

Estimating relative efficacy in acute postoperative pain: network meta-analysis is consistent with indirect comparison to placebo alone

R. Andrew Moore^{a,*}, Sheena Derry^a, Philip J. Wiffen^a, Samik Banerjee^b, Rajesh Karan^c, Ekkehard Glimm^c, Anna Wiksten^c, Dominic Aldington^d, Christopher Eccleston^{e,f}

Abstract

Network meta-analysis uses direct comparisons of interventions within randomized controlled trials and indirect comparisons across them. Network meta-analysis uses more data than a series of direct comparisons with placebo, and theoretically should produce more reliable results. We used a Cochrane overview review of acute postoperative pain trials and other systematic reviews to provide data to test this hypothesis. Some 261 trials published between 1966 and 2016 included 39,753 patients examining 52 active drug and dose combinations (27,726 given active drug and 12,027 placebo), in any type of surgery (72% dental). Most trials were small; 42% of patients were in trials with arms <50 patients, and 27% in trials with arms ≥100 patients. Response to placebo in third molar extraction fell by half in studies over 30 to 40 years (171 trials, 7882 patients given placebo). Network meta-analysis and Cochrane analyses provided very similar results (average difference 0.04 number needed to treat units), with no significant difference for almost all comparisons apart from some with small patient numbers or small effect size, or both. Network meta-analysis did not detect significant differences between effective analgesics. The similarity between network meta-analysis and Cochrane indirect analyses probably arose from stringent quality criteria in trials accepted in Cochrane reviews (with consequent low risk of bias) and consistency in methods and outcomes. Network meta-analysis is a useful analytical tool that increases our confidence in estimates of efficacy of analgesics in acute postoperative pain, in this case by providing similar results.

Keywords: Network meta-analysis, Postoperative pain, Systematic review, Cochrane

1. Introduction

Clinical trials of single-dose oral analgesics in acute postoperative pain provide the foundations of analgesic testing.^{18,28} The basic design was worked out in the 1950s and 1960s.¹⁸ It was rigorously tested at the time and validated subsequently in several individual patient analyses.^{2,3,38,40,42}

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

^a Pain Research and Nuffield Division of Anaesthetics, Nuffield Department of Clinical Neurology, University of Oxford, Oxford, United Kingdom, ^b Novartis Healthcare Pvt Ltd, Salarpuria Sattva Knowledge City, Hyderabad, India, ^c Novartis Pharma AG, Postfach, Basel, Switzerland, ^d Royal Hampshire County Hospital, Winchester, United Kingdom, ^e Centre for Pain Research, University of Bath, Bath, United Kingdom, ^f Department of Experimental, Clinical, and Health Psychology, Faculty of Psychology and Educational Sciences, Ghent University, Ghent, Belgium

*Corresponding author. Address: Pain Research and Nuffield Division of Anaesthetics, Nuffield Department of Clinical Neurology, University of Oxford, The Churchill, Oxford OX3 7LE, United Kingdom. Tel.: +44 1865 225674. E-mail address: andrew.moore@ndcn.ox.ac.uk (R.A. Moore).

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.painjournalonline.com).

PAIN 159 (2018) 2234–2244

Copyright © 2018 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the International Association for the Study of Pain. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

<http://dx.doi.org/10.1097/j.pain.0000000000001322>

Original outcomes were the average summed pain-intensity difference (SPID) or total pain relief (TOTPAR) over 4 to 6 hours. The emphasis has changed to the individual patient's response, particularly the proportion of patients achieving the outcome of at least 50% of the maximum possible pain relief (PR).²⁸ Unsurprisingly, patient satisfaction is highly correlated with good PR,³⁰ and the 50% maximum PR cutoff point is acceptable by both scientists²⁴ and patients.⁴¹

The original design demonstrated whether a drug was better than placebo, and therefore powered to examine the *direction* of the result. The clinical practice issue is somewhat different: we ask not only whether a drug is an analgesic, but also *how good* an analgesic it is, comparing it with other analgesics. Considerably more information is needed to demonstrate the size of effect,³⁹ so systematic reviews typically provide better information than individual studies.

