REVIEW

Annals of Internal Medicine

# Comparative Effectiveness of Pharmacologic Interventions for Knee Osteoarthritis

## A Systematic Review and Network Meta-analysis

**Raveendhara R. Bannuru, MD; Christopher H. Schmid, PhD; David M. Kent, MD; Elizaveta E. Vaysbrot, MD; John B. Wong, MD; and Timothy E. McAlindon, MD**

**Background:** The relative efficacy of available treatments of knee osteoarthritis (OA) must be determined for rational treatment algorithms to be formulated.

**Purpose:** To examine the efficacy of treatments of primary knee OA using a network meta-analysis design, which estimates relative effects of all treatments against each other.

**Data Sources:** MEDLINE, EMBASE, Web of Science, Google Scholar, Cochrane Central Register of Controlled Trials from inception through 15 August 2014, and unpublished data.

**Study Selection:** Randomized trials of adults with knee OA comparing 2 or more of the following: acetaminophen, diclofenac, ibuprofen, naproxen, celecoxib, intra-articular (IA) corticosteroids, IA hyaluronic acid, oral placebo, and IA placebo.

**Data Extraction:** Two reviewers independently abstracted study data and assessed study quality. Standardized mean differences were calculated for pain, function, and stiffness at 3-month follow-up.

**Data Synthesis:** Network meta-analysis was performed using a Bayesian random-effects model; 137 studies comprising 33 243 participants were identified. For pain, all interventions significantly outperformed oral placebo, with effect sizes from 0.63 (95% credible interval [CrI], 0.39 to 0.88) for the most efficacious treatment (hyaluronic acid) to 0.18 (CrI, 0.04 to 0.33) for the least efficacious treatment (acetaminophen). For function, all interventions except IA corticosteroids were significantly superior to oral placebo. For stiffness, most of the treatments did not significantly differ from one another.

**Limitation:** Lack of long-term data, inadequate reporting of safety data, possible publication bias, and few head-to-head comparisons.

**Conclusion:** This method allowed comparison of common treatments of knee OA according to their relative efficacy. Intra-articular treatments were superior to nonsteroidal anti-inflammatory drugs, possibly because of the integrated IA placebo effect. Small but robust differences were observed between active treatments. All treatments except acetaminophen showed clinically significant improvement from baseline pain. This information, along with the safety profiles and relative costs of included treatments, will be helpful for individualized patient care decisions.

**Primary Funding Source:** Agency for Healthcare Research and Quality.

*Ann Intern Med.* 2015;162:46-54. doi:10.7326/M14-1231 **www.annals.org**
For author affiliations, see end of text.

Knee osteoarthritis (OA) is a common and progressive joint disease affecting more than 250 million people worldwide (1). It has significant effects on function (2) and considerable societal costs in terms of work loss (3), early retirement, and joint replacement (4). Osteoarthritis is a leading indication for use of prescription drugs, which costs about $3000 per year per patient (5). In the absence of effective disease-modifying medical treatments, a range of symptomatic treatments is available. Nonsteroidal anti-inflammatory drugs (NSAIDs) are the most frequently prescribed medicines for OA yet have significant toxicity, especially among the demographic groups in which the disorder is most prevalent (6). Intra-articular (IA) treatments are widely used, although their efficacy and safety remain in question. More knowledge about the comparative efficacy and toxicity of these compounds, which would be helpful for patients, physicians, payers, and policymakers, is needed to formulate rational treatment algorithms for OA. The relative effectiveness of OA treatments is difficult to discern from the literature, in part because few head-to-head comparison studies are available and traditional pairwise meta-analysis cannot integrate all of the evidence from several comparators. Therefore, our goal was to comprehensively review the literature and determine the relative efficacy of the primary knee OA treatments using a network meta-analysis design.

**See also:**

Editorial comment . . . . . 71

*Web-Only*

Supplement
CME quiz

## METHODS

### Data Sources and Searches

We searched MEDLINE, EMBASE, Web of Science, Google Scholar, and the Cochrane Central Register of Controlled Trials from inception to 15 August 2014 (**Supplement Table 1**, available at www.annals.org). All searches were limited to randomized, controlled trials in humans. No limits were applied for language, publication date, or publication status, and foreign-language papers were translated. We also hand-searched the reference lists of all retrieved studies and conference proceedings of the American Association of Orthopedic

Surgeons, American College of Rheumatology, British Society for Rheumatology, European League Against Rheumatism, International League of Associations of Rheumatology, and Osteoarthritis Research Society International. The conference proceedings were searched from January 1990 to August 2014. We attempted to identify unpublished data by searching the Food and Drug Administration registry, ClinicalTrials.gov, product inserts, and pharmaceutical company Web sites, and by contacting experts, study authors, manufacturers, and primary authors of abstracts with incomplete data.

### Study Selection

We included all randomized, controlled trials involving adult human participants with clinical or radiologic diagnosis of symptomatic primary knee OA that compared at least 2 interventions of interest and reported extractable data for at least 1 measure of pain, function, or stiffness. On the basis of the treatment recommendations from the latest clinical practice guidelines for knee OA (7, 8) and the current prescription patterns worldwide (6), we included the following interventions and comparators: acetaminophen, diclofenac, ibuprofen, naproxen, celecoxib, IA corticosteroids, IA hyaluronic acid, oral placebo, and IA placebo. We did not include nonrandomized studies because they generally lacked the high quality of the randomized evidence; without individual-participant data, we could not properly adjust effect estimates for potential confounders.

Two reviewers independently screened all titles and abstracts identified by the searches. Full manuscripts of studies screened as potentially relevant by either reviewer were obtained and assessed by 2 independent reviewers according to the aforementioned criteria. Any discrepancies were resolved by consensus.

### Data Extraction and Quality Assessment

After developing a data extraction form, we tested it on 10 randomly selected, included studies and refined accordingly. After completing an a priori training exercise, 2 reviewers independently extracted data from each study. The data were reviewed for consistency between the 2 extractors, and any disagreements were resolved by consensus. For each study, data extraction details included design, selection criteria, population characteristics, treatments, outcome measures, length of follow-up, and results. The outcome measures of interest were change from baseline in pain, function, and stiffness scores reported at 3-month follow-up. If 3-month data were not available, we used data from 2 to 6 months (the data point closest to 3 months was given preference). Intention-to-treat analysis data were used whenever available. When an article provided data on more than 1 outcome scale or a different outcome from the same construct, we extracted data from the scale that was highest on the hierarchy of suggested outcomes for meta-analysis of knee OA trials (9). Two independent reviewers assessed individual study quality using the Cochrane risk-of-bias tool, with any discrepancies resolved by consensus (10). We investigated the effect of study quality on results in the sensitivity analysis.

