celecoxib group (21/31 due to treatment failure or adverse event) and 35% (34/96) in the placebo group (26/34 due to treatment failure or adverse event) as shown in Figure 1 (p. 3104). Analyses of the primary efficacy measures were based on the evaluable cohort, with supporting analyses based on the intention-to-treat (ITT) cohort. The evaluable cohort included all treated participants except those with no post-baseline data and those with major protocol deviations. The ITT cohort included all randomized participants who received at least one dose of study medication. Analyses of secondary efficacy measures were based on the ITT cohort. Any missing observations were imputed using the last observation carried forward (LOCF) method (p. 3105)
Selective reporting (reporting bias)	Low risk	All outcomes specified in the methods are also reported in the results section. Adverse events were compared across treatment groups and included all participants who received at least one dose of study drugs and consisted of laboratory findings and the adverse event (AE) incidences
Other bias	Unclear risk	Participants were allowed low-dose aspirin (≤ 325 mg/d, for cardiovascular prophylaxis); occasional use of acetaminophen (use of acetaminophen had to be discontinued for 48 hours prior to the arthritis assessments) or antacids (p. 3104). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Hochberg 2011 Study 307

Methods	Study design: randomized, double-blind, placebo and active controls, 3 parallel groups, multicenter study Duration: 12 weeks Study dates: April to December 2008 Locations: 157 centers in the USA
Participants	Randomized N = 619

Celecoxib for osteoarthritis (Review)

Hochberg 2011 Study 307 (Continued)

Completed N = 521

Mean age: 61.9 years (range: 49 to 90 years)

Female: 64%

Inclusion: age $\geq$ 50 years, osteoarthritis (OA) of the knee class I-III according to the American College of Rheumatology (ACR) criteria

Interventions	<p>Celecoxib 200 mg (N = 247)</p> <p>Naproxen 1000 mg plus esomeprazole 40 mg magnesium tablets daily (500 mg + 20 mg twice daily) (N = 248)</p> <p>Placebo (N = 124)</p>
Outcomes	<p>Primary efficacy outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain, WOMAC function, and PGA-VAS scores</p> <p>Secondary efficacy outcomes: American Pain Society, the Patient Outcome Questionnaire (APS-POQ), WOMAC pain (VAS 100 mm); Multi-Dimensional Health Assessment Questionnaire (MDHAQ), Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trial Response Criteria Initiative (OMERACT-OARSI) Responder Index</p> <p>Harms outcomes: adverse events (AEs), clinical laboratory tests and vital signs and physical examination findings at screening and final visit</p>
Funding	POZEN Inc
Declaration of interest	4/5 authors were employees of pharmaceutical companies POZEN Inc and AstraZeneca. Authors also reported stock ownership, research grants and consulting fees from pharmaceutical companies producing study drugs
Notes	Trial registration: ClinicalTrials.gov, http://clinicaltrials.gov/ , NCT00664560

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomization was conducted via the interactive web response system according to unique participant numbers and a randomisation list provided by the sponsor (p. 1244)
Allocation concealment (selection bias)	Low risk	Central randomization (p. 1244)
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Participants, investigators, study site staff, those performing study assessments, Pozen staff and analysts remained blinded to treatment identity throughout the study, except in the case of emergency (p. 1244)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Low risk	Investigators and assessors were blinded to treatment identity during the outcome assessment (p. 1244)

Celecoxib for osteoarthritis (Review)

Hochberg 2011 Study 307 (Continued)

Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Attrition rates were 16% (39/243) in the celecoxib group (16/39 due to adverse events) and 15% (19/124) in the placebo group (7/19 due to adverse events) as shown in Figure 1. All efficacy analyses were based on modified intention-to-treat (mITT) population (all randomized participants who took at least one dose of study medication and had data available from at least one post-baseline efficacy assessment). Missing data were imputed using LOCF analysis (p. 1246)
Selective reporting (reporting bias)	Unclear risk	Adverse events reported if their incidence was above 2% in any treatment group (Table 3)
Other bias	Unclear risk	Participants were allowed to take oral prednisone (≤ 7.5 mg/day), low-dose aspirin (≤ 325 mg/day), and antiplatelet agents (non-concomitant with low-dose aspirin); supplemental use of antacid (≤ 3 g/d, except within 48 hours before efficacy assessments) (p. 1245). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Hochberg 2011 Study 309

Methods	Study design: randomized, double-blind, placebo and active controls, 3 parallel groups, multicenter study Duration: 12 weeks Study dates: April to December 2008 Locations: 157 centers in the USA
Participants	Randomized N = 615 Completed N = 488 Mean age: 61.9 years (range: 50 to 89 years) Female: 64% Inclusion: age ≥ 50 years, osteoarthritis (OA) of the knee class I-III according to the American College of Rheumatology (ACR) criteria
Interventions	Celecoxib 200 mg (N = 247) Naproxen 1000 mg plus esomeprazole 40 mg magnesium tablets daily (500 mg + 20 mg twice daily) (N = 244) Placebo (N = 124)
Outcomes	Primary efficacy outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain, WOMAC function, and PGA-VAS scores. Secondary efficacy outcomes: American Pain Society, the Patient Outcome Questionnaire (APS-POQ), WOMAC pain (VAS 100 mm); Multi-Dimensional Health Assessment Questionnaire (MDHAQ), Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trial Response Criteria Initiative (OMERACT-OARSI) Responder Index Harms outcomes: adverse events (AEs), clinical laboratory tests and vital signs and physical examination findings at screening and final visit
Funding	POZEN Inc

Celecoxib for osteoarthritis (Review)

Hochberg 2011 Study 309 (Continued)

Declaration of interest 4/5 authors were employees of pharmaceutical companies POZEN Inc and AstraZeneca. Authors also reported stock ownership, research grants and consulting fees from pharmaceutical companies producing study drugs

Notes Trial registration: ClinicalTrials.gov, <http://clinicaltrials.gov/>, NCT00665431

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Randomization was conducted via the interactive web response system according to unique participant numbers and a randomisation list provided by the sponsor (p. 1244)
Allocation concealment (selection bias)	Low risk	Central randomization (p. 1244)
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Participants, investigators, study site staff, those performing study assessments, Pozen staff and analysts remained blinded to treatment identity throughout the study, except in the case of emergency (p. 1244)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Low risk	Investigators and assessors were blinded to treatment identity during the outcome assessment (p. 1244)
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	High attrition rates were reported, with 24% (59/247) in the celecoxib group (22/59 due to adverse events) and 21% (26/124) in the placebo group (6/26 due to adverse events) groups as shown in Figure 1. All efficacy analyses were based on modified intention-to-treat (mITT) population (all randomized participants who took at least one dose of study medication and had data available from at least one post-baseline efficacy assessment). Missing data were imputed using last observation carried forward (LOCF) analysis (p. 1246)
Selective reporting (reporting bias)	Low risk	All outcomes specified in the methods are also reported in the results section
Other bias	Unclear risk	Participants were allowed to take oral prednisone (≤ 7.5 mg/day), low-dose aspirin (≤ 325 mg/day), and antiplatelet agents (non-concomitant with low-dose aspirin); supplemental use of antacid (≤ 3 g/d, except within 48 hours before efficacy assessments) (p. 1245). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Kivitz 2001

Methods Study design: randomized, double-blind, placebo and active controls, 5 parallel groups, multicenter study

Duration: 12 weeks

Celecoxib for osteoarthritis (Review)

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Kivitz 2001 (Continued)

Study dates: not reported

Locations: 176 sites in the USA and Canada

Participants	Randomized N = 1061 Completed N = 538 Mean age: 62.6 years (range: 30 to 93 years) Female: 66% Inclusion: osteoarthritis (OA) of the hip, American College of Rheumatology (ACR) clinical and radiographic criteria, functional class I-III, symptomatic OA flare at the baseline visit Taking up to 325 mg/day aspirin for non-arthritic reasons was allowed
Interventions	Celecoxib 100 mg (N = 216) Celecoxib 200 mg (N = 207) Celecoxib 400 mg (N = 213) Naproxen 1000 mg (N = 207) Placebo (N = 218)
Outcomes	Primary efficacy outcomes: Patient's and Physician's Global Assessments, arthritis pain using visual analog scale (VAS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain, joint stiffness and physical function Secondary efficacy outcomes: incidence of withdrawal due to lack of arthritis efficacy and time to withdrawal due to lack of arthritis efficacy Harms outcomes: adverse events and changes from baseline in clinical laboratory tests, vital signs, and physical examinations
Funding	Pharmacia Corporation and Pfizer, Inc
Declaration of interest	Not reported According to the author affiliation, 5/7 authors were employees of Pharmacia Corporation

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Random assignment was stratified by site in blocks of 10 (p. 469). Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	This was a double-blind trial (p. 468)
Blinding of outcome assessment (detection bias):	Low risk	Participants blinded

Celecoxib for osteoarthritis (Review)

Kivitz 2001 (Continued)

Patient-reported out-comes

Blinding of outcome assessment (detection bias): Non-patient-reported out-comes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition rates were high with attrition of 64% (139/218) in the placebo group (129/139 due to treatment failure or adverse events), 46% (96/207) in the celecoxib 200 mg/day group (88/96 due to treatment failure or adverse events) and 42% (89/207) in the naproxen group (80/89 due to treatment failure or adverse events) as shown in Table 2. All efficacy analyses were performed on the intention-to-treat (ITT) cohort, which was defined as all participants who were randomly assigned and took one dose or more of study medication. Missing data values were assigned the participant's most recent observation for that outcome, including those for participants with treatment failure (p. 470)
Selective reporting (reporting bias)	Unclear risk	For arthritis pain outcome measured by visual analog scale (VAS) no statistical measure of dispersion provided. Only adverse events affecting 3% or more of any treatment group reported (Table 2)
Other bias	Unclear risk	Participants were allowed to take aspirin (up to 325 mg/d); and acetaminophen (up to 2 g/d, for up to 3 consecutive days, when absolutely necessary for non-arthritic conditions, the use of acetaminophen was prohibited within 48 hours of an assessment of arthritis efficacy) (p. 469). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Lehmann 2005

Methods	Study design: randomized, double blind, placebo and active controls, 4 parallel groups, multicenter study Duration: 13 weeks Study dates: not reported Locations: 116 centers in 15 countries
Participants	Randomized N = 1684 Completed N = 1488 Mean age: 62.4 ± 10.1 years Female: 70% Inclusion: age ≥18 years, osteoarthritis (OA) of the knee, meeting American College of Rheumatology (ACR) criteria and expected to require chronic non-steroidal anti-inflammatory drug (NSAID) or other analgesic therapy for ≥ 3 months
Interventions	Celecoxib 200 mg (N = 420) Lumiracoxib 100 mg (N = 420) Lumiracoxib 100 mg with an initial (loading) dose of 200 mg for the first two weeks of the study (N = 420)

Lehmann 2005 (Continued)

Placebo (N = 424)

Outcomes	<p>Primary efficacy outcomes: OA pain intensity in the target knee at week 13 (VAS mm), patient’s global assessment of disease activity at week 13 (VAS mm), total score of Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questionnaire</p> <p>Secondary efficacy outcomes: OA pain intensity in the target knee by visit using visual analog scale (VAS) mm, patient’s and physician’s global assessment of disease activity by visit (VAS mm), WOMAC total and sub scale scores, Short-Form Health Survey (SF-36), response to treatment using Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trial Response Criteria Initiative (OMERACT-OARSI) criteria by visit</p> <p>Harms outcomes: reporting of adverse events and serious adverse events, laboratory testing-liver function parameters; vital signs</p>	
Funding	Novartis Pharma AG	
Declaration of interest	Not reported. According to the author affiliation, 4/8 authors were employees of Novartis	
Notes		
Risk of bias		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Participants were randomized using a validated automated system (p. 519)
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	This was a double-blind, double-dummy study (p. 518)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Suspected GI, cardiovascular, cerebrovascular and hepatic events were assessed in a blinded fashion by independent safety committees according to prespecified criteria (p. 519, 520). For other outcomes, the process of outcome assessment is not described
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition was 12% (52/420) in the celecoxib group (33/52 due to adverse events or unsatisfactory therapeutic effect) and 15% (64/424) in the placebo group (11%/64 due to adverse events or unsatisfactory therapeutic effect) as shown in Figure 1 (p. 521). All efficacy evaluations were performed on the intention-to-treat (ITT) population which included all randomized participants who had been exposed to study medication. All safety evaluations were performed on the safety population which was identical to the ITT population. Last observation carried forward (LOCF) techniques were used to impute missing data (p. 520)
Selective reporting (reporting bias)	Low risk	All primary and secondary outcomes specified in the methods are included in the report of results, except for SF-36, for which authors provided general description of findings, without any numerical results (p. 522)

Celecoxib for osteoarthritis (Review)

Lehmann 2005 (Continued)

Other bias	Unclear risk	Participants were allowed paracetamol (≤ 2 g/day, as rescue medication throughout the study, including during the screening/washout period); low-dose aspirin (≤ 325 mg/d, for prophylaxis against a cardio- or cerebrovascular event in participants considered at increased risk); chondroitin sulfate and/or glucosamine sulfate (if established stable dosage and regimen), corticosteroids (topical, ophthalmic, nasal, or inhaled, at the usual labelled doses), H2-receptor antagonists, proton pump inhibitors, antacids and cytoprotective agents (taken at the usual labelled doses); and physiotherapy as prescribed by the physician (only when ongoing with a stable regimen) (p. p. 518). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results
------------	--------------	--