Overview reviews of Cochrane reviews in acute postoperative pain provide a wealth of information on different drugs and doses.^{36,37} They indicate where there are no data, and where the data are inadequate or cannot be trusted because of susceptibility to publication bias. They also provide data on over 50 drug and dose combinations where the efficacy estimate can be trusted.³⁷

The overview can be used to assess the relative efficacy of oral drugs in acute postoperative pain, but only indirectly; each analgesic is compared directly with placebo, but not with other analgesics in head-to-head comparisons. Although this may not be a problem in looking at relative efficacy,⁴⁹ it remains as a possible limitation because information from head-to-head comparisons between active drugs is effectively ignored.

The solution may be to perform a network meta-analysis (NMA, also called mixed treatments comparison or multiple treatments comparison meta-analysis) using all the available comparisons with both placebo and active comparators. A number of NMAs have been conducted in pain, most notably for drugs in arthritis,^{8,53} fibromyalgia,⁴⁴ and migraine.⁵¹ This approach has not previously been attempted for acute postoperative pain.

Our aim was to compare the relative efficacy of analgesic drugs in acute postoperative pain in the Cochrane overview with that from an NMA. Because the trials in acute postoperative pain use standard methods, with a standard efficacy outcome, measured over the same period, in relatively similar patients, it is expected that results from both analyses might produce very similar relative efficacy. That would help confirm the validity of either approach. Any major disagreement would be a matter of concern.

2. Methods

2.1. Data sources

A Cochrane overview review for acute postoperative pain in adults categorised data from different drug-evidence combinations³⁷:

- (1) Drugs for which Cochrane reviews found no information.
- (2) Drugs for which Cochrane reviews found inadequate information (fewer than 200 participants in comparisons in at most 2 studies).
- (3) Drugs for which Cochrane reviews found no evidence of effect or evidence of no effect.
- (4) Combinations of drug and dose for which Cochrane reviews found evidence of effect, but where the addition of hypothetical studies without a treatment effect, representing fewer than 400 patients combined, would increase the number needed to treat (NNT) of treatment vs placebo to a clinically irrelevant value above 10.³⁵
- (5) Combinations of drug and dose for which Cochrane reviews found evidence of effect, and where the addition of hypothetical studies without a treatment effect, representing fewer than 400 patients in total, would *not* increase the NNT of treatment vs placebo to a clinically irrelevant value above 10.³⁵

Note that “hypothetical studies” refer to studies unavailable to the Cochrane review, for example, because they were not published. In this analysis, we used drugs and doses from category 5 only. Information from the individual trials in individual systematic reviews was tabulated, extracting reference, date, average age, and number and percentage of women. Type of surgery was categorised as dental for third molar extraction or other oral surgery, abdominal, thoracic, orthopaedic, or gynaecological.

Data from trials were available in the body of the text or more helpfully in appendices where all comparators and outcomes were usually detailed. Two investigators (S.D. and R.A.M.) extracted data and entered them into a table, with a third (P.J.W.) acting as arbiter in case of disagreement. The final table was checked for potential duplication because some studies may have been used in more than one review; typically, acute pain studies used a placebo, the active drug under the test, and an active comparator.

The reviews included only primary studies that were both randomised and double-blind, so minimising the risk of bias. They did not use studies with fewer than 10 patients in treatment arms. All studies in all reviews included only participants who experienced at least moderate pain intensity (PI) at baseline, providing a sensitive assay of analgesic efficacy.⁶ All included studies used standard methods and report standard outcomes,

or provided data from which they could be calculated using validated methods. For studies in acute pain lasting up to 6 hours (all the studies in the overview and reviews), it has been shown that imputation of data after remedication using last observation carried forward rather than baseline observation carried forward does not significantly influence results.³⁸ The methods used in individual reviews were consistent and can be found in any of the individual reviews.¹⁶

We also used additional data from 4 other sources. Data from a pooled analysis of tramadol were included in the Cochrane overview as part of the discussion, based on an individual patient-level analysis.⁴⁰ Since the overview, 3 further Cochrane reviews had been published and were included, a new fixed-dose combination of dexketoprofen and tramadol,¹¹ an updated analysis of dexketoprofen and ketoprofen,¹⁴ and an updated review of dipyrrone.¹⁶ Most other reviews had been stabilized, indicating that no important additional information has been found since the most recent update.