### Data Synthesis and Analysis

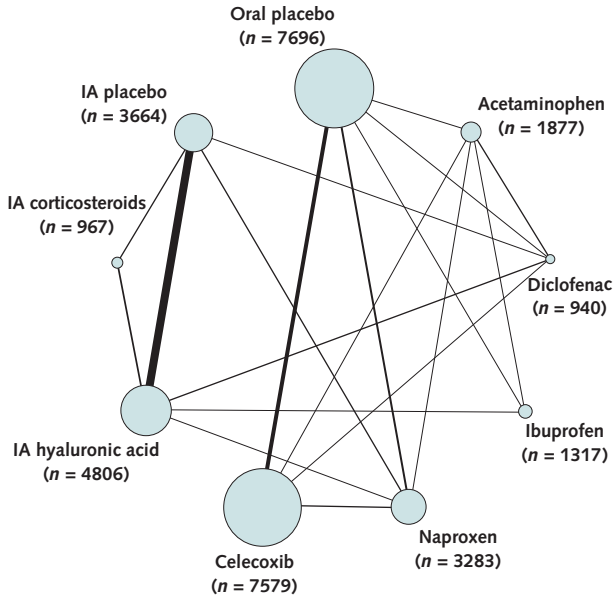
Because the studies used different outcome measures, the change from baseline Western Ontario and McMaster Universities OA Index (WOMAC), visual analogue scale (VAS), and Likert scale scores in each study was translated into Hedges *g* effect sizes (11). Hedges *g* is defined as the difference in change from baseline between 2 interventions divided by the pooled SD of the differences, with corrections for small sample sizes. To assess potential heterogeneity among the studies, we calculated the between-study variance and examined baseline characteristics of participants, interventions, outcomes, and study quality.

### Network Meta-analysis

A network meta-analysis synthesizes all available evidence within a consistent framework, thereby fully preserving the randomization within each trial (12). It accounts for multiple comparisons within a trial when there are more than 2 treatment groups (13–15). This method considers all trials simultaneously and enables integration of direct evidence from head-to-head trials (when they exist) with indirect evidence (obtained from comparisons of treatments through their common reference) (16). To account for the expected clinical and methodological heterogeneity, we used Bayesian hierarchical random-effects models for mixed multiple-treatment comparisons with noninformative prior distributions (**Supplement** [Data Synthesis and Analysis], available at www.annals.org) (17, 18). The model contained parameters that described the relative treatment effect of each intervention compared with a common comparator (oral placebo). Other treatment comparisons were derived as differences between model parameters. We assumed a normal likelihood distribution for the effect size.

The main assumption behind the validity of network meta-analysis is transitivity (13). This assumption requires that a valid synthesis of studies indirectly comparing 2 treatments (for example, A with C) by way of 2 direct comparisons (for example, A with B and B with C) must include studies that are sufficiently similar in important clinical and methodological characteristics (potential effect modifiers) (19). The populations within the included studies were similar and could be eligible for any of the treatments considered here based on the distributions of effect modifiers (mean age, percentage of women, baseline disease severity, baseline pain scores, duration of disease, and study quality) and inclusion and exclusion criteria of the studies. Another key assumption in a network meta-analysis is consistency—the notion that the direct and indirect estimates of the treatment effects are the same (20). Consistency was assessed using the node-splitting method (**Supplement** [Data Synthesis and Analysis]) (21). The results were presented graphically to visually assess the agreement between direct and indirect estimates. A value near 0 indicated that the comparisons in the network were consistent.

***Figure 1.*** Network of treatment comparisons for pain.




| Comparisons | Trials, *n* |
|---|---|
| Oral placebo vs. acetaminophen | 6 |
| Oral placebo vs. diclofenac | 6 |
| Oral placebo vs. ibuprofen | 5 |
| Oral placebo vs. naproxen | 14 |
| Oral placebo vs. celecoxib | 28 |
| Acetaminophen vs. diclofenac | 2 |
| Acetaminophen vs. ibuprofen | 2 |
| Acetaminophen vs. naproxen | 1 |
| Acetaminophen vs. celecoxib | 4 |
| Diclofenac vs. celecoxib | 1 |
| Diclofenac vs. IA hyaluronic acid | 2 |
| Diclofenac vs. IA placebo | 2 |
| Ibuprofen vs. IA hyaluronic acid | 1 |
| Naproxen vs. celecoxib | 7 |
| Naproxen vs. IA haluronic acid | 1 |
| Naproxen vs. IA placebo | 2 |
| IA Hyaluronic acid vs. IA corticosteroids | 12 |
| IA Hyaluronic acid vs. IA placebo | 52 |
| IA Corticosteroids vs. IA placebo | 7 |

Circle size reflects number of participants, and the line width reflects the number of direct comparisons. No connecting line between 2 treatments indicates that there was no direct comparison. IA = intra-articular.

Results were presented as median effect sizes for pain, function, and stiffness along with 95% central credible intervals (CrIs). For improving the clinical interpretability, they were converted back to the natural units of the most commonly used scale (WOMAC VAS, 0 to 100) (22). On the basis of the Osteoarthritis Research Society International–Outcome Measures in Rheumatology responder criteria, we prespecified an absolute change of 20 points on a scale of 0 to 100 as clinically significant improvement (23).

We performed several sensitivity analyses on the primary outcome of pain to explore potential causes for heterogeneity. Multiple-treatment meta-regression analysis and subgroup analyses were done to assess the effect of study quality, sample size, and type of outcome scale used (WOMAC vs. other) (24). To examine the potential effect of reporting bias, we analyzed pain outcomes in trials reporting only pain; those reporting both pain and function; and those reporting pain, function, and stiffness. We also compared the baseline characteristics and study quality measures of these subsets of trials. The Supplement (Data Synthesis and Analysis) provides additional details of the statistical methods used.

### Role of the Funding Source

The Agency for Healthcare Research and Quality had no role in study design, data collection, analysis or interpretation, preparation, review, or approval of the manuscript. The funding agency had no access to the data and did not perform any of the study analyses.

