TITLE PAGE

Title: Does flare trial design affect the effect size of non-steroid anti-inflammatory drugs in symptomatic osteoarthritis? A systematic review and meta-analysis.

Authors: Smith TO, Zou K, Abdullah N, Chen X, Kingsbury SR, Doherty M, Zhang W, Conaghan PG.

Affiliations:

Toby O Smith BSc (Hons), MSc, MA, PhD, MCSP – University Lecturer, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, UK

Kun Zou, PhD – Post-Doctorial Researcher, Division of Rheumatology, Orthopaedics and Dermatology, University of Nottingham, UK

Natasya Abdullah, MD - Post-Doctorial Researcher, Division of Rheumatology, Orthopaedics and Dermatology, University of Nottingham, UK

Xi Chen, PhD - Post-Doctorial Researcher, Division of Rheumatology, Orthopaedics and Dermatology, University of Nottingham, UK

Sarah R Kingsbury BSc (Hons), PhD - Osteoarthritis Strategic Project Lead, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, UK.

Michael Doherty MA MD FRCP - Professor of Rheumatology and Head of Division of Rheumatology, Orthopaedics and Dermatology, University of Nottingham, UK

Weiya Zhang BMed BPH, MEpi, PhD - Professor of Epidemiology, Faculty of Medicine & Health Sciences, University of Nottingham, UK.

Philip G Conaghan MB BS PhD FRACP FRCP – Professor of Musculoskeletal Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, UK.

Corresponding Author: Professor Philip Conaghan, NIHR Leeds Musculoskeletal Biomedical Research Unit, Section of Musculoskeletal Disease, Chapel Allerton Hospital, Chapeltown Road, Leeds, LS7 4SA, United Kingdom. Email: p.conaghan@leeds.ac.uk; Tele +44 (0) 113 3924883

ABSTRACT

Objectives: It is thought that the clinical trial benefits of oral non-steroid anti-inflammatory drugs (NSAIDs) may relate to flare designs. The aim of this study was to examine the difference in NSAID (including COX-2 inhibitors) response in osteoarthritis (OA) trials based on different designs.

Methods: Systematic review was undertaken of the databases MEDLINE, EMBASE, AMED, CINAHL and the Cochrane library to February 2015. Randomised controlled trials assessing pain, function and/or stiffness following commencement of NSAIDs in flare and non-flare designs were eligible. Trials were assessed using the Cochrane Risk of Bias tool. Meta-analyses were conducted to assess the effect sizes of NSAIDs for OA with flare versus non-flare trial designs.

Results: Fifty-seven studies including 33,263 participants assessing 26 NSAIDs were included. Twenty-two (39%) were flare design, 24 (42%) were non-flare designs, 11 (19%) were possible flare designs. On meta-analysis, there was no statistically significant difference in effect size of NSAIDs versus placebo between flare and non-flare trial designs for absolute pain and function or stiffness at immediate (1 week), short (2 to 4 week) or longer (12 to 13 week) follow-up periods ($p>0.05$). However there was a lower effect size for mean change in pain in flare and possible flare trials compared to non-flare trials at short-term follow-up (0.36 versus 0.69; $p=0.05$).

Conclusions: Contrary to previous understanding, flare trial designs do not result in an increased treatment effect for NSAIDs in people with OA compared to non-flare design. Whether flare design influences other outcomes such as joint effusion remains unknown.

Keywords: Randomised Controlled Trial; NSAIDs; Osteoarthritis; Effect Size; Methodology

INTRODUCTION

Osteoarthritis (OA) is a debilitating musculoskeletal disorder which symptomatically affects approximately 10% of the population aged over 60 years, and increases with age [1,2]. The most commonly affected joints are the hands, feet, knees and hips, with principle manifestations being pain, stiffness and resultant loss of function and independence [3]. The optimal treatment for people with OA combines both pharmacological and non-pharmacological treatments [4]. Nonsteroidal anti-inflammatory drugs (NSAIDs and selective cyclooxygenase-2 inhibitors) are the most commonly used painkillers for people with OA in Europe and the USA with 20% to 35% of the OA population reporting their use [5,6].

Flare design trials have been commonly used to assess the efficacy of NSAIDs. They are defined as trials which have recruited patients with increased pain after ceasing their usual pharmacological treatment [7]. Accordingly, these participants may respond differently to the general OA population with respect to the therapy under investigation. This may be of particular importance if only those who have previously responded to a NSAID are recruited to a trial investigating NSAID efficacy, inflating the effect size compared to an unselected OA group.

Trijau et al [8] previously presented a well-designed meta-analysis comparing the efficacy of NSAIDs in flare and non-flare design trials. They reported that flare trials evaluating NSAIDs resulted in a higher magnitude of treatment effect compared to non-flare trials. However, a large number of relevant papers have been published since the March 2009 search date in that publication. Our aim was therefore to conduct a contemporary systematic literature review investigating the effects of flare design trials on the efficacy of NSAIDs for people with OA and then to perform a meta-regression to examine the effects of other possible factors including study setting, allocation concealment and sample size on outcomes.

METHODS

Search Strategy

A search strategy was undertaken of the published databases: MEDLINE, EMBASE, AMED, CINAHL and the Cochrane library. The search was undertaken from database inception to 1st February 2015. A review of the potentially included papers' reference lists and previous review articles was undertaken to identify any additional studies. The search terms for the MEDLINE search are presented in Supplementary Table 1. These were amended for the other search databases. We did not exclude papers based on year or language of publication.

Identification of Studies

All randomised placebo controlled trials assessing the efficacy of NSAIDs in people with OA were included. Flare design was defined as trials where participants were only eligible when they had increased pain after ceasing their usual treatment before entering the trial [7]. Where there was uncertainty regarding the magnitude of this increased pain but there was sufficient evidence to suggest that these could have been flare design trials, the studies were included and termed 'possible flare design' trials. Where there was no reference to 'flare trial design' and it was clear a non-flare trial design was adopted, these were defined as 'non-flare design' trials. Participants with OA of any joint or multiple joints were included. OA was defined according to the American College of Rheumatology (ACR) criteria, radiological and/or clinical diagnosis [9]. The interventions included all NSAIDs (conventional and COX-2 inhibitors).

Outcomes

The primary outcome was pain. Pain could be measured by visual analogue scale (VAS) or numerical rating scale (NRS) methods, or as a sub-domain of an overall scoring system such as the Western Ontario and McMaster Universities Arthritis Index (WOMAC)[10]. Where pain was assessed using a number of different measures, we selected the scale according to the hierarchy of the outcomes suggested by Juhl et al [11]. Secondary outcome measures were function and stiffness.

Outcomes were assessed at specific follow-up periods. These were classified *a priori* as: immediate, short or longer-term. Immediate term outcomes were defined as outcomes within the first week of commencing the trial; short term was defined as two to four weeks following commencement, and longer-term outcomes were defined as six weeks and over.

Data Extraction

Data were extracted by one reviewer (KZ) and validated by three others (NA, XC, TS). Any disagreements were resolved through discussion with a fifth reviewer (WZ). Data were extracted onto a pre-defined database and included: country of origin, sample size, gender, age, BMI, setting (community or hospital-based), NSAID medication (type, dose, frequency, duration, route of delivery), placebo comparison, follow-up intervals and period, baseline and follow-up outcomes.

Critical Appraisal

Each included trial was assessed for methodological quality using the Cochrane Risk of Bias tool [12]. Trial design was assessed using the five criteria: random sequence generation, allocation concealment, blinding to participants, blinding to outcome assessment, withdrawals (attrition bias) and selective reporting (reporting bias).

Statistical Methods

Study heterogeneity was assessed through visual assessment of the participant characteristics, trial design, NSAIDs and placebo approaches and outcome measures. Where there was evidence of trial homogeneity, a meta-analysis was undertaken.

Heterogeneity was measured using I^2 index and Chi-squared test. Where I^2 was 30% or above and Chi-squared $p \leq 0.10$, a random-effects meta-analysis was undertaken. When I^2 was less than 30% and Chi-squared $p > 0.10$, a fixed effects meta-analysis was undertaken. All meta-analyses were undertaken by two reviewers (TS, KZ) and interpreted by four reviewers. Through this we assessed the effect size (ES) (standard mean difference between NSAID versus placebo interventions) overall and at each time point (immediate, short, longer-term). Clinically, an effect size of 0.2 suggested a small effect, 0.5 meant a

moderate effect and 0.8 and over indicated a large effect. The analysis of flare versus non-flare trial design was then made to assess for differences between these two subgroups of the NSAID data, presenting this with Chi-square p-values and I^2 statistics between the two pooled effect sizes. A sensitivity analysis was also undertaken to compare 'flare design' or 'possible flare designs' for each time point.

A meta-regression analysis (random-effects model) was undertaken to confirm whether flare design affected pain and other clinical outcomes given the adjustment for setting (community-based), allocation concealment, intention-to-treat analysis and whether there was more than or less than 100 participants per study arm, as suggested by Nüesch et al [13]. These are the common factors that may affect the results from RCTs and that may confound the difference between flare and non-flare designs. The partial regression coefficient (β) was used to present the contribution of each variable. A funnel plot was constructed to assess for publication bias [14].

All data were presented with 95% confidence intervals (CI) and with forest-plots. A two-sided p-value of < 0.05 was considered statistically significant. Analyses were undertaken using RevMan (Review Manager). Version 5.1. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2011) and STATA (version 14.0, STATA Corp, Dallas, Texas, USA).

RESULTS

Characteristics of included studies

A total of 8,592 citations were identified from the search. Of these 57 were eligible and included in the meta-analysis (Supplementary Figure 1). Of the 57 trials, 22 (39%) were 'flare designs', 24 (42%) were 'non-flare designs' and 11 (19%) were 'possible flare designs' (Supplementary Tables 2 and 3). Of the 22 flare designs, 20 (91%) were funded by industry and two were unclear about funding source. In the 24 non-flare trial designs, 10 (42%) were funded by industry, two were funded by public funding (8%), and 13 remained unclear about funding source.

A total of 33,263 participants were included in the review (10,480 men/21,877 women). Two studies did not provide the gender composition of their cohort [15,16]. Mean age ranged from 53.5 [17] to 83 years [18]. The duration of NSAID/placebo intervention ranged from one week [19,20] to 26 weeks [21]. Thirty studies were for knee OA, four for hip OA, 21 for hip and knee OA, and two for hand OA. The mean duration of OA was documented in 31 papers. This ranged from 2.2 years [22] to 15 years [18]. Trials were conducted in a hospital setting in 36 studies, in the community in two studies, whilst unclear in 19 papers.

Critical Appraisal

In general, the quality of studies was higher in the flare designs than the non-flare or possible flare design trials (Supplementary Table 4). There was a higher proportion of papers which clearly designed the randomisation, blinded of their assessors, and assessed a minimum of 85% of their cohort in the flare compared to non/possible flare trials. The quality of studies was comparable between the non-flare and possible flare trials.

Publication Bias

As presented in Supplementary Figure 2 there was some evidence of small sample size publication bias in the non-flare designs but not in flare designs. That is, studies with smaller sample size were more likely to produce larger effect size and they were more likely to be published in non-flare trials.

Primary outcome: Pain

There was no statistically significant difference in effect size between flare and non-flare trial design for pain as measured by mean change in pain score ($p=0.08$; $I^2=66.4\%$; Figure 1; Table 1) or absolute pain score ($p=0.23$; $I^2=29.4\%$; Figure 2) These findings remained when the data were analysed by follow-up period for the flare versus non-flare trials (Table 2). There were two exceptions to this. There was a lower effect size in flare and possible flare trial designs in mean change in pain score at short-term follow-up compared to non-flare trial designs (ES: 0.36 vs. 0.69; $p=0.05$; $I^2=73.3\%$; Table 3), although this presented with high statistical heterogeneity. Conversely there was a statistically

significant difference between flare and possible flare trials for absolute pain score at longer-term follow-up, being greater in flare trial designs (ES: 0.85 vs. 0.40; p=0.05; I^2=74.0%), and in the flare and non-flare trials for the same outcome at the same follow-up period (ES: 0.44 vs. 0.00; p<0.01; I^2=90.2%).

Secondary outcomes: Function

There was no statistically significant difference in effect size between flare and non-flare trial design for function as measured by mean change (p=0.54; I^2=0%; Table 1) or absolute functional scores (p=0.08; I^2=67.4%). However, when assessed by follow-up period, there was a statistical difference for immediate-term follow-up analysis with greater effect sizes in mean change in functional scores for non-flare trial designs compared to flare trial papers (ES: 0.26 vs. 0.47; p=0.04; I^2=75.6%; Table 2). This was also evident for the short-term follow-up in the flare and possible flare trial designs versus non-flare trial designs (ES: 0.28 vs. 0.68; p<0.01; I^2=93.5%; Table 3), and in the longer-term follow-up (ES: 0.35 vs. 0.55; p=0.01; I^2=84.9%; Table 3). There was no statistically significant difference between flare versus non-flare (Table 2) and flare/possible flare trial designs versus non-flare trial design for absolute functional score (Table 3).