McKenna 2001a

Methods	Study design: randomized, double blind, placebo and active controls, 3 parallel groups, multicenter study Duration: 6 weeks Study dates: February and May 2000 Locations: 20 investigational sites in the USA
Participants	Randomized N = 182 Completed N = 142 Mean age: 62.2 ± 10.3 years Female: 71% Inclusion: age ≥ 40 years, osteoarthritis (OA) of the knee according to the American College of Rheumatology (ACR) criteria, functional class I-III, worsening signs and symptoms
Interventions	Celecoxib 200 mg (N = 63) Rofecoxib 25 mg (N = 59) Placebo (N = 60)
Outcomes	Primary efficacy outcomes: mean changes from baseline in arthritis pain using visual analog scale (VAS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), proportion of participants who improved in the patient's global assessment Secondary efficacy outcomes: mean changes from baseline in the VAS for pain on walking Harms outcomes: adverse events (AEs) and serious adverse events (SAEs)
Funding	Pharmacia and Pfizer Inc
Declaration of interest	Not reported. According to the author affiliation, 1/5 authors was an employee of Pharmacia Corporation
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
------	--------------------	-----------------------

Celecoxib for osteoarthritis (Review)

McKenna 2001a (Continued)

Random sequence generation (selection bias)	Low risk	Participants were randomly assigned to treatment groups, and separate computer-generated randomisation schedules were prepared for participants with different baseline characteristics
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	This was a double-blind, double-dummy study (p. 152)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Attrition was 22% (14/63) in celecoxib group (9/14 due to treatment failure or adverse events) and 27% (16/60) in the placebo group (13/16 due to treatment failure or adverse event) as shown in Figure 1 (p. 154). All efficacy and safety analyses were conducted using the intention-to-treat (ITT) cohort, which included any participant who received at least one dose of study medication (p. 153). ITT cohort was not defined and methods for handling missing values were not described
Selective reporting (reporting bias)	Unclear risk	All outcomes specified in the methods are also reported in results section. Arthritis pain on visual analog scale (VAS) reported without statistical measure of dispersion (Table 2). Adverse events reported if they occurred in 5% or more participants in one study group (Table 3)
Other bias	Unclear risk	Participants were allowed to take occasional acetaminophen (for non arthritic pain, use was prohibited during the 48 h before the follow-up visits at Weeks 3 and 6); low-dose aspirin (≤ 325 mg/d, for cardiovascular prophylaxis); occasional antacid (p. 153). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

McKenna 2001b

Methods	Study design: randomized, double-blind, placebo and active controls, 3 parallel groups, multicenter study Duration: 6 weeks Study dates: not reported Locations: 54 institutions across the USA
Participants	Randomized N = 600 Completed N = 450 Mean age: 61.7 years (range: 32 to 88 years)

Celecoxib for osteoarthritis (Review)

McKenna 2001b (Continued)

Female: 65%

Inclusion: osteoarthritis (OA) of the knee according to the American College of Rheumatology (ACR) criteria, symptomatic

Interventions	Celecoxib 200 mg (100 mg twice daily) (N = 199) Diclofenac 150 mg (50 mg three times a day) (N = 199) Placebo (N = 200)
Outcomes	Primary efficacy outcomes: patient's assessment of arthritis pain using a visual analogue scale (VAS) where 0 mm = no pain, 100 mm = very severe pain, Patient's Global Assessment of Arthritis and the Physician's Global Assessment of Arthritis (both using five grades from 'very good' to 'very poor'), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) Secondary efficacy outcomes: incidence of withdrawal due to lack of efficacy of the study medication, time to withdrawal due to lack of efficacy, and the OA Severity Index (range from 1 to 24 with the lower score as better); American Pain Society (APS) pain measure questionnaires Harms outcomes: spontaneously reported adverse events and adverse event related withdrawals. Safety assessments included recording of serious adverse events, evaluating changes in clinical laboratory tests pre- and post-treatment, and monitoring changes in vital signs
Funding	Pharmacia Corporation
Declaration of interest	Not reported. According to the author affiliation, five out of six authors were employees of Pharmacia Corporation

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	This was a double-blind study (p. 12)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	High attrition disparity reported: 36% (71/200) in placebo group, 21% (42/201) in celecoxib group and 19% (37/199) in diclofenac group (active treatments vs placebo: $P < 0.002$) (p. 13). It is indicated that the majority of these discontinuations were due to treatment failure or adverse events, but precise num-

Celecoxib for osteoarthritis (Review)

McKenna 2001b (Continued)

ber of patients withdrawn for various reasons is not described (p. 13). Analyses were performed on the intention-to-treat (ITT) population defined as all s that took at least one dose of study medication (p. 12). Methods for handling missing values were not described. Adverse events reported if their incidence was above 3% of participants in any treatment group (Table 4)

Selective reporting (reporting bias)	Low risk	All outcomes specified in the methods are also reported in results section
Other bias	Unclear risk	Participants were allowed to take occasional acetaminophen (for non arthritic pain, use was prohibited during the 48 h before the follow-up visits at Weeks 3 and 6); low-dose aspirin (≤ 325 mg/d, for cardiovascular prophylaxis); occasional antacid (p. 153). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Pincus 2004 PACES-a

Methods	<p>Study design: randomized, double blind, placebo and active controls, 3 parallel groups, crossover multicenter clinical study</p> <p>Study acronym: PACES (Patient Preference for Placebo, Acetaminophen or Celecoxib Efficacy Studies)</p> <p>Duration: 14 weeks</p> <p>Study dates: not reported</p> <p>Locations: Finland and different sites in the USA</p>
Participants	<p>Randomized N = 524</p> <p>Completed N = not reported</p> <p>Mean age: 63.6 years</p> <p>Female: 62%</p> <p>Inclusion: age ≥ 45 years, osteoarthritis (OA) of knee or hip, radiographic Kellgren-Lawrence grade 2–4, pain intensity on visual analog scale (VAS) 0–100: 40–90 mm</p>
Interventions	<p>Participants were randomized to receive a sequence of two of three treatments, each for 6 weeks. Only results of the first sequence were included in the analysis.</p> <p>Celecoxib 200 mg (N = 181)</p> <p>Acetaminophen 1000 mg (four times a day) (N = 171)</p> <p>Placebo (N = 172)</p>
Outcomes	<p>Primary efficacy outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain, stiffness and function, Multi-Dimensional Health Assessment Questionnaire (MDHAQ) using VAS 100 mm, Short-Form Health Survey (SF-36)</p> <p>Secondary efficacy outcomes: patient global assessment, also measured on a 100 mm visual analogue scale, SF-36 pain scores, MDHAQ activities of daily living scale scores, investigator assessment of patient global status, and investigator assessment of patient change in global status</p> <p>Harms: adverse events (AEs) and serious adverse events (SAEs)</p>
Funding	Pfizer Inc

Celecoxib for osteoarthritis (Review)

Pincus 2004 PACES-a (Continued)

Declaration of interest Not reported. According to the author affiliation, 4/12 authors were employees of Pfizer Inc

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	These were double-blind, double-dummy trials (p. 932)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Completion rates are reported only for the both period of this cross-over study, but no data are available on attrition in the first period, or reasons for attrition. The last observation carried forward (LOCF) procedure was used to manage missing data. Analyses of total Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) score, pain score, and paired patient preference were conducted on three groups: intention-to-treat (ITT) group (all participants who received a single dose of the study drug in period I); protocol adherent (omitting participants with major protocol violations such as noncompliance); all completers (participants with complete data for the relevant period) (p. 933)
Selective reporting (reporting bias)	High risk	Data were not shown for the analyses of the patient global scale, Multi-Dimensional Health Assessment Questionnaire (MDHAQ) activities of daily living scale, investigator assessment of patient global status, investigator assessment of patient change in global status, and Short-Form Health Survey (SF-36) pain scores disclosed patterns similar to those of the primary end points. It is indicated that these data are available on request (p. 935). This was crossover trial, but adverse events were reported as combined for both treatment periods (Table 5)
Other bias	Unclear risk	Participants were allowed to take propoxyphene (Darvon) 65 mg up to four times per day was given as a rescue analgesic drug; codeine 60 mg or tramadol (Ultram) 100 mg, up to four times per day, were provided as alternatives to fewer than 5% of participants if propoxyphene was poorly tolerated or ineffective. Participants were instructed not to take any rescue drug within 12 hours of any visit (p. 932). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Pincus 2004 PACES-b

Methods	<p>Study design: randomized, double blind, placebo and active controls, 3 parallel groups, crossover multicenter clinical study</p> <p>Study acronym: PACES (Patient Preference for Placebo, Acetaminophen or Celecoxib Efficacy Studies)</p> <p>Duration: 14 weeks</p> <p>Study dates: not reported</p> <p>Locations: Finland and different sites in USA</p>
Participants	<p>Randomized N = 524</p> <p>Completed N = not reported</p> <p>Mean age: 63.6 years</p> <p>Female: 62%</p> <p>Inclusion: age $\geq$ 45 years, osteoarthritis (OA) of knee or hip, radiographic Kellgren-Lawrence grade 2–4, pain intensity on visual analog scale (VAS) 0–100: 40–90 mm</p>
Interventions	<p>Participants were randomized to receive a sequence of 2 of 3 treatments, each for 6 weeks. Only results of the first sequence were included in the analysis.</p> <p>Celecoxib 200 mg (N = 189)</p> <p>Acetaminophen 1000 mg (4 times a day) (N = 185)</p> <p>Placebo (N = 182)</p>
Outcomes	<p>Primary efficacy outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain, stiffness and function, Multi-Dimensional Health Assessment Questionnaire (MDHAQ) using VAS 100 mm, Short-Form Health Survey (SF-36)</p> <p>Secondary efficacy outcomes: patient global assessment, also measured on a 100 mm visual analogue scale, SF-36 pain scores, MDHAQ activities of daily living scale scores, investigator assessment of patient global status, and investigator assessment of patient change in global status</p> <p>Harms: adverse events (AEs) and serious adverse events (SAEs)</p>
Funding	Pfizer Inc
Declaration of interest	Not reported. According to the author affiliation, four out of twelve authors were employees of Pfizer Inc
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described

Pincus 2004 PACES-b (Continued)

Blinding of participants and personnel (performance bias) All outcomes	Low risk	These were double-blind, double-dummy trials (p. 932)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Completion rates are reported only for the both period of this cross-over study, but no data are available on attrition in the first period, or reasons for attrition. The last observation carried forward (LOCF) procedure was used to manage missing data. Analyses of total Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) score, pain score, and paired patient preference were conducted on three groups: intention-to-treat (ITT) group (all participants who received a single dose of the study drug in period I); protocol adherent (omitting participants with major protocol violations such as noncompliance); all completers (participants with complete data for the relevant period) (p. 933)
Selective reporting (reporting bias)	High risk	Data were not shown for the analyses of the patient global scale, Multi-Dimensional Health Assessment Questionnaire (MDHAQ) activities of daily living scale, investigator assessment of patient global status, investigator assessment of patient change in global status, and Short-Form Health Survey (SF-36) pain scores disclosed patterns similar to those of the primary end points. It is indicated that these data are available on request (p. 935). This was crossover trial, but adverse events were reported as combined for both treatment periods (Table 5)
Other bias	Unclear risk	Participants were allowed to take propoxyphene (Darvon) 65 mg up to four times per day was given as a rescue analgesic drug; codeine 60 mg or tramadol (Ultram) 100 mg, up to four times per day, were provided as alternatives to fewer than 5% of participants if propoxyphene was poorly tolerated or ineffective. Participants were instructed not to take any rescue drug within 12 hours of any visit (p. 932). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Rother 2007

Methods	Study design: randomized, double-blind, placebo and active controls, 4 parallel groups, multicenter study Duration: 6 weeks Study dates: July 2003 and January 2004 Locations: 30 outpatient units in Germany
Participants	Randomized N = 397 Completed N = 324

Celecoxib for osteoarthritis (Review)

Rother 2007 (Continued)

Mean age: 62.8 ± 9.8 years

Women: 60%

Inclusion: osteoarthritis (OA) of the knee ≥ 6 months, meeting two of the following three clinical criteria: (1) morning stiffness of 30 minutes duration, crepitus on motion and age > 40 years; (2) rating their pain in the index knee as > 3 on a five-point Likert scale; and (3) taking oral non-steroidal anti-inflammatory drugs (NSAIDs) at least 3 days per week for the past 3 months or for > 25 of the past 30 days. Participants had to meet three osteoarthritis flare criteria: (1) pain in the index knee on walking > 40 mm on a visual analog scale (VAS); (2) increased by > 15 mm compared with pain on previous treatment (screening); and patient global assessment (PGA) score for osteoarthritis of 3 to 5 and at least one grade increase from screening

Interventions	<p>Celecoxib 200 mg (100 mg twice daily) (N = 132)</p> <p>Epicutaneous ketoprofen 110 mg in 4.8 g Transfersome (a semi-solid formulation, IDEA-033) (N = 138)</p> <p>Both placebo formulations (N = 127)</p>
Outcomes	<p>Primary efficacy outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain and physical function, PGA response</p> <p>Secondary efficacy outcomes: changes from baseline to end of study in PGA of osteoarthritis (5-point classification), WOMAC stiffness sub scale, use of rescue medication, discontinuation due to lack of efficacy, and primary outcome measures after 2 and 4 weeks of treatment</p> <p>Harms outcomes: physical examination, standard haematology and biochemistry tests, adverse events (AEs), plasma concentrations of ketoprofen, scoring of erythema and oedema of the skin area exposed to Transfersome</p>
Funding	IDEA AG and McNeill Consumer & Specialty Pharmaceuticals
Declaration of interest	Not reported. 2/5 authors were employees of the pharmaceutical companies producing study drugs
Notes	Trial registration: ClinicalTrials.gov, http://clinicaltrials.gov/ , NCT00317733