## RESULTS

Of the 4122 citations identified through our literature search, 3625 were excluded through title and abstract screening. Among the 497 full-text reports, 137 studies met inclusion criteria for the network meta-analysis (Appendix Figure, available at www.annals.org). Figure 1 shows the network of all treatment comparisons analyzed for pain; the networks for function and stiffness are shown in the Supplement. Thirteen trials had 3 study groups, and the rest had 2 study groups. Many treatments within the network were never actively compared. Data were available for only 19 of 36 possible comparisons. In particular, there were few direct comparisons between the IA and oral agents. All included trials were published between 1980 and 2014 and involved a total of 33 243 randomly assigned participants. Because reported outcomes differed in these trials, 129 trials (32 129 participants) contributed to the analyses of pain-related outcomes, 76 trials (24 059 participants) contributed to the analyses of physical function outcomes, and 55 trials (18 267 participants) contributed to the analyses of stiffness outcomes. In each trial, the average age of the participants ranged from 45 to 76 years (median, 62; interquartile range, 60 to 64) and the proportion of women ranged from 3% to 100% of participants (median, 67%; interquartile range, 62% to 72%) (details of included trials are shown in Supplement Table 2).

The quality of included trials is presented in Supplement Table 3. The sample size varied between 24 and 779 participants. Forty-five trials (34%) had fewer than 100 participants. Participant baseline characteristics were similar in 118 trials (86%), randomization sequences were adequate and clearly reported in 75 trials (55%), reporting of allocation concealment was assessed as adequate in 68 trials (50%), and blinding of both participants and assessors occurred in 110 trials (80%). Ninety-two trials (67%) reported intention-to-treat analyses for pain outcomes. Figure 2 presents a plot of quality of direct versus indirect evidence for each pairwise comparison. Comparisons with 5 or more trials providing direct evidence are plotted. The overall quality of evidence is moderate, with the newer

trials (published after 2000) reporting better methods. Trials involving celecoxib were larger, and 60% had low risk of bias. The quality of direct and indirect evidence did not differ much, except for comparisons in which celecoxib was involved either directly or indirectly. Only 4 (6%) trials comparing oral therapies and 9 (13%) trials comparing IA therapies were independently funded. All 4 trials involving oral therapies compared 2 active agents, whereas 8 of the IA therapy trials compared them with IA placebo. Because 90% of the trials were industry-funded, it was challenging to predict the overall direction of sponsorship bias in this network.

## Pain

A total of 129 trials (32 129 participants) contributed to the analyses of pain-related outcomes (**Figure 1**). All interventions were statistically significantly better than oral placebo (**Table 1**), with effect sizes ranging from 0.18 (CrI, 0.04 to 0.33) for the least efficacious treatment (acetaminophen) to 0.63 (CrI, 0.39 to 0.88) for the most efficacious treatment (IA hyaluronic acid). All treatments except acetaminophen met the prespecified criteria for clinically significant improvement (**Supplement Figure 1** and **Supplement Table 4**). Naproxen, ibuprofen, diclofenac, IA hyaluronic acid, and IA corticosteroids were statistically significantly superior to acetaminophen. Intra-articular placebo was statistically significantly better than oral placebo (effect size, 0.29 [CrI, 0.04 to 0.54]). Intra-articular treatments were more effective than oral treatments (**Supplement Figure 2** and **Figure 3**). No significant disagreement between direct and indirect estimates (inconsistency) was detected (**Supplement Figure 4**).

## Function

Seventy-six trials (24 059 participants) contributed to the analyses of physical function outcomes (**Supplement Figure 5**). All interventions except IA corticosteroids were statistically significantly superior to oral placebo (**Supplement Table 5**), with effect sizes ranging from 0.15 to 0.45. Naproxen, ibuprofen, diclofenac, and celecoxib were statistically significantly better than acetaminophen. Intra-articular hyaluronic acid was statistically significantly better than IA placebo and IA corticosteroids. Intra-articular placebo was not significantly better than oral placebo (effect size, 0.15 [CrI, −0.22 to 0.53]). No significant inconsistency was detected between the direct and indirect estimates (**Supplement Figure 6**).

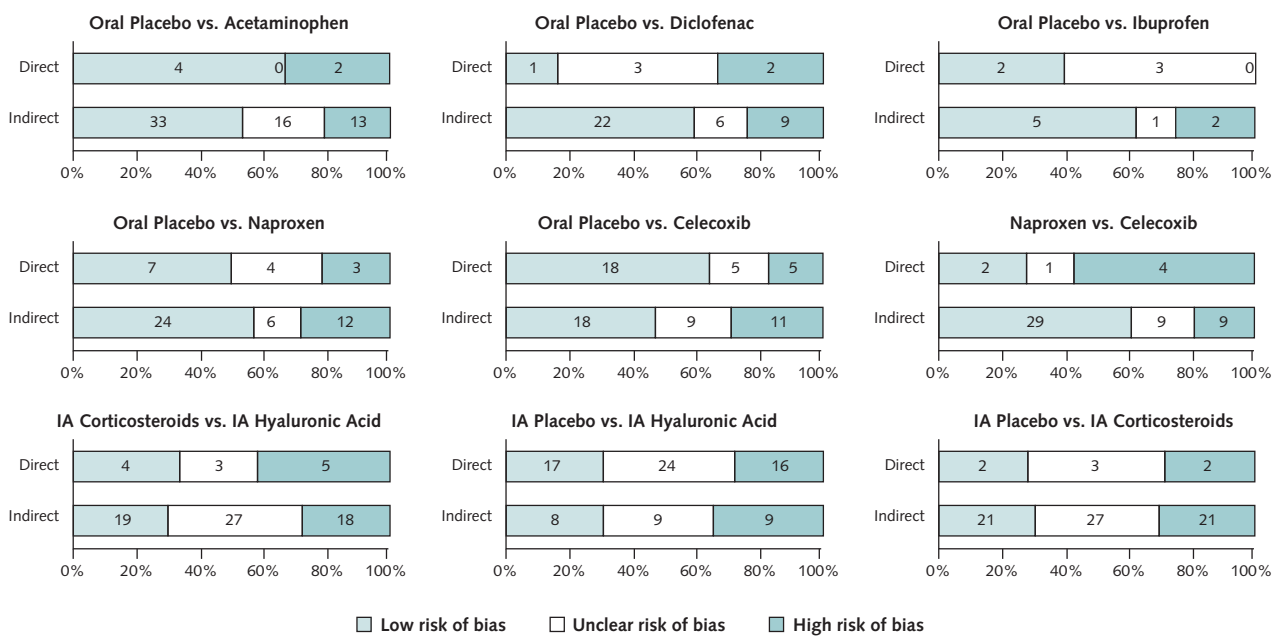
## Stiffness

Fifty-five trials (18 267 participants) contributed to the analyses of stiffness outcomes (**Supplement Figure 7**). Naproxen, ibuprofen, diclofenac, and celecoxib were statistically significantly better than oral placebo and acetaminophen (**Supplement Table 6**). Intra-articular hyaluronic acid was statistically significantly better than IA placebo. Intra-articular placebo was not significantly better than oral placebo (effect size, 0.10 [CrI, −0.26 to 0.46]). No significant disagreement between direct and indirect estimates (inconsistency) was detected (**Supplement Figure 8**).