Secondary outcomes: Stiffness

There was no statistically significant difference in effect size between flare and non-flare trial design for stiffness as measured with mean change in functional scores (p=0.75; I^2=0%; Table 1) or absolute stiffness scores (p=1.00; I^2=0%). There was no statistically significant difference in effect size between flare and non-flare or flare and possible flare compared to non-flare trial designs for stiffness as measured with mean change from baseline to any follow-up interval or absolute score (Table 2; Table 3). The only exception was for non-flare trial designs which demonstrated a greater effect size for absolute stiffness score compared to flare and possible flare trial designs on immediate-term follow-up analysis (ES: 0.22 vs. 0.84; p=0.01; I^2=86.3%; Table 3).

Meta-Regression

The results of the meta-regression are presented in Table 4. This analysis confirmed that flare design had similar results as non-flare designs, given the adjustment for the five major study-level confounding factors (study setting, allocation concealment, ITT, blinding to participants and ≥100 participants per trial arm).

DISCUSSION

The findings of this paper indicate that there is no significant difference between flare and non-flare trial designs for NSAIDs versus placebo when assessed in people with OA. Mean change in pain at short-term follow-up was significantly higher in non-flare than flare and possible flare trial designs. These results differ to previous findings [8]. The current study included an increased number of trials: whilst the earlier paper assessed 33 studies, all of which were included in the current analysis, an additional 24 trials contributed to our analysis. Furthermore, we conducted a meta-regression analysis to adjust for other variables that may have influenced the outcome and confirmed that the flare design had indeed no impact on results. Both our analysis and Trijiau et al's [8] adopted a similar definition of flare trial design; hence this was not a potential source of difference between the analyses. Similarly, the new trials included since Trijau et al's [8] meta-analysis did not differ in terms of duration, patient numbers or characteristics.

Previous studies have suggested that flare study designs may be a more efficient trial design when investigating NSAIDs in people with OA [7,8]. This has been justified through reported higher treatment effect conferred through flare designs. It was suggested that flare trial design may be valuable to assess the efficacy of a NSAID without the additional effects of other analgesics (current or recently previous) affecting outcome, to provide higher discriminant capacity, thus allowing sufficiently powerful analyses from smaller sample sizes [8]. Accordingly, such NSAIDs may be more likely to provide change in pain scores ranging from 30% to 70% which is the most sensitive change on the "S" curve of pain response. However, the current results question the value of the flare design.

The statistical analysis indicated small sample size publication bias, especially for non-flare designs, which tended to have smaller sample size, therefore more likely to produce larger effect sizes (Supplementary Figure 2). This may partially explain the reason why non-flare designs had larger pain reduction than flare designs in the short-term. Should this publication bias be excluded, it is likely that flare and non-flare designs have no difference in the short-term.

One explanation for trials which found a difference between flare and non-flare trial design may be attributed to recruitment or trial selection bias. Consideration should be given to whether flare trial designs recruit a certain phenotype of patient. It may be that flare trial designs recruit NSAID responders with a more 'inflammatory' phenotype of OA. In such instances, these participants, when ceasing their usual medications, and particularly NSAIDs, would be recruited as their pain could flare within the specified wash-out period. Conversely, those with more mechanically-related OA pain may not have the same change in pain scores on discontinuing NSAIDs, and therefore be excluded. However, they may also increase the chances of detecting a 'regression to the mean' as even if no treatment is provided, pain which has 'flared up' could naturally subside. This may therefore be considered a substantial limitation to this study design.

A second possible explanation for our findings is that participants whose pain increases following cessation of current analgesia may gain more pain relief not just from their NSAID but also from the placebo intervention. This is conceivable since participants in both trial arms in the flare study design might have an increased expectancy, a major driver of placebo/contextual response [23] through previous experience of the positive effects from their medications. Consequently there would be no difference between the two trial arms for flare-trials, i.e. no inflation of effect size calculated on the separation of treatment from the placebo intervention, compared to the difference between treatment and placebo arms in non-flare trials.

The included papers poorly documented the frequency to which their participants presented with joint effusion, or how the presence of effusion changed with stopping treatment. Maricar et al [24] found mixed results about whether clinically-detected joint effusion is a significant predictor of pain outcome

following intra-articular steroid injection in people with knee OA. Modern imaging studies suggest clinical detection of synovitis at the knee is not very accurate [25], and since synovitis is extremely common in knee OA [25] (the most prevalent joint in this analysis), it is very likely that most participants in the included studies had synovitis, though of varying degree. The accurate detection of synovitis volume or activity may in future identify a responsive subgroup to anti-inflammatory therapy within the OA population [26].

There are limitations to this work which should be considered when interpreting these findings. Firstly, the analysis was based on study-level analysis. Accordingly it was not possible to account for potential variation between patients at an individual patient data level. Secondly, whilst we adopted a clear definition of flare design based on current recommendations [7], the exact nature of the trial design was unclear in 11 papers (19%). To adjust for this potential classification-based uncertainty, we analysed 'possible flare designs' separately in a sensitivity analysis, which did not change the overall findings. Thirdly data in this analysis were only based on NSAIDs. It is therefore not possible to generalise these findings to other analgesics, which may have a different response to pain and inflammatory components to specific patient's presenting OA. Fourthly the included trials did not state which medications their participants stopped at study entry, that is, whether they stopped NSAIDs or other analgesics. If the majority of participants stopped NSAIDs, the implication is that stopping NSAID response predicts NSAID response in flare trials. This possibility is supported by a recent European survey suggesting that NSAIDs are used in nearly 60% of the OA population [5]. However, conversely a large proportion of participants (40%) would have stopped other analgesics. This subgroup may therefore have not been eligible for flare-trials, thereby potentially accounting for a difference between flare/non-flare trials. Limited information on which medications were ceased on study entry, precludes this analysis. Finally, the analyses were based on randomised controlled trial cohorts, and therefore homogenous, self-selecting populations. This loses diversity of the wider, general public, which may reduce the clinical sensitivity and generalisability to answer the research question.

To conclude, the results from this meta-analysis suggest there is no statistically significant difference in effect size in pain, function or stiffness for flare compared to non-flare trials in the assessment of

NSAID efficacy for people with OA, with some evidence indicating an increase in treatment effect detected in non-flare trial designs. Consideration should be made by industrial and non-industrial researchers on their rationale for using flare trial design, based on these results.

DECLARATION OF SOURCES OF FUNDING AND ACKNOWLEDGEMENTS

Acknowledgement: None.

Contributors: PC and WZ conceptualised the study. PC, WZ and TS contributed to the design of the study. TS, KZ and WZ conducted the analysis. All authors contributed to data collection, interpretation of research findings and manuscript writing.

Sources of Funding and **Conflicts of interest**: PGC and SRK are part-funded through the National Institute for Health Research (NIHR) though the Leeds Musculoskeletal Biomedical Research Unit. WZ had grants from Nottingham-China Scholarship, during the conduct of the study; MD reports personal fees from Ad hoc advisory boards for osteoarthritis and gout for AstraZeneca, Menarini, Nordic Biosciences, Pfizer, outside the submitted work; no other relationships or activities that could appear to have influenced the submitted work.

Role of Funding Source: Dr Zou was supported by the International Scholarship for Research Excellence from the University of Nottingham. Dr Chen was funded by the Nottingham Arthritis Research UK Pain Centre. The funding organisations did not participant in the study design, data collection, analysis, interpretation of data and report.

REFERNCES

1. Menz HB, Roddy E, Marshall M, et al. Demographic and clinical factors associated with radiographic severity of first metatarsophalangeal joint osteoarthritis: cross-sectional findings from the Clinical Assessment Study of the Foot. *Osteoarthritis Cartilage* 2015;23:77-82.

2. Turkiewicz A, Gerhardsson de Verdier M, et al. Prevalence of knee pain and knee OA in southern Sweden and the proportion that seeks medical care. *Rheumatology* 2014;54:827-835

3. March L, Smith EU, Hoy DG, et al. Burden of disability due to musculoskeletal (MSK) disorders. *Best Pract Res Clin Rheumatol* 2014;28:353-366.

4. National Institute for Health and Care Excellence (NICE). Osteoarthritis: Care and Management in Adults - 2014. Accessed on 09.12.2014. Available at: http://www.nice.org.uk/guidance/CG177

5. Kingsbury SR, Gross HJ, Isherwood G, et al. Osteoarthritis in Europe: impact on health status, work productivity and use of pharmacotherapies in five European countries. *Rheumatology* 2014; 53:937-947.

6. Kingsbury SR, Hensor EM, Walsh CA, et al. How do people with knee osteoarthritis use osteoarthritis pain medications and does this change over time? Data from the Osteoarthritis Initiative. *Arthritis Res Ther* 2013;15:R106.

7. Cooper C, Adachi JD, Bardin T, et al. How to define responders in osteoarthritis. *Curr Med Res Opin* 2013;29:719–729.

8. Trijau S, Avouac J, Escalas C, et al. Influence of flare design on symptomatic efficacy of non-steroidal anti-inflammatory drugs in osteoarthritis: a meta-analysis of randomized placebo-controlled trials. *Osteoarthritis Cartilage* 2010;18:1012-1018.

9. Altman R, Alarcón G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;34:505-514.

10. Bellamy N. WOMAC Osteoarthritis Index User Guide. Version V. Brisbane, Australia 2002

11. Juhl C, Lund H, Roos EM, et al. A hierarchy of patient-reported outcomes for meta-analysis of knee osteoarthritis trials: empirical evidence from a survey of high impact journals. *Arthritis* 2012;136245.

12. Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S (eds) Cochrane Handbook for systematic reviews of interventions. Wiley-Blackwell, Chichester, West Sussex, 2008;187-242.

13. Nüesch E, Trelle S, Reichenbach S, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ* 2010;341:c3515.

14. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-634.

15. Pincus T, Koch G, Lei H, et al. Patient Preference for Placebo, Acetaminophen (paracetamol) or Celecoxib Efficacy Studies (PACES): Two randomised, double blind, placedo controlled, crossover clinical trials in patients with knee or hip osteoarthritis. *Ann Rheum Dis* 2004;63:931-939.

16. Bourgeois P, Dreiser RL, Lequesne MG, et al. Multi-centre double-blind study to define the most favourable dose of nimesulide in terms of efficacy/safety ratio in the treatment of osteoarthritis. *Europ J Rheumatol Inflam* 1994;2:39-50.

17. Paul S, Das N, Ghosh S. The effects of aceclofenac and nabumetone in osteoarthritis. *JNMA J Nepal Med Assoc* 2009;48:121-125.

18. Truitt KE, Sperling RS, Ettinger WH, Jr., et al. A multicenter, randomized, controlled trial to evaluate the safety profile, tolerability, and efficacy of rofecoxib in advanced elderly patients with osteoarthritis. *Aging Clin* 2001;13:112-121.

19. Bensen WG, Fiechtner JJ, McMillen JI, et al. Treatment of osteoarthritis with celecoxib, a cyclooxygenase-2 inhibitor: a randomized controlled trial. *Mayo Clinic Proceedings.* 1999;74:1095-1105.

20. Wittenberg RH, Schell E, Krehan G, et al. First-dose analgesic effect of the cyclo-oxygenase-2 selective inhibitor lumiracoxib in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled comparison with celecoxib [NCT00267215]. *Arthritis Res Therap* 2006;8:R35.

21. Altman RD, Moskowitz R. Intraarticular sodium hyaluronate (Hyalgan) in the treatment of patients with osteoarthritis of the knee: A randomized clinical trial. *J Rheumatol* 1998;25:2203-2212.

22. Nunes CP, De Oliveira PC, De Oliveira JM, et al. A double-blind, comparative, placebo-controlled study in two arms of the safety and efficacy of the anti-inflammatory and analgesic action of the association of cyanocobalamin, pyridoxine chlorihydrate, thiamine mononitrate and diclofenac sodium in tablets, in patients with osteoarthritis. *Revista Brasil Med* 2005;62:486-491.

23. Doherty M, Dieppe PA. The "placebo" response in osteoarthritis and its implications for clinical practice. *Osteoarthritis Cartilage* 2009; 17:1255-62.

24. Maricar N, Callaghan MJ, Felson DT, et al Predictors of response to intra-articular steroid injections in knee osteoarthritis--a systematic review. *Rheumatology* 2013;52:1022-1032.

25. Conaghan P, D'Agostino MA, Ravaud P, et al. EULAR report on the use of ultrasonography in painful knee osteoarthritis. Part 2: exploring decision rules for clinical utility. *Ann Rheum Dis* 2005;64:1710-1714.

26. Keen HI, Hensor EM, Wakefield RJ et al. Ultrasound assessment of response to intra-articular therapy in osteoarthritis of the knee. *Rheumatology* 2015;54:1385-1391.

27. Roemer FW, Kassim Javaid M, Guermazi A, et al. Anatomical distribution of synovitis in knee osteoarthritis and its association with joint effusion assessed on non-enhanced and contrast-enhanced MRI. *Osteoarthritis Cartilage* 2010;18:1269-1274.

28. Bocanegra TS, Weaver AL, Tindall EA, et al. Diclofenac/misoprostol compared with diclofenac in the treatment of osteoarthritis of the knee or hip: a randomized, placebo controlled trial. Arthrotec Osteoarthritis Study Group. *J Rheumatol* 1998;25:1602-1611.