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	A computer-generated centralized randomisation list was used (p. 1179)
Allocation concealment (selection bias)	Low risk	Randomisation was performed by a centralized telephone procedure (p. 1179)
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Treatment assignments were double-blind
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described

Celecoxib for osteoarthritis (Review)

Rother 2007 (Continued)

Incomplete outcome data (attrition bias) All outcomes	Low risk	Total attrition and attrition was 17% (23/132) in the celecoxib group (21/23 due to adverse events and lack of efficacy) and 19% (25/127) in the placebo group (23/25 due to adverse events and lack of efficacy) as shown in Figure 1 (p. 1180). The primary efficacy analysis was performed on all randomized participants who used at least one dose of study medication in intention to treat (ITT) population. All efficacy outcome measures were also analysed in the per protocol (PP) population, which included participants who did not have significant protocol deviations (p. 1179)
Selective reporting (reporting bias)	Low risk	All outcomes specified in the methods are also reported in results section
Other bias	Unclear risk	Participants were allowed to take paracetamol (up to 2 g/d, as rescue medication for knee pain for 3 days in any week, apart from the 48 hours preceding a study visit) (p. 1179). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Schnitzer 2011

Methods	<p>Study design: randomized, double-blind, double-dummy, placebo and positive-interval (celecoxib)-controlled, 3 parallel groups, multicenter study</p> <p>Duration: 13 weeks</p> <p>Study dates: not reported</p> <p>Locations: 162 centers in 5 countries; USA (7 centers), Canada (42 centers), Germany (30 centers), Italy (12 centers), and UK (7 centers)</p>
Participants	<p>Randomized N = 1262</p> <p>Completed N = 951</p> <p>Mean age: Lumiracoxib 100 mg once daily group 61.7 years (range 40 to 90 years); celecoxib 200 mg once daily group 61.7 years (range 30 to 86 years); and placebo group 61.4 years (range 41 to 90 years)</p> <p>Female: 61.7 %</p> <p>Inclusion: symptomatic primary osteoarthritis (OA) of the hip according to the American College of Rheumatology (ACR) criteria, who had experienced symptoms for at least 3 months prior screening</p>
Interventions	<p>Lumiracoxib 100 mg (N = 427)</p> <p>Celecoxib 200 mg (N = 419)</p> <p>Placebo (N = 416)</p>
Outcomes	<p>Efficacy outcomes: the patient's global assessment of disease activity measured on a 100-mm visual analog scale (VAS); the patient's functional status during the previous 48 h using the Western Ontario and McMaster Universities Osteoarthritis Index Likert version 3.1 (WOMAC LK 3.1) sub scale scores (pain, difficulty in performing daily activities (DPDA) as a measure of function, stiffness), and total score; the patient's overall OA pain intensity in the target hip (measured on a 100-mm VAS); the physician's global assessment of disease activity (measured on a 100-mm VAS); actual OA pain intensity in the target joint in the evening during screening and at 12 h post dose for 1 week immediately following the baseline (day 0), visit 3 (day 28), and visit 4 (day 56) using a 0-100-mm VAS; and rescue medication use.</p>

Schnitzer 2011 (Continued)

Harms outcomes: recording adverse events (AEs) and serious adverse events (SAEs); physical examinations, laboratory tests (hematology, blood chemistry, and urinalysis), and vital sign assessments; pre-specified cardiovascular (CV)/cerebrovascular, liver, and GI safety events were adjudicated, blinded to treatment, by the appropriate independent CV, liver or gastrointestinal (GI) safety committee

Funding	"Novartis designed the study in collaboration with the independent authors. Both Novartis and the independent authors were involved in the collection, analysis, and interpretation of the data, as well as in the writing of the report and in the decision to submit the paper for publication. Novartis was the sponsor funding the role of the medical editor/writer" (p. 1442)
Declaration of interest	According to the author affiliations, four out of 9 authors were employees of Novartis. For other authors the following was reported in the conflict of interest section: receiving research grants or honoraria and having ownership in relevant pharmaceutical companies (p. 1442)
Notes	Trial registration: ClinicalTrials.gov, http://clinicaltrials.gov/ , NCT00154219

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Patients were randomized according to a randomization list, which was computer generated by Novartis Drug Supply Management (Basel, Switzerland) using a validated system (Almedica Drug Label System) (p. 1434)
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Patients, investigator staff, and data analysts remained blind to the identity of the treatment from the time of randomization until database lock. Blinding was maintained by the use of study drugs that were identical in packaging, labelling, schedule of administration, appearance, and odor (p. 1435)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Low risk	Investigator staff and data analysts were blinded (p. 1435)
Incomplete outcome data (attrition bias) All outcomes	High risk	High attrition was reported; 92/419 (22%) treated with celecoxib 200 mg (unsatisfactory therapeutic effect: 32 (7.6%), adverse events (AEs): 22 (5.3%), protocol violation(s): 12 (2.9%), withdrew consent: 14 (3.3%), administrative problems: 6 (1.4%), lost to follow-up: 3 (0.7%), abnormal lab value(s) or test results: 1 (0.2%), deaths: 2 (0.5%)) and 129/416 (31%) receiving placebo (unsatisfactory therapeutic effect: 80 (19.2%), AEs: 18 (4.3%), protocol violation(s): 11 (2.6%), withdrew consent: 11 (2.6%), administrative problems: 2 (0.5%), lost to follow-up: 6 (1.4%), abnormal lab value(s) or test results: 1 (0.2%), deaths: 0 (0.0%)) (p. 1437)
Selective reporting (reporting bias)	Low risk	All primary outcomes were included in report of results. (pp. 1436, 1438-9). Secondary outcomes data shown. (pp. 1439-40). Adverse events were defined and reported. (pp. 1440-1).
Other bias	Unclear risk	The use of rescue medication acetaminophen (paracetamol) was permitted up to a maximum of 3 g/day during the study. Patients who took > 3 g of acetaminophen daily for four or more continuous days during the study were to be

Celecoxib for osteoarthritis (Review)

Schnitzer 2011 (Continued)

discontinued from the trial due to “unsatisfactory therapeutic effect.” No analgesic rescue medications were to be taken in 24 h before a study visit for evaluation (p. 1435). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Sheldon 2005

Methods	<p>Study design: randomized, double-blind, placebo and active controls, 4 parallel groups, multicenter study</p> <p>Duration: 13 weeks</p> <p>Study dates: not reported</p> <p>Locations: 38 centers in Canada and 137 centers in the USA</p>
Participants	<p>Randomized N = 1551</p> <p>Completed N = 1182</p> <p>Mean age: 60.5 ± 10.8 years</p> <p>Women: 62%</p> <p>Inclusion: age ≥18 years, osteoarthritis (OA) of the knee according to the American College of Rheumatology (ACR) criteria with symptoms requiring NSAID treatment for ≥ 3 months before study entry</p>
Interventions	<p>Celecoxib 200 mg (N = 393)</p> <p>Lumiracoxib 100 mg (4 times a day) (N = 391)</p> <p>Lumiracoxib 100 mg 4 times a day, with a loading dose of 200 mg 4 times daily for the first 2 weeks of the study (N = 385)</p> <p>Placebo (N = 382)</p>
Outcomes	<p>Primary efficacy outcomes: OA pain intensity in the target knee (VAS 100 mm), the patient's global assessment of disease activity using visual analog scale (VAS) 100 mm, Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) total score</p> <p>Secondary efficacy outcomes: OA pain intensity in the target knee, the patient's and physician's global assessments of disease activity, WOMAC pain and physical function, use of rescue medication</p> <p>Harms outcomes: adverse events (AEs) and serious adverse events (SEAs), physical examination, vital signs at each clinical visit, standard laboratory test</p>
Funding	Not reported
Declaration of interest	Not reported. According to the author affiliations, 3/6 authors were employees of Novartis Corporation and lumiracoxib by Novartis was used in the study
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described

Celecoxib for osteoarthritis (Review)

Sheldon 2005 (Continued)

Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding was maintained through use of the double-dummy technique (p. 66)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Unblinded safety data were reviewed on a quarterly basis by an independent Data Safety Monitoring Board. In addition, independent safety committees reviewed individual cases of suspected events in a blinded fashion (p 67). For other outcomes, the process of outcome assessment not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition was high with 20% (80/393) of participants lost in the celecoxib group (58/80 due to adverse events and unsatisfactory therapeutic effect) and 34% (132/382) in the placebo group (94/132 due to adverse events and unsatisfactory therapeutic effect) as shown in Figure 1 (p. 68). The intention-to-treat (ITT) population, defined as all randomized participants who took at least one dose of study medication, was used in all efficacy analyses and was identical to the safety population, which was used in all safety analyses. Last observation carried forward (LOCF) techniques were used to account for missing data (p. 67)
Selective reporting (reporting bias)	Unclear risk	All outcomes specified in the methods are also reported in results section. Adverse events reported if their incidence was above 3% in any treatment group (Table 5)
Other bias	Unclear risk	Participants were allowed to take low-dose aspirin (≤ 325 mg/d, for CV prophylaxis in participants at increased cardiovascular risk); acetaminophen (500 mg tablets, as rescue medication throughout the study, including during the screening period); chondroitin sulphate and/or glucosamine sulphate (if the dose and regimen were established and stable, including during the screening period), corticosteroids (topical, ophthalmic, nasal, or inhaled, at the usual labelled doses), histamine-2-receptor antagonists, proton pump inhibitors, antacids and cytoprotective agents (taken at the usual labelled doses), and physiotherapy as prescribed by the physician (only if an ongoing, stable regimen) (p. 66). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Smugar 2006 Study 1

Methods	Study design: randomized, double-blind, placebo and active controls, 4 parallel groups, multicenter study
	Duration: 6 weeks
	Study dates: October 1999 and September 2000
	Locations: USA
Participants	Randomized N = 1521
	Completed N = 1248

Smugar 2006 Study 1 (Continued)

Mean age: 61.5 years (range: 39 to 92 years)

Women: 68%

Inclusion: age $\geq$ 40 years, osteoarthritis (OA) of the knee or hip $\geq$ 6 months, class I-III according to the American College of Rheumatology (ACR) criteria, visual analog scale (VAS) 0-100 mm pain on walking on flat surface score on Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) at least 40 mm, VAS 0-100 mm patient assessed pain at night at least 40 mm, an increase of 15 mm from baseline (on therapy) on patient assessed pain on walking on flat surface VAS, and worsening in Investigator Global Assessment of Disease Status (IGADS) of at least one point on five-point Likert scale compared with the screening visit.

Interventions	<p>Celecoxib 200 mg (N = 456)</p> <p>Rofecoxib 12.5 mg (N = 456)</p> <p>Rofecoxib 25 mg (N = 459)</p> <p>Placebo (N = 150)</p>
Outcomes	<p>Primary efficacy outcomes: WOMAC (VAS 100 mm), Patient Global Assessment of Response to Therapy (PGART) (Likert 0-4), Investigator's global assessment of response to therapy (IGART) (Likert 0-4), Patient Global Assessment of Disease status (PGADS) (VAS 100 mm)</p> <p>Harms outcomes: adverse events (AEs) and serious adverse events (SAEs), physical examination, vital signs (blood pressure and pulse), blood and urine tests</p>
Funding	Merck & Co. Inc
Declaration of interest	3/6 authors were employees of Merck & Co., Inc. The other authors reported financial ties with pharmaceutical companies, including consultant fees, clinical research support, equity interest, medical advisory bureaus, speaker bureaus, board of directors
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind study (p. 1355)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Independent, blinded adjudication committees evaluated cardiovascular and gastrointestinal adverse events (p. 1356). The outcome assessment process was not described for other outcomes

Celecoxib for osteoarthritis (Review)

Smugar 2006 Study 1 (Continued)

Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition was 18% (80/456) in celecoxib group (68/80 due to lack of efficacy and adverse events) and 38% (57/150) in placebo group (48/57 due to lack of efficacy and adverse events) as shown in Figure 1A (p. 1357). A modified intention-to-treat (mITT) analyses method was employed for the analyses of efficacy and safety data. For efficacy analyses, all participants with a baseline value and at least one on-treatment value were included, with the last observation carried forward for those with missing on-treatment values (pp. 1356-7)
Selective reporting (reporting bias)	High risk	Data are missing for the outcome of Investigator's global assessment of response to therapy (IGART)
Other bias	Unclear risk	Participants were allowed to take after the first 7 days acetaminophen (max. 2600 mg/d, as rescue therapy for osteoarthritis pain, which was documented and discontinued at least 24 h before each visit); low-dose aspirin (81 mg or less daily, for cardioprotective effects); glucosamine and chondroitin sulfate (if taken for longer than 6 months, at the same stable dose for the duration of study) (p. 1355). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Smugar 2006 Study 2