## Sensitivity Analysis

Several sensitivity analyses for pain confirmed the robustness of our results. Studies reporting WOMAC pain scales showed higher effect sizes than the studies

***Figure 2.*** Summary of risk-of-bias assessment: direct and indirect evidence for each pairwise comparison.




Only comparisons with at least 5 direct head-to-head trials were plotted. The numbers inside the bars indicate the number of trials within each comparison that had low, unclear, or high risk of bias. The *x*-axis illustrates the percentage of trials with low, unclear, or high risk of bias for each comparison. IA = intra-articular.

*Table 1.* Effect Sizes for Pain at 3 Months*

| Comparator | Effect Sizes (95% CrI), by Treatment | | | |
|---|---|---|---|---|
| | Acetaminophen | IA Placebo | Celecoxib | Naproxen |
| Oral placebo | 0.18 (0.04 to 0.33)† | 0.29 (0.04 to 0.54)† | 0.33 (0.25 to 0.42)† | 0.38 (0.27 to 0.49)† |
| Acetaminophen | – | 0.11 (−0.17 to 0.38) | 0.15 (0.00 to 0.30) | 0.20 (0.03 to 0.37)† |
| IA placebo | – | – | 0.04 (−0.21 to 0.30) | 0.09 (−0.15 to 0.34) |
| Celecoxib | – | – | – | 0.05 (−0.08 to 0.17) |
| Naproxen | – | – | – | – |
| Ibuprofen | – | – | – | – |
| Diclofenac | – | – | – | – |
| IA corticosteroids | – | – | – | – |

CrI = credible interval; IA = intra-articular.
* Standardized mean differences adjusted for small samples (Hedges *g*).
† Statistically significant effect sizes. Effect sizes favor the above (column heading) intervention in each comparison.

reporting pain on other scales, which was more evident for the IA treatments (Supplement Table 7). In an analysis of the 86 trials that randomly assigned 50 or more participants per group, the effect sizes were consistent with those in the primary analysis (Supplement Table 7). However, in another subgroup analysis of the 56 trials that randomly assigned 100 or more participants per group, the effect sizes for the IA treatments were smaller than those in the primary analysis results, although IA hyaluronic acid was still more effective than all treatments except for diclofenac. We could not rule out reporting bias because the results of studies reporting pain alone differed from those reporting function and stiffness for some treatments (Supplement Table 8). In particular, the effects for diclofenac, celecoxib, IA corticosteroids, and acetaminophen were stronger than oral placebo in the studies reporting pain alone.

### Safety

A summary of adverse events reported in the included trials is presented in Supplement Table 9. In general, oral nonselective NSAIDs led to more gastrointestinal adverse events and withdrawals due to adverse events than oral placebo and acetaminophen, whereas these events were similar between acetaminophen and celecoxib. Fewer than 50% of trials involving NSAIDs and celecoxib reported on cardiovascular adverse events. The short exposure time of 2 to 3 months might be one of the reasons for the very few reported events in these trials. Withdrawals due to adverse events were more common among oral treatments (acetaminophen, nonselective NSAIDs, and celecoxib) than IA therapies. The most commonly reported adverse events among the IA therapies were transient local reactions, such as pain, swelling, and arthralgia, which usually subsided in a few days. These events were reported to be similar between different IA therapies (corticosteroids and hyaluronic acid). Among the 29 trials reporting on septic arthritis, 1 patient who received IA placebo had a septic joint out of 3152 patients who had approximately 9500 IA injections.

## DISCUSSION

Our network meta-analysis provides evidence-based estimates of the relative efficacy of widely used treatments for knee OA. We found that for pain outcomes, all interventions were significantly better than oral placebo. Except for celecoxib, all active interventions were significantly better than acetaminophen. Intra-articular placebo was significantly better than oral placebo, and active IA treatments were more effective than active oral treatments. Most treatments did not significantly differ in function and stiffness outcomes, in part because of the limited number of trials investigating many of the studied interventions (inevitably resulting in wider CrIs).

One striking aspect of our results is that IA therapies were the most effective treatments for knee OA pain. This result is especially salient for hyaluronic acid, which is a treatment generally considered by expert panels to be minimally effective (25, 26). For example, a recent meta-analysis found a nonclinically important benefit of hyaluronic acid in relation to an IA placebo (effect size, 0.37 [CI, 0.28 to 0.46]) (25). Compared with traditional meta-analyses, the larger effect of hyaluronic acid generated by our network meta-analysis seems to derive from the use of the IA delivery method itself, which we found to have a significant effect (effect size, 0.29 [CrI, 0.04 to 0.54]) compared with oral placebo—an effect inherently unobservable in a traditional meta-analysis. Similarly, we found that none of the oral NSAIDs were significantly superior to IA placebo. Thus, in a network meta-analysis comparing all interventions with an oral placebo, the effect size estimates for IA therapies were boosted by a benefit derived from the effect of the intervention's delivery. Indeed, our estimate of the effect of IA hyaluronic acid versus IA placebo is nearly identical to that observed in a conventional meta-analysis (25, 27).

