Osteoarthritis and Cartilage

Comparative pain reduction of oral non-steroidal anti-inflammatory drugs and opioids for knee osteoarthritis: systematic analytic review

S.R. Smith †, B.R. Deshpande †, J.E. Collins ††, J.N. Katz †††‖, E. Losina †††‖

† Orthopaedic and Arthritis Center for Outcomes Research, Department of Orthopedic Surgery, Brigham and Women’s Hospital, Boston, MA, USA
‖ Harvard Medical School, Brigham and Women’s Hospital, Boston, MA, USA
‖ Department of Epidemiology and Environmental Health, Harvard T. H. Chan School of Public Health, Boston, MA, USA
‖ Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

Objective: Summarize the comparative effectiveness of oral non-steroidal anti-inflammatory drugs (NSAIDs) and opioids in reducing knee osteoarthritis (OA) pain.

Methods: Two reviewers independently screened reports of randomized controlled trials (RCTs), published in English between 1982 and 2015, evaluating oral NSAIDs or opioids for knee OA. Included studies were at least 8 weeks duration, conducted in Western Europe, the Americas, New Zealand, or Australia, and reported baseline and follow-up pain using the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) Pain subscale (0–100, 100-worst). Effectiveness was evaluated as reduction in pain, accounting for study dropout and heterogeneity.

Results: Twenty seven treatment arms (nine celecoxib, four non-selective NSAIDs [diclofenac, naproxen, piroxicam], 11 less potent opioids [tramadol], and three potent opioids [hydromorphone, oxycodone]) from 17 studies were included. NSAID and opioid studies reported similar baseline demographics and efficacy withdrawal rates; NSAID studies reported lower baseline pain and toxicity withdrawal rates. Accounting for efficacy-related withdrawals, all drug classes were associated with similar pain reductions (NSAIDs: −18; less potent opioids: −18; potent opioids: −19). Meta-regression did not reveal differential effectiveness by drug class but found that study cohorts with a higher proportion of male subjects and worse mean baseline pain had greater pain reduction. Similarly, results of the network meta-analysis (NMA) did not find a significant difference in WOMAC Pain reduction for the three analgesic classes.

Conclusion: NSAIDs and opioids offer similar pain relief in OA patients. These data could help clinicians and patients discuss likely benefits of alternative analgesics.

© 2016 Published by Elsevier Ltd on behalf of Osteoarthritis Research Society International.

Introduction

Knee osteoarthritis (OA) affects millions of American adults and is characterized by joint pain, joint stiffness, and functional limitations. Although over half of all knee OA patients eventually undergo total knee replacement, nearly all will require at least some amount of long-term pain control.

Standard treatment begins with non-pharmacologic approaches to symptom relief and functional restoration, including weight reduction, orthotic devices, exercise, and physical therapy. Because these treatments often provide limited pain relief, pharmacologic analgesics are frequently also employed. Many professional societies suggest the use of non-steroidal anti-inflammatory drugs (NSAIDs) or tramadol, a lower potency opioid, for primary pharmacologic management of knee OA. Recommendations on the use of more potent opioids remain conflicted for this population.

Both NSAIDs and opioids are associated with a wide variety of adverse effects, and there are no long-term trials of knee OA

Please cite this article in press as: Smith SR, et al., Comparative pain reduction of oral non-steroidal anti-inflammatory drugs and opioids for knee osteoarthritis: systematic analytic review, Osteoarthritis and Cartilage (2016), http://dx.doi.org/10.1016/j.joca.2016.01.135
patients comparing their efficacies. Given the limited comparative
evaluation of oral NSAIDs and opioids in the knee OA population as
well as the importance of understanding the long-term effective-
ness of the each of the classes of drugs through decision modeling
and formal comparative-effectiveness analyses, we employed a
systematic analytic review to evaluate both classes of analgesics in
reducing pain among persons with OA.

Methods

We conducted our analysis according to the principles of
Preferred Reporting Items for Systematic Reviews and Meta-Ana-
lyses (PRISMA) guidelines7. Our analysis was not preregistered.

Identification of studies

We conducted a search of all articles available in PubMed, Web
of Science – Science Citation Index Expanded, EMBASE, and the
Cochrane Central Register of Controlled Trials. Our search identi-
fied all articles including OA and any of the following terms in the
title: NSAID(s), ibuprofen, celecoxib, diclofenac, naproxen, melox-
icam, nabumetone, etodolac, indomethacin, piroxicam, sulindac, salsulinate, flurbiprofen, ketoprofen, oxycodon, hydrocodone, hydro-
morphone, fentanyl, methadone, morphine, tramadol, or codeine.
These search terms reflect NSAIDs and opioids commonly pre-
scribed to US Medicare beneficiaries with knee OA as of 20098,
excluding the since-withdrawn propoxophene9. Two reviewers
(SRS and BRD) independently screened the abstract of each article
to determine whether it was a randomized controlled trial (RCT)
conducted in humans and published between January 1, 2000 and
March 6, 2015.