29. Ehrich EW, Schnitzer TJ, McIlwain H, et al. Effect of specific COX-2 inhibition in osteoarthritis of the knee: A 6 week double blind, placebo controlled pilot study of rofecoxib. *J Rheumatol* 1999;26:2438-2447.

30. Day R, Morrison B, Luza A, et al. A randomized trial of the efficacy and tolerability of the COX-2 inhibitor rofecoxib vs ibuprofen in patients with osteoarthritis. *Arch Intern Med* 2000;160:1781-1787.

31. Ehrich EW, Bolognese JA, Watson DJ, et al. Effect of rofecoxib therapy on measures of health-related quality of life in patients with osteoarthritis. *Am J Managed Care* 2001;7:609-616.

32. McKenna F, Borenstein D, Wendt H, et al. Celecoxib versus diclofenac in the management of osteoarthritis of the knee: A placebo-controlled, randomised, double-blind comparison. *Scand J Rheumatol* 2001;30:11-18.

33. Williams GW, Hubbard RC, Yu SS, et al. Comparison of once-daily and twice-daily administration of celecoxib for the treatment of osteoarthritis of the knee. *Clin Therap* 2001;23:213-227.

34. Leung AT, Malmstrom K, Gallacher AE, et al. Efficacy and tolerability profile of etoricoxib in patients with osteoarthritis: A randomized, double-blind, placebo and active-comparator controlled 12-week efficacy trial. *Curr Med Res Opin* 2002;18:49-58.

35. Grifka JK, Zacher J, Brown JP, et al. Efficacy and tolerability of lumiracoxib versus placebo in patients with osteoarthritis of the hand. *Clin Experiment Rheumatol* 2004;22:589-596.

36. Fleischmann R, Sheldon E, Maldonado-Cocco J, et al. Lumiracoxib is effective in the treatment of osteoarthritis of the knee: a prospective randomized 13-week study versus placebo and celecoxib. *Clin Rheumatol* 2005;25:42-53.

37. Wiesenhutter CW, Boice JA, Ko A, et al. Evaluation of the comparative efficacy of etoricoxib and ibuprofen for treatment of patients with osteoarthritis: A randomized, double-blind, placebo-controlled trial. *Mayo Clinic Proceedings* 2005;80:470-479.

38. Bingham ICO, Sebba AI, Rubin BR, 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;46:496-507.

39. Wittenberg RH, Schell E, Krehan G, et al. First-dose analgesic effect of the cyclo-oxygenase-2 selective inhibitor lumiracoxib in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled comparison with celecoxib [NCT00267215]. *Arthritis Res Therap* 2006;8:R35.

40. Puopolo A, Boice JA, Fidelholtz JL, et al. A randomized placebo-controlled trial comparing the efficacy of etoricoxib 30 mg and ibuprofen 2400 mg for the treatment of patients with osteoarthritis. *Osteoarthritis Cartilage* 2007;15:1348-1356.

41. Reginster Y, Malmstrom K, Mehta A, et al. Evaluation of the efficacy and safety of etoricoxib compared with naproxen in two, 138-week randomised studies of patients with osteoarthritis. *Ann Rheum Dis* 2007;66:945-951.

42. Rother M, Lavins BJ, Kneer W, et al. Efficacy and safety of epicutaneous ketoprofen in Transfersome (IDEA-033) versus oral cefecoxib and placebo in osteoarthritis of the knee: Multicentre randomised controlled trial. *Ann Rheum Dis* 2007;66:1178-1183.

43. Karlsson J, Pivodic A, Aguirre D, et al. Efficacy, safety, and tolerability of the cyclooxygenase-inhibiting nitric oxide donator naproxcinod in treating osteoarthritis of the hip or knee. *J Rheumatol* 2009;36:1290-1297.

44. Simon LS, Grierson LM, Naseer Z, et al. Efficacy and safety of topical diclofenac containing dimethyl sulfoxide (DMSO) compared with those of topical placebo, DMSO vehicle and oral diclofenac for knee osteoarthritis. *Pain* 2009;143:238-245.

45. Baerwald C, Verdecchia P, Duquesroix B, et al. Efficacy, safety, and effects on blood pressure of naproxcinod 750 mg twice daily compared with placebo and naproxen 500 mg twice daily in patients with osteoarthritis of the hip: a randomized, double-blind, parallel-group, multicenter study. *Arthritis Rheum* 2010;62:3635-3644.

46. Cryer BL, Sostek MB, Fort JG, et al. A fixed-dose combination of naproxen and esomeprazole magnesium has comparable upper gastrointestinal tolerability to celecoxib in patients with osteoarthritis of the knee: results from two randomized, parallel-group, placebo-controlled trials. *Ann Med* 2011;43:594-605.

47. Schnitzer TJ, Dattani ID, Seriolo B, et al. A 13-week, multicenter, randomized, double-blind study of lumiracoxib in hip osteoarthritis. *Clin Rheumatol* 2011;30:1433-1446.

48. Essex MN, O'Connell M, Bhadra Brown P. Response to nonsteroidal antiinflammatory drugs in African Americans with osteoarthritis of the knee. *J Internat Med Res* 2012;40:2251-2266.

49. El-Mehairy MM, Shaker A, Bahgat NE, et al. A double-blind comparison of niflumic acid with phenylbutazone, oxyphenylbutazone and placebo in the treatment of osteoarthrosis. *Rheumatol Rehabil* 1974;13:198-203.

50. Shipley M, Berry H, Broster G, et al. Controlled trial of homoeopathic treatment of osteoarthritis. *Lancet* 1983;1:97-98.

51. Lee P, Davis P, Prat A. The efficacy of diflunisal in osteoarthritis of the knee. A Canadian multicenter study. *J Rheumatol* 1985;12:544-548.

52. Dreiser RL, Gersberg M, Thomas F, et al. Ibuprofen 800 mg for the treatment of osteoarthritis of the interphalangeal joints of the hand or trapezo metacarpal joint. *Revue du Rhumatisme* 1993;60:719-724.

53. Nguyen M, Dougados M, Berdah L, et al. Diacerhein in the treatment of osteoarthritis of the hip. *Arthritis Rheumat* 1994;37:529-536.

54. Sandelin J, Harilainen A, Crone H, et al. Local NSAID gel (Eltenac) in the treatment of osteoarthritis of the knee: A double blind study comparing eltenac with oral diclofenac and placebo gel. *Scand J Rheumatol* 1997;26:287-292.

55. Lund B, Distel M, Bluhmki E. A double-blind, randomized, placebo-controlled study of efficacy and tolerance of meloxicam treatment in patients with osteoarthritis of the knee. *Scand J Rheumatol* 1998; 27:32-37.

56. Davies GM, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the Western Ontario and McMaster Universities Osteoarthritis Index and the Short-Form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. *Arthritis Care Res* 1999;12:172-179.

57. Schmitt W, Walter K, Kurth HJ. Clinical trial on the efficacy and safety of different diclofenac formulations: Multiple-unit formulations compared to enteric coated tablets in patients with activated osteoarthritis. *Inflammopharmacol* 1999;7:363-375.

58. Goldstein DJ, Wang O, Todd LE, et al. Study of the analgesic effect of lanepitant in patients with osteoarthritis pain. *Clin Pharmacol Therapeut* 2000;67:419-426.

59. Scott DL, Berry H, Capell H, et al. The long-term effects of non-steroidal anti-inflammatory drugs in osteoarthritis of the knee: a randomized placebo-controlled trial. *Rheumatology* 2000;39:1095-1101.

60. Dickson DJ, Hosie G, English JR. A double-blind, placebo-controlled comparison of hylan G-F 20 against diclofenac in knee osteoarthritis. *J Drug Assess* 2001;4(Part 3):179-190.

61. Uzun H, Tuzun S, Ozaras N, et al. The effect of flurbiprofen and tiaprofenic acid on serum cytokine levels of patients with osteoarthrosis. *Acta Orthop Scand* 2001;72:499-502.

62. Kivitz A, Eisen G, Zhao WW, et al. Randomized placebo-controlled trial comparing efficacy and safety of valdecoxib with naproxen in patients with osteoarthritis. *J Family Pract* 2002:530-537.

63. Case JP, Baliunas AJ, Block JA. Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis: a randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium. *Arch Intern Med* 2003;163:169-178.

64. Gibofsky A, Williams GW, McKenna F, et al. Comparing the efficacy of cyclooxygenase 2-specific inhibitors in treating osteoarthritis: appropriate trial design considerations and results of a randomized, placebo-controlled trial. *Arthritis Rheum* 2003;48:3102-3111.

65. Biegert C, Wagner I, Lüdtke R, et al. Efficacy and safety of willow bark extract in the treatment of osteoarthritis and rheumatoid arthritis: results of 2 randomized double-blind controlled trials. *J Rheumat* 2004;31:2121-2130.

66. Schnitzer TJ, Beier J, Geusens P, et al. Efficacy and safety of four doses of lumiracoxib versus diclofenac in patients with knee or hip primary osteoarthritis: a phase II, four-week, multicenter, randomized, double-blind, placebo-controlled trial. *Arthritis Rheum 2*004;51:549-557.

67. Tannenbaum H, Berenbaum F, Reginster JY, et al. Lumiracoxib is effective in the treatment of osteoarthritis of the knee: a 13 week, randomised, double blind study versus placebo and celecoxib. *Ann Rheum Dis* 2004:1419-1426.

68. Haghighi M, Khalvat A, Toliat T, et al. Comparing the effects of ginger (Zingiber officinale) xtract and ibuprofen on patients with osteoarthritis. *Arch Iran Med* 2005;4:267-271.

69. Sheldon E, Beaulieu A, Paster Z, et al. Efficacy and tolerability of lumiracoxib in the treatment of osteoarthritis of the knee: a 13-week, randomized, double-blind comparison with celecoxib and placebo. *Clin Therapeut* 2005;27:64-77.

70. Tuzun S, Uzun H, Aydin S, et al. Effects of flurbiprofen and tiaprofenic acid on oxidative stress markers in osteoarthritis: A prospective, randomized, open-label, active- and placebo-controlled trial. *Curr Therapeut Res-Clin Experiment* 2005;66:335-344.

71. Svensson O, Malmenas M, Fajutrao L, et al. Greater reduction of knee than hip pain in osteoarthritis treated with naproxen, as evaluated by WOMAC and SF-36. *Ann Rheum Dis* 2006;65:781-784.

72. Dougados M, Moore A, Yu S, et al. 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 Res Therap* 2007;9:R11.

73. Kruger K, Klasser M, Mossinger J, et al. Oxaceprol--a randomised, placebo-controlled clinical study in osteoarthritis with a non-conventional non-steroidal anti-inflammatory drug. *Clin Exp Rheumatol* 2007;25:29-34.

74. Ding M-h, Zhang H, Li Y. A randomized controlled study on warming needle moxibustion for treatment of knee osteoarthritis. *Zhongguo zhenjiu* 2009;29:603-607.

75. DeLemos BP, Xiang J, Benson C, 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. *Am J Therapeut* 2011;18:216-226.

76. Petersen SG, Beyer N, Hansen M, et al. Nonsteroidal Anti-Inflammatory Drug or Glucosamine Reduced Pain and Improved Muscle Strength With Resistance Training in a Randomized Controlled Trial of Knee Osteoarthritis Patients. *Arch Phys Med Rehabil* 2011;92:1185-1193.

77. Schnitzer TJ, Hochberg MC, Marrero CE, et al. Efficacy and safety of naproxcinod in patients with osteoarthritis of the knee: a 53-week prospective randomized multicenter study. *Sem Arthritis Rheum* 2011;40:285-297.

FIGURE AND TABLE LEGENDS

Figure 1: Forest plot of mean change in pain score for NSAID versus Placebo for flare versus non-flare trial design.

Figure 2: Forest plot of absolute pain score for NSAID versus Placebo for flare versus non-flare trial design.

Table 1: Flare versus non-flare trial design meta-analysis results by outcome measure.

Table 2: Flare versus non-flare trial design meta-analysis results as assessed by immediate, short- and longer-term follow-up intervals.

Table 3: Flare and possible flare versus non-flare trial design meta-analysis results as assessed by immediate, short, and longer-term follow-up intervals.

Table 4: Meta-regression of effect size of NSAIDs for osteoarthritis pain (number of observation=131)

Supplementary Figure 1: PRISMA flow-chart.

Supplementary Figure 2: Funnel plot assessing small sample size publication bias for primary outcome measure (mean change in pain score) for flare versus non-flare trial design.

Supplementary Table 1: MEDLINE search strategy

Supplementary Table 2: Study Characteristics (Study design)

Supplementary Table 3: Participant characteristics of the included studies (medications and demographics)

Supplementary Table 4: Summary of the included study quality assessment results

Figure 1: Forest plot of mean change in pain score for NSAID versus Placebo for flare versus non-flare trial design.