This observation raises many important philosophical and therapeutic questions about the extent to which this benefit is attributable to a true placebo response or physiologic effects after injecting a fluid by means of a needle into the knee joint. The latter could potentially influence nociceptive response through effects on the peripheral nervous system and have pathophysiologic benefits, especially if any fluid were aspirated (28–30). However, regardless of the mechanism of the apparent benefit attributable to needle placement in the knee, the practical reality is that this

*Table 1*–Continued

| Effect Sizes (95% CrI), by Treatment | | | |
|---|---|---|---|
| **Ibuprofen** | **Diclofenac** | **IA Corticosteroids** | **IA Hyaluronic Acid** |
| 0.44 (0.25 to 0.63)† | 0.52 (0.34 to 0.69)† | 0.61 (0.32 to 0.89)† | 0.63 (0.39 to 0.88)† |
| 0.26 (0.05 to 0.47)† | 0.33 (0.12 to 0.54)† | 0.42 (0.12 to 0.73)† | 0.45 (0.18 to 0.72)† |
| 0.15 (−0.13 to 0.44) | 0.23 (−0.03 to 0.49) | 0.32 (0.16 to 0.47)† | 0.34 (0.26 to 0.42)† |
| 0.11 (−0.10 to 0.31) | 0.18 (−0.01 to 0.37) | 0.27 (−0.02 to 0.56) | 0.30 (0.04 to 0.55)† |
| 0.06 (−0.15 to 0.27) | 0.13 (−0.07 to 0.33) | 0.22 (−0.06 to 0.51) | 0.25 (0.01 to 0.49)† |
| – | 0.07 (−0.17 to 0.32) | 0.16 (−0.15 to 0.48) | 0.19 (−0.09 to 0.47) |
| – | – | 0.09 (−0.20 to 0.38) | 0.11 (−0.14 to 0.37) |
| – | – | | 0.02 (−0.12 to 0.17) |

procedure contributes to the overall benefit conferred in clinical practice. Elimination of these effects when evaluating the overall effect of IA therapies thus underestimates their clinical benefit. The effect size produced by the multiple-treatment comparison approach is therefore closer to that which would be seen in clinical practice. Results of other comparisons were generally as expected and consistent with prior direct comparisons (**Table 2**) (25, 27, 31–38).

Contrary to popular belief, our results showed that celecoxib was not superior to acetaminophen (effect size, 0.15 [CrI, 0.00 to 0.30]). The robustness of our results was confirmed by performing a direct pairwise meta-analysis in which we analyzed data from 4 randomized, controlled trials that compared acetaminophen ($n$ = 712) with celecoxib ($n$ = 978). Although the effect size from this direct comparison favored celecoxib (effect size, 0.19 [0.09 to 0.029]), it was not very different from our primary estimate. When doing pairwise or network analyses with random-effects models, different between-study heterogeneities may be estimated, which potentially explains the apparent disagreement in conclusions. Moreover, celecoxib's lower performance in other trials, especially versus placebo, may have contributed to the wider CrIs of the network's estimates. The acetaminophen comparison trials may have elicited better response rates in the control group than the placebo-controlled trials because of patients' expectations that they were receiving an active treatment regardless of treatment allocation. Of note, celecoxib was not significantly inferior to other oral NSAIDs. Our results do not completely discredit the efficacy of celecoxib, yet they may call into question the rationale for its use and prompt caution in estimating the risk–benefit ratios for patients with multiple comorbid conditions.

We chose to compare these treatments on the basis of recommendations from the latest clinical practice guidelines for knee OA (7, 8) and current prescription patterns worldwide (6). Acetaminophen is the most commonly used over-the-counter pain medication for knee OA. We chose ibuprofen and diclofenac because they are the most commonly prescribed drugs by primary care physicians (6). We chose naproxen because it is one of the most highly prescribed drugs worldwide for knee OA (6). Moreover, naproxen is favored by health insurance providers as the cheapest generic NSAID, and it is considered to be a benchmark drug by the pharmaceutical industry. We chose celecoxib as a representative of the cyclooxygenase-2 inhibitor NSAID family because it is a widely prescribed drug in this group (6) and is considered a reference drug of the group by the pharmaceutical industry. Corticosteroids have been widely used for decades to treat knee OA, yet questions remain about their efficacy and safety. Hyaluronic acid is a widely used newer product for which convincing evidence of efficacy has been lacking. This model can easily be expanded in the future to include all OA remedies. All of the OA clinical practice guidelines recommend nonpharmacologic therapies, but several methodological challenges prevented us from incorporating these into our network. Such challenges were that most of these trials involved patients with any type of arthritis or patients with OA in any joint site (such as knee, hip, or hand); most of the trials examining nonpharmacologic therapies were either open-label or single-blind and were of lower quality; and almost all of these trials were compared against differing controls, such as no intervention or standard of care, which cannot feasibly be compared with pharmacologic treatments. All of these factors affect the underlying assumptions of transitivity and heterogeneity of the network meta-analysis.

Although our review included 137 studies, limitations and differences in the reporting of data (selective outcome reporting bias) restricted the data available for our analyses (129 for pain, 76 for function, and 55 for stiffness). The baseline characteristics among all trials reporting pain; those reporting both pain and function; and those reporting pain, function, and stiffness were similar, but the trial quality measures differed. Trials reporting outcomes other than pain more thoroughly reported methods (**Supplement Table 10**). One possible reason for better reporting among the trials reporting function, stiffness, and pain could be that these trials were conducted between 1999 and 2013, which is when the CONSORT (Consolidated Standards of Reporting Trials) guidelines for reporting randomized trials were established. Of note, all of the trials reporting stiffness and 91% of the trials reporting function used WOMAC scales to measure outcomes. The subset of trials reporting pain alone had a higher rate of pain reduction than other subsets (**Supplement Table 8**). Intra-articular hyaluronic acid and IA placebo were the only interventions whose results did not differ substantially among these subsets of trials, suggesting

*Table 2.* Comparison of Effect Sizes Between Other Meta-analyses and the Network Meta-analysis for Pain