Inclusion and exclusion criteria

We included clinical trials of predominantly knee OA patients
of at least 8 weeks duration that evaluated efficacy of oral anal-
gesics using the Western Ontario and McMaster Universities
Osteoarthritis Index (WOMAC) Pain subscale10; study arms that
combined patients with OA and patients with other forms of
arthritis were excluded. We included only reports of RCTs, as other
study designs do not restrict concurrent treatment utilization and
would not provide measures of pain severity pre- and post-
treatment initiation. As more invasive placeboes have been asso-
ciated with greater effectiveness11,12, we excluded studies
employing combination therapies of oral analgesics and non-oral
placeboes to provide a more homogeneous basis for our direct and
indirect comparisons. Studies not reporting group mean and
standard deviation (SD) values for baseline pain and either change
from baseline or follow-up pain were excluded for insufficient
data. When possible, we calculated SDs from reported standard
errors or confidence intervals. Additional exclusion criteria elimi-
nated studies that were not published in English or were pri-
marily conducted outside of developed countries (defined as
Western Europe, the Americas; New Zealand, or Australia). For
studies that reported multiple follow-up time points, we selected
the one nearest 12 weeks.

Data abstraction and quality assessment

From the reports meeting all inclusion criteria, we obtained the
following data: identity and dose of drugs evaluated; funding
source; geographic location of the study; sample size; discontinu-
tions due to loss of efficacy, adverse events (clinical adverse
event, laboratory adverse event, and fatali), and other reasons;
cohort characteristics (age, gender, height, weight, body mass
index (BMI), time since OA diagnosis, race/ethnicity, and primary
joint affected [knee vs hip]); study duration; baseline WOMAC
Pain (mean, SD); and either change in WOMAC Pain (mean, SD),
follow-up WOMAC Pain (mean, SD) or both. If the range or
directionality of the scale was ambiguous, we contacted the au-
thors for clarification. Except where noted, all abstracted data
were obtained from the intention-to-treat (ITT) analysis popu-
lation. Included articles were evaluated for quality using the Jadad
assessment tool, a five-point scoring system assessing reports of
RCTs based on appropriate methods of randomization, blinding,
and withdrawal reporting.13

The two reviewers independently completed all screening, data
extraction, and quality assessment. Cases of disagreement were
discussed and resolved by the two reviewers, consulting other
authors if necessary. In a sensitivity analysis, we excluded study
arms evaluating 100 mg and 400 mg tramadol, which are not
representative of contemporary clinical practice.

Statistical analysis

We converted pain data to a 0–100 (100 worst) scale by arith-
metic transformation and evaluated cohort differences using the
t-testing method. Analyses were performed using the WinCom test
of the Social Science Statistics program.

For studies providing only baseline and final pain scores, we
converted mean change by subtracting final pain from baseline
pain. To calculate the SD of change we first calculated both the
correlation between baseline pain and change in pain as well as
the correlation between baseline pain and follow-up pain for the
studies that reported all three time points. We calculated the SD of
change for those studies that did not explicitly report it, using
properties of variances (the variance of the sum of two distribu-
tions is the sum of the distributions’ variances plus twice their
covariance) and assuming that the correlations derived among
studies that reported all three time points would also apply for
those that reported only two time points15. Finally, we modified
the mean change in pain accounting for withdrawals due to
insufficient efficacy by assuming, conservatively, that these sub-
jects would report no change from baseline. While we abstracted
data from ITT analyses, which employ methods to handle missing
data, the specific methods used were heterogeneous, including
strategies such as last observation carried forward, baseline
observation carried forward, and imputation utilizing dropout
reason. Thus, we modified mean change in pain to account for
inefficacy withdrawals to produce a more conservative estimate.
In sensitivity analyses, we used the unadjusted mean change in
pain as reported in the literature. The standard error of change
was calculated by dividing the SD by the square root of the sample
size, using the intention to treat population when reported and
the number randomized otherwise. Mean changes were com-
pared into a weighted average, weighted by the precision (the
reciprocal of the variance) of the each estimate. Separate analyses
were performed for three analgesic classes: NSAIDs, less potent
opioids, and potent opioids.

We used funnel plots and Egger’s linear regression test to
investigate publication bias. We chose Egger’s test over the rank
correlation test because the rank test has been shown to have low
power when the number of studies is small16,17. When publication
bias was suspected, we used the trim and fill method as a sensitivity
analysis18. The trim and fill is a non-parametric method to correct
for publication bias. It uses rank-based augmentation techniques to
impute potential missing studies in order to make the funnel plot
symmetric. Outcomes are re-estimated on the augmented data. To
determine if the results were robust to assumptions of the meta-
analysis, we performed heterogeneity analyses and report the H
and I² statistics for each analysis19. The contribution of each study

Please cite this article in press as: Smith SR, et al., Comparative pain reduction of oral non-steroidal anti-inflammatory drugs and opioids for
to the overall heterogeneity was assessed by the Q-term and influence. The Q-term is the contribution of the study to Cochrane’s Q statistic, and the influence is computed by comparing the overall pooled estimate with and without the study included. We used a random effects analysis using restricted maximum likelihood to calculate a final combined estimate of change in pain in order to account for heterogeneity. Finally, we used meta-regression to determine factors systematically associated with greater change in pain. We included a three-level analog class variable (NSAIDs vs less potent opioids vs potent opioids) as the primary independent variable of interest and adjusted for mean baseline pain, percent of the cohort with knee OA (vs hip), percent of the cohort that was female, study year, and country (exclusively US-based vs all other).