2.2. Responder rates

Responder rates over 4 to 6 hours had been determined in each review in a standard manner using the same methods. Briefly:

For efficacy analyses, the numbers of participants in each treatment group who were randomised, received medication, and provided at least one postbaseline assessment were used to determine the total in each group.

The mean values for total pain relief (TOTPAR), SPID, visual analogue scale TOTPAR, or visual analogue scale SPID for the active and placebo groups were converted to %maxTOTPAR or %maxSPID by division into the calculated maximum value.⁶ The proportions of participants in each treatment group who achieved at least 50%maxTOTPAR were calculated using verified equations.^{32–34} These proportions were converted into the number of participants achieving at least 50%maxTOTPAR by multiplying by the total number of participants in the treatment group.

The following pain measures were accepted for the calculation of TOTPAR or SPID, in order of priority:

- (1) 5-point categorical PR scales with comparable wording to “none,” “slight,” “moderate,” “good,” and “complete.”
- (2) 4-point categorical PI scales with comparable wording to “none,” “mild,” “moderate,” and “severe.”
- (3) Visual analogue scale for PR.
- (4) Visual analogue scale for PI.

“Response” was also calculated using the number of participants reporting “very good or excellent” on a 5-point categorical global scale with the wording “poor,” “fair,” “good,” “very good,” and “excellent” for the number of participants achieving at least 50% PR.⁵

2.3. Network meta-analysis

Network meta-analysis simultaneously analyses both direct comparisons of interventions within randomized controlled trials (A vs B) and indirect comparisons across trials based on a common comparator such as placebo (A vs placebo and B vs placebo).^{4,21,23,25,31,47} They use all available high-quality trials, producing a network of randomized clinical trials evaluated in an NMA (A vs B vs C vs placebo, etc). Theoretically, this produces insights into the comparative effectiveness of interventions that are more informed than the common representation of an indirect comparison of a series of direct comparisons of active drugs each with a common comparator such as placebo.

For avoidance of doubt, we use these definitions of terms:

- (1) Direct comparison—comparison in the same clinical trial of 2 active drugs or 1 active drug against placebo.
- (2) Network meta-analysis indirect comparison—estimation of relative efficacy using comparisons of active drug with placebo and different active drugs.
- (3) Cochrane indirect comparison—estimation of relative efficacy using direct comparisons of active drugs with placebo and excluding direct comparisons between active drugs.

The primary analysis was an NMA of responder rates using a logistic regression model with treatment and trial as categorical covariates. Between-trial variation in the treatment effect was estimated nonparametrically by quasi-likelihood methods.²⁷ This method does not require the assumption of a specific form of distribution for the treatment-by-trial interaction and could, in theory, also detect “underdispersion” (ie, it would detect whether there is less variation in the trial-specific treatment effects than that would be expected by chance, and—in contrast to maximum-likelihood-based estimation with a normally distributed random variation of the trial-specific treatment effect—it would not fail to converge), but this did not happen. The model was fitted using SAS PROC GLIMMIX using the software’s default assumptions and fitting methods (nonparametric estimation of overdispersion is performed by using the option RANDOM_RESIDUAL_).

To test the main model, we performed sensitivity analyses by fitting the following additional models:

- (1) a model that includes average age, percentage of female patients, and type of surgery as additional covariates;
- (2) a model that includes type of surgery as a stratification factor (with and without percentage of female patients and average age as additional covariates);
- (3) separate analyses for the different types of surgery (with and without percentage of female patients and average age as additional covariates).

In addition, we compared the results from the NMA with logistic regression models on subsets of trials that are restricted to direct comparisons to assess any potential discrepancies between direct and indirect comparisons.

We summarised results using effect estimates from logistic regression models, the corresponding odds ratios (ORs), their confidence intervals, and *P* values of tests for significant differences between the treatments, and used forest plots for visualisation. Number needed to treat values compared with placebo were calculated from the results of the meta-analysis to detect any contrast between the results of NMA indirect comparison and Cochrane indirect comparison.