Methods	<p>Study design: randomized, double-blind, placebo and active controls, 3 parallel groups, multicenter study</p> <p>Duration: 6 weeks</p> <p>Study dates: October 1999 and September 2000</p> <p>Locations: USA</p>
Participants	<p>Randomized N = 1082</p> <p>Completed N = 897</p> <p>Mean age: 62.1 years (range: 40 to 91 years)</p> <p>Women: 66%</p> <p>Inclusion: age $\geq$ 40 years, osteoarthritis (OA) of the knee or hip $\geq$ 6 months, class I-III according to the American College of Rheumatology (ACR) criteria, visual analog scale (VAS) 0-100 mm pain on walking on flat surface score on Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) at least 40 mm, VAS 0-100 mm patient assessed pain at night at least 40 mm, an increase of 15 mm from baseline (on therapy) on patient assessed pain on walking on flat surface VAS, and worsening in Investigator Global Assessment of Disease Status (IGADS) of at least one point on five-point Likert scale compared with the screening visit</p>
Interventions	<p>Celecoxib 200 mg (N = 460)</p> <p>Rofecoxib 25 mg (N = 471)</p> <p>Placebo (N = 151)</p>
Outcomes	<p>Primary efficacy outcomes: WOMAC (VAS 100 mm), Patient Global Assessment of Response to Therapy (PGART) (Likert 0-4), Investigator's global assessment of response to therapy (IGART) (Likert 0-4), Patient Global Assessment of Disease status (PGADS) (VAS 100 mm)</p> <p>Harms outcomes: adverse events (AEs) and serious adverse events (SAEs), physical examination, vital signs (blood pressure and pulse), blood and urine tests</p>

Smugar 2006 Study 2 (Continued)

Funding	Merck & Co. Inc	
Declaration of interest	3/6 authors were employees of Merck & Co., Inc. The other authors reported financial ties with pharmaceutical companies, including consultant fees, clinical research support, equity interest, medical advisory bureaus, speaker bureaus, board of directors	
Notes		
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind study (p. 1355)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Independent, blinded adjudication committees evaluated cardiovascular and gastrointestinal adverse events (p. 1356). For other outcomes, the process of outcome assessment not
Incomplete outcome data (attrition bias) All outcomes	High risk	Attrition was 15% (in celecoxib group (56/68 due to lack of efficacy or adverse events) and 38% (58/151) in placebo group (50/58 due to lack of efficacy and adverse events) as shown in Figure 1B (p. 1357). A modified intention-to-treat (mITT) analyses method was employed for the analyses of efficacy and safety data. For efficacy analyses, all participants with a baseline value and at least one on-treatment value were included, with the last observation carried forward for those with missing on-treatment values (pp. 1356-7)
Selective reporting (reporting bias)	High risk	Data are missing for the outcome of investigator’s global assessment of response to therapy (IGART)
Other bias	Unclear risk	Participants were allowed to take after the first 7 days acetaminophen (max. 2600 mg/d, as rescue therapy for osteoarthritis pain, which was documented and discontinued at least 24 h before each visit); low-dose aspirin (81 mg or less daily, for cardioprotective effects); glucosamine and chondroitin sulfate (if taken for longer than 6 months, at the same stable dose for the duration of study) (p. 1355). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Sowers 2005

Methods	Study design: randomized, double-blind, active controls, 3 parallel groups, multicenter study
---------	---

Sowers 2005 (Continued)

Duration: 12 weeks

Study dates: May 2001 and April 2002

Locations: 65 centers in 7 countries participated in this trial

Participants	Randomized N = 404 Completed N = 323 Mean age: 63 years Women: 60% Inclusion: osteoarthritis (OA) of the knee or hip according to the American College of Rheumatology (ACR) criteria Ongoing treatment with 81 mg to 325 mg of aspirin daily (if used for at least 30 days before randomization) was permitted if the treatment could be maintained at a stable dose throughout the study
Interventions	Celecoxib 200 mg (N = 136) Rofecoxib 25 mg (N = 138) Naproxen 1000 mg (500 mg twice daily) (N = 130)
Outcomes	Efficacy outcomes: patient's assessment of arthritis pain (VAS 0-100 mm), patient's global assessment of arthritis; Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) scale, WOMAC pain, WOMAC physical function Harms outcomes: AEs
Funding	Pharmacia Corporation and Pfizer Inc
Declaration of interest	2/9 authors were Pfizer Inc. employees
Notes	

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	There was a single, computer-generated randomisation code, and each site was given block sizes of 3 (p. 162)
Allocation concealment (selection bias)	Low risk	Single (central) randomisation code
Blinding of participants and personnel (performance bias) All outcomes	Low risk	This was a double-blind study (p. 162)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias):	Unclear risk	Not described

Celecoxib for osteoarthritis (Review)

Sowers 2005 (Continued)

Non-patient-reported outcomes

Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Attrition was 16% (22/136) in the celecoxib group (7/136 due to adverse events) and 21% (27/128) in the naproxen group (11/128 due to adverse events) (p. 163). Detailed reasons for attrition were not reported. For osteoarthritis assessment modified intention-to-treat (mITT) analysis was performed (all participants who took at least 1 dose of study medication and had a baseline efficacy assessment) (p. 164)
Selective reporting (reporting bias)	High risk	Adverse events not reported
Other bias	Unclear risk	Participants were allowed ongoing treatment with estrogen, progesterone, and/or testosterone (if used for at least 2 months before randomization); aspirin (81-325 mg/d, if used for at least 30 days before randomization) (p. 162). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Tannenbaum 2004

Methods	<p>Study design: randomized, double blind, placebo and active controls, 4 parallel groups, multicenter study</p> <p>Study dates: not reported</p> <p>Duration: 13 weeks</p> <p>Locations: Canada, France, Belgium, Germany, UK, Hungary, Denmark, Switzerland, Italy, Spain</p>
Participants	<p>Randomized N = 1702</p> <p>Completed N = 1423</p> <p>Mean age: 64.3 ± 10.4 years</p> <p>Female: 68%</p> <p>Inclusion: age ≥18 years, osteoarthritis (OA) of the knee according to the American College of Rheumatology (ACR) criteria, pain intensity on visual analog scale (VAS) 0-100 mm in the affected knee > 40 mm in the past 24 h</p>
Interventions	<p>Celecoxib 200 mg (N = 481)</p> <p>Lumiracoxib 200 mg (N = 487)</p> <p>Lumiracoxib 400 mg (N = 491)</p> <p>Placebo (N = 243)</p>
Outcomes	<p>Primary efficacy outcomes: OA pain intensity in the target knee (VAS 100 mm), patient's global assessment of disease activity (VAS 100 mm), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain and total score</p> <p>Secondary efficacy outcomes: OA pain intensity in the target knee, patient's and physician's global assessment of disease activity, WOMAC pain and physical function</p> <p>Harms outcomes: adverse events (AEs) and serious adverse events (SEAs), physical examinations, vital signs, and standard laboratory tests, electrocardiogram (ECG) measurements</p>

Celecoxib for osteoarthritis (Review)

Tannenbaum 2004 (Continued)

Funding	Novartis Pharma	
Declaration of interest	3/13 authors were employees of Novartis Pharma. Other potential conflicts not reported	
Notes		
<i>Risk of bias</i>		
Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Blinding was maintained by a double-dummy technique (p. 1420)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	Low risk	Attrition was 17% (80/481) in the celecoxib group (62/80 due to unsatisfactory therapeutic effect and adverse events) and 18% (43/243) in the placebo group (36/43 due to unsatisfactory therapeutic effect and adverse events) as shown in Figure 1 (p. 1420). All safety and efficacy evaluations were performed using the intention-to-treat (ITT) population. The safety and ITT populations were identical and included all participants randomized to treatment who had been exposed to the study drug. Conventional last observation carried forward (LOCF) methodology was used when data were missing (p. 1422)
Selective reporting (reporting bias)	High risk	All outcomes specified in the methods are also reported in results section. Adverse events reported if their incidence was above 3% in any treatment group (Table 5)
Other bias	Unclear risk	Participants were allowed to take paracetamol (≤ 2 g/d, supplied by the investigator, as a rescue drug throughout the trial; however, they were asked to refrain from using the rescue drug from midnight before each clinic visit); low dose aspirin (≤ 325 mg/d, for cardiovascular indication) (p. 1421). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Williams 2000

Methods	Study design: randomized, double-blind, placebo controls, 3 parallel groups, multicenter study
	Duration: 6 weeks

Celecoxib for osteoarthritis (Review)

Williams 2000 (Continued)

Study dates: not reported

Locations: 50 clinical sites in the USA

Participants	Randomized N = 686 Completed N = 522 Mean age: 62.8 ± 10.9 years Women: 67% Inclusion: osteoarthritis (OA) of the knee, class I-III according to the American College of Rheumatology (ACR) criteria, in a flare state Participants taking a stable dose of aspirin ≤ 325 mg/day for non-arthritis conditions were eligible
Interventions	Celecoxib 200 mg (100 mg twice daily) (N = 231) Celecoxib 200 mg (once daily) (N = 223) Placebo (N = 232)
Outcomes	Primary efficacy outcomes: the Patient's and Physician's Global Assessments of Arthritis; Lequesne Osteoarthritis Severity Index Secondary efficacy outcomes: Patient's Assessment of Arthritis Pain using visual analog scale (VAS) 100 mm; Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain, stiffness and physical function Harms outcomes: adverse events (AEs) and serious adverse events (SAEs), physical examination, laboratory tests
Funding	GD Searle & Co. and Pfizer Inc
Declaration of interest	Not reported. According to the author affiliations, 5/8 authors were employees of GD Searle & Co

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind study (p. 66)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded

Celecoxib for osteoarthritis (Review)

Williams 2000 (Continued)

Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Total attrition was 37% (86/232) in placebo group and 16% (37/231) and 18% (41/223) in the two celecoxib groups. Attrition due to treatment failure was 24% (56/86) in placebo group, while 8% (18/37) and 9% (21/41) in celecoxib groups. Attrition due to adverse events was 9% (20/86) in placebo group, while 5% (11/37) and 4% (9/41) in celecoxib groups (Figure 1, p. 69). All randomized participants were included in the baseline analyses. All efficacy analyses were based on the ITT cohort, which included all participants who were enrolled and took at least one dose of the study medication. All missing observations were extrapolated by carrying forward the last observation (p. 67)
Selective reporting (reporting bias)	High risk	All outcomes specified in the methods are also reported in results section. Adverse events reported if their incidence was above 3% in any treatment group (Table 4)
Other bias	Unclear risk	Participants were allowed to take aspirin (≤ 325 mg/d, for conditions other than arthritis, for at least 30 days before the first dose of study medication, at the same dose regimen); acetaminophen (up to 2 g/d, for reasons other than relief of arthritis symptoms and for no more than 3 consecutive days, it must have been avoided within 48 hours before arthritis assessments performed at any visit) (p. 67). Amount of co-interventions consumed in each groups was not reported and it is unclear whether this could have affected study results

Williams 2001

Methods	Study design: randomized, double-blind, placebo controls, 3 parallel groups, multicenter study Duration: 6 weeks Study dates: not reported Locations: 98 clinical sites in the USA
Participants	Randomized N = 718 Completed N = 549 Mean age: 61.5 ± 8.5 years Women: 70% Inclusion: osteoarthritis (OA) of the knee, rheumatology and radiographic criteria, class I-III according to the American College of Rheumatology (ACR) criteria, in a flare state Participants taking a stable dose of aspirin ≤ 325 mg/day for non-arthritis conditions were eligible
Interventions	Celecoxib 200 mg (100 mg twice daily) (N = 243) Celecoxib 200 mg (once daily) (N = 231) Placebo (N = 244)
Outcomes	Primary efficacy outcomes: the Patient's and Physician's Global Assessments of Arthritis, the Lequesne Osteoarthritis Severity Index

Celecoxib for osteoarthritis (Review)

Williams 2001 (Continued)

Secondary efficacy outcomes: Patient's Assessment of Arthritis Pain-Visual Analog Scale using visual analog scale (VAS) 0-100 mm, Western Ontario and McMaster Universities Osteoarthritis Index (WOM-AC) pain and physical function, Patient's and Physician's Global Assessment of Arthritis, Lequesne Osteoarthritis Index

Harms outcomes: adverse events (AEs) and serious adverse events (SAEs), physical examination, laboratory tests

Funding	Pharmacia Corporation and Pfizer Inc
Declaration of interest	Not reported. According to the author affiliations, 4/5 authors were employees of Pharmacia Corporation

Notes

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of random sequence generation not described
Allocation concealment (selection bias)	Unclear risk	Not described
Blinding of participants and personnel (performance bias) All outcomes	Low risk	Double-blind study (p. 215)
Blinding of outcome assessment (detection bias): Patient-reported outcomes	Low risk	Participants blinded
Blinding of outcome assessment (detection bias): Non-patient-reported outcomes	Unclear risk	Not described
Incomplete outcome data (attrition bias) All outcomes	High risk	Total attrition was 33% (80/244) in placebo group, while 20% (49/243) and 17% (40/231) in celecoxib groups (pp. 217-8). Reasons for attrition not provided. All randomized participants were included in the baseline analyses. All efficacy analyses were based on intention-to-treat (ITT) cohort, which included all participants who were enrolled and took at least one dose of study medication. All missing observations were extrapolated using the last observation carried forward (LOCF) approach. Every randomized participant receiving at least one dose of study drug was included in the safety assessment (p. 217)
Selective reporting (reporting bias)	Unclear risk	All outcomes specified in the methods are also reported in results section. Adverse events reported if their incidence was above 3% in any treatment group (Table 4)
Other bias	Unclear risk	Participants were allowed to take aspirin (≤ 325 mg/d, for reasons other than arthritis, for ≥ 30 days before the first dose of study medication, were permitted to continue with the same dosing regimen); acetaminophen (up to 2 g/d, for 3 consecutive days, for reasons other than relief of arthritis symptoms, it must not have been taken within 48 hours before OA assessments were performed at any visit) (p. 216). Amount of co-interventions consumed in each

Celecoxib for osteoarthritis (Review)

Williams 2001 (Continued)

groups was not reported and it is unclear whether this could have affected study results