We conducted a secondary network meta-analysis (NMA), using the NMA framework to evaluate both direct and indirect comparisons between NSAIDs and opioids. We used a random-effects model with Gaussian quadrature to fit the model. All statistical analyses were conducted using SAS 9.4 (SAS Institute Inc., Cary, NC).

Results

Study selection

Figure 1 summarizes the article selection process. Our initial search identified 1688 unique articles: 1535 evaluating NSAIDs and 153 evaluating opioids. Upon screening the abstracts of the 940 articles published post-1999, we identified 247 articles for full text review (209 and 38 for NSAIDs and opioids, respectively). Of those, 24% (60/247) were excluded for not utilizing the WOMAC Pain subscale. Additionally, 10% (25/247) measured pain via the WOMAC Pain subscale but did not report sufficient data to be included in our analysis. Six articles (2%) were excluded for their use of non-oral placebo. We identified 17 studies meeting all outlined inclusion criteria: 11 examining just NSAIDs (celecoxib, diclofenac, naproxen, piroxicam), three examining just less potent opioids (tramadol, tramadol/acetaminophen), one examining both NSAIDs and less potent opioids (celecoxib, tramadol), and two examining potent opioids (hydromorphone, oxycodone). This resulted in 27 active treatment arms to be included in our analysis (celecoxib [9], diclofenac [1], naproxen [2], piroxicam [1], tramadol [10], tramadol/acetaminophen [1], hydromorphone [2], and oxycodone [1]).

Table I describes the included studies with selected abstracted data. Trial duration ranged from 8 to 52 weeks; median duration was 13 weeks for NSAID treatment arms and 12 weeks for opioid treatment arms (P < 0.01). The size of the treatment arms varied from 25 to 481 (median 236) persons for NSAID arms and from 60 to 202 (median 176) persons for opioid arms (P < 0.01). Several baseline patient demographics did not vary substantially from a clinical perspective for NSAID and opioid studies, with a median age of 62 years for the NSAID arms compared to 60 years for the opioid arms. Mean baseline WOMAC Pain was somewhat lower for NSAID arms (52 points) than opioid arms (60 points, P = 0.04). Treatment arms evaluating NSAIDs reported shorter median time since-diagnosis (5 years and 8 years, respectively; P = 0.03), as well as a lower median BMI (31.0 kg/m² and 32.4 kg/m² respectively; P = 0.07) and proportion of subjects withdrawing due to toxicity compared to opioid studies (7% vs 24%, respectively, P < 0.01). NSAIDs and opioid studies presented a similar median proportion of subjects withdrawing due to insufficient efficacy (7% vs 11%, respectively, P = 0.10) or any other reason (5% vs 8%, respectively, P = 0.40).

Table II shows change in WOMAC Pain values modified to account for subjects withdrawing due to insufficient efficacy, by assuming these subjects would report no change from baseline. Among the 17 included articles, 12 (71%) had a Jadad quality score of 4 or 5 (maximum score 5), and the remainder had a score of 3. All articles detailed the withdrawals and dropouts during the trials, and only two articles did not report any funding from the pharmaceutical industry.

Heterogeneity and effectiveness

NSAIDs (celecoxib, diclofenac, naproxen, piroxicam)

NSAIDs studies exhibited a large amount of heterogeneity (I² = 95.9, H = 4.58). The estimate from Hochberg et al. was the most influential (Influence = 9.35) and contributed substantial weight to the heterogeneity score (Q-term = 150). We investigated whether dropping the Hochberg study would reduce heterogeneity, as it had the largest pain decrement among NSAID studies (adjusted WOMAC Pain change = −35.6) and the highest baseline pain (74.1). Heterogeneity remained high after excluding Hochberg et al. (I² = 98.8, H = 2.90); thus, we included this study and further investigated sources of heterogeneity in metaregression.

The high heterogeneity suggested that a fixed effects approach was inappropriate, and we therefore used a random effects analysis. The random effects model, accounting for between-observation and between-study variability, produced a combined estimate of −18 (SE 1.9) [Fig. 3(a)]. In a sensitivity analysis using reported unadjusted mean change in pain, we estimated a pain decrement of −20 (SE 2.1).