Risk differences (RDs) with 95% confidence intervals were calculated by using the multivariate delta method on the estimated parameters and their estimated covariance matrix from the logistic regression model. Numbers needed to treat were calculated as 1/(RD). All point estimates of RD vs placebo are positive. For a few drugs, because the lower limit of the confidence interval for the RD was negative, we set the corresponding boundary of the confidence interval for NNT to infinity.

For the main body of the text, we show 28 commonly used analgesics, including fixed-dose combinations, with all analyses presented in Supplementary file 1 (available online as supplemental digital content at <http://links.lww.com/PAIN/A611>). Many of these have licensed indications for acute or postoperative pain by regulators such as the U.S. Food and Drug Administration (FDA). Some drugs may not have a license for acute pain but are widely available analgesics often used for a range of pain conditions.

2.4. Analysis and comparison

We set out to analyse results in 4 stages, and we report results in this order.

- (1) Stage 1: review of available data.
- (2) Stage 2: assessment of validity of available data.
- (3) Stage 3: results of NMA.
- (4) Stage 4: contrast NMA and Cochrane indirect comparisons.

3. Results

3.1. Review of available data

We used data from 261 individual randomised, double-blind trials in acute postoperative pain from Cochrane reviews and an equivalent review of tramadol. In the individual reviews, all patients had initial PI of moderate or severe, and each individual review reported the number and proportion of patients achieving at least 50% of maximum possible PR using standard, validated methods, or results that allowed their calculation.

The earliest included trial was published in 1966 and the latest in 2016. Most (240 trials, 92%) were published between 1980 and 2010, with 5% published earlier and 3% published later. The 261 trials had 753 relevant treatment arms (**Table 1**) with a median of 41 patients per arm. The bulk of the treatment arms (496, 66%) involved fewer than 50 patients, with 43% of the total number of patients. A smaller number (191 trials, 25%) involved 50 to 99 patients, with 30% of the total number of patients. Sixty-six treatment arms involved 100 patients or more; this was 9% of treatment arms but 27% of patients. The newer drugs and drug combinations tended to have larger treatment arms for active components (ibuprofen + oxycodone, median 169 patients; ibuprofen + paracetamol, 146; rofecoxib, 91; and etoricoxib, 89). Median treatment arm size was little changed over time; for placebo for example, median treatment arm size in years before 2000 was 39, rising to 50 in years after 1999.

The trials involved 39,753 patients, of whom 27,726 were given an active analgesic and 12,027 were given placebo. **Table 2** shows the distribution of the number of trials and patients in each definition of surgery type as provided by the original studies. Dental surgery, all third molar extraction, provided 66% of the trials and 72% of patients. Together with mixed surgery (any surgical procedure), episiotomy, and orthopaedic and gynaecological surgery, these categories provided 97% of trials and 98% of patients.

Mean age was reported in 233 trials. The average mean age was 28 years and the median was 25 years. Only 7/233 trials reported a mean age of over 50 years. The sex distribution was reported in 241 trials. The average proportion of patients who were women was 64% (median 60%). Two trials involved only men and 44 only women.

3.2. Assessment of validity of available data

All included trials met the fundamental requirements for validity for acute pain analgesic assessment (randomised, double-blind, adequate initial PI, standard pain measurement methods, 4–6 hours of duration, oral drug administration, adequate outcome reporting, plus checking for reporting inconsistencies, errors, or duplication). Two main questions regarding their validity for inclusion in the NMA remained. These were the effects of small study size, and whether there had been any temporal change in studies over the 5 decades covering the studies.

Table 1
Distribution of 753 treatment arms and patient numbers by group size.