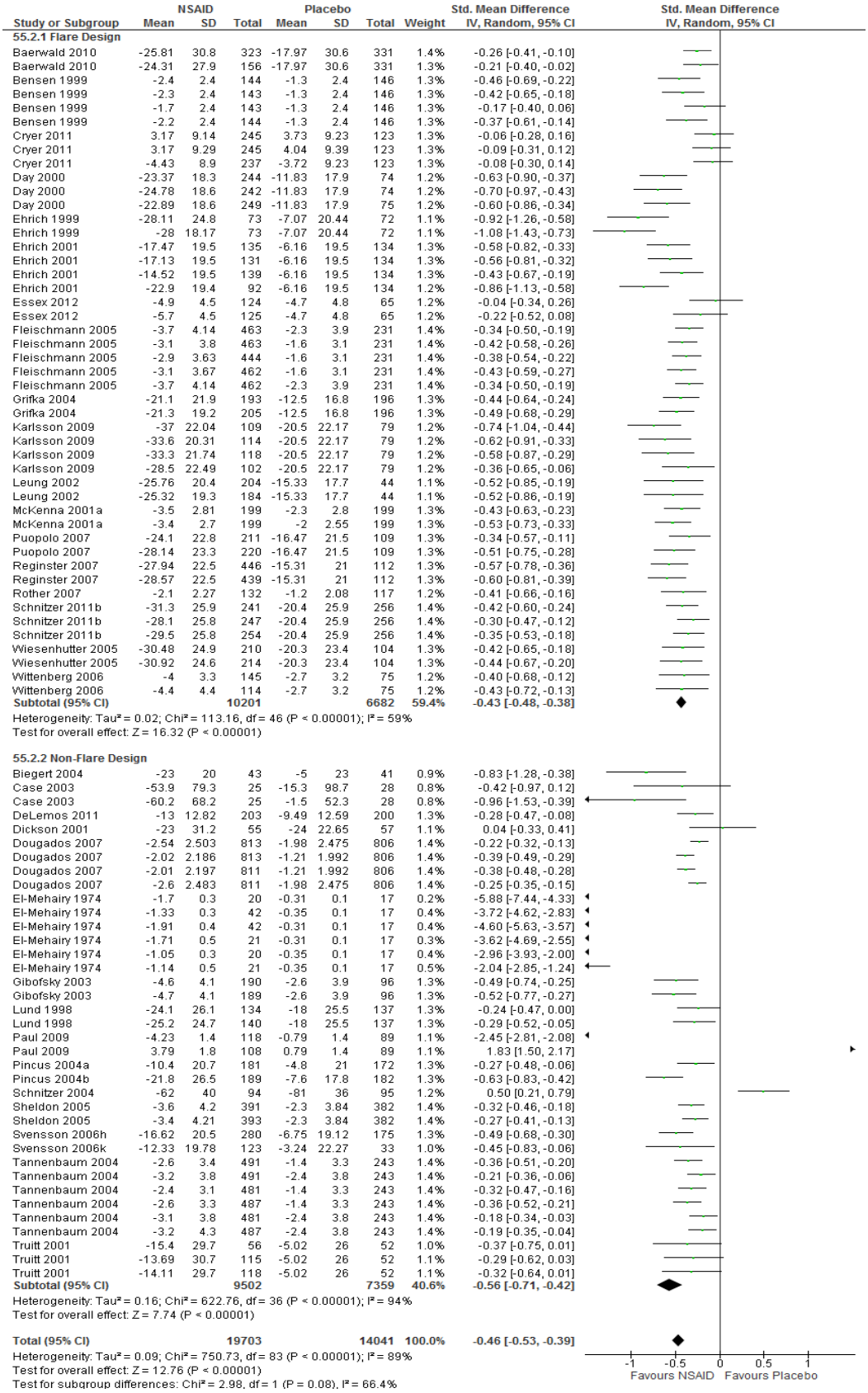

NSAID
Placebo
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
55.2.1 Flare Design
Baerwald 2010 -25.81 30.8 323 -17.97 30.6 331 1.4% -0.26 [-0.41, -0.10]
Baerwald 2010 -24.31 27.9 156 -17.97 30.6 331 1.3% -0.21 [-0.40, -0.02]
Bensen 1999 -2.4 2.4 144 -1.3 2.4 146 1.3% -0.46 [-0.69, -0.22]
Bensen 1999 -2.3 2.4 143 -1.3 2.4 146 1.3% -0.42 [-0.65, -0.18]
Bensen 1999 -1.7 2.4 143 -1.3 2.4 146 1.3% -0.17 [-0.40, 0.06]
Bensen 1999 -2.2 2.4 144 -1.3 2.4 146 1.3% -0.37 [-0.61, -0.14]
Cryer 2011 3.17 9.14 245 3.73 9.23 123 1.3% -0.06 [-0.28, 0.16]
Cryer 2011 3.17 9.29 245 4.04 9.39 123 1.3% -0.09 [-0.31, 0.12]
Cryer 2011 -4.43 8.9 237 -3.72 9.23 123 1.3% -0.08 [-0.30, 0.14]
Day 2000 -23.37 18.3 244 -11.83 17.9 74 1.2% -0.63 [-0.90, -0.37]
Day 2000 -24.78 18.6 242 -11.83 17.9 74 1.2% -0.70 [-0.97, -0.43]
Day 2000 -22.89 18.6 249 -11.83 17.9 75 1.2% -0.60 [-0.86, -0.34]
Ehrich 1999 -28.11 24.8 73 -7.07 20.44 72 1.1% -0.92 [-1.26, -0.58]
Ehrich 1999 -28 18.17 73 -7.07 20.44 72 1.1% -1.08 [-1.43, -0.73]
Ehrich 2001 -17.47 19.5 135 -6.16 19.5 134 1.3% -0.58 [-0.82, -0.33]
Ehrich 2001 -17.13 19.5 131 -6.16 19.5 134 1.3% -0.56 [-0.81, -0.32]
Ehrich 2001 -14.52 19.5 139 -6.16 19.5 134 1.3% -0.43 [-0.67, -0.19]
Ehrich 2001 -22.9 19.4 92 -6.16 19.5 134 1.2% -0.86 [-1.13, -0.58]
Essex 2012 -4.9 4.5 124 -4.7 4.8 65 1.2% -0.04 [-0.34, 0.26]
Essex 2012 -5.7 4.5 125 -4.7 4.8 65 1.2% -0.22 [-0.52, 0.08]
Fleischmann 2005 -3.7 4.14 463 -2.3 3.9 231 1.4% -0.34 [-0.50, -0.19]
Fleischmann 2005 -3.1 3.8 463 -1.6 3.1 231 1.4% -0.42 [-0.58, -0.26]
Fleischmann 2005 -2.9 3.63 444 -1.6 3.1 231 1.4% -0.38 [-0.54, -0.22]
Fleischmann 2005 -3.1 3.67 462 -1.6 3.1 231 1.4% -0.43 [-0.59, -0.27]
Fleischmann 2005 -3.7 4.14 462 -2.3 3.9 231 1.4% -0.34 [-0.50, -0.19]
Grifka 2004 -21.1 21.9 193 -12.5 16.8 196 1.3% -0.44 [-0.64, -0.24]
Grifka 2004 -21.3 19.2 205 -12.5 16.8 196 1.3% -0.49 [-0.68, -0.29]
Karlsson 2009 -37 22.04 109 -20.5 22.17 79 1.2% -0.74 [-1.04, -0.44]
Karlsson 2009 -33.6 20.31 114 -20.5 22.17 79 1.2% -0.62 [-0.91, -0.33]
Karlsson 2009 -33.3 21.74 118 -20.5 22.17 79 1.2% -0.58 [-0.87, -0.29]
Karlsson 2009 -28.5 22.49 102 -20.5 22.17 79 1.2% -0.36 [-0.65, -0.06]
Leung 2002 -25.76 20.4 204 -15.33 17.7 44 1.1% -0.52 [-0.85, -0.19]
Leung 2002 -25.32 19.3 184 -15.33 17.7 44 1.1% -0.52 [-0.86, -0.19]
McKenna 2001a -3.5 2.81 199 -2.3 2.8 199 1.3% -0.43 [-0.63, -0.23]
McKenna 2001a -3.4 2.7 199 -2 2.55 199 1.3% -0.53 [-0.73, -0.33]
Puopolo 2007 -24.1 22.8 211 -16.47 21.5 109 1.3% -0.34 [-0.57, -0.11]
Puopolo 2007 -28.14 23.3 220 -16.47 21.5 109 1.3% -0.51 [-0.75, -0.28]
Reginster 2007 -27.94 22.5 446 -15.31 21 112 1.3% -0.57 [-0.78, -0.36]
Reginster 2007 -28.57 22.5 439 -15.31 21 112 1.3% -0.60 [-0.81, -0.39]
Rother 2007 -2.1 2.27 132 -1.2 2.08 117 1.2% -0.41 [-0.66, -0.16]
Schnitzer 2011b -31.3 25.9 241 -20.4 25.9 256 1.4% -0.42 [-0.60, -0.24]
Schnitzer 2011b -28.1 25.8 247 -20.4 25.9 256 1.4% -0.30 [-0.47, -0.12]
Schnitzer 2011b -29.5 25.8 254 -20.4 25.9 256 1.4% -0.35 [-0.53, -0.18]
Wiesenhutter 2005 -30.48 24.9 210 -20.3 23.4 104 1.3% -0.42 [-0.65, -0.18]
Wiesenhutter 2005 -30.92 24.6 214 -20.3 23.4 104 1.3% -0.44 [-0.67, -0.20]
Wittenberg 2006 -4 3.3 145 -2.7 3.2 75 1.2% -0.40 [-0.68, -0.12]
Wittenberg 2006 -4.4 4.4 114 -2.7 3.2 75 1.2% -0.43 [-0.72, -0.13]
Subtotal (95% CI) 10201 6682 59.4% -0.43 [-0.48, -0.38]
Heterogeneity: Tau² = 0.02; Chi² = 113.16, df = 46 (P < 0.00001); I² = 59%
Test for overall effect: Z = 16.32 (P < 0.00001)
55.2.2 Non-Flare Design
Biegert 2004 -23 20 43 -5 23 41 0.9% -0.83 [-1.28, -0.38]
Case 2003 -53.9 79.3 25 -15.3 98.7 28 0.8% -0.42 [-0.97, 0.12]
Case 2003 -60.2 68.2 25 -1.5 52.3 28 0.8% -0.96 [-1.53, -0.39]
DeLemos 2011 -13 12.82 203 -9.49 12.59 200 1.3% -0.28 [-0.47, -0.08]
Dickson 2001 -23 31.2 55 -24 22.65 57 1.1% 0.04 [-0.33, 0.41]
Dougados 2007 -2.54 2.503 813 -1.98 2.475 806 1.4% -0.22 [-0.32, -0.13]
Dougados 2007 -2.02 2.186 813 -1.21 1.992 806 1.4% -0.39 [-0.49, -0.29]
Dougados 2007 -2.01 2.197 811 -1.21 1.992 806 1.4% -0.38 [-0.48, -0.28]
Dougados 2007 -2.6 2.483 811 -1.98 2.475 806 1.4% -0.25 [-0.35, -0.15]
El-Mehairy 1974 -1.7 0.3 20 -0.31 0.1 17 0.2% -5.88 [-7.44, -4.33]
El-Mehairy 1974 -1.33 0.3 42 -0.35 0.1 17 0.4% -3.72 [-4.62, -2.83]
El-Mehairy 1974 -1.91 0.4 42 -0.31 0.1 17 0.4% -4.60 [-5.63, -3.57]
El-Mehairy 1974 -1.71 0.5 21 -0.31 0.1 17 0.3% -3.62 [-4.69, -2.55]
El-Mehairy 1974 -1.05 0.3 20 -0.35 0.1 17 0.4% -2.96 [-3.93, -2.00]
El-Mehairy 1974 -1.14 0.5 21 -0.35 0.1 17 0.5% -2.04 [-2.85, -1.24]
Gibofsky 2003 -4.6 4.1 190 -2.6 3.9 96 1.3% -0.49 [-0.74, -0.25]
Gibofsky 2003 -4.7 4.1 189 -2.6 3.9 96 1.3% -0.52 [-0.77, -0.27]
Lund 1998 -24.1 26.1 134 -18 25.5 137 1.3% -0.24 [-0.47, 0.00]
Lund 1998 -25.2 24.7 140 -18 25.5 137 1.3% -0.29 [-0.52, -0.05]
Paul 2009 -4.23 1.4 118 -0.79 1.4 89 1.1% -2.45 [-2.81, -2.08]
Paul 2009 3.79 1.8 108 0.79 1.4 89 1.1% 1.83 [1.50, 2.17]
Pincus 2004a -10.4 20.7 181 -4.8 21 172 1.3% -0.27 [-0.48, -0.06]
Pincus 2004b -21.8 26.5 189 -7.6 17.8 182 1.3% -0.63 [-0.83, -0.42]
Schnitzer 2004 -62 40 94 -81 36 95 1.2% 0.50 [0.21, 0.79]
Sheldon 2005 -3.6 4.2 391 -2.3 3.84 382 1.4% -0.32 [-0.46, -0.18]
Sheldon 2005 -3.4 4.21 393 -2.3 3.84 382 1.4% -0.27 [-0.41, -0.13]
Svensson 2006h -16.62 20.5 280 -6.75 19.12 175 1.3% -0.49 [-0.68, -0.30]
Svensson 2006k -12.33 19.78 123 -3.24 22.27 33 1.0% -0.45 [-0.83, -0.06]
Tannenbaum 2004 -2.6 3.4 491 -1.4 3.3 243 1.4% -0.36 [-0.51, -0.20]
Tannenbaum 2004 -3.2 3.8 491 -2.4 3.8 243 1.4% -0.21 [-0.36, -0.06]
Tannenbaum 2004 -2.4 3.1 481 -1.4 3.3 243 1.4% -0.32 [-0.47, -0.16]
Tannenbaum 2004 -2.6 3.3 487 -1.4 3.3 243 1.4% -0.36 [-0.52, -0.21]
Tannenbaum 2004 -3.1 3.8 481 -2.4 3.8 243 1.4% -0.18 [-0.34, -0.03]
Tannenbaum 2004 -3.2 4.3 487 -2.4 3.8 243 1.4% -0.19 [-0.35, -0.04]
Truitt 2001 -15.4 29.7 56 -5.02 26 52 1.0% -0.37 [-0.75, 0.01]
Truitt 2001 -13.69 30.7 115 -5.02 26 52 1.1% -0.29 [-0.62, 0.03]
Truitt 2001 -14.11 29.7 118 -5.02 26 52 1.1% -0.32 [-0.64, 0.01]
Subtotal (95% CI) 9502 7359 40.6% -0.56 [-0.71, -0.42]
Heterogeneity: Tau² = 0.16; Chi² = 622.76, df = 36 (P < 0.00001); I² = 94%
Test for overall effect: Z = 7.74 (P < 0.00001)
Total (95% CI) 19703 14041 100.0% -0.46 [-0.53, -0.39]
Heterogeneity: Tau² = 0.09; Chi² = 750.73, df = 83 (P < 0.00001); I² = 89%
Test for overall effect: Z = 12.76 (P < 0.00001)
Test for subgroup differences: Chi² = 2.98, df = 1 (P = 0.08), I² = 66.4%
-1 -0.5 0 0.5 1
Favours NSAID Favours Placebo

Figure 2: Forest plot of absolute pain score for NSAID versus Placebo for flare versus non-flare trial design.