Less potent opioids (tramadol, tramadol/acetaminophen)

The analysis of heterogeneity suggested moderate to high inconsistency and heterogeneity (I² = 0.71, H = 1.85). The most influential study arms were the 100 mg tramadol dose in Delemos et al.24, the tramadol/acetaminophen treatment group Emkey et al.,25, and the 300 mg tramadol dose in Fishman et al. (Influence = 1.12, 0.75, and 0.59). These studies contributed substantial weight to the heterogeneity score (Q-term = 101, 4.8 and 10.5). We investigated whether excluding Emkey et al., the only study not evaluating tramadol exclusively, would reduce heterogeneity and continued to find moderate to high inconsistency and heterogeneity (I² = 0.69, H = 1.8). Using a random effects analysis we found a combined estimate of effectiveness of −18 (SE 1.0) for less potent opioids [Fig. 3(b)]. In sensitivity analysis using reported unadjusted mean change in pain, we estimated a pain decrement of −21 (SE 1.0).

In another sensitivity analysis, we excluded the 100 mg and 400 mg doses of tramadol, as they represent doses not regularly used in clinical practice, along with the combined tramadol/acetaminophen regimen, which produced an overall pain reduction estimate of −19 (SE 1.3).

The funnel plot for less potent opioids exhibited asymmetry, suggesting that there may be missing studies which would have reported less change [Fig. 2]. Egger’s test was not statistically significant (P = 0.09). The trim and fill method was used to impute hypothetical missing publications. A funnel plot with imputed trim and fill values is shown in Fig. 4. While the peak remained uncentered, the plot was more symmetric and Egger’s test was no longer statistically significant (P = 0.57). After the trim and fill imputation, the combined estimate for change in pain from baseline decreased to −17 (SE 1.0).
Potent opioids (oxycodone, hydromorphone)

With only three studies, heterogeneity was difficult to assess for potent opioids. The corresponding statistics indicated low to mild inconsistency and heterogeneity ($I^2 = 0$, $H = 0.6$), though these measures may be inflated due to the limited number of studies\(^\text{19}\). The estimate of the change in pain obtained from both the random and fixed effects models for potent opioids was $-19$ (SE 1.3) \[\text{Fig. 3(c)}\]. In a sensitivity analysis using reported unadjusted mean change in pain, we estimated a pain decrement of $-20$ (SE 1.3).

**Meta-regression**

Results of the meta-regression analysis did not suggest clinically important or statistically significant difference among the drug classes under consideration (NSAIDs, less potent opioids, potent opioids, $P = 0.22$). We found that worse mean WOMAC Pain score was significantly associated with greater amount of change in pain score ($P < 0.001$); specifically, a 10 point higher pain score at baseline was associated with an additional five point decrement in WOMAC Pain score at the end of the study. Greater proportion of patients with knee (as opposed to hip) OA was associated with a greater change in WOMAC Pain, after adjusting for baseline pain ($P < 0.01$); for example, an increase in the proportion of knee OA patients by 10% resulted in an additional two point decrement in WOMAC Pain.

**Secondary analysis: NMA**

Direct and indirect treatment comparisons are shown in Fig. 5. The mean treatment effect across the nine comparisons of placebo