Abbreviations, in alphabetical order: ACR - American College of Rheumatology, AE - adverse event, ARA - American Rheumatology Association, ARCI - Addiction Research Center Inventory, APS - American Pain Society, APS-POQ - American Pain Society, the Patient Outcome Questionnaire, BOCF - baseline observation carried forward, CHF - congestive heart failure, CV - cardiovascular, GI - gastrointestinal, HAQ - Health Assessment Questionnaire, IGADS - Investigator Global Assessment of Disease Status, IGART - Investigator's global assessment of response to therapy, ITT - intention-to-treat analysis, LOCF - last observation carried forward, MDHAQ - Multi-Dimensional Health Assessment Questionnaire, mITT - modified intention-to-treat analysis, NSAID = non-steroidal anti-inflammatory drug, OA - osteoarthritis, OMERCT-OARSI - Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trial Response Criteria Initiative, PaGAA - Patient's Global Assessment of Arthritis, PDQ - Physical Dependence Questionnaire, PGA-VAS - Patient Global Assessment of osteoarthritis using a visual analog scale, PGAC - Patient Global Assessment of Change, PGADS - Patient Global Assessment of Disease status, PGART - Patient Global Assessment of Response to Therapy, PhGAA - Physician's Global Assessment of Arthritis, PHQ-9 - Patient Health Questionnaire-9, SAE - serious adverse event, SF-36 - Short Form Health Survey, SODA - Severity of Dyspepsia Assessment, TWA - time-weighted average, UGI - upper gastrointestinal, VAS - visual analog scale, WOMAC - Western Ontario and McMaster Universities Osteoarthritis Index

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Adegbehingbe 2008	Included participants with post-traumatic osteoarthritis
Andrews 1999	Review article; not a randomized controlled trial
Angiolillo 2014	Safety of chronic nonsteroidal anti-inflammatory drugs and concomitant low-dose aspirin; pooling results from 5 different studies, not a randomized controlled trial
Ashcroft 2001	Systematic review of published and unpublished randomized controlled trials on celecoxib, not a randomized controlled trial
Battisti 2004	Acetaminophen was the comparator; no comparators of interest
Bensen 2000	Review article; not a randomized controlled trial
Bensen 2000b	Analysis of 5 trials; three unpublished, one included in this review (Bensen 1999), one excluded from this review (Simon 1998); not a randomized controlled trial
Bianchi 2003	Fewer than 50 participants in each arm
Bianchi 2007	Duration less than four weeks and fewer than 50 participants in each arm
Bingham 2008	Secondary pooled analysis of trials; not a randomized controlled trial
Bingham 2009	Secondary pooled analysis of trials; not a randomized controlled trial
Bingham 2011	Secondary pooled analysis of trials; not a randomized controlled trial
Brereton 2014	Not a randomized controlled trial
Castgillo 2015	Comparing efficacy and safety of chondroitin sulfate plus glucosamine hydrochloride versus celecoxib; no comparators of interest
Castillo 2014	Comparison of a combination of glucosamine hydrochloride and chondroitin sulfate versus celecoxib; no comparators of interest

Study	Reason for exclusion
Chan 2015	Combination of celecoxib or naproxen combined with proton pump inhibitors; no intervention/comparator of interest
Detrembleur 2005	Duration less than four weeks and fewer than 50 participants in each arm
Dougados 2007	Pooled analysis of two studies (Lehmann 2005 ; Sheldon 2005) that are included in this review; not a randomized controlled trial
Essex 2014b	Comparison of continuous versus intermittent celecoxib only; no comparators of interest
Gallelli 2013	Fewer than 50 participants in each arm; duration less than four weeks (NCT01860833)
Hawel 2003	Duration less than four weeks
Henrotin 2015	Comparison of a combination chondroitin sulfate and glucosamine versus celecoxib; no comparators of interest
Hirayama 2014	Not a randomized controlled trial
Hochberg 2014a	Comparison of combined glucosamine hydrochloride and chondroitin sulfate versus celecoxib; no comparators of interest
Hochberg 2014b	Comparison of combined glucosamine hydrochloride and chondroitin sulfate versus celecoxib; no comparators of interest
Johnson 2003	Practice recommendation based on a systematic review by Deeks 2002 ; not a randomized controlled trial
Leeb 2004	Fewer than 50 participants in each arm
Luyten 2007	Comparison between continuous and intermittent treatment with celecoxib; results for placebo group are not shown
Moore 2009	Pooled data from 7 placebo-controlled, 12-week studies utilizing etoricoxib 30 or 60 mg, celecoxib 200 mg, naproxen 1000 mg, or ibuprofen 2400 mg; not a randomized controlled trial
NCT02192190	Randomized controlled trial that included osteoarthritis patients and compared celecoxib 200 mg, LY2951742, placebo subcutaneous injections and oral placebo. On the ClinicalTrials.gov (NCT02192190) it is indicated that the trial was terminated because interim assessment showed lack of efficacy. Full study results are not available. Two conference abstract with study results have been published (Jin 2016 , Smith 2016 ; available as secondary references for this study in the list of references). They indicate that "LY was generally well tolerated but failed to demonstrate adequate efficacy for the treatment of signs and symptoms of OA in patients with moderate to severe knee pain due to OA." Jin 2016 abstract indicates that at the time of the interim analysis there were only 36 participants in the celecoxib group. Sponsor: Eli Lilly and Company. Therefore, the study was excluded because there were less than 50 participants per study arm
Nissen 2016	Patients with osteoarthritis and rheumatoid arthritis were not analyzed separately
Ogunlade 2005	Not a randomized controlled trial
Ozgocmen 2005	Fewer than 50 participants in each arm
Pelletier 2016	Chondroitin sulfate vs. celecoxib; no comparators of interest
Rozenberg 2008	Not not a randomized controlled trial

Study	Reason for exclusion
Sakamoto 2011	Review article; not a randomized controlled trial
Sampalis 2012	Fewer than 50 participants in each arm
Sands 2013	Comparison between continuous and intermittent treatment with celecoxib; no comparators of interest
Schnitzer 2015	Randomized controlled trial comparing tanezumab (5 or 10 mg) or placebo every 8 weeks with or without oral naproxen 500 mg twice daily or celecoxib 100 mg twice daily. There were no arms of interest for this systematic review
Silverstein 2000	Randomized controlled trial about gastrointestinal toxicity of celecoxib at 400mg a day. Methods to assess disease severity or progression were not used
Simon 1998	Duration less than four weeks
Stam 2012	Systematic review
Stengaard-Pedersen 2004	Comparison of different doses and timing of celecoxib treatment; no comparators of interest
Strand 2011	Comparison of continuous and intermittent celecoxib use; no comparators of interest (NCT00139776)
Taşcioğlu 2004	Fewer than 50 participants in each arm
Tive 2015	Pooled study of three controlled trials, where some groups of participants took naproxen 500 mg twice daily and celecoxib 100 mg twice daily; not a randomized controlled trial
Tran 2004	Comparison of celecoxib with rofecoxib and valdecoxib; no comparators of interest
Trudeau 2015	Fewer than 50 participants in each arm
Tsvetkova 2001	Not a randomized controlled trial
Wittenberg 2006	Duration less than four weeks
Wolfe 2004	Not a randomized controlled trial
Yoo 2014a	Randomized controlled trial comparing etoricoxib and celecoxib; no comparators of interest
Yoo 2014b	Compared PG201 (Layla), an ethanol extract, versus celecoxib; no comparators of interest

Characteristics of studies awaiting assessment *[ordered by study ID]*

[Essex 2016](#)

Methods	Randomized, double-blind, parallel-group trial Duration: 6 weeks Study dates: not indicated Locations: 31 centers in the US
Participants	Asian patients aged ≥ 45 years with defined criteria for knee osteoarthritis (OA)

Essex 2016 (Continued)

Interventions	<p>Celecoxib 200 mg</p> <p>Naproxen 1000 mg (500 mg twice daily)</p> <p>Placebo</p>
Outcomes	<p>Primary outcome: change from baseline to week 6 in the Patient's Assessment of Arthritis Pain, measured using a standard visual analog scale (VAS) ranging from 0 mm (no pain) to 100 mm (worst pain)</p> <p>Secondary outcomes: change from baseline the Patient's and Physician's Global Assessments of Arthritis, Western Ontario and McMaster Universities OA Index (WOMAC), Pain Satisfaction Scale, and the Patient Health Questionnaire (PHQ-9) (to week 6/early termination), and the American Pain Society (APS) pain score (to day 7). The WOMAC total domain score (range 0–96) was the sum of the pain, stiffness and physical function domain scores. The tolerability of celecoxib versus placebo was also evaluated, by comparing treatment-emergent treatment-related adverse events (TEAEs) and measurements of upper gastrointestinal (UGI) tolerability. This evaluation was carried out in the safety population (randomized patients receiving at least one dose of study medication)</p>
Notes	<p>Study registration: not indicated</p> <p>Study funding: Pfizer</p> <p>Authors' conflict of interest: all four authors employees of Pfizer</p> <p>The study was found as eligible in the updated search performed on April 11, 2017 and it will be included in the next update of this review</p>

EUCTR2005-002772-14-GB

Methods	<p>Randomized, double-blind, parallel-group trial</p> <p>Duration: 13 weeks</p> <p>Date of first enrolment: October 11, 2005</p> <p>Global completion date: October 19, 2006</p>
Participants	<p>Male or female outpatients 40 years of age or older who have the highest pain intensity occurring in the target hip joint, relative to other osteoarthritis (OA) joints (including the contralateral hip), with a diagnosis of primary hip osteoarthritis which meets American College of Rheumatology (ACR) criteria and have symptoms present for at least 3 months prior to screening. Diagnosis can be made at screening if symptoms have been present for 3 months by history</p>
Interventions	<p>Lumiracoxib 100 mg</p> <p>Celecoxib 200 mg</p> <p>Placebo</p>
Outcomes	<p>Primary outcomes: 1. Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain sub-scale score at 13 weeks, 2. Patient's global assessment of disease activity at 13 weeks (0–100 mm VAS), 3. WOMAC DPDA (difficulty in performing daily activities; a measure of function) sub-scale score at 13 weeks</p> <p>Secondary outcomes: none</p>
Notes	<p>Study identified at the WHO ICTRP registry. Results available online, but incomplete information available to judge risk of bias (https://www.clinicaltrialsregister.eu/ctr-search/tri-</p>

Celecoxib for osteoarthritis (Review)

EUCTR2005-002772-14-GB (Continued)

al/2005-002772-14/results). Query about trials on celecoxib was sent to email address media.relations@novartis.com, with no response. We did not find results of this study published in a peer-reviewed journal

Gordo 2017

Methods	Randomized, double-blind, parallel-group trial Duration: 6 weeks Study dates: February 2003 to January 2004 Locations: 34 centres in Spain, Germany and the UK
Participants	Patients diagnosed with osteoarthritis (OA) in a flare state and with a Functional Capacity Classification of I to III were eligible for study participation
Interventions	Celecoxib 200 mg Ibuprofen 2400 mg (800 mg three times a day) Placebo
Outcomes	Primary outcome: non-inferiority of celecoxib to ibuprofen in Patient's Assessment of Arthritis Pain (scored 0–100) Secondary outcomes: Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index, Pain Satisfaction Scale, and upper gastrointestinal tolerability
Notes	Study registration: ClinicalTrials.gov identifier: NCT00630929 Study funding: Pfizer Authors' conflict of interest: all four authors employees of Pfizer The study was found as eligible in the updated search performed on April 11, 2017 and it will be included in the next update of this review

Iyengar 2013

Methods	Manuscript reports data from 5 previously unpublished post-approval randomized studies, placebo and active controls, multicenter studies Duration: 6 weeks Study dates: 2001 to 2003
Participants	Included a total of 1497 participants Inclusion: aged at least 40 years, diagnosed with active and symptomatic osteoarthritis (OA), with a Functional Capacity Classification of I-III, including people in a flare state
Interventions	Celecoxib 200 mg Ibuprofen 2400 mg (800 mg three times a day) or Naproxen 1000 mg (500 mg twice daily) Placebo

Celecoxib for osteoarthritis (Review)

Iyengar 2013 (Continued)

Outcomes	Patient Health Questionnaire-9 (PHQ-9) depression score, pain on visual analog scale (VAS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)
Notes	Corresponding author was contacted regarding missing information. The corresponding author requested to see a protocol for this systematic review. The protocol was sent, but no further response was received

Lisse 2001

Methods	<p>Study design: 3 randomized studies, double-blind, placebo and active controls, 4 parallel groups, multicenter studies</p> <p>Duration: 12 weeks</p> <p>Study dates: not indicated</p> <p>Locations: various sites in the USA and Canada</p>
Participants	Numbers of participants who were randomized and completed study, their aged and percentage of women, were not provided for any of the 3 separate studies. Total number of participants in all 3 studies was 768
Interventions	<p>Celecoxib 200 mg</p> <p>Celecoxib 400 mg</p> <p>Naproxen 1000 mg</p> <p>Placebo</p>
Outcomes	<p>Benefit outcomes: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) total score and pain, stiffness and physical function sub scales; Short-Form Health Survey (SF-36)</p> <p>Harms outcomes: adverse events, rate of withdrawal due to adverse events</p>
Notes	<p>The study reports data pooled from 3 randomized controlled trials. The data were not presented separately for each trial separately and methods for each RCT were not described separately, which precludes risk of bias assessment for each study. Although it is indicated that these studies were unpublished at the moment of manuscript publication, it is unclear whether they were published subsequently, since there is no clinical trial registration number. A request for data was submitted to Pfizer in March 2014, the manufacturer and sponsor of the study. The response from Pfizer was received on April 7, 2016, indicating that they will not provide the requested data.</p> <p>This response was received from Daireen Garcia, MPH:</p> <p>"Thank you for your submission entitled "Functional Status and Health-Related Quality of Life of Elderly Osteoarthritic Patients Treated With Celecoxib." We regret to inform you that the data sets you requested are not available and we are therefore unable to support your request. Your request falls outside of the scope of Pfizer's clinical trial data (CTD) policy and while we make every effort to support requests that fall outside of our policy, we cannot guarantee that the data sets from clinical trials pre-dating 2007 will be available. We are committed to supporting important scientific research proposals such as yours and allowing secure access to Pfizer's anonymized patient-level data. We appreciate your interest in Pfizer's clinical trial data access program and would be pleased to review another submission from you in the future."</p>