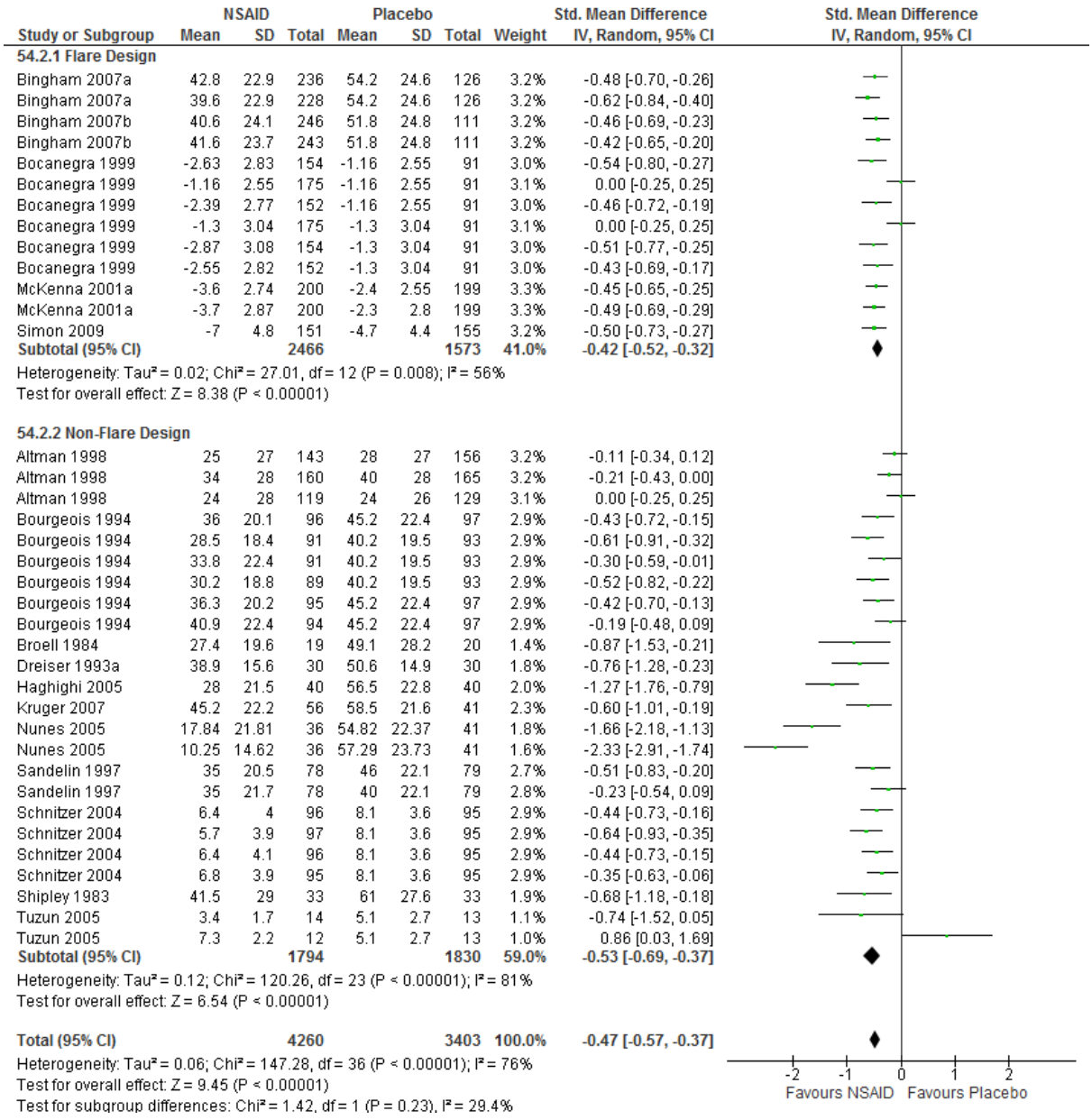


Table 1: Flare versus non-flare trial design meta-analysis results by outcome measure.

Outcome	**N**	**Flare Trial Design**		**N**	**Non-Flare Trial Design**		**Difference between flare and non-flare (Chi^2; I^2)**
		Effect size; 95% CI	**Statistical Heterogeneity (Chi^2 P-value/I^2%)**		**Effect size; 95% CI**	**Statistical Heterogeneity (Chi^2 P-value/I^2%)**	
Absolute Pain	13	-0.42 [-0.52, -0.32]	P = 0.008; I^2 = 56%	24	-0.53 [-0.69, -0.37]	P < 0.00001; I^2 = 81%	P = 0.23; I^2 = 29.4%
Mean change in pain score	48	-0.43 [-0.48, -0.38]	P < 0.00001; I^2 = 59%	37	-0.56 [-0.71, -0.42]	P < 0.00001; I^2 = 94%	P = 0.08; I^2 = 66.4%
Absolute functional score	10	-0.40 [-0.47, -0.33]	P = 0.28; I^2 = 18%	8	-0.13 [-0.43, 0.18]	P < 0.00001; I^2 = 85%	P = 0.08; I^2 = 67.4%
Mean change in functional score	38	-0.51 [-0.61, -0.41]	P < 0.00001; I^2 = 87%	39	-0.58 [-0.67, -0.48]	P < 0.00001; I^2 = 86%	P = 0.34; I^2 = 0%
Absolute stiffness score	2	-0.48 [-0.67, -0.30]	P = 0.23; I^2 = 31%	7	-0.49 [-0.66, -0.31]	P = 0.22; I^2 = 27%	P = 1.00; I^2 = 0%
Mean change in stiffness score	27	-0.38 [-0.44, -0.31]	P = 0.0005; I^2 = 54%	16	-0.35 [-0.50, -0.20]	P < 0.00001; I^2 = 84%	P = 0.75; I^2 = 0%

CI – confidence intervals; I^2 – inconsistency value; vs. - versus

Table 2: Flare versus non-flare trial design meta-analysis results as assessed by immediate, short, and longer-term follow-up intervals.

Outcome	Follow-up interval (weeks)	N	Flare Trial Design		N	Non-Flare Trial Design		Difference between flare and non-flare (Chi^2; I^2)
			Effect size [95% CI]	Statistical Heterogeneity (Chi^2 P-value/I^2%)		Effect size [95% CI]	Statistical Heterogeneity (Chi^2 P-value/I^2%)	
Absolute Pain	0 - 1	0	N/E	N/E	8	-0.56 [-0.82, -0.31]	P < 0.0001; 78%	N/E
	2 – 4	4	-0.36 [-0.59, -0.13]	P = 0.01; 72%	15	-0.55 [-0.77, -0.34]	P < 0.00001; 82%	P = 0.24; 28.9%
	6 - over	9	-0.44 [-0.55, -0.34]	P = 0.06; 47%	2	-0.15 [-0.48, 0.18]	P = 0.11; 60%	P = 0.10; 62.1%
Mean change in pain score	0 - 1	4	-0.35 [-0.48, -0.23]	P = 0.31; 16%	0	N/E	N/E	N/E
	2 – 4	8	-0.44 [-0.50, -0.37]	P = 0.97; 0%	14	-0.69 [-1.01, -0.37]	P < 0.00001; 97%	P = 0.13; 56.5%
	6 - over	35	-0.44 [-0.51, -0.37]	P < 0.00001; 68%	23	-0.47 [-0.60, -0.34]	P < 0.00001; 88%	P = 0.71; 0.0%
Absolute functional score	0 - 1	0	N/E	N/E	2	0.77 [-1.25, 2.79]	P < 0.00001; 96%	N/E
	2 – 4	2	-0.34 [-0.47, -0.21]	P = 0.76; 0%	6	-0.32 [-0.54, -0.10]	P = 0.01; 67%	P = 0.85; 0.0%
	6 - over	8	-0.43 [-0.51, -0.34]	P = 0.20; 28%	0	N/E	N/E	N/E
Mean change in functional score	0 - 1	4	-0.26 [-0.38, -0.15]	P = 0.59; 0%	3	-0.47 [-0.64, -0.31]	P = 0.40; 0%	P = 0.04; 75.6%*
	2 – 4	7	-0.86 [-1.29, -0.42]	P < 0.00001; 97%	14	-0.68 [-0.85, -0.50]	P < 0.00001; 87%	P = 0.45; 0.0%
	6 - over	27	-0.46 [-0.52, -0.39]	P = 0.0003; 56%	22	-0.55 [-0.68, -0.42]	P < 0.0001; 86%	P = 0.23; 31.4%
Absolute stiffness score	0 - 1	0	N/E	N/E	1	-0.84 [-1.50, -0.18]	P = 0.01; N/E	N/E
	2 – 4	0	N/E	N/E	6	-0.46 [-0.63, -0.29]	P = 0.23; 27%	N/E
	6 - over	2	-0.48 [-0.67, -0.30]	P = 0.23; 31%	0	N/E	N/E	N/E
Mean change in stiffness score	0 - 1	0	N/E	N/E	0	N/E	N/E	N/E
	2 – 4	7	-0.31 [-0.38, -0.24]	P = 0.94; 0%	4	-0.11 [-0.42, 0.20]	P < 0.0001; 86%	P = 0.21; 37.2%

	6 - over	20	-0.41 [-0.50, -0.32]	P < 0.0001; 64%	12	-0.43 [-0.61, -0.26]	P < 0.0001; 83%	P = 0.81; 0.0%

* - signified analysis reach a statistical significant difference; CI – confidence intervals; I2 – inconsistency value; N – number of studies; N/E – Not estimatable; vs. - versus

Table 3: Flare and possible flare versus non-flare trial design meta-analysis results as assessed by immediate, short and longer-term follow-up intervals.