---

**Fig. 1.** One study included both NSAIDs and opioids treatment groups and is represented in both arms in this figure.
Fig. 2. This figure displays the funnel plot of precision (reciprocal of the variance) by mean change from baseline, modified for efficacy-related withdrawals, in WOMAC Pain for all included studies of (a) NSAIDs and (b) opioids. The NSAIDs funnel plot appears fairly symmetrical, and Egger's test was not statistically significant ($P = 0.50$). The opioids funnel plot is asymmetrical, with more studies reporting more change and fewer studies with lower precision reporting less change; Egger's test was borderline statistically significant ($P = 0.05$). The dashed lines represent combined efficacy estimates from a random effects model of the two classes of analgesics (NSAIDs: $-0.30$; opioids: $-0.10$).
Table I
Characteristics of included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Analgesic evaluated (daily total dose)</th>
<th>n ITT</th>
<th>Mean age (years)</th>
<th>% Female</th>
<th>Mean BMI</th>
<th>Mean OA duration (years)</th>
<th>% Knee OA</th>
<th>Baseline WOMAC Pain (Mean, SD)</th>
<th>Study duration (weeks)</th>
<th>Jadad score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NSAIDs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bingham et al. (2007a)</td>
<td>Celecoxib (200 mg)</td>
<td>236 63</td>
<td>70</td>
<td>80</td>
<td>68 (16)</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bingham et al. (2007b)</td>
<td>Celecoxib (200 mg)</td>
<td>246 62</td>
<td>62</td>
<td>78</td>
<td>67 (19)</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conaghan et al. (2013)</td>
<td>Celecoxib (200 mg)</td>
<td>233 62</td>
<td>67</td>
<td>100</td>
<td>47 (10)</td>
<td>12</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delenmos et al. (2011)</td>
<td>Celecoxib (200 mg)</td>
<td>202 60</td>
<td>65</td>
<td>8</td>
<td>74</td>
<td>57 (19)</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fleischmann et al. (2006)</td>
<td>Celecoxib (200 mg)</td>
<td>444 61</td>
<td>67</td>
<td>32</td>
<td>100</td>
<td>52 (17)</td>
<td>13</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hochberg et al. (2016)</td>
<td>Celecoxib (200 mg)</td>
<td>282* 63</td>
<td>81</td>
<td>31</td>
<td>100</td>
<td>74 (8)</td>
<td>26</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lehmann et al. (2005)</td>
<td>Celecoxib (200 mg)</td>
<td>420 63</td>
<td>68</td>
<td>30</td>
<td>100</td>
<td>51 (16)</td>
<td>13</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sheldon et al. (2005)</td>
<td>Celecoxib (200 mg)</td>
<td>393 63</td>
<td>63</td>
<td>33</td>
<td>100</td>
<td>54 (16)</td>
<td>13</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tannenbaum et al. (2004)</td>
<td>Celecoxib (200 mg)</td>
<td>481 64</td>
<td>69</td>
<td>30</td>
<td>5</td>
<td>100</td>
<td>51 (17)</td>
<td>13</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Case et al. (2003)</td>
<td>Diclofenac (150 mg)</td>
<td>25* 63</td>
<td>60</td>
<td>27</td>
<td>100</td>
<td>40 (20)</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kriegel et al. (2001)</td>
<td>Naproxen (750 mg)</td>
<td>187 65</td>
<td>75</td>
<td>4</td>
<td>71</td>
<td>48 (19)</td>
<td>26</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raynauld et al. (2009)</td>
<td>Naproxen (1000 mg)</td>
<td>154 60</td>
<td>65</td>
<td>31</td>
<td>100</td>
<td>58 (12)</td>
<td>26</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aryal et al. (2003)</td>
<td>Piroxicam (20 mg)</td>
<td>221 61</td>
<td>61</td>
<td>31</td>
<td>100</td>
<td>40 (19)</td>
<td>52</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Less potent opioids</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delenmos et al. (2011)</td>
<td>Tramadol (100 mg)</td>
<td>201 60</td>
<td>58</td>
<td>8</td>
<td>74</td>
<td>60 (20)</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delenmos et al. (2011)</td>
<td>Tramadol (200 mg)</td>
<td>199 60</td>
<td>62</td>
<td>9</td>
<td>73</td>
<td>61 (19)</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delenmos et al. (2011)</td>
<td>Tramadol (300 mg)</td>
<td>199 60</td>
<td>62</td>
<td>9</td>
<td>73</td>
<td>61 (19)</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fishman et al. (2007a)</td>
<td>Tramadol (100 mg)</td>
<td>103 60</td>
<td>31</td>
<td>100</td>
<td>58 (16)</td>
<td>12</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fishman et al. (2007b)</td>
<td>Tramadol (200 mg)</td>
<td>105 60</td>
<td>66</td>
<td>31</td>
<td>100</td>
<td>63 (19)</td>
<td>12</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gana et al. (2006)</td>
<td>Tramadol (100 mg)</td>
<td>204 58</td>
<td>62</td>
<td>34</td>
<td>8</td>
<td>75</td>
<td>63 (18)</td>
<td>12</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Gana et al. (2006)</td>
<td>Tramadol (200 mg)</td>
<td>201 59</td>
<td>64</td>
<td>34</td>
<td>8</td>
<td>75</td>
<td>63 (18)</td>
<td>12</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Gana et al. (2006)</td>
<td>Tramadol (300 mg)</td>
<td>201 59</td>
<td>59</td>
<td>33</td>
<td>8</td>
<td>75</td>
<td>59 (19)</td>
<td>12</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Gana et al. (2006)</td>
<td>Tramadol (400 mg)</td>
<td>202 58</td>
<td>58</td>
<td>34</td>
<td>8</td>
<td>74</td>
<td>60 (19)</td>
<td>12</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Emkley et al. (2004)</td>
<td>Tramadol (150–300 mg)</td>
<td>153 60</td>
<td>65</td>
<td>82</td>
<td>54 (14)</td>
<td>13</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Potent opioids</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hale et al. (2007)</td>
<td>Hydromorphone (8–64 mg)</td>
<td>64 63</td>
<td>77</td>
<td>34</td>
<td>84</td>
<td>63 (16)</td>
<td>8</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vojta et al. (2011)</td>
<td>Hydromorphone (4 mg)</td>
<td>138 65</td>
<td>77</td>
<td>32</td>
<td>76</td>
<td>59 (13)</td>
<td>16</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hale et al. (2007)</td>
<td>Oxycodone (20–160 mg)</td>
<td>60 64</td>
<td>64</td>
<td>31</td>
<td>75</td>
<td>61 (15)</td>
<td>8</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** ITT, modified intention-to-treat population; OA duration, years since diagnosed with osteoarthritis.

* Data for ITT population not available; total randomized population evaluated instead.