Mastbergen 2010

Methods	Randomized, blinded, trial design not indicated Duration: “at least 4 weeks” Study dates: not indicated Locations: not indicated
Participants	Patients with severe knee osteoarthritis prior to joint replacement surgery
Interventions	celecoxib 2dd200mg celecoxib 2dd200mg stopped 3 days prior to surgery, naproxen 2dd250mg No treatment Total N = 172. Number of participants per group not indicated
Outcomes	Western Ontario and McMaster Universities (WOMAC) questionnaires were used to evaluate pain, stiffness and function before and after treatment
Notes	Study registration: not indicated Study funding: not indicated This was a conference abstract. We are trying to locate the full text. It is unclear from the abstract how many participants were in each group

Reginster 2016

Methods	Blinded, design not indicated Duration: 182 days Study dates: not indicated Locations: “5 European countries”
Participants	Participants with knee OA in accordance with the ACR criteria
Interventions	Celecoxib 200 mg Chondroitin Sulfate (CS) 800 mg/day Placebo Total N = 604. Number of participants per group not indicated
Outcomes	pain intensity on visual analog scale (VAS), Lequesne index measuring pain and function, Minimum Clinically Important Improvement (MCII), Patient Acceptable Symptoms State (PASS), safety analyses
Notes	Study registration: not indicated Study funding: not indicated This was a conference abstract. It is unclear from the abstract whether the study was randomized and how many participants were in each group. In communication with the lead author (Reginster IY) on April 21, 2017 we found out that the study was “just accepted for publication” in Annals of

Reginster 2016 (Continued)

the Rheumatic Diseases. The study will be analyzed in full text upon publication and screened for potential inclusion

Singh 2006

Methods	<p>Study design: randomized, double-blind, placebo and active controls, 3 parallel groups, multicenter study</p> <p>Duration: 12 weeks</p> <p>Study dates: not indicated</p> <p>Locations: Europe, South Africa, Latin America, USA, Canada, Asia, Australia and New Zealand</p>
Participants	<p>Number of participants that were randomized and completed study, as well as their age and percent of women, were not provided for each of the separate groups</p>
Interventions	<p>Celecoxib 200 mg</p> <p>Celecoxib 400 mg</p> <p>Non-specific non-steroidal anti-inflammatory drug (NSAID) therapy - diclofenac 100 mg or naproxen 1000 mg</p>
Outcomes	<p>Benefit outcomes: Patients' assessment of arthritis pain (VAS), Patient's Global Assessment of Arthritis, Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) Index (pain, joint stiffness, physical function)</p> <p>Harms outcomes: adjudication of serious upper gastro-intestinal events, overall evaluation of investigator-reported adverse events including cardiovascular events</p>
Notes	<p>The study included 13,274 participants from 39 countries. Two randomization schedules were used: one for celecoxib treatment groups and naproxen, and the other for the celecoxib treatment groups and diclofenac. However, data were pooled together and shown as NSAID groups and celecoxib groups. Since data were not presented as randomized, they could not be used. The first author was contacted and indicated that Pfizer, the manufacturer and sponsor of the study, owns the data. A request was submitted to Pfizer to provide summary data for this systematic review. The request for data was denied on October 18, 2013. Details about this communication are presented in Table 1.</p>

Abbreviations, in alphabetical order: APS - American Pain Society, DPDA - difficulty in performing daily activities, OMERCT-OARSI - Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trial Response Criteria Initiative, INSAID - non-steroidal anti-inflammatory drug, OA - osteoarthritis, PHQ-9 - Patient Health Questionnaire, VAS - visual analog scale, WOMAC - Western Ontario and McMaster Universities OA Index

Characteristics of ongoing studies [ordered by study ID]

EUCTR2011-005398-22-ES

Trial name or title	<p>Clinical study comparing a new single dose chewable tablet as treatment knee osteoarthritis with moderate to severe pain, compared to placebo and celecoxib</p>
Methods	<p>A randomized, multicenter, double-blind, double dummy and parallel study to evaluate the efficacy of the combination of chondroitin sulfate and glucosamine hydrochloride in a single dose chewable tablet versus placebo, using celecoxib as an active control in patients with knee osteoarthritis with moderate to severe pain</p>

EUCTR2011-005398-22-ES (Continued)

Duration: 6 months

Participants	Osteoarthritis of the knee diagnosed by the diagnostic criteria of the ACR; pain with moderate to severe score between 301 and 400 in the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) sub scale pain, or well with a score greater than 4 cm visual analogue scale of 10 cm; diagnosis clinical pain and functional limitation in the 6 months prior to the inclusion; male or female patients, aged over 40 years.
Interventions	Chondroitin sulfate Celecoxib Placebo
Outcomes	Primary outcome: Decrease of 26% in the total score of the questionnaire WOMAC score at 6 months of treatment with initial baseline. Secondary outcomes: 1. Improvement in pain in the WOMAC sub scale. 2. Improvement in stiffness in the WOMAC sub scale. 3. Improvement in the function in the WOMAC sub scale. 4. Evaluation of the safety of each treatment group. 5. Assessment by the patient global evolution of osteoarthritis. 6. Use of rescue medication. 7. Use of other concomitant medications: gastroprotective, anxiety and/or antidepressants. 8. Presence of swelling or effusion, or both. 9. Quality of life according to the Health Assessment Questionnaire (HAQ) form. 10. Valuation global medical 11. Responder Patients according to Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trial Response Criteria Initiative (OMERACT-OARSI) criteria 2004.
Starting date	Date of first enrolment: 11/10/2012
Contact information	Name: Clinical Operations Department Address: C/ Pamplona 92-94, 08018, Barcelona, Spain Telephone: 34933005218 Email: cod@recercaclinica.es Affiliation: Recerca Clinica SL
Notes	Latest study status checked on: December 14, 2016. The status of the study was: "Recruitment status: Authorised - recruitment may be ongoing or finished"

NCT01765296

Trial name or title	A Double-blind, Randomized, Multicenter, Active- and Placebo-Controlled Phase III Study to Evaluate the Efficacy and Safety of CG100649 in Osteoarthritis Patients
Methods	Allocation: Randomized Endpoint Classification: Safety/Efficacy Study Intervention Model: Parallel Assignment Masking: Double Blind (Subject, Caregiver, Investigator, Outcomes Assessor)

Celecoxib for osteoarthritis (Review)

NCT01765296 (Continued)

	Primary Purpose: Treatment
	Duration: 6 weeks
Participants	Participants with osteoarthritis of the hip or knee
Interventions	Drug: CG100649 Drug: Celecoxib Placebo
Outcomes	Primary Outcome Measures: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) Pain Subscale in the index joint Secondary Outcome Measures: WOMAC index, WOMAC Subscales, Subject's Global Assessment (SGA), Physician's Global Assessment (PGA), Withdrawals [Time Frame: Withdrawals due to lack of analgesic efficacy. Other Outcome Measures: Safety Measures of Extended Safety Study (24 weeks), Adverse events (AEs), Vital signs (blood pressure, pulse rate, body temperature), 12-lead ECG clinical laboratory tests (hematology, chemistry, coagulation, urinalysis), physical examination
Starting date	March 2013
Contact information	Su Jeong Yang, sjyang@cgxinc.com Seonggu Ro, PhD, sgro@cgxinc.com
Notes	Estimated Study Completion Date: March 2014 URL: https://clinicaltrials.gov/ct2/show/NCT01765296 Latest study status checked on: December 14, 2016. The status of the study was: "The recruitment status of this study is unknown because the information has not been verified recently." The last update of the entry was April 23, 2013.

NCT01768520

Trial name or title	Evaluation of the Efficacy and Safety of Entelon Tab. 150 mg in Patients With Osteoarthritis of Knee
Methods	Allocation: Randomized Endpoint Classification: Efficacy Study Intervention Model: Parallel Assignment Masking: Double Blind (Subject, Investigator) Primary Purpose: Treatment Duration: 12 weeks
Participants	Patients with a diagnosis of osteoarthritis (OA) of the knee within 3 months prior to study participation, as determined by the American College of Rheumatology (ACR) clinical and radiographic criteria, Kellgren and Lawrence Scale Grade II~III, both gender, 35 years ≤ age ≤ 75 years
Interventions	Drug: Enteron tab. 150mg (vitis vinifera extract 150mg): twice daily for 12 weeks Drug: Celebrex cap. (celecoxib 200 mg): once daily, for 12 weeks

Celecoxib for osteoarthritis (Review)

NCT01768520 (Continued)

	Placebo
Outcomes	Primary Outcome Measures: the change of total sum of K-WOMAC (Korean The Western Ontario and McMaster Universities Arthritis Index) Secondary Outcome Measures: Number of participants with adverse events
Starting date	July 2012
Contact information	Not provided
Notes	Estimated Study Completion Date: June 2014 URL: https://clinicaltrials.gov/ct2/show/NCT01768520 Latest study status checked on: December 14, 2016. The status of the study was: "The recruitment status of this study is unknown. The completion date has passed and the status has not been verified in more than two years." The last update of the entry was January 2013.

NCT02079727

Trial name or title	A Multicentre, Comparative, randomized, Double-blind, Double-dummy Clinical Trial on the Efficacy and Safety of Condrosulf Versus Celebrex and Versus a Placebo in the Treatment of Knee Osteoarthritis
Methods	Allocation: Randomized Endpoint Classification: Safety/Efficacy Study Intervention Model: Parallel Assignment Masking: Double Blind (Subject, Caregiver, Investigator, Outcomes Assessor) Primary Purpose: Treatment Duration: 6 months
Participants	Outpatients of either sex, aged ≥ 50 years, Participants affected by primary knee osteoarthritis of the medial or lateral femoro-tibial compartment, Diagnosis according to the American College of Rheumatology (ACR) criteria, Kellgren & Lawrence grade I-III, Knee osteoarthritis evolving for more than 6 months
Interventions	Drug: Chondroitin sulphate Drug: Celecoxib Placebo
Outcomes	Primary Outcome Measures: Lequesne's Index, Pain on visual analog scale (VAS) in mm Secondary Outcome Measures: Lequesne's Index, Pain (VAS in mm), minimal clinically important improvement (MCI), patient acceptable symptom state (PASS), Consumption of paracetamol Global efficacy assessment, Global efficacy assessed by the patient and the Investigator by means of a semi-quantitative verbal scale, Number of adverse events related to the treatments, Number of drop-outs due to adverse event (AE) related to the treatment Other Outcome Measures: Treatment compliance
Starting date	April 2014

Celecoxib for osteoarthritis (Review)

NCT02079727 (Continued)

Contact information	Jean-Yves Reginster, Prof. MD PhD, Unité d'Exploration du Métabolisme Osseux, CHU Centre ville, Polycliniques Universitaires L-Brull, Liège, Belgium
Notes	<p>Estimated Study Completion Date: June 2016</p> <p>URL: https://clinicaltrials.gov/ct2/show/NCT02079727</p> <p>Latest study status checked on: December 14, 2016. The status of the study was "This study is on-going, but not recruiting participants."</p> <p>The last update of the entry was February 23, 2016.</p>

Abbreviations, in alphabetical order: ACR - American College of Rheumatology, AE - adverse events, HAQ - Health Assessment Questionnaire, K-WOMAC - Korean The Western Ontario and McMaster Universities Arthritis Index, MCII - minimal clinically important improvement, OMERACT-OARSI - Outcome Measures for Rheumatology Committee and Osteoarthritis Research Society International Standing Committee for Clinical Trial Response Criteria Initiative, PASS - patient acceptable symptom state, PGA - Physician's Global Assessment, SGA - Subject's Global Assessment, VAS - visual analog scale, WOMAC - Western Ontario and McMaster Universities Arthritis Index

DATA AND ANALYSES

Comparison 1. Celecoxib versus placebo

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain	4	1622	Std. Mean Difference (IV, Random, 95% CI)	-0.22 [-0.32, -0.12]
2 Physical function	4	1622	Std. Mean Difference (IV, Random, 95% CI)	-0.17 [-0.27, -0.07]
3 Number withdrawn due to adverse events	28	12965	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.99 [0.85, 1.15]
4 Number experiencing any serious adverse events	28	13393	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.95 [0.66, 1.36]
5 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds)	8	3263	Peto Odds Ratio (Peto, Fixed, 95% CI)	1.91 [0.24, 14.90]
6 Number experiencing cardiovascular events (myocardial infarction, stroke)	5	2947	Peto Odds Ratio (Peto, Fixed, 95% CI)	3.40 [0.73, 15.88]

Analysis 1.1. Comparison 1 Celecoxib versus placebo, Outcome 1 Pain.

Study or subgroup	Celecoxib 200 mg		Placebo		Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
	N	Mean(SD)	N	Mean(SD)			
Clegg 2006	318	-99.2 (74.3)	313	-86.1 (74.2)		41%	-0.18[-0.33,-0.02]

Analysis 1.2. Comparison 1 Celecoxib versus placebo, Outcome 2 Physical function.

Analysis 1.3. Comparison 1 Celecoxib versus placebo, Outcome 3 Number withdrawn due to adverse events.