Outcome	Follow-up interval (weeks)	N	Flare and Possible Flare Trial Design		N	Not Flare Trial Design		Difference between flare and possible flare to non-flare (Chi^2; I^2)
			Effect size [95% CI]	Statistical Heterogeneity (Chi^2 P-value/I^2%)		Effect size [95% CI]	Statistical Heterogeneity (Chi^2 P-value/I^2%)	
Absolute Pain	0 - 1	2	-0.25 [-1.21, 0.70]	P = 0.004; 88%	8	-0.56 [-0.82, -0.31]	P < 0.0001; 78%	P = 0.54; 0%
	2 – 4	11	-0.39 [-0.59, -0.20]	P < 0.0001; 75%	15	-0.55 [-0.77, -0.34]	P < 0.00001; 82%	P = 0.28; 14.4%
	6 - over	6	-0.85 [-1.15, -0.55]	P < 0.00001; 97%	2	-0.40 [-0.74, -0.06]	P = 0.91; 0%	P = 0.05; 74.6%*
Mean change in pain score	0 - 1	1	-0.52 [-0.99, -0.05]	P = 0.03; N/E	0	N/E	N/E	N/E
	2 – 4	8	-0.36 [-0.46, -0.26]	P = 0.11; 41%	14	-0.69 [-1.01, -0.37]	P < 0.00001; 97%	P = 0.05; 73.3%*
	6 - over	20	-0.34 [-0.40, -0.27]	P = 0.02; 44%	23	-0.47 [-0.60, -0.33]	P < 0.00001; 89%	P = 0.08; 67.5%
Absolute functional score	0 - 1	1	0.20 [-0.20, 0.61]	P = 0.32	2	0.77 [-1.25, 2.79]	P < 0.00001; 96%	P = 0.59; 0%
	2 – 4	5	-0.30 [-0.55, -0.06]	P = 0.06; 56%	6	-0.32 [-0.54, -0.10]	P = 0.01; 67%	P = 0.95; 0%
	6 – over	5	-0.28 [-0.57, 0.00]	P = 0.02; 65%	0	N/E	N/E	N/E
Mean change in functional score	0 - 1	0	N/E	N/E	3	-0.47 [-0.64, -0.31]	P = 0.40; 0%	N/E
	2 – 4	4	-0.28 [-0.38, -0.19]	P = 0.49; 0%	14	-0.68 [-0.85, -0.50]	P < 0.00001; 87%	P < 0.01; 93.5%*
	6 - over	16	-0.35 [-0.42, -0.28]	P = 0.01; 50%	22	-0.55 [-0.68, -0.42]	P < 0.00001; 86%	P = 0.01; 84.9%*
Absolute stiffness score	0 - 1	1	0.22 [-0.18, 0.62]	P = 0.28	1	-0.84 [-1.50, -0.18]	P = 0.01	P = 0.01; 86.3%*
	2 – 4	4	-0.61 [-1.47, 0.25]	P < 0.0001; 87%	6	-0.46 [-0.63, -0.29]	P = 0.23; 27%	P = 0.74; 0%
	6 - over	0	N/E	N/E	0	N/E	N/E	N/E
Mean change in stiffness score	0 - 1	0	N/E	N/E	0	N/E	N/E	N/E
	2 – 4	7	-0.31 [-0.38, -0.24]	P = 0.94; 0%	4	-0.11 [-0.42, 0.20]	P < 0.0001; 86%	P = 0.21; 37.2%

	6 - over	8	-0.61 [-0.74, -0.48]	P < 0.00001; 84%	12	-0.59 [-0.73, -0.45]	P < 0.00001; 89%	P = 0.83; 0%

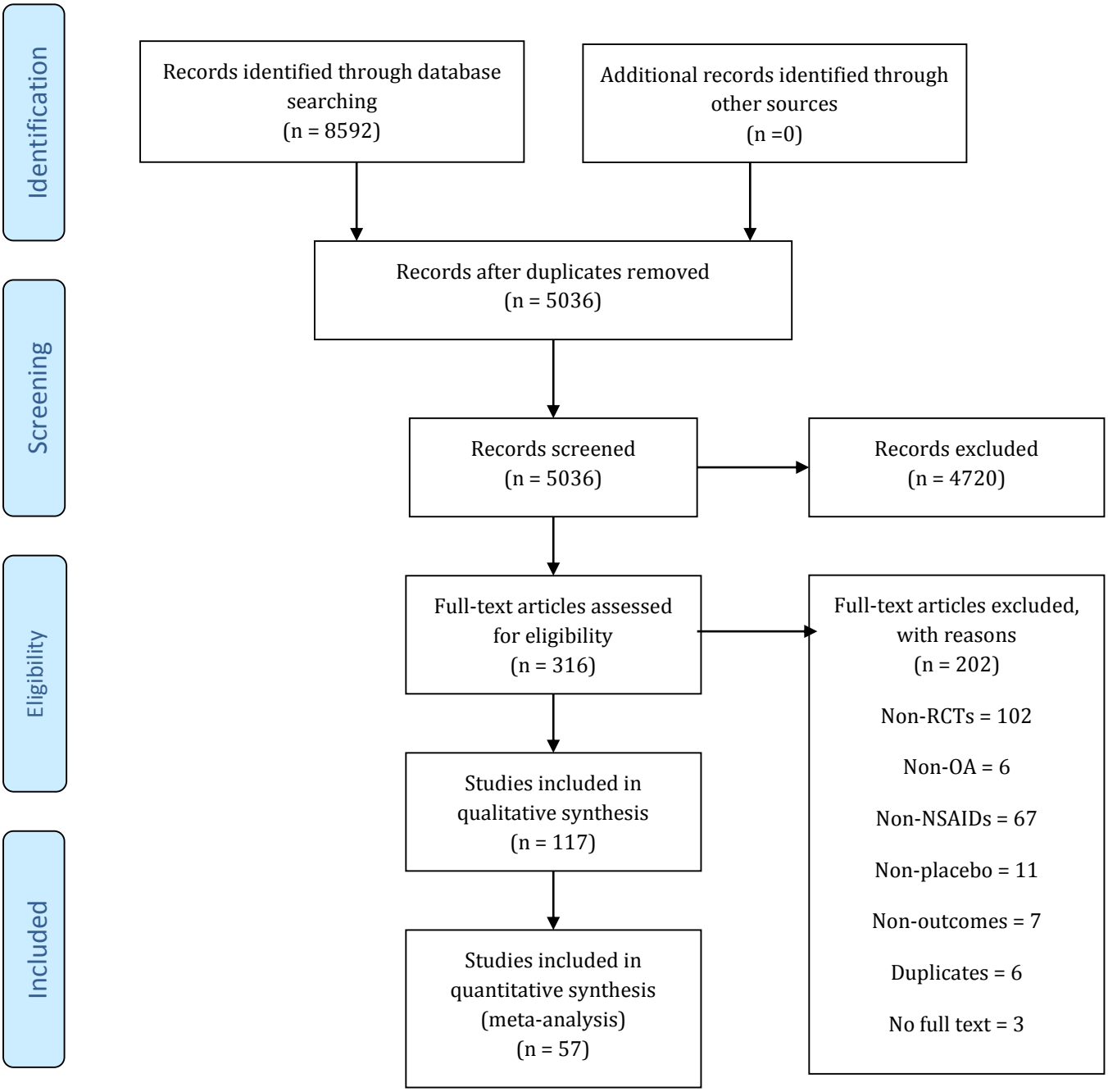
* - signified analysis reach a statistical significant difference; CI – confidence intervals; I2 – inconsistency value; N/E – Not estimatable; vs. - versus

Table 4: Meta-regression of effect size of NSAIDs for osteoarthritis pain (number of observation=125)

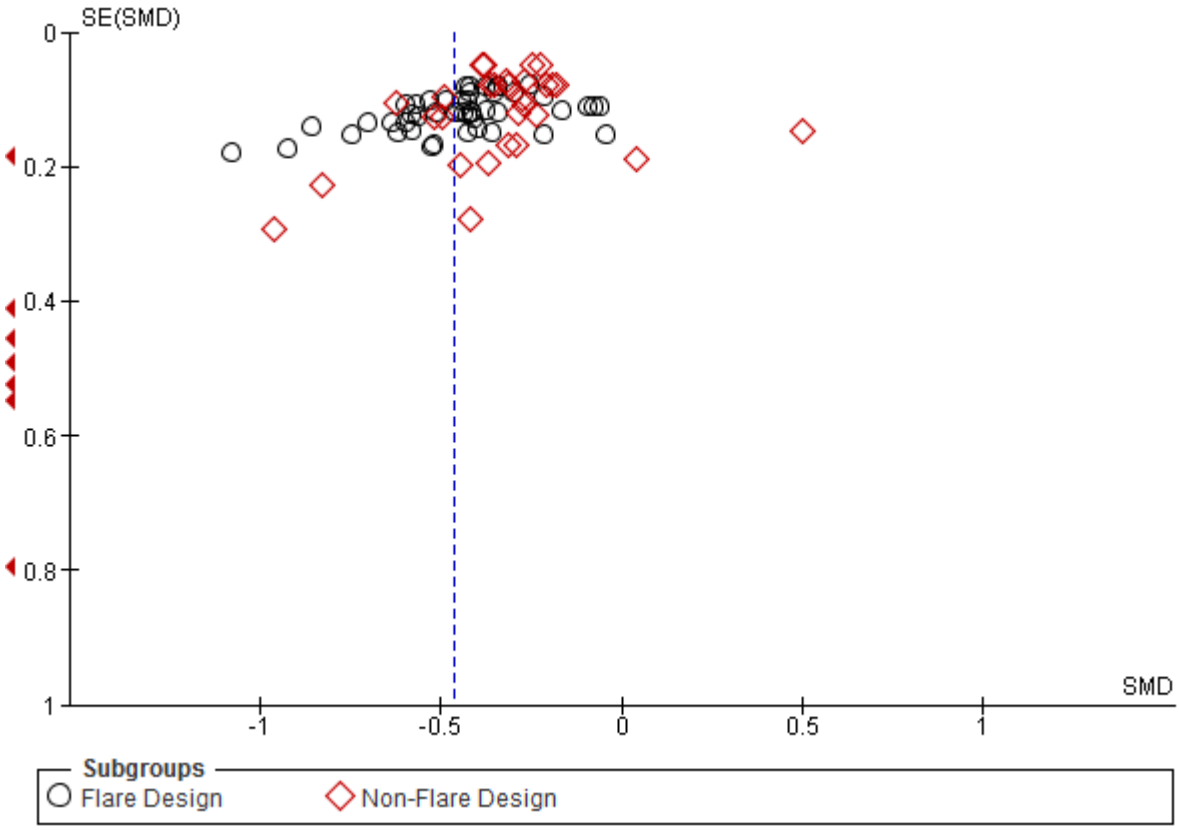
	B	**(95% Confidence Intervals)**	**P-value**
Flare design (yes=1, all others=0)	0.033	(-0.184, 0.251	0.763
Setting (community yes=1, all others=0)	0.324	(-0.154, 0.802)	0.182
Allocation concealment (yes=1, all others=0)	-0.030	(-0.260, 0.200)	0.798
Blinding to participants (yes=1, all others=0)	-0.056	(-0.704, 0.592)	0.864
Intent to treat analysis (yes=1, no=0)	0.076	(-0.167, 0.319)	0.538
Sample size (>=100 per arm vs. <100 per arm)	0.205	(-0.015, 0.424)	0.067
_cons	-0.617	(-1.238, 0.004)	0.051

β – meta-regression value; p-value – probability value

Supplementary Figure 1: PRISMA flow-chart



Supplementary Figure 2: Funnel plot assessing small sample size publication bias for primary outcome measure (mean change in pain score) for flare versus non-flare trial design.



Supplementary Table 1: MEDLINE search strategy

1. osteoarthritis.mp. or exp osteoarthritis/
2. arthrosis.mp.
3. osteoarthr$.mp.
4. (degenerative adj2 arthritis).mp.
5. gonarthrosis.mp.
6. coxarthrosis.mp.
7. or/1-6
8. randomized controlled trial.pt.
9. controlled clinical trial.pt.
10. randomized.ab.
11. placebo.ab.
12. clinical trials as topic.sh.
13. randomly.ab.
14. trial.ti.
15. or/8-14
16. (animals not (humans and animals)).sh.
17. 15 not 16
18. exp Anti-Inflammatory Agents, Non-Steroidal/ 164
19. NSAIDs.mp.
20. cyclooxygenase.mp.
21. cox* inhibitor.mp.
22. *coxib/
23. Lodine.mp.
24. celecoxib.mp.
25. Celebrex.mp.
26. rofecoxib.mp.
27. Vioxx.mp.
28. meloxicam.mp.
29. Mobic.mp.
30. *Naprosyn/
31. Anaprox*.mp.
32. Naprapac Aleve.mp.
33. (Cataflam or Voltaren or Arthrotec or Pennsaid).mp.
34. lumiracoxib.mp.
35. etoricoxib.mp.
36. Motrin.mp.
37. Profen.mp.
38. Vicoprofen.mp.
39. Combunox.mp.
40. Advil.mp.
41. Dolobid.mp.
42. Nalfon.mp.
43. Ansaid.mp.
44. indometacin.mp.
45. Indocin.mp.
46. Indo-Lemmon.mp.
47. Indomethagan.mp.
48. Oruvail.mp.
49. Toradol.mp.