| Comparison* | Other Meta-analyses | | | Network Meta-analysis | | Are the Effect Sizes Comparable? | |
|---|---|---|---|---|---|---|---|
| | Study, Year (Reference) | Effect Size (95% CI) | Trials, *n* | Effect Size (95% CrI) | Trials, *n* | Overlapping Studies, %† | Overlapping CrIs |
| Acetaminophen vs. oral placebo | Zhang et al, 2004 (31) | 0.21 (0.02 to 0.41) | 2 | 0.18 (0.04 to 0.33) | 6 | 100 | Yes |
| NSAIDs (pooled) vs. oral placebo | Lee et al, 2005 (32) | 0.37 (0.26 to 0.49) | 10 | 0.44 (0.34 to 0.55)‡ | 25 | 100 | Yes |
| Diclofenac vs. oral placebo | Stam et al, 2012 (33) | 0.49 (0.31 to 0.67) | 3 | 0.52 (0.34 to 0.69) | 6 | 100 | Yes |
| Ibuprofen vs. oral placebo | Stam et al, 2012 (33) | 0.41 (0.18 to 0.63) | 2 | 0.44 (0.25 to 0.63) | 5 | 100 | Yes |
| Naproxen vs. oral placebo | Stam et al, 2012 (33) | 0.39 (0.26 to 0.53) | 5 | 0.38 (0.27 to 0.49) | 14 | 100 | Yes |
| Celecoxib vs. oral placebo | Lee et al, 2005 (32) | 0.26 (0.15 to 0.37) | 4 | 0.33 (0.25 to 0.42) | 28 | 75 | Yes |
| | Stam et al, 2012 (33) | 0.34 (0.27 to 0.41) | 13 | | | 100 | Yes |
| NSAIDs vs. acetaminophen | Zang et al, 2004 (31) | 0.20 (0.10 to 0.30) | 8§ | 0.26 (0.10 to 0.42)‡ | 5 | 100 | Yes |
| NSAIDs vs. IA hyaluronic acid | Bannuru et al, 2014 (34) | 0.07 (−0.10 to 0.24) | 5 | 0.19 (−0.03 to 0.42)‡ | 5 | 100 | Yes |
| IA hyaluronic acid vs. IA placebo | Rutjes et al, 2012 (25) | 0.37 (0.28 to 0.46) | 71‖ | 0.34 (0.26 to 0.42) | 52 | 100 | Yes |
| | Bannuru et al, 2011 (27) | 0.34 (0.22 to 0.46) | 45 | | | 100 | Yes |
| IA corticosteroids vs. IA placebo | Bellamy et al, 2006 (35) | 21.91 (13.89 to 29.93)¶ | 4 | 0.32 (0.16 to 0.47) | 7 | 75 | Not accessible |
| | Arrol et al, 2004 (36) | 16.47 (10.03 to 22.92)¶ | 5 | | | 60 | Not accessible |
| IA hyaluronic acid vs. IA corticosteroids | | | | | | | |
| 4-wk follow-up | Bannuru et al, 2009 (37) | −0.01 (−0.23 to 0.21) favoring IA corticosteroids | 7 | 0.02 (−0.12 to 0.17) favoring IA hyaluronic acid | 12 | 100 | Yes |
| 8-wk follow-up | Bannuru et al, 2009 (37) | 0.22 (−0.05 to 0.49) favoring IA hyaluronic acid | 5 | | | | No |

CI = confidence interval; CrI = credible interval; IA = intra-articular; NSAID = nonsteroidal anti-inflammatory drug.
* Positive effect sizes favor the first comparator listed; negative effect sizes favor the second comparator.
† Percentage of trials from the previous pairwise meta-analysis that were eligible to be part of the network.
‡ Results from sensitivity analysis.
§ Effect size was generated from nonselective and cyclooxygenase-2–selective NSAIDs.
‖ Effect size was generated from studies using heterogeneous comparators (IA placebo, NSAIDs, or combination treatments).
¶ Weighted mean difference (scale, 0 to 100).

that reporting bias among the trials comparing these 2 interventions is minimal.

A wide range of scales and pain activities were used in the studies in our review; focusing the analysis on a single measure (for example, the WOMAC–Likert scale) would have seriously limited the amount of evidence that could be combined in an analysis. In addition, transforming the 5-point Likert scale and 100-mm VAS to the same scale may be an extreme assumption given that their SDs may not be transformable on the same scale. In these situations, converting treatment effects to standardized mean differences is a standard approach (10). In a subgroup analysis, trials reporting pain on the WOMAC scale showed higher effect sizes than those reporting pain on other scales (**Supplement Table 7**). For example, the effect size for IA hyaluronic acid was 0.80 as measured by the WOMAC scale versus 0.54 on the VAS. Acetaminophen had an effect size of 0.21 on the WOMAC scale versus 0.14 on the VAS.

One of the key assumptions in a network meta-analysis is that of transitivity between trials (13). We sought to minimize possible concerns about within- or between-intervention heterogeneity by using strict inclusion criteria and by excluding interventions with more than 1 treatment. The patient characteristics seemed broadly similar across interventions. Some clinical heterogeneity is inevitable in a wide-ranging study such as this, but baseline pain did not seem to vary systematically between interventions as far as it was possible to tell, given the wide variation of scales used. Most included trials enrolled patients with Kellgren–Lawrence grade 2 or 3 knee OA. Because there is no step-up or step-down care model for knee OA, patients randomly assigned in these trials could be eligible for any of the treatments considered here. We conducted sensitivity analyses excluding trials causing heterogeneity. Our analyses used a random-effects model to incorporate heterogeneity, and we evaluated levels of inconsistency and model fit (39). We did not find inconsistency in any of our models. Despite this precaution, it might be possible that unmeasured effect modifiers influenced the results of indirect comparisons.

Our rigorous search strategy minimized the risk for missing eligible studies, and an active and extensive search for unpublished literature yielded 13 unpublished trials (3 involving celecoxib, 2 involving naproxen, and 9 involving hyaluronic acid). Nevertheless, we cannot completely rule out publication bias, so our results must be interpreted with caution. Higher efficacy is more commonly found in smaller trials than in

larger ones for various reasons (10). For example, drugs with little efficacy will need trials with larger numbers of participants to attain statistical significance compared with more effective drugs. This may have been the case in trials of celecoxib versus placebo, which randomly assigned an average of about 410 participants (range, 123 to 844). Celecoxib had an effect size of 0.33 (0.25 to 0.42) and was less effective than all therapies but acetaminophen. Our sensitivity analysis based on trial size (small study bias) showed that trials that randomly assigned 50 or more participants per group had effect sizes similar to the primary analysis. However, trials that randomly assigned 100 or more participants per group had smaller effect sizes for IA treatments than the primary analysis results. Despite the lower effect sizes for IA treatments, IA hyaluronic acid was still more effective than all treatments but diclofenac. We could not analyze the trials that randomly assigned fewer than 50 participants per group separately because of the lack of a well-connected network.

Our study has several limitations. It should be noted when interpreting our results that trials investigating combination therapies were excluded. Although combination therapy is common in practice, the large number of possible therapy combinations and scarcity of trials comparing these treatments prompted us to restrict our analysis to monotherapies. The decision to exclude combination treatments was also intended to minimize heterogeneity among interventions because it was believed that combination therapies would probably vary considerably, which would preclude meaningful pooling of data. Further, the lack of long-term data (1 or 2 years) limits the interpretation of our results to only the short-term effects of the included treatments. Regardless, the knowledge of which treatments would probably work better for a moderate treatment duration (such as 3 months), as opposed to a shorter one, was believed to be advantageous for clinicians because of the chronic course of OA. None of the studies provided data on the tolerability of the included interventions, and too few studies provided data on quality-of-life measurements to reach meaningful conclusions. Inadequate reporting of adverse event data, heterogeneity among reported data, and short trial duration precluded us from doing a quantitative safety review of the interventions in this study.