† Tramadol in conjunction with acetaminophen (1300–2600 mg).

and NSAIDs was −8 (range −4 to −15), as compared to −6 (range 1 to −11) across the 11 comparisons of placebo and less potent opioids. The only direct comparison of placebo and potent opioids had a treatment effect of −1 (reduction of 17 points in placebo compared to 18 in potent opioids). One study directly compared less potent opioids and NSAIDs; the mean decrement in WOMAC Pain for the NSAID arm was 22, compared to 12, 15, and 21 in the Tramadol 100 mg, 200 mg, and 300 mg arms, respectively. The NMA suggested a trend for NSAIDs to result in larger WOMAC Pain changes than opioids; however, these differences did not reach statistical significance: NSAIDs vs less potent opioids (Δ = −3.0, P = 0.13), NSAIDs vs potent opioids (Δ = −7.5, P = 0.08), less potent vs potent opioids (Δ = −4.4, P = 0.31).

**Discussion**

We used meta-analytic techniques to evaluate pain reduction in persons with OA treated with NSAIDs and opioids as reported in RCTs. Our results suggest that the mean decrement in WOMAC Pain achieved by NSAIDs (−18 points), less potent opioids (−18) and potent opioids (−19) are all comparable. Opioid-treated subjects generally had higher pain; adjusting for this difference, we nonetheless observed comparable pain reduction across the three analgesic classes. Clinicians must consider differences in patient populations when discussing the pain relief one can expect from NSAIDs or opioids. As it is likely that most patients considering opioids have previously taken NSAIDs, our analyses provide a practical way of describing the extent of pain relief a patient can expect with opioids.

There exists literature summarizing the effectiveness of NSAIDs and opioids in OA management; however, this is the first to focus on analgesics commonly employed in knee OA treatment and evaluate effectiveness using WOMAC Pain, the most commonly used pain instrument in an OA population. Four previous reviews of these analgesics in the OA literature were identified: Verkleij et al., evaluating short-term effects of NSAIDs and acetaminophen; Bjordal et al., evaluating short-term effects of NSAIDs and opioids; Myers et al., evaluating longer-term effects of NSAIDs and opioids compared to duloxetine; and Bannuru et al., assessing the relative efficacy of analgesics for knee OA. Due to a differences in drugs of interest, pain measurement instrument, and primary outcome, few studies included in our analyses were also included in the aforementioned reviews (one, four, nine, and eight studies overlapped with our analysis and those of Verkleij et al., Bjordal et al., Myers et al., and Bannuru et al., respectively). Verkleij et al. and Bjordal et al. did not restrict studies according to the instrument used to measure pain, while Myers et al. limited analyses to studies reporting WOMAC composite scores, which includes subscales for function and stiffness along with a subscale for pain. Bannuru et al. included all studies utilizing any measure of pain, function, or stiffness, and through NMA, derived effect sizes for each analgesic, which cannot be directly compared to the absolute WOMAC Pain reductions we present. Bjordal et al. reported 10 mm pain decreases for both NSAIDs and opioids over placebo on the 100 mm Visual-Analog Scale over a 1 month horizon; however, the VAS cannot be directly compared to the WOMAC Pain subscale. The meta-regression conducted by Myers et al. suggested a similar
association between baseline and change from baseline in WOMAC composite score as we report for WOMAC Pain. Prior to the development of those results should be interpreted with caution.

Both NSAIDs and opioids present non-trivial risks of significant adverse events, leading to contrasting views on their appropriate use. The American College of Rheumatology and European League Against Rheumatism conditionally recommend the use of NSAIDs or tramadol as primary analgesic agents and suggest using potent opioids only when all previous treatments have failed 4,5. The Osteoarthritis Research Society International takes a more conservative stance, stating that the appropriateness of any opioid prescription is uncertain 6. Though differing in their recommendations, professional societies consistently stress the paucity of long-term data on efficacy and adverse effects of many analgesics, particularly potent opioids.

We acknowledge several limitations of this analysis. Restricting the literature to studies published in English may have biased our evaluation; however, studies evaluating the effects of language restrictions in systematic reviews have not found any biases commonly associated with these restrictions 31. As cultural, ethnic, and psychosocial factors have been suggested to be important influences on pain perception and response to pain stimuli, we limited our analyses to studies primarily performed in developed countries to limit the heterogeneity of study populations 19,32,33. Although topical and oral NSAIDs appear to have similar efficacies 34, there are few topical opioid formulations and none are commonly used for arthritis pain management 35. Thus, we limited our analyses to oral formulations to examine medications with comparable delivery mechanisms.

We included only reports of RCTs. While observational studies and pragmatic trials can be employed evaluate analgesic effectiveness, they do not restrict concomitant utilization of additional treatments, thereby not allowing for an estimate of pain reduction attributable to the analgesic of interest. We ultimately excluded 246 articles because they were not RCTs; of those excluded, only 11 were cohort studies, none of which evaluated the outcome of interest.

Publication bias can be a significant problem for assessing the quality of the clinical trials literature, particularly when analyzing data from small cohorts 36. The asymmetry in the funnel plots led us to suspect publication bias in these data, particularly for less potent opioids. We attempted to adjust for publication bias using the trim and fill method; however these results should be interpreted as a sensitivity analysis rather than a corrected estimate, as we cannot ensure that funnel plot asymmetry is caused exclusively by publication bias.