Supplementary Table 1: MEDLINE search strategy (cont)

50. Mefenamic Acid.mp. or exp Mefenamic Acid/
51. Ponstel.mp.
52. Nabumetone.mp.
53. Relafen.mp.
54. Oxaprozin.mp.
55. Daypro.mp.
56. Piroxicam.mp. or exp Piroxicam/
57. Feldene.mp.
58. Sulindac.mp. or exp Sulindac/
59. Clinoril.mp.
60. Tolmetin.mp. or exp Tolmetin/
61. Tolectin.mp.
62. Valdecoxib.mp.
63. Bextra.mp.
64. Diacerein.mp.
65. Diacerhein.mp.
66. Rhein.mp.
67. Anthraquinones.mp. or exp Anthraquinones/
68. Diacetyilrhein.mp.
69. ART 50.mp.
70. Cyclooxygenase Inhibitors.mp. or exp Cyclooxygenase Inhibitors/
71. exp Cyclooxygenase Inhibitors/ or exp Cyclooxygenase 2 Inhibitors/
72. exp Aspirin/ or Aspirin.mp.
73. Etodolac.mp. or exp Etodolac/
74. naproxen.mp. or exp Naproxen/
75. Diclofenac.mp. or exp Diclofenac/
76. Ibuprofen.mp. or exp Ibuprofen/
77. Diflunisal.mp. or exp Diflunisal/
78. Fenoprofen.mp. or exp Fenoprofen/
79. Flurbiprofen.mp. or exp Flurbiprofen/
80. Indomethacin.mp. or exp Indomethacin/
81. Ketoprofen.mp. or exp Ketoprofen/
82. exp Ketorolac/ or Ketorolac.mp.
83. or/18-82
84. and/7,17,83

Supplementary Table 2: Study Characteristics (Study design)

Study	Funding (Public/ Industry/ Unclear)	Washout period specified	Flare Design Clearly Stated	Definition of Flare Design
Altman [21]	Unclear	✓	X	No information provided.
Baerwald [45]	Industry	✓	✓	To be eligible for inclusion, patients had to have experienced a flare of pain at the baseline visit (defined as a score of 50 mm for question 1 of the WOMAC pain subscale [17] that was increased by 15 mm as compared with the screening visit).”
Bensen [19]	Industry	✓	✓	OA was considered symptom-active if the patient's and physician's global assessment scores were "fair," "poor," or "very poor" and if 3 of the following 4 criteria were present: (1) a patient's assessment of arthritis pain (VAS) measurement of 40 mm or higher, (2) an increase of 2 points or more in the OA Severity Index from the screening assessment; (3) an increase from the screening visit of 1 grade or more in the patient's global assessment; and (4) an increase from the screening visit of 1 grade or more in the physician's global assessment. For patients not receiving NSAID or analgesic therapy and who had uncontrolled OA, 3 of the following 4 conditions were necessary for randomization at the baseline visit: (1) a patient's assessment of arthritis pain (VAS) measurement of 40 mm or higher, (2) an OA Severity Index score of 7 or more; (3) a patient's global assessment grade of poor or very poor, and (4) a physician's global assessment grade of poor or very poor
Biegert [65]	Public	✓	X	No information provided.
Bingham [38]	Industry	✓	✓	NSAID users had to demonstrate a minimum score of 40mm with an increase of 15mm on patient-assessed pain walking on a flat surface, and Eligibility required patients to meet specific flare criteria upon medication washout. IGADS worsening of at least one point on a 5-point Likert scale. Acetaminophen users had to demonstrate a minimum of 40mm of patient-assessed pain walking on a flat surface, fair, poor or very poor on IGADS, and a minimum of 40mm on PGADS.
Bocanegra [28]	Industry	✓	✓	Worsening of the OA symptoms was defined as at least 2 of the following 3: (1) an increase of one grade or more since screening, or a score of "poor" or "very poor," on the physician's Global Assessment"; (2) an increase of at least one grade since screening, or a score of "poor" or "very poor," on the patient's Global Assessment"; and (3) an increase of at least 2 points since screening, or a score of 7 or higher, on the Osteoarthritis Severity Index
Bourgeois [16]	Unclear	✓	X	No information provided.
Case [63]	Public	✓	X	No information provided.
Cryer [46]	Industry	✓	✓	Patients had a wash-out period of 7-14 days, a baseline visit (following a flare in OA pain)
Davies [56]	Industry	✓	Possible	No information provided.
Day [30]	Industry	✓	✓	patients were randomized to the study if they reported a minimum of 40mm and an increase of 15 mm on the VAS compared with the value at the screening visit
DeLemos [75]	Industry	✓	X	No information provided.
Dickson [60]	Industry	✓	X	No information provided.

Ding [74]	Public	Unclear	Possible	No information provided.
Dougados [72]	Industry	✓	X	No information provided.
Dreiser [52]	Unclear	✓	X	No information provided.
Ehrich [29]	Industry	✓	✓	Patients were randomized to the study if they reported a minimum of 40 mm on a 100 mm OA pain VASVAS (0 mm = no pain; 100 mm = extreme pain) after discontinuation of NSAID therapy, and an increase of 15 mm compared with the value recorded at the screening visit
Ehrich [31]	Unclear	✓	✓	To be eligible, patients had to demonstrate worsening in pain after discontinuation of previous therapy with NSAIDs
El-Mehairy [49]	Unclear	✓	X	No information provided.
Essex [48]	Industry	✓	✓	African American patients aged ≥ 45 years, with OA of the knee (diagnosed according to American College of Rheumatology guidelines30) in a flare state, and with a functional capacity classification of I – III were eligible for study participation." "For patients receiving NSAID or analgesic therapy, a flare was demonstrated if the physician's Global Assessment of Arthritis and the patient's Global Assessment of Arthritis were both 'fair', 'poor' or 'very poor' at the baseline visit, and if the baseline Patient's Assessment of Arthritis Pain VAS measurement was between 40 and 90 mm (out of 100 mm; 0 representing no pain and 100 representing very severe pain), the patient's Global Assessment of Arthritis showed an increase of one or more grades and the physician's Global Assessment of Arthritis showed an increase of one or more grades." "For patients who were not receiving treatment, a flare was defined if the Patient's Assessment of Arthritis Pain VAS was between 40 and 90 mm, the patient's and physician's Global Assessment of Arthritis was 'poor' or 'very poor', and the Global Assessment of Arthritis was 'poor' or 'very poor'
Fleischmann [36]	Industry	✓	✓	At the end of the screening period, patients with pain intensity (during the last 24 hours) in the targeted knee >=40 mm on a 100 mm VAS were eligible for entry into the treatment phase
Gibofsky [64]	Industry	X	X	No information provided.
Goldstein [58]	Unclear	Unclear	Possible	No information provided.
Grifka [35]	Industry	✓	✓	Patients were required to have pain intensity a 40 mm on a 100 mm VAS(most pain) in the target hand during the 24 hours prior to baseline. An increase in pain intensity in the target hand of either a 20% or a 10 mm VAS at the baseline visit com-pared with screening values (whichever was greater) was required to assess those patients who required analgesia
Haghighi [68]	Unclear	✓	X	No information provided.
Karlsson [43]	Industry	✓	✓	Patients were also required to experience a pain flare within 3–14 days of discontinuing all pain medications during a washout phase (between screening and baseline). The VAS pain score for pain on walking on a flat surface at baseline was required to be ≥ 40 mm, with an increase of at least 15 mm compared to screening
Kivitz [62]	Industry	✓	Possible	No information provided.
Kruger [73]	Unclear	✓	X	No information provided.
Lee [51]	Industry	✓	Possible	No information provided.
Leung [34]	Industry	✓	✓	The flare criteria were: 40 mm and an increase of 15mm compared with screening values on question 1 of WOMAC questionnaire and a worsening on the investigator's global assessment of disease status by 1 point on a

				5-point Likert scale. Pre-study paracetamol (acetaminophen) users had to demonstrate reproducible symptoms on the screening and randomization visits: of 40 mm pain while walking on a flat surface and patient's global assessment of disease status
Lund [55]	Unclear	✓	X	No information provided.
McKenna [32]	Industry	✓	✓	OA evidenced by a defined worsening of the signs and symptoms of the disease following discontinuation of treatment with NSAIDs for other analgesic medications.
Nguyen [53]	Industry	✓	Possible	No information provided.
Nunes [22]	Unclear	X	X	No information provided.
Paul [17]	Unclear	Unclear	X	No information provided.
Petersen [76]	Public	✓	Possible	No information provided.
Pincus [15]	Unclear	✓	X	No information provided.
Puopolo [40]	Industry	✓	✓	A sufficient flare within the washout period was defined as a patient-reported pain score of at least 40 mm while the patient walked on a flat surface, and was at least 15 mm greater than that recorded at the pre-study visit as well as a worsening of at least one point (0- to 5-point Likert scale) for IGADS
Reginster [41]	Unclear	✓	✓	Pre-study NSAID users were required to demonstrate worsening of pain (flare) after a pre-specified washout period based on the half-life of the drug
Rother [42]	Industry	✓	✓	Patients had to meet three osteoarthritis flare criteria
Sandelin [54]	Unclear	X	X	No information provided.
Schmitt [57]	Unclear	✓	Possible	No information provided.
Schnitzer [47]	Industry	✓	✓	Patient had experienced a flare of pain (Baseline WOMAC: question 1 of pain subscale value of 50 mm, with an increase of 15 mm compared with screening) after discontinuing all analgesic therapy at screening (for at least 5 half-lives of the prior analgesic or anti-inflammatory therapy before the baseline visit).”
Schnitzer [66]	Industry	✓	X	No information provided.
Schnitzer [77]	Public	✓	Possible	No information provided.
Scott [59]	Unclear	✓	Possible	No information provided.
Sheldon [69]	Unclear	✓	X	No information provided.
Shipley [50]	Industry	X	X	No information provided.
Simon [44]	Industry	✓	✓	Patients had to meet the osteoarthritis flare criteria
Svensson [71]	Industry	X	X	No information provided.
Tannenbaum [65]	Industry	✓	X	No information provided.
Truitt [18]	Industry	✓	X	No information provided.
Tuzun [70]	Unclear	Unclear	X	No information provided.
Uzun [61]	Not stated	Unclear	Possible	No information provided.
Wiesenhutter [37]	Industry	✓	✓	A flare was classified as sufficient if the minimum patient-reported pain score was 40mm while the patient waked on a flat surface
Williams [33]	Industry	✓	✓	All patients included in this study experienced an OA flare at the baseline visit (day 0, within 24 hours before the first dose of study medication).”

				"Patients were considered to have an OA flare if baseline scores on both the Patient's and Physician's Global Assessments of Arthritis indicated that their condition was fair, poor, or very poor. Furthermore, baseline assessments had to meet the following criteria: Patient's Assessment of Arthritis Pain-VAS measurement of 240 mm; an increase of 22 points on the Lequesne Osteoarthritis Severity Index versus values at the screening visit; and an increase of 21 grade on the Patient's or Physician's Global Assessment of Arthritis versus values at the screening visit." "Patients with uncontrolled OA who were not receiving NSAIDs or analgesics before the study were considered to be experiencing an OA flare and therefore eligible for enrolment if they satisfied the following criteria: Patient's Assessment of Arthritis Pain-VAS measurement of 240 mm, a Lequesne Osteoarthritis Severity Index score of 27, and a score on the Patient's or Physician's Global Assessment of Arthritis of 4 (poor) or 5 (very poor)."
Wittenberg [39]	Industry	✓	✓	Patients were required to have VAS actual pain intensity at baseline of ≥50 mm for the most severely affected (target) knee joint after activity. (The pain requirement at baseline following washout [≥50 mm] was greater than at screening [≥40 mm]; thus, an increase in pain from screening to baseline was required for study entry.)

✓ - Yes; X – No; IGADS - Investigator Global Assessment of Disease Status; mm – millimetres; NSAIDs – non-steroidal anti-inflammatory drugs; OA – osteoarthritis; VAS – visual analogue scale; WOMAC - Western Ontario and McMaster Universities Osteoarthritis Index.

Supplementary Table 3: Participant characteristics of the included studies (medications and demographics)

Study	NSAIDs and Dose	Duration of NSAID (weeks)	Contaminant	N	Mean Age	Gender (M/F)	Joint Affected	Mean disease duration (years)	Setting (Hospital/Community)
Altman [21]	Naproxen 500 mg BID	26	Acetaminophen	333	64	143/190	Knee	Unclear	Hospital
Baerwald [45]	Naproxcinod 750 mg BID	13	Acetaminophen	810	63	279/531	Hip	Unclear	Hospital
	Naproxen 500 mg BID								
Bensen [19]	Celecoxib, 50 mg BID	1	Acetaminophen, aspirin	1003	62.2	281/722	Knee	9.8	Hospital
	Celecoxib, 100 mg BID								
	Celecoxib, 200 mg BID								
	Celecoxib, 500 mg BID								
Biegert [65]	Diclofenac, 2 tablets BID100 mg/day	6	Aspirin, physical therapy	84	61.8	53/31	Knee/Hip	Unclear	Hospital
Bingham [38]	Etoricoxib 30 mg QD	12	Acetaminophen	599	62.4	195/404	Knee/Hip	Unclear	Unclear
	Celecoxib 200 mgQD								
	Eetoricoxib 30 mg QD	12	Acetaminophen	608	61.8	209/399	Knee/Hip	Unclear	Unclear
	Celecoxib 200 mg QD								
Bocanegra [28]	Diclofenac sodium 75 mg BID	6	Unclear	572	62.5	180/392	Knee/Hip	11.2	Hospital
	Diclofenac/misoprostol D50/M200 TID								
	Diclofenac/misoprostol D75/M200 BID								

Bourgeois [16]	Nimesulide 50 mg BID Nimesulide 100 mg BID Nimesulide 200 mg BID	4	Paracetamol	382	u	Unclear	Knee	Unclear	Unclear
Case [63]	Diclofenac sodium 75 mg BID	12	Unclear	82	62.2	41/41	Knee	Unclear	Hospital
Cryer [46]	Naproxen/esomeprazel magnesium tablets BID Celecoxib 200mg capsules QD	12	Prednisone, antiplatelet agents , antacid, acetaminophen	612	61.6	221/391	Knee	Unclear	Hospital
Davies [56]	Ibuprofen 800 mg TID	4	Acetaminophen	104	61.5	38/66	Knee/Hip/ Spine	7.9	Hospital
Day [30]	Rofecoxib 12.5 mg QD Rofecoxib 25 mg QD Ibuprofen 800 mg TID	6	Acetaminophen	809	63.6	162/647	Knee/Hip	8.7	Hospital
DeLemos [75]	Celecoxib 200 mg QD	12	Aspirin, acetaminophen	1001	58	369/632	Knee/Hip	8.1	Hospital
Dickson [60]	Diclofenac 100mg/day, 3 weekly arthrocenteses	3	Acetaminophen	165	64.5	73/92	Knee	Unclear	Community