Our results might be more applicable to localized knee OA than to multijoint OA. In general, NSAIDs are used to treat generalized OA involving multiple joints and IA treatments are used locally to treat 1 or 2 large joints. The oral treatments, including acetaminophen, have well-known risks, such as gastrointestinal, cardiac, and renal adverse events. In older populations in which OA is prevalent and multiple comorbid conditions and polypharmacy are common, local treatments not associated with such risks are thus preferable.

This network meta-analysis compared the most commonly used pharmacologic interventions for knee OA–related pain at 3 months and concluded that all treatments except acetaminophen showed clinically significant improvement in pain. Intra-articular treatments were more effective than NSAIDs for pain, which is possibly due to the contribution of the integrated IA placebo effect. This information, along with the safety profiles and relative costs of included treatments, should be helpful to clinicians when making care decisions tailored to individual patient needs.

From Center for Treatment Comparison and Integrative Analysis, Predictive Analytics and Comparative Effectiveness Center, and Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, and Sackler School of Graduate Biomedical Sciences, Tufts University, Boston, Massachusetts, and Center for Evidence Based Medicine and Brown University School of Public Health, Providence, Rhode Island.

**Disclaimer:** The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality (AHRQ).

**Acknowledgment:** The authors thank Matthew Sullivan for his contributions to this work.

**Grant Support:** By the AHRQ (grants F32HS021396 and R01-HS018574; Drs. Bannuru and Schmid, respectively).

**Disclosures:** Dr. Bannuru reports grants from AHRQ during the conduct of the study. Dr. Wong reports grants from Patient-Centered Outcomes Research Institute and National Institutes of Health National Center for Complementary and Alternative Medicine and nonfinancial support from European League Against Rheumatism outside the submitted work. Dr. McAlindon reports grants from Croma, Flexion Therapeutics, National Institutes of Health, and AHRQ; and personal fees from Bioventus outside the submitted work. Authors not named here have disclosed no conflicts of interest. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M14-1231.

**Requests for Single Reprints:** Raveendhara R. Bannuru, MD, Center for Treatment Comparison and Integrative Analysis, Division of Rheumatology, Tufts Medical Center, 800 Washington Street, Box 406, Boston, MA 02111; e-mail, rbannuru@tuftsmedicalcenter.org.

Current author addresses and author contributions are available at www.annals.org.

## References

1. **Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al.** Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380:2163-96. [PMID: 23245607] doi:10.1016/S0140-6736(12)61729-2

2. **Guccione AA, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, et al.** The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. Am J Public Health. 1994;84:351-8. [PMID: 8129049]

3. **Kosorok MR, Omenn GS, Diehr P, Koepsell TD, Patrick DL.** Restricted activity days among older adults. Am J Public Health. 1992; 82:1263-7. [PMID: 1503169]

4. The incidence and prevalence database for procedures. Sunnyvale ,CA: Timely Data Resources. 1995:592-6.

5. **White AG, Birnbaum HG, Janagap C, Buteau S, Schein J.** Direct and indirect costs of pain therapy for osteoarthritis in an insured pop-