Comorbidities are frequently associated with poorer symptom management and thus are important factors in assessing analgesic effectiveness. Studies of analgesics, however, frequently exclude persons with clinically significant comorbidities and do not systematically present the distribution of comorbidities within the study population. Similarly, more than one-third of included studies failed to report BMI or duration of OA diagnosis. We were unable to adjust for these factors in the meta-regression, and thus, those results should be interpreted with caution.

Our analyses focused on the WOMAC Pain subscale. The WOMAC is contained within the Knee Injury and Osteoarthritis Outcome Score (KOOS), which could have been incorporated as an outcome measure; however, no identified studies reported KOOS Pain instead of WOMAC Pain. Prior to the development of the WOMAC, numerous measures for pain and function among OA patients were and continue to be commonly used. We ultimately excluded a substantial proportion of otherwise eligible studies due to the use of another pain assessment measure. Expanding our analyses to include additional pain metrics would...

Please cite this article in press as: Smith SR, et al., Comparative pain reduction of oral non-steroidal anti-inflammatory drugs and opioids for knee osteoarthritis: systematic analytic review, Osteoarthritis and Cartilage (2016), http://dx.doi.org/10.1016/j.joca.2016.01.135

Table II

<table>
<thead>
<tr>
<th>Study</th>
<th>Analgesic evaluated</th>
<th>Mean change (SD)</th>
<th>Adjusted mean change*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NSAIDs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bingham et al. (2007a)</td>
<td>Celecoxib (200 mg)</td>
<td>–25 (25)</td>
<td>–22</td>
</tr>
<tr>
<td>Bingham et al. (2007b)</td>
<td>Celecoxib (200 mg)</td>
<td>–27 (27)</td>
<td>–24</td>
</tr>
<tr>
<td>Conaghan et al. (2013)</td>
<td>Celecoxib (200 mg)</td>
<td>–19 (16)</td>
<td>–19</td>
</tr>
<tr>
<td>DeLemos et al. (2013)</td>
<td>Celecoxib (200 mg)</td>
<td>–22 (26)</td>
<td>–22</td>
</tr>
<tr>
<td>Fleischmann et al. (2006)</td>
<td>Celecoxib (200 mg)</td>
<td>–19 (21)</td>
<td>–16</td>
</tr>
<tr>
<td>Hochberg et al. (2016)</td>
<td>Celecoxib (200 mg)</td>
<td>–37 (24)</td>
<td>–36</td>
</tr>
<tr>
<td>Lehmann et al. (2005)</td>
<td>Celecoxib (200 mg)</td>
<td>–17 (18)</td>
<td>–17</td>
</tr>
<tr>
<td>Sheldon et al. (2005)</td>
<td>Celecoxib (200 mg)</td>
<td>–17 (21)</td>
<td>–15</td>
</tr>
<tr>
<td>Tannenbaum et al. (2004)</td>
<td>Celecoxib (200 mg)</td>
<td>–16 (19)</td>
<td>–15</td>
</tr>
<tr>
<td>Case et al. (2003)</td>
<td>Diclofenac (150 mg)</td>
<td>–11 (16)</td>
<td>–10</td>
</tr>
<tr>
<td>Kriegel et al. (2001)</td>
<td>Naproxen (750 mg)</td>
<td>–11 (27)</td>
<td>–10</td>
</tr>
<tr>
<td>Raynauld et al. (2009)</td>
<td>Naproxen (1000 mg)</td>
<td>–25 (20)</td>
<td>–21</td>
</tr>
<tr>
<td>Aryal et al. (2003)</td>
<td>Piroxicam (20 mg)</td>
<td>–11 (22)</td>
<td>–10</td>
</tr>
<tr>
<td><strong>Less potent opioids</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeLemos et al. (2011)</td>
<td>Tramadol (100 mg)</td>
<td>–19 (25)</td>
<td>–12</td>
</tr>
<tr>
<td>DeLemos et al. (2011)</td>
<td>Tramadol (200 mg)</td>
<td>–18 (25)</td>
<td>–15</td>
</tr>
<tr>
<td>DeLemos et al. (2011)</td>
<td>Tramadol (300 mg)</td>
<td>–24 (25)</td>
<td>–21</td>
</tr>
<tr>
<td>Fishman et al. (2007)</td>
<td>Tramadol (100 mg)</td>
<td>–24 (29)</td>
<td>–19</td>
</tr>
<tr>
<td>Fishman et al. (2007)</td>
<td>Tramadol (200 mg)</td>
<td>–24 (26)</td>
<td>–22</td>
</tr>
<tr>
<td>Fishman et al. (2007)</td>
<td>Tramadol (300 mg)</td>
<td>–29 (25)</td>
<td>–26</td>
</tr>
<tr>
<td>Gana et al. (2006)</td>
<td>Tramadol (100 mg)</td>
<td>–21 (24)</td>
<td>–18</td>
</tr>
<tr>
<td>Gana et al. (2006)</td>
<td>Tramadol (200 mg)</td>
<td>–22 (25)</td>
<td>–19</td>
</tr>
<tr>
<td>Gana et al. (2006)</td>
<td>Tramadol (300 mg)</td>
<td>–21 (25)</td>
<td>–19</td>
</tr>
<tr>
<td>Gana et al. (2006)</td>
<td>Tramadol (400 mg)</td>
<td>–22 (25)</td>
<td>–19</td>
</tr>
<tr>
<td>Emkey et al. (2004)</td>
<td>Tramadol (150–300 mg)</td>
<td>–16 (19)</td>
<td>–15</td>
</tr>
<tr>
<td><strong>Potent opioids</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hale et al. (2007)</td>
<td>Hydromorphone (8–64 mg)</td>
<td>–21 (20)</td>
<td>–21</td>
</tr>
<tr>
<td>Vojtassak et al. (2011)</td>
<td>Hydromorphone (4 mg)</td>
<td>–19 (22)</td>
<td>–18</td>
</tr>
<tr>
<td>Hale et al. (2007)</td>
<td>Oxycodone (20–160 mg)</td>
<td>–20 (20)</td>
<td>–19</td>
</tr>
</tbody>
</table>