Analysis 1.4. Comparison 1 Celecoxib versus placebo, Outcome 4 Number experiencing any serious adverse events.

Analysis 1.5. Comparison 1 Celecoxib versus placebo, Outcome 5 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).

Analysis 1.6. Comparison 1 Celecoxib versus placebo, Outcome 6 Number experiencing cardiovascular events (myocardial infarction, stroke).

Comparison 2. Celecoxib versus tNSAIDs

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain	2	1180	Mean Difference (IV, Random, 95% CI)	-4.52 [-10.65, 1.61]
2 Physical function	1	264	Mean Difference (IV, Random, 95% CI)	-4.00 [-11.40, -0.60]
3 Number withdrawn due to adverse events	8	3150	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.97 [0.74, 1.27]
4 Number experiencing any serious adverse events	5	2404	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.92 [0.66, 1.28]
5 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds)	4	1755	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.61 [0.15, 2.43]
6 Number experiencing cardiovascular events (myocardial infarction, stroke)	1	916	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.47 [0.17, 1.25]

Analysis 2.1. Comparison 2 Celecoxib versus tNSAIDs, Outcome 1 Pain.

Analysis 2.2. Comparison 2 Celecoxib versus tNSAIDs, Outcome 2 Physical function.

Analysis 2.3. Comparison 2 Celecoxib versus tNSAIDs, Outcome 3 Number withdrawn due to adverse events.

Analysis 2.4. Comparison 2 Celecoxib versus tNSAIDs, Outcome 4 Number experiencing any serious adverse events.

Analysis 2.5. Comparison 2 Celecoxib versus tNSAIDs, Outcome 5 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).

Analysis 2.6. Comparison 2 Celecoxib versus tNSAIDs, Outcome 6 Number experiencing cardiovascular events (myocardial infarction, stroke).

Comparison 3. Celecoxib versus placebo (all eligible studies)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 All pain < 24 weeks	31	13069	Std. Mean Difference (IV, Random, 95% CI)	-0.41 [-0.49, -0.32]
2 Pain VAS at 6 weeks	11	3722	Std. Mean Difference (IV, Random, 95% CI)	-0.49 [-0.64, -0.34]
3 Pain VAS at 12 weeks	3	1226	Mean Difference (IV, Random, 95% CI)	-11.09 [-12.68, -9.50]
4 Pain VAS at 13 weeks	5	3853	Mean Difference (IV, Random, 95% CI)	-6.35 [-6.00, -4.70]
5 Pain on walking VAS at 6 weeks	4	1572	Mean Difference (IV, Random, 95% CI)	-12.42 [-14.83, -10.01]
6 Pain on walking WOMAC at 12 weeks	1	402	Mean Difference (IV, Random, 95% CI)	7.0 [1.73, 12.27]
7 Pain WOMAC at 6 weeks	12	4178	Mean Difference (IV, Random, 95% CI)	-3.28 [-4.51, -2.06]
8 Pain WOMAC at 12 weeks	9	3328	Std. Mean Difference (IV, Random, 95% CI)	-0.32 [-0.40, -0.23]
9 Pain WOMAC at 13 weeks	5	3853	Mean Difference (IV, Random, 95% CI)	-1.06 [-1.31, -0.80]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
10 Pain WOMAC at 24 weeks	1	631	Mean Difference (IV, Random, 95% CI)	-13.10 [-24.69, -1.51]
11 All physical function < 24 weeks	27	11940	Std. Mean Difference (IV, Random, 95% CI)	-0.35 [-0.43, -0.27]
12 Physical function WOMAC at 6 weeks	12	4069	Std. Mean Difference (IV, Random, 95% CI)	-0.31 [-0.41, -0.22]
13 Physical function WOMAC at 12 weeks	9	3185	Std. Mean Difference (IV, Random, 95% CI)	-0.46 [-0.65, -0.26]
14 Physical function WOMAC at 13 weeks	5	3853	Mean Difference (IV, Random, 95% CI)	-3.70 [-4.67, -2.74]
15 Physical function WOMAC at 24 weeks	1	631	Mean Difference (IV, Random, 95% CI)	-32.60 [-81.07, 15.87]
16 Quality of life: SF-36 physical component scores	1	402	Mean Difference (IV, Random, 95% CI)	-2.2 [-3.86, -0.54]
17 Quality of life: SF-36 mental component scores	1	402	Mean Difference (IV, Random, 95% CI)	0.20 [-1.46, 1.86]
18 Number of responders with at least 50% improvement in WOMAC pain	4	1816	Risk Ratio (M-H, Random, 95% CI)	1.47 [1.16, 1.87]
19 Number withdrawn due to adverse events	28	12965	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.99 [0.85, 1.15]
20 Number experiencing any serious adverse events	28	13393	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.95 [0.66, 1.36]
21 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds)	8	3263	Peto Odds Ratio (Peto, Fixed, 95% CI)	1.91 [0.24, 14.90]
22 Number experiencing cardiovascular events (myocardial infarction, stroke)	5	2947	Peto Odds Ratio (Peto, Fixed, 95% CI)	3.40 [0.73, 15.88]

Analysis 3.1. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 1 All pain < 24 weeks.

Study or subgroup	Celecoxib 200 mg		Placebo		Std. Mean Difference Random, 95% CI	Weight	Std. Mean Difference Random, 95% CI
	N	Mean(SD)	N	Mean(SD)			
Asmus 2014 Study 1	186	-27.3 (28.6)	184	-14.9 (27.1)	+	3.32%	-0.44[-0.65,-0.24]
Asmus 2014 Study 2	194	-28 (29.2)	186	-24.6 (27.3)	+	3.35%	-0.12[-0.32,0.08]
Bensen 1999	197	-27.5 (9.9)	203	-17.4 (14.2)	+	3.33%	-0.82[-1.02,-0.61]
Bingham 2007 Study 1	192	-24.9 (20.8)	85	-15.4 (22.1)	+	2.99%	-0.45[-0.7,-0.19]
Bingham 2007 Study 2	202	-27.8 (23.5)	60	-16.2 (19.4)	+	2.78%	-0.51[-0.8,-0.22]
					-5 -2.5 0 2.5 5	Favors celecoxib Favors placebo

Analysis 3.2. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 2 Pain VAS at 6 weeks.

Analysis 3.3. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 3 Pain VAS at 12 weeks.

Analysis 3.4. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 4 Pain VAS at 13 weeks.

Analysis 3.5. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 5 Pain on walking VAS at 6 weeks.

Analysis 3.6. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 6 Pain on walking WOMAC at 12 weeks.

Analysis 3.7. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 7 Pain WOMAC at 6 weeks.

Analysis 3.8. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 8 Pain WOMAC at 12 weeks.

Analysis 3.9. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 9 Pain WOMAC at 13 weeks.

Analysis 3.10. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 10 Pain WOMAC at 24 weeks.

Analysis 3.11. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 11 All physical function < 24 weeks.

Analysis 3.12. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 12 Physical function WOMAC at 6 weeks.

Analysis 3.13. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 13 Physical function WOMAC at 12 weeks.

Analysis 3.14. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 14 Physical function WOMAC at 13 weeks.

Analysis 3.15. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 15 Physical function WOMAC at 24 weeks.

Analysis 3.16. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 16 Quality of life: SF-36 physical component scores.

Analysis 3.17. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 17 Quality of life: SF-36 mental component scores.

Analysis 3.18. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 18 Number of responders with at least 50% improvement in WOMAC pain.

Analysis 3.19. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 19 Number withdrawn due to adverse events.

Analysis 3.20. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 20 Number experiencing any serious adverse events.

Analysis 3.21. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 21 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).

Analysis 3.22. Comparison 3 Celecoxib versus placebo (all eligible studies), Outcome 22 Number experiencing cardiovascular events (myocardial infarction, stroke).

Comparison 4. Celecoxib versus tNSAIDs (all eligible studies)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 All pain under 24 weeks	8	2277	Std. Mean Difference (IV, Random, 95% CI)	0.00 [-0.10, 0.10]
2 All pain over 24 weeks	1	916	Mean Difference (IV, Random, 95% CI)	-2.0 [-5.32, 1.32]
3 All physical function under 24 weeks	7	2176	Std. Mean Difference (IV, Random, 95% CI)	0.06 [-0.09, 0.21]
4 Number withdrawn due to adverse events	9	3739	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.85 [0.68, 1.07]
5 Number experiencing any serious adverse events	5	2404	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.92 [0.66, 1.28]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
6 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds)	4	1755	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.61 [0.15, 2.43]
7 Number experiencing cardiovascular events (myocardial infarction, stroke)	1	916	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.47 [0.17, 1.25]

Analysis 4.1. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 1 All pain under 24 weeks.

Analysis 4.2. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 2 All pain over 24 weeks.

Analysis 4.3. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 3 All physical function under 24 weeks.

Analysis 4.4. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 4 Number withdrawn due to adverse events.

Analysis 4.5. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 5 Number experiencing any serious adverse events.

Analysis 4.6. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 6 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).

Analysis 4.7. Comparison 4 Celecoxib versus tNSAIDs (all eligible studies), Outcome 7 Number experiencing cardiovascular events (myocardial infarction, stroke).

Comparison 5. Celecoxib versus naproxen 1000 mg

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 All pain $\leq$ 24 weeks	6	1781	Std. Mean Difference (IV, Random, 95% CI)	-0.04 [-0.14, 0.05]
2 Pain VAS at 6 weeks	2	398	Mean Difference (IV, Random, 95% CI)	0.55 [-3.97, 5.07]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3 Pain VAS at 12 weeks	3	975	Mean Difference (IV, Random, 95% CI)	-0.97 [-2.69, 0.76]
4 Pain WOMAC at 6 weeks	2	503	Mean Difference (IV, Random, 95% CI)	0.35 [-0.54, 1.23]
5 Pain WOMAC at 12 weeks	3	1073	Mean Difference (IV, Random, 95% CI)	-0.03 [-0.65, 0.60]
6 Pain WOMAC at 6 months	1	310	Mean Difference (IV, Random, 95% CI)	0.10 [-0.46, 0.66]
7 All physical function $\leq$ 24 weeks	6	1817	Std. Mean Difference (IV, Random, 95% CI)	-0.01 [-0.18, 0.16]
8 Physical function WOMAC at 6 weeks	2	313	Mean Difference (IV, Random, 95% CI)	-2.99 [-6.43, 0.44]
9 Physical function WOMAC at 12 weeks	3	1073	Mean Difference (IV, Random, 95% CI)	-1.27 [-4.32, 1.78]
10 Physical function WOMAC at 6 months	1	301	Mean Difference (IV, Random, 95% CI)	0.30 [-1.94, 2.54]
11 Number withdrawn due to adverse events	6	2173	Odds Ratio (M-H, Random, 95% CI)	0.81 [0.54, 1.23]
12 Number experiencing any serious adverse events	2	841	Peto Odds Ratio (Peto, Fixed, 95% CI)	1.11 [0.45, 2.75]
13 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds)	2	587	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.37 [0.05, 2.62]

Analysis 5.1. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 1 All pain $\leq$ 24 weeks.

Analysis 5.2. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 2 Pain VAS at 6 weeks.

Analysis 5.3. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 3 Pain VAS at 12 weeks.

Analysis 5.4. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 4 Pain WOMAC at 6 weeks.

Analysis 5.5. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 5 Pain WOMAC at 12 weeks.

Analysis 5.6. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 6 Pain WOMAC at 6 months.

Analysis 5.7. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 7 All physical function ≤ 24 weeks.

Analysis 5.8. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 8 Physical function WOMAC at 6 weeks.

Analysis 5.9. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 9 Physical function WOMAC at 12 weeks.

Analysis 5.10. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 10 Physical function WOMAC at 6 months.

Analysis 5.11. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 11 Number withdrawn due to adverse events.

Analysis 5.12. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 12 Number experiencing any serious adverse events.

Analysis 5.13. Comparison 5 Celecoxib versus naproxen 1000 mg, Outcome 13 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).

Comparison 6. Celecoxib versus diclofenac 100 mg

Outcome or subgroup title	No. of studies	No. of partici- pants	Statistical method	Effect size
1 Pain VAS at 1 year	1	916	Mean Difference (IV, Ran- dom, 95% CI)	-2.0 [-5.32, 1.32]
2 Number withdrawn due to adverse events	1	916	Peto Odds Ratio (Peto, Fixed, 95% CI)	1.44 [0.80, 2.61]
3 Number experiencing any serious ad- verse events	1	916	Odds Ratio (M-H, Random, 95% CI)	0.90 [0.62, 1.30]
4 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds)	1	916	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.14 [0.01, 2.16]
5 Number experiencing cardiovascular events (myocardial infarction, stroke)	1	916	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.47 [0.17, 1.25]

Analysis 6.1. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 1 Pain VAS at 1 year.

Analysis 6.2. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 2 Number withdrawn due to adverse events.

Analysis 6.3. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 3 Number experiencing any serious adverse events.

Analysis 6.4. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 4 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).

Analysis 6.5. Comparison 6 Celecoxib versus diclofenac 100 mg, Outcome 5 Number experiencing cardiovascular events (myocardial infarction, stroke).