Ding [74]	Ibuprofen, 0.3g, BID	2	Health education	90	56	25/65	Knee	4.7	Hospital
Dougados [72]	Lumiracoxib 100 mg QD	13	Sheldon 2005: acetaminophen				Knee	5.5	Unclear
	Lumiracoxib 100 mg QDwith initial dose								
	Celecoxib 200 mg QD			3235	61.5	1097/2138			
Dreiser [52]	Ibuprofen 800 mg QD	2	Unclear	60	59.4	9/51	Hand	Unclear	Hospital
Ehrich [31]	Rofecoxib 5 mg QD	6	None	672	61.7	195/477	Knee/Hip	10.9	Unclear
	Rofecoxib 12.5 mg QD								
	Rofecoxib 25 mg QD								
	Rofecoxib 50 mg QD								
Ehrich [29]	Rofecoxib 25 mg QD	6	Acetaminophen	219	63.5	63/156	Knee	11.9	Hospital
	Rofecoxib 125 mg QD								
El-Mehairy [49]	Nifiumic acid 250 mg TID	8	Unclear				Knee/Hip	Unclear	Hospital
	Phenylbutazone (100 mg/capsule), NSAIDs, TID								
	Oxyphenylbutazon 100 mg, NSAIDs, TID			100	54.6	14/86			
Essex [48]	Celecoxib 200 mg QD	6	Aspirin, acetaminophen	322	58	64/258	Knee	5.4	Hospital
	Naproxen 500 mg BID						Knee/Hip		
Fleischmann [36]	Lumiracoxib 200 mg oQD	13	Paracetamol	1600	61.1	539/1061	Knee	6.4	Hospital
	Lumiracoxib 400 mg QD								

	Celecoxib 200 mg QD								
Gibofsky [64]	Celecoxib 200 mg/day	6	Aspirin, acetaminophen	477	62.9	157/320	Knee	8.6	Unclear
	Rofecoxib 25 mg/day								
Goldstein [58]	Naproxen 375 mg daily week 1, week 2-3: naproxen 375 mg BID	3	Acetaminophen	194	61.2	72/122	Knee/Hip	Unclear	Hospital
Grifka [35]	Lumiracoxib 200 mg QD	4	Paracetamol	594	61.9	104/490	Hand	5.3	Hospital
	Lumiracoxib 400 mg QD								
Haghighi [68]	Ibuprofen three 400 mg tablets daily	4	Acetaminophen	120	56.8	89/31	Knee/Hip	Unclear	Hospital
Karlsson [43]	Naproxcinod 750mg QD	6	Paracetamol, antihypertensive drugs	543	61.5	177/366	Knee/Hip	Unclear	Hospital
	Naproxcinod 750mg BID								
	Naproxcinod 1125mg BID								
	Rofecoxib 25 mg QD								
Kivitz [62]	Valdecoxib 5 mg QD	12	Unclear	1015	59.7	356/659	Knee	9.1	Community
	Valdecoxib 10 mg QD								
	Valdecoxib 20 mg QD								
	Naproxen 500 mg BID								
Kruger [73]	Oxaceprol 200 mg TID	3	Acetaminophen	97	59.6	31/66	Knee/Hip	Unclear	Hospital
Lee [51]	Diflunisal 500 mg BID	6	None	422	61.3	139/283	Knee	5	Unclear
	Diflunisal 375 mg BID								
Leung [34]	Etoricoxib 60 mg QD	12	Paracetamol	501	63.2	109/392	Knee/Hip	6.1	Hospital

	Naproxen 500 mg BID								
Lund [55]	Meloxicam 7.5 mg	3	Paracetamol, message, exercise	411	68.5	112/299	Knee	Unclear	Hospital
McKenna [32]	Celecoxib 100 mg BID	6	Aspirin	600	61.7	208/392	Knee	8.6	Unclear
	Diclofenac 50 mg TID								
Nguyen [53]	Tenoxicam 20 mg QD	8	Paracetamol	145	62.6	62/83	Hip	5.6	Hospital
Nunes [22]	Alginac TIDvitamin b12, b6, b1	2	Unclear	80	42.1	42/38	Knee/Hip	2.2	Unclear
Paul [17]	Aceclofenac (100 mg) BID	4	Paracetamol				Knee	4.3	Hospital
	Nabumetone (750 mg) BID			423	53.5	188/235			
Petersen [76]	Ibuprofen 600mg BID	12	Quadriceps muscle strength, acupuncture	35	62.4	14/21	Knee	Unclear	Hospital
Pincus [15]	Celecoxib 200 mg/day	6	Propoxyphene; codeine 60 mg or tramadol rescue medication	524	U	Unclear	Knee/Hip	Unclear	Unclear
Puopolo [40]	Etoricoxib 30 mg QD	12	Acetaminophen	816	62.6	198/618	Knee/Hip	6.6	Hospital
	Ibuprofen 800 mg TID								
Reginster [41]	Etoricoxib 60 mg QD Naproxen 500 mg BID	12	Paracetamol	997	62.8	279/718	Knee/Hip	Unclear	Unclear
Rother [42]	Celecoxib 100 mg oral and placebo gel	6	Paracetamol	397	62.8	160/237	Knee	Unclear	Hospital
Sandelin [54]	Diclofenac 50 mg BID	4	None	281	61	92/189	Knee	Unclear	Hospital

Schmitt [57]	Diclofenac sodium 150 mg dual release capsules (DRC150) QD Diclofenac sodium 75 mg QD Voltaren 50 mg enteric coated tablet (EC50) TID	12	None	393	60.9	63/330	Knee/Hip	8.8	Hospital
Schnitzer [47]	Naproxcinod 750 mg BID Naproxcinod 375 mg BID Naproxcinod 500 mg BID	13	Acetaminophen	1000	59.8	291/709	Knee	Unclear	Unclear
Schnitzer [66]	Lumiracoxib 50 mg BID Lumiracoxib 100 mg BID Lumiracoxib 200 mg BID Lumiracoxib 400 mg BID	4	Acetaminophen	583	60.3	187/396	Knee/Hip	6.9	Unclear
Schnitzer [77]	Lumiracoxib 100 mg QD Celecoxib 200 mg QD	13	Acetaminophen	1262	61.6	485/777	Hip	Unclear	Unclear
Scott [59]	Tiaprofenic acid 300 mg BID Indomethacin 25 mg TID	4	Acetaminophen	812	61	240/572	Knee	5	Unclear
Sheldon [69]	Lumiracoxib 100 mg QD	13	Acetaminophen	1551	60.5	583/968	Knee	6.9	Hospital
Shipley [50]	Fenoprofen 600 mg TID Rhus tox	2	Paracetamol; homeopathy therapy	33	65	9/24	Knee/Hip	Unclear	Hospital
Simon [44]	oral diclofenac tablets 100 mg	12	glucosamine, chondroitin, paracetamol	772	61.6	289/483	Knee	Unclear	Hospital

Svensson [71]	Naproxen 500 mg BID	8	Unclear	511	59.7	151/360	Knee	Unclear	Unclear
	Naproxen 500 mg BID	8	Unclear	511	59.7	151/360	Hip	Unclear	Unclear
Tannenbaum [67]	Lumiracoxib 200 mg, QD	13	Paracetamol				Knee	4.8	Unclear
	Lumiracoxib 400 mg, QD								
	Celecoxib 200 mg, QD			1702	64.2	536/1166			
Truitt [18]	Rofecoxib 12.5 mg QD	6	Acetaminophen				Knee/Hip	15	Hospital
	Rofecoxib 25 mg QD								
	Nabumetone 1500 mg QD			341	83	124/217			
Tuzun [70]	Flurbiprofen 100 mg PO (tablets) BID	3	Unclear				Knee	5.1	Hospital
	Tiaprofenic acid 300 mg PO (tablets) BID			39	59.1	19/20			
Uzun [61]	Flurbiprofen 100 mg BID	3	Unclear				Knee	5.1	Hospital
	Tiaprofenic acid 300 mg BID			39	59.1	20/19			
Wiesenhutter [37]	Etoricoxib 30 mg/d	12	Acetaminophen, aspirin, stable glucosamine or chondroitin	528	61.5	156/372	Knee/Hip	7.8	Unclear
	Ibuprofen 2400mg/d								
Williams [33]	Celecoxib 100 mg BID	6	Aspirin, acetaminophen	718	61.5	214/504	Knee	Unclear	Unclear
	Celecoxib 200 mg QD								
Wittenberg [39]	Lumiracoxib 400 mg QD	1	Acetaminophen	334	65	123/211	Knee	7.5	Unclear
	Celecoxib 200 mg BID								

BID – twice a day; F – female; M- Male; mg – milligrams; mg/d – milligrams per day; N – number of participants; PO – orally taken; QD – once a day; TID – three times a day

Supplementary Table 4: Summary of the included trial quality assessment results

Study	Randomisation Defined	Allocation concealment	Blinding of participants	Blinding of clinicians	Blinding of assessors	Follow-up >85%	ITT Analysis Performed
Altman [21]	Unclear	Unclear	✓	X	✓	X	Unclear
Baerwald [45]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Bensen [19]	✓	Unclear	✓	✓	✓	X	✓
Biegert [65]	✓	Unclear	✓	Unclear	✓	X	✓
Bingham [38]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Bocanegra [28]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Bourgeois [16]	✓	Unclear	✓	✓	✓	X	X
Case [63]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Cryer [46]	✓	✓	✓	✓	✓	✓	✓
Davies [56]	Unclear	Unclear	✓	Unclear	Unclear	✓	X
Day [30]	✓	✓	✓	✓	✓	✓	X
DeLemos [75]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Dickson [60]	Unclear	Unclear	✓	X	✓	X	✓
Ding [74]	✓	✓	X	X	X	✓	✓
Dougados [72]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Dreiser [52]	Unclear	Unclear	✓	Unclear	Unclear	✓	✓
Ehrich [29]	✓	✓	✓	✓	✓	X	✓
Ehrich [31]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
El-Mehairy [49]	Unclear	Unclear	✓	✓	✓	✓	✓
Essex [48]	✓	Unclear	✓	Unclear	Unclear	X	✓
Fleischmann [36]	Unclear	Unclear	✓	✓	Unclear	X	✓
Gibofsky [64]	✓	✓	✓	✓	✓	X	✓
Goldstein [58]	Unclear	Unclear	✓	Unclear	Unclear	✓	X
Grifka [35]	Unclear	Unclear	✓	Unclear	Unclear	✓	✓
Haghighi [68]	Unclear	Unclear	✓	Unclear	Unclear	✓	✓
Karlsson [43]	✓	✓	✓	✓	✓	X	✓
Kivitz [62]	✓	✓	✓	✓	✓	X	✓

Kruger [73]	✓	✓	✓	✓	✓	X	✓
Lee [51]	Unclear	Unclear	✓	Unclear	Unclear	X	X
Leung [34]	✓	✓	✓	✓	Unclear	X	X
Lund [55]	Unclear	Unclear	✓	Unclear	Unclear	✓	✓
McKenna [32]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Nguyen [53]	✓	Unclear	✓	✓	✓	✓	✓
Nunes [22]	Unclear	Unclear	✓	Unclear	Unclear	✓	Unclear
Paul [17]	Unclear	Unclear	✓	✓	✓	X	X
Petersen [76]	✓	✓	✓	✓	✓	✓	X
Pincus [15]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Puopolo [40]	✓	✓	✓	✓	✓	X	✓
Reginster [41]	✓	✓	✓	Unclear	Unclear	X	X
Rother [42]	✓	✓	✓	✓	Unclear	X	✓
Sandelin [54]	✓	✓	✓	Unclear	Unclear	✓	✓
Schmitt [57]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Schnitzer [47]	Unclear	Unclear	✓	✓	✓	X	✓
Schnitzer [66]	Unclear	Unclear	✓	Unclear	Unclear	✓	✓
Schnitzer [77]	✓	Unclear	✓	✓	✓	X	✓
Scott [59]	Unclear	Unclear	✓	X	X	Unclear	Unclear
Sheldon [69]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Shipley [50]	Unclear	Unclear	✓	✓	Unclear	✓	X
Simon [44]	✓	✓	✓	✓	✓	X	✓
Svensson [69]	Unclear	Unclear	✓	Unclear	Unclear	Unclear	Unclear
Tannenbaum [65]	Unclear	Unclear	✓	Unclear	Unclear	X	✓
Truitt [18]	✓	✓	✓	✓	✓	X	✓
Tuzun [70]	Unclear	Unclear	X	X	X	Unclear	Unclear
Uzun [61]	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Wiesenhutter [37]	Unclear	Unclear	✓	✓	Unclear	X	✓
Williams [33]	Unclear	Unclear	✓	✓	✓	X	✓
Wittenberg [39]	Unclear	Unclear	✓	✓	Unclear	✓	✓

✓ - satisfied; X – not satisfied; ITT – intention-to-treat analysis