Abbreviations: WOMAC Pain, Western Ontario and McMaster Universities Osteoarthritis Index pain subscale; SD, standard deviation

* Modified for withdrawals due to insufficient efficacy.
increase the number of eligible studies, potentially reducing heterogeneity and increasing the generalizability of our results. However, while various measures of OA pain are correlated, there are no direct methods to transform a non-WOMAC measure into a validated score standardized with the WOMAC Pain subscale. Our analyses focused on the absolute pain decrements achieved from analgesics. We recognize that established methods such as the standard mean difference or effect size can be used to synthesize data from studies that use distinct metrics to assess a common outcome such as pain. However, these methods yield unitless measures of effect, which are not useful for estimating absolute differences.

The results of the exploratory NMA did not show a significant difference in WOMAC Pain reduction for the three comparisons of interest: NSAIDs vs less potent opioids, NSAIDs vs potent opioids, less potent vs potent opioids. These results should be interpreted with caution, as there were no direct comparisons between potent opioids and either less potent opioids or NSAIDs, and there was only one indirect comparison through placebo. However, there was a trend for NSAIDs to have a larger WOMAC Pain change than opioids. This finding warrants future investigation, particularly of the consistency assumption implicit in NMA that states that direct and indirect evidence must be in agreement. This assumption could be threatened by differences in populations, treatments, and outcome ascertainment. Additionally, we found that placebo effects may be greater in studies evaluating opioids. Further studies should examine the consistency of the oral placebo effects in studies of pharmacologic regimens with hypothesized differential analgesic potency.

These analyses offer important implications for research, policy, and clinical care. Studies assessing the comparative effectiveness of opioids and NSAIDs are central to clarifying the proper role of these agents for chronic OA pain. While there have been various RCTs evaluating the efficacy of analgesics, reporting has not been standardized, producing literature that is difficult to compare. Although long-term RCTs comparing effectiveness of these analgesics remain the gold standard, such studies are presently unavailable. Our results suggest that opioids provide similar levels of analgesia as NSAIDs; moreover, similar pain relief is observed for less potent and potent opioids. In addition to giving clinicians a practical way to consider the effectiveness of these analgesics, the results we present can also be used in decision analysis modeling to help policymakers understand the role of these analgesics in the treatment of knee OA and prioritize future data acquisition.
**Author contributions**

Drs. Losina and Collins had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conception and design: Smith, Deshpande, Losina.

Collection and assembly of data: Smith, Deshpande.

Analysis and interpretation of the data: Smith, Deshpande, Collins, Katz, Losina.

Statistical expertise: Collins, Losina.

Drafting of the article: Smith, Deshpande, Collins.

Critical revision of the article for important intellectual content: Smith, Deshpande, Collins, Katz, Losina.

Final approval of the article: Smith, Deshpande, Collins, Katz, Losina.

Obtaining of funding: Katz, Losina.

**Competing interests**

Dr. Katz is the President-Elect of the Osteoarthritis Research Society International. Drs. Katz and Losina are Deputy Editors for Biostatistics and Methodology for the *Journal of Bone and Joint Surgery*. All other authors have no conflicts of interest to report.

**Funding**

This study was supported by grants from the National Institutes of Health/National Institute of Arthritis and Musculoskeletal and Skin Diseases: R01 AR064320 (Losina), K24 AR057827 (Losina), and P60 AR047782 (Katz).

**Previous presentation**

A portion of the study was presented at the May 2015 World Congress of the Osteoarthritis Research Society International.

**Role of the sponsors**

The funding organization had no role in the design or conduct of the study, the collection, management, analysis, or interpretation of data, the preparation, review, or approval of the manuscript, or the decision to submit the manuscript for publication.
Acknowledgments
The authors wish to thank Heidi Y. Yang, MS, MPH, for her assistance with statistical analyses.

Supplementary data
Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.joca.2016.01.135.

References