Comparison 7. Celecoxib versus diclofenac 150 mg

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Pain VAS at 6 weeks	1	398	Mean Difference (IV, Random, 95% CI)	1.90 [-3.68, 7.48]
2 Pain WOMAC at 6 weeks	1	398	Mean Difference (IV, Random, 95% CI)	0.30 [-0.52, 1.12]
3 Pain on walking at 12 weeks	1	98	Mean Difference (IV, Random, 95% CI)	13.0 [3.11, 22.89]
4 Physical function WOMAC at 6 weeks	1	398	Mean Difference (IV, Random, 95% CI)	1.90 [-0.72, 4.52]
5 Number withdrawn due to adverse events	2	650	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.78 [0.46, 1.32]
6 Number experiencing any serious adverse events	2	647	Peto Odds Ratio (Peto, Fixed, 95% CI)	0.79 [0.21, 2.93]
7 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds)	1	252	Peto Odds Ratio (Peto, Fixed, 95% CI)	7.45 [0.46, 119.74]

Analysis 7.1. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 1 Pain VAS at 6 weeks.

Analysis 7.2. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 2 Pain WOMAC at 6 weeks.

Analysis 7.3. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 3 Pain on walking at 12 weeks.

Analysis 7.4. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 4 Physical function WOMAC at 6 weeks.

Analysis 7.5. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 5 Number withdrawn due to adverse events.

Analysis 7.6. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 6 Number experiencing any serious adverse events.

Analysis 7.7. Comparison 7 Celecoxib versus diclofenac 150 mg, Outcome 7 Number experiencing gastro-intestinal events (perforation, ulcer, bleeds).

ADDITIONAL TABLES

Table 1. Detailed account of data request from Pfizer regarding Singh 2006 study

March 8, 2013: L Puljak (review author) contacted Prof Gurkirpal Singh to request missing data, because data were not shown as randomized
March 8, 2013: Prof Singh responds that the study protocol prespecified data analysis
March 8, 2013: L Puljak asked Prof Singh if raw data can be obtained
March 8, 2013: Prof Singh responds that data are owned by Pfizer and copies the message to 2 Pfizer employees, Manuela Berger and Gail Cawkwell
March 12, 2013: L Puljak writes to Manuela Berger and Gail Cawkwell, repeating the request
March 13, 2013: Gail Cawkwell responds, thanks for the interest, directs communication to Dr Peter Park, medical lead for Celebrex, who will "follow up [...] as soon as he returns from a business trip"
March 25, 2013: Since there was no response, L Puljak emailed Peter Park
March 25, 2013: Automatic reply received, indicating Peter Park is away until March 24
March 25, 2013: Peter Park responds, asks L Puljak to file for "independent data grant application" – Investigator Initiated Research (IIR) request
March 29, 2013: L Puljak filed the IIR request in order to obtain summary data (means and standard deviations) for each randomized group separately
Shortly afterwards, a gentleman from Croatian Pfizer (Dr Lado Uglesic) called L Puljak on the office phone to ask about the identity of authors, purpose of doing this
April 30, 2013: Peter Park responds that a committee will meet in May to decide about the request
May 8, 2013: Lado Uglesic from Croatian Pfizer emails L Puljak confirmation of the IIR
May 8, 2013: Lado Uglesic asks L Puljak to anonymize her CV submitted with IIR application
May 13, 2013: Lado Uglesic confirms that the CV was successfully anonymized
October 18, 2013: Lado Uglesic sends email containing the following decision: "Thank you for your submission entitled <i>Celecoxib for osteoarthritis</i> . After careful consideration by the CELEBREX IIR Grant Review Committee, we regret to inform you that we are unable to support it at this time. We receive many promising requests and unfortunately cannot respond favourably to all of them. [...] We appreciate your interest in Pfizer's IIR program and would be pleased to review another submission from you in the future."

Table 2. Standardised mean difference (SMD) for pain versus placebo with different instruments and follow-up times

Pain instrument	6 weeks	12 and 13 weeks	24 weeks	All follow-up times combined
Pain VAS	SMD = -0.49 (95% CI -0.56 to -0.43), $P < 0.001$, heterogeneity: 79% (11 studies, 3722 participants)	SMD = -0.37 (95% CI -0.43 to -0.30), $P < 0.001$, heterogeneity: 93% (6 studies, 3842 participants)	No studies reported pain VAS outcome	SMD = -0.43 (95% CI -0.48 to -0.38), $P < 0.001$, heterogeneity: 88% (17 studies, 7564 participants)
Pain WOMAC	SMD = -0.40 (95% CI -0.46 to -0.33), $P < 0.001$, heterogeneity: 71% (12 studies, 4718 participants)	SMD = -0.28 (95% CI -0.33 to -0.23), $P < 0.001$, heterogeneity: 19% (12 studies, 5944 participants)	SMD = -0.18 (95% CI -0.33 to -0.02), $P = 0.03$, heterogeneity: not applicable	SMD = -0.32 (95% CI -0.36 to -0.28), $P < 0.001$, heterogeneity: 62% (25 studies, 10,753 participants)

Table 2. Standardised mean difference (SMD) for pain versus placebo with different instruments and follow-up times (Continued)

	(1 study, 631 participants)			
Both instruments	SMD = -0.43 (95% CI -0.53 to -0.34), $P < 0.001$, heterogeneity: 76% (17 studies, 7900 participants)	SMD = -0.35 (95% CI -0.44 to -0.25), $P < 0.001$, heterogeneity: 81% (12 studies, 9786 participants)	SMD = -0.18 (95% CI -0.33 to -0.02), $P = 0.03$, heterogeneity: not applicable (1 study, 631 participants)	SMD = -0.39 (95% CI -0.46 to -0.32), $P < 0.001$, heterogeneity: 80% (30 studies, 18,317 participants)

Each cell contains the following information: Standardized mean difference (SMD), 95% confidence interval (95% CI), heterogeneity, number of studies, number of participants.

Abbreviations: VAS: visual analogue scale; WOMAC: Western Ontario and McMaster Universities Arthritis Index

APPENDICES

Appendix 1. CENTRAL search strategy

1 (cyclooxygenase-2 or cyclooxygenase 2 or cyclooxygenase-II or cyclooxygenaseII).ti,ab.
2 (cyclo oxygenase-2 or cyclo oxygenase 2 or cyclo oxygenase-II or cyclo oxygenaseII).ti,ab.
3 cyclooxygenase.kw.
4 (cox-2 or cox2 or cox-II or coxII).ti,ab.
5 cox 2.kw.
6 (celecoxib or celebrex or SC-58635).ti,ab.
7 celecoxib.kw.
8 exp Cyclooxygenase Inhibitors/
9 or/1-8
10 exp Osteoarthritis/
11 osteoarthritis.kw.
12 (arthrit\$ or osteoarthritis\$).ti,ab.
13 or/10-12
14 9 and 13

Appendix 2. MEDLINE (Ovid) search strategy

1 (cyclooxygenase-2 or cyclooxygenase 2 or cyclooxygenase-II or cyclooxygenaseII).ti,ab.
2 (cyclo oxygenase-2 or cyclo oxygenase 2 or cyclo oxygenase-II or cyclo oxygenaseII).ti,ab.
3 (cox-2 or cox2 or cox-II or coxII).ti,ab.
4 (celecoxib or celebrex or SC-58635).af
5 exp Cyclooxygenase inhibitors/
6 or/1-5
7 exp Osteoarthritis/
8 (arthrit\$ or osteoarthritis\$).ti,ab.
9 7 or 8
10 6 and 9

Appendix 3. Embase search strategy

1 (cyclooxygenase-2 or cyclooxygenase2 or cyclooxygenase-II or cyclooxygenaseII).ti,ab.
2 (cyclo oxygenase-2 or cyclo oxygenase2 or cyclo oxygenase-II or cyclo oxygenaseII).ti,ab.

3 (cox-2 or cox2 or cox-II or coxII).ti,ab.

4 (celecoxib or celebrex or SC-58635).af.

5 Cyclooxygenase 2 inhibitor/

6 Cyclooxygenase 2/

7 Celecoxib/

8 or/1-7

9 (arthrit\$ or osteoarthrit\$).ti,ab.

10 exp Arthritis/

11 9 or 10

12 8 and 11

13 random\$.tw.

14 factorial\$.tw.

15 crossover\$.tw.

16 cross over.tw.

17 cross-over.tw.

18 placebo\$.tw.

19 (doubl\$ adj blind\$).tw.

20 (singl\$ adj blind\$).tw.

21 assign\$.tw.

22 allocat\$.tw.

23 volunteer\$.tw.

24 crossover procedure/

25 double blind procedure/

26 randomized controlled trial/

27 single blind procedure/

28 or/13-27

29 12 and 28

FEEDBACK

Data for 50% pain relief, 9 June 2017

Summary

Feedback submitted: 31 June, 2017

Prof. Andrew Moore

Comment: Thank you for the work on this review, which will take a while to digest.

I am particularly interested in the results for pain responder – at least 50% reduction in pain – because that is the outcome desired by people with pain, including arthritis (Anaesthesia. 2013 Apr;68(4):400-12.). Unfortunately it seems that the numbers in analysis 3.18 are entered

incorrectly – placebo numbers for celecoxib and vice versa. That is why the risk ratio is less than 1 – indicating a reduction in responders with celecoxib compared with placebo. Celecoxib may have limited efficacy at the 200 mg dose, but it probably is better than placebo.

There is one other published responder analysis for celecoxib, which you have not captured, though it may have had some useful information, at least for comparison purposes (Ann Rheum Dis 2010;69:374–379). It provided an individual participant analysis, inter alia, of celecoxib versus placebo in 700 or so participants, with data analysis by level of pain relief and duration of observation. The 12-week NNT for at least 50% pain reduction was 5.8 (4.2 to 9.5), somewhat different from the value of 14 (8 to 73) that I calculated from data extracted from the same two papers as you use.

The fact of the difference is interesting, and is largely down to a 42% placebo response rate in the 24 weeks comparison in Clegg 2006. While we expect high withdrawal rates in 12-week studies in chronic pain, and high placebo response rates often, this is very different from what I would expect in most OA studies, though such high placebo response rates are also seen in studies of topical NSAIDs for OA, which also share large amounts of unpublished data (Cochrane Database of Systematic Reviews 2016, Issue 4. Art. No.: CD007400).

There is much we need learn, and more uncertainty than we would like.

Can I also say how refreshing it is, in the light of current evidence, to see a Cochrane review using a cut-off of 50 participants per treatment arm for included studies. There were battles enough with 10, so 50 is a real, and in many cases sensible, policy.

Reply

Reply submitted: 09 June, 2017

Dear Prof. Moore,

thank you very much for taking time to write to us about the review, suggestions for its improvement, and we appreciate your kind words about the aspects that you found commendable. Prof. Peter Tugwell suggested in the protocol phase to limit the studies to those that had more than 50 patients and duration of more than 4 weeks so he has credit for that idea. Based on the feedback, we did the following changes: In the 50% pain relief outcome we erroneously switched the numbers, and this has been corrected now, both in the analysis 3.18 and in the text of review.

Furthermore, we looked into the Ann Rheum Dis 2010;69:374–379 (Moore 2010) manuscript again. We excluded that manuscript long time ago at the screening level because it was meta-analysis and because in our original protocol 50% pain relief was not among the planned outcomes. In the final stages of the review preparation experts from the Cochrane Musculoskeletal Group suggested that we might also add that outcome, which was available in only few studies, and is now considered of great importance for patients. At that stage we went only through already included studies and extracted data for 50% pain relief where available.

After obtaining permission from you for using data presented in the Moore 2010 manuscript, we added data from two more studies to Analysis 3.18., added Moore 2010 as secondary studies in Bingham 2017 Study 1 and Study 2, and made corresponding changes in the text of review. We also added a short comment in discussion and a note in the recommendations for future research to include 50% pain relief outcome.

Contributors

Livia Puljak

WHAT'S NEW

Date	Event	Description
9 June 2017	Feedback has been incorporated	Analysis 3.18 was corrected based on the feedback and two more studies were added to this analysis.

HISTORY

Protocol first published: Issue 5, 2012

Review first published: Issue 5, 2017

Date	Event	Description
14 November 2008	Amended	Converted to new review format.

CONTRIBUTIONS OF AUTHORS

LP, AM, DV and FM screened initial search results, identified studies that fulfilled the inclusion criteria, extracted data and performed 'Risk of bias' assessment. AU searched the literature. PT contributed to major decisions about study design. LP entered data in RevMan, conducted analyses and wrote the first draft. All authors contributed to the interpretation of the review and approved the final draft.

DECLARATIONS OF INTEREST

Livia Puljak: None known.

Ana Marin: None known.

Davorka Vrdoljak: None known.

Filipa Markotic: None known.

Ana Utrobicic: None known.

Peter Tugwell: None known.

SOURCES OF SUPPORT

Internal sources

- University of Split School of Medicine, Croatia.

External sources

- No sources of support supplied

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

In the review, studies were excluded if they had fewer than 50 participants in each arm or involved treatment of less than four weeks duration. Studies were included if any accepted method to assess disease severity or progression was used. Only participants receiving celecoxib 200 mg were included because this is the dose recommended for people with osteoarthritis. A new outcome was included: numbers of responders with at least 50% improvement in pain score; CMSG editorial team recommended adding this outcome to keep in line with the latest guidance on reporting pain outcomes ([Busse 2015](#)).

As recommended by the CMSG, we included a search of adverse events on regulatory agencies.

INDEX TERMS

Medical Subject Headings (MeSH)

Anti-Inflammatory Agents, Non-Steroidal [adverse effects] [*therapeutic use]; Celecoxib [adverse effects] [*therapeutic use]; Diclofenac [therapeutic use]; Naproxen [therapeutic use]; Osteoarthritis, Hip [*drug therapy]; Osteoarthritis, Knee [*drug therapy]; Pain Measurement [methods]; Placebos [adverse effects] [therapeutic use]; Quality of Life; Randomized Controlled Trials as Topic

MeSH check words

Adult; Female; Humans; Male; Middle Aged