ulation in the United States. J Occup Environ Med. 2008;50:998-1005. [PMID: 18784547] doi:10.1097/JOM.0b013e3181715111
6. **McGettigan P, Henry D.** Use of non-steroidal anti-inflammatory drugs that elevate cardiovascular risk: an examination of sales and essential medicines lists in low-, middle-, and high-income countries. PLoS Med. 2013;10:e1001388. [PMID: 23424288] doi:10.1371/journal.pmed.1001388
7. **Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al; American College of Rheumatology.** American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken). 2012;64:465-74. [PMID: 22563589]
8. **McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al.** OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage. 2014;22:363-88. [PMID: 24462672] doi:10.1016/j.joca.2014.01.003
9. **Juhl C, Lund H, Roos EM, Zhang W, Christensen R.** A hierarchy of patient-reported outcomes for meta-analysis of knee osteoarthritis trials: empirical evidence from a survey of high impact journals. Arthritis. 2012;2012:136245. [PMID: 22792458] doi:10.1155/2012/136245
10. **Higgins JPT, Green S, eds.** Cochrane Handbook for Systematic Reviews of Interventions. Version 5.10. The Cochrane Collaboration; 2011. Accessed at http://handbook.cochrane.org on 13 November 2014.
11. **Cooper H, Hedges LV.** The Handbook of Research Synthesis. New York: Russell Sage Foundation; 1994.
12. **Dias S, Welton NJ, Sutton AJ, Ades AE.** Evidence synthesis for decision making 1: introduction. Med Decis Making. 2013;33:597-606. [PMID: 23804506] doi:10.1177/0272989X13487604
13. **Salanti G.** Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Research Synthesis Methods. 2012;3:80-97. doi: 10.1002/jrsm.1037
14. **Dias S, Sutton AJ, Ades AE, Welton NJ.** Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making. 2013;33:607-17. [PMID: 23104435] doi:10.1177/0272989X12458724
15. **Franchini A, Dias S, Ades A, Jansen J, Welton N.** Accounting for correlation in network meta-analysis with multi-arm trials. Research Synthesis Methods. 2012;3:142-60. doi: 10.1002/sim.6321
16. **Higgins JP, Whitehead A.** Borrowing strength from external trials in a meta-analysis. Stat Med. 1996;15:2733-49. [PMID: 8981683]
17. **Smith TC, Spiegelhalter DJ, Thomas A.** Bayesian approaches to random-effects meta-analysis: a comparative study. Stat Med. 1995;14:2685-99. [PMID: 8619108]
18. **Lu G, Ades AE.** Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004;23:3105-24. [PMID: 15449338]
19. **Baker SG, Kramer BS.** The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C? BMC Med Res Methodol. 2002;2:13. [PMID: 12429069]
20. **Cipriani A, Higgins JP, Geddes JR, Salanti G.** Conceptual and technical challenges in network meta-analysis. Ann Intern Med. 2013;159:130-7. [PMID: 23856683] doi:10.7326/0003-4819-159-2-201307160-00008
21. **Dias S, Welton NJ, Caldwell DM, Ades AE.** Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010;29:932-44. [PMID: 20213715] doi:10.1002/sim.3767
22. **Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH.** Pooling health-related quality of life outcomes in meta-analysis–a tutorial and review of methods for enhancing interpretability. Research Synthesis Methods. 2011;2:188-203. doi: 10.1002/jrsm.46
23. **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 Cartilage. 2004;12:389-99. [PMID: 15094138]
24. **Dias S, Sutton AJ, Welton NJ, Ades AE.** Evidence synthesis for decision making 3: heterogeneity–subgroups, meta-regression, bias, and bias-adjustment. Med Decis Making. 2013;33:618-40. [PMID: 23804507] doi:10.1177/0272989X13485157
25. **Rutjes AW, Jüni P, da Costa BR, Trelle S, Nüesch E, Reichenbach S.** Viscosupplementation for osteoarthritis of the knee: a systematic review and meta-analysis. Ann Intern Med. 2012;157:180-91. [PMID: 22868835] doi:10.7326/0003-4819-157-3-201208070-00473
26. **American Academy of Orthopaedic Surgeons.** Treatment of osteoarthritis of the knee: evidence-based guideline. 2nd edition. Rosemont, IL: American Academy of Orthopaedic Surgeons; 2013. Accessed at www.aaos.org/research/guidelines/treatmentofOsteoarthritisoftheKneeGuideline.pdf on 13 November 2014.
27. **Bannuru RR, Natov NS, Dasi UR, Schmid CH, McAlindon TE.** Therapeutic trajectory following intra-articular hyaluronic acid injection in knee osteoarthritis–meta-analysis. Osteoarthritis Cartilage. 2011;19:611-9. [PMID: 21443958] doi:10.1016/j.joca.2010.09.014
28. **Zhang RX, Ren K, Dubner R.** Osteoarthritis pain mechanisms: basic studies in animal models. Osteoarthritis Cartilage. 2013;21:1308-15. [PMID: 23973145] doi:10.1016/j.joca.2013.06.013
29. **Gulen H, Ataoglu H, Haliloglu S, Isik K.** Proinflammatory cytokines in temporomandibular joint synovial fluid before and after arthrocentesis. Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 2009;107:e1-4. [PMID: 19426899] doi:10.1016/j.tripleo.2009.02.006
30. **Kirwan JR, Rankin E.** Intra-articular therapy in osteoarthritis. Baillieres Clin Rheumatol. 1997;11:769-94. [PMID: 9429736]
31. **Zhang W, Jones A, Doherty M.** Does paracetamol (acetaminophen) reduce the pain of osteoarthritis? A meta-analysis of randomised controlled trials. Ann Rheum Dis. 2004;63:901-7. [PMID: 15020311]
32. **Lee C, Hunsche E, Balshaw R, Kong SX, Schnitzer TJ.** Need for common internal controls when assessing the relative efficacy of pharmacologic agents using a meta-analytic approach: case study of cyclooxygenase 2-selective inhibitors for the treatment of osteoarthritis. Arthritis Rheum. 2005;53:510-8. [PMID: 16082648]
33. **Stam W, Jansen J, Taylor S.** Efficacy of etoricoxib, celecoxib, lumiracoxib, non-selective NSAIDs, and acetaminophen in osteoarthritis: a mixed treatment comparison. Open Rheumatol J. 2012;6:6-20. [PMID: 22582102] doi:10.2174/1874312901206010006
34. **Bannuru RR, Vaysbrot EE, Sullivan MC, McAlindon TE.** Relative efficacy of hyaluronic acid in comparison with NSAIDs for knee osteoarthritis: a systematic review and meta-analysis. Semin Arthritis Rheum. 2014;43:593-9. [PMID: 24216297] doi:10.1016/j.semarthrit.2013.10.002
35. **Bellamy N, Campbell J, Robinson V, Gee T, Bourne R, Wells G.** Intraarticular corticosteroid for treatment of osteoarthritis of the knee. Cochrane Database Syst Rev. 2006:CD005328. [PMID: 16625636]
36. **Arroll B, Goodyear-Smith F.** Corticosteroid injections for osteoarthritis of the knee: meta-analysis. BMJ. 2004;328:869. [PMID: 15039276]
37. **Bannuru RR, Natov NS, Obadan IE, Price LL, Schmid CH, McAlindon TE.** Therapeutic trajectory of hyaluronic acid versus corticosteroids in the treatment of knee osteoarthritis: a systematic review and meta-analysis. Arthritis Rheum. 2009;61:1704-11. [PMID: 19950318] doi:10.1002/art.24925
38. **Bellamy N, Campbell J, Robinson V, Gee T, Bourne R, Wells G.** Viscosupplementation for the treatment of osteoarthritis of the knee. Cochrane Database Syst Rev. 2006:CD005321. [PMID: 16625635]
39. **Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ.** Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Stat Med. 2009;28:1861-81. [PMID: 19399825] doi:10.1002/sim.3594

**Current Author Addresses:** Drs. Bannuru, Vaysbrot, and McAlindon: Center for Treatment Comparison and Integrative Analysis, Division of Rheumatology, Tufts Medical Center, 800 Washington Street, Boston, MA 02111.
Drs. Kent and Wong: Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, 800 Washington Street, Boston, MA 02111.
Dr. Schmid: Brown University School of Public Health, 121 South Main Street, Providence, RI 02912.

**Author Contributions:** Conception and design: R.R. Bannuru, C.H. Schmid, T.E. McAlindon.
Analysis and interpretation of the data: R.R. Bannuru, C.H. Schmid, D.M. Kent, J.B. Wong, T.E. McAlindon.
Drafting of the article: R.R. Bannuru, C.H. Schmid, J.B. Wong.
Critical revision of the article for important intellectual content: R.R. Bannuru, C.H. Schmid, D.M. Kent, E.E. Vaysbrot, J.B. Wong, T.E. McAlindon.
Final approval of the article: R.R. Bannuru, C.H. Schmid, D.M. Kent, E.E. Vaysbrot, J.B. Wong, T.E. McAlindon.
Provision of study materials or patients: R.R. Bannuru.
Statistical expertise: R.R. Bannuru, C.H. Schmid.
Obtaining of funding: R.R. Bannuru.
Administrative, technical, or logistic support: R.R. Bannuru, E.E. Vaysbrot.
Collection and assembly of data: R.R. Bannuru, E.E. Vaysbrot.

***Appendix Figure.*** Summary of evidence search and selection.