Group size (no. of patients per arm)	No. of		Percentage of total	
	Treatment arms	Patients	Treatment arms	Patients
10-19	34	512	4.5	1.3
20-29	99	2504	13.1	6.3
30-39	186	6346	24.7	16.0
40-49	177	7679	23.5	19.3
50-59	105	5477	13.9	13.8
60-69	34	2136	4.5	5.4
70-79	22	1642	2.9	4.1
80-89	18	1494	2.4	3.8
90-99	12	1122	1.6	2.8
>100	66	10,841	8.8	27.3

3.2.1. Small study size

The impact of small study size was investigated by examination of placebo responses in the most clinically homogeneous data set, the 168 individual studies that involved third molar extraction (and omitting 3 pooled analyses). **Figure 1** shows the distribution of the percentage with at least 50% maximum PR related to the number of patients in the placebo arm. Overall, the proportion obtaining this level of PR with placebo was 12%, but with small individual studies, values could vary between 0% and 70%.

This finding is not new,³⁹ but underscores the dangers when making comparisons with small amounts of data. This degree of numerical heterogeneity despite substantial clinical homogeneity is explained by the random play of chance with small numbers of actual events.^{9,10,19,39,43,50}

3.2.2. Temporal changes

Again, the most clinically homogeneous third molar extraction data set was used to assess the presence and/or impact of any

Table 2
Distribution of trials and patients by type of surgery.

Surgery type	Trial		Patients	
	Frequency	Percent	Frequency	Percent
Dental (third molar extraction)	171	65.5	28,773	72.4
Mixed surgery	25	9.6	3131	7.9
Episiotomy	25	9.6	2793	7.0
Orthopaedic	17	6.5	2138	5.4
Gynaecological	11	4.2	1498	3.8
Caesarian	3	1.2	410	1.0
Bunionectomy	2	0.8	355	0.9
Gynae	1	0.4	200	0.5
Urogenital	2	0.8	120	0.3
Meniscectomy	1	0.4	88	0.2
Tonsillectomy	1	0.4	83	0.2
Hernia repair	1	0.4	82	0.2
Oral surgery	1	0.4	82	0.2

temporal changes. Looking at all the placebo responses (171 trials, 7882 patients given placebo), there was a reduction from over 15% in studies performed before 1990 to around 7% to those performed since 2010 (**Fig. 2**).

Data for analgesics were not as readily available, as many different drugs and doses had been used in these trials. There were, however, sufficient data for ibuprofen acid 400 mg to examine whether there was any temporal change to the effect size. **Table 3** shows the NNT calculations for ibuprofen acid 400 mg compared with placebo using only the studies with the direct comparison. It can be seen that results were consistent, with no trend of change in effect size.

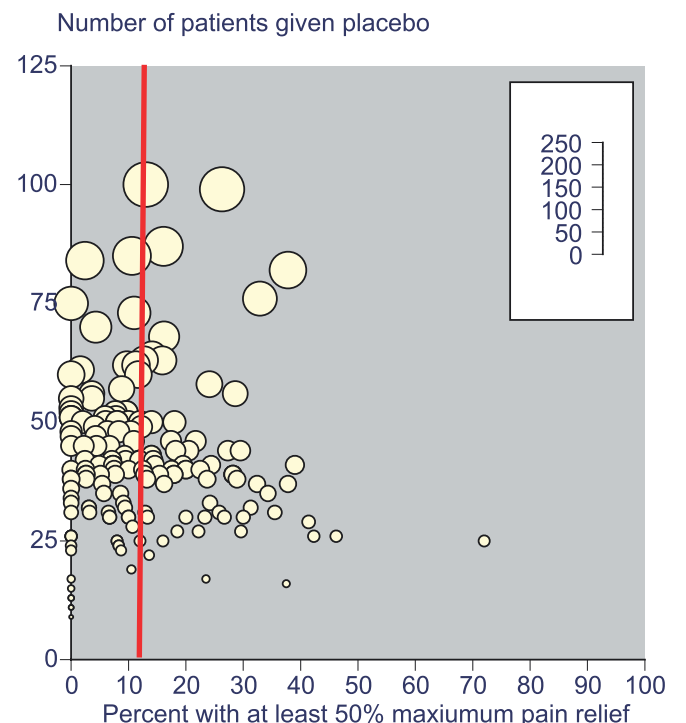


Figure 1. Distribution of the percentage with at least 50% maximum pain relief related to the number of patients given placebo (red line shows overall average of 12%, N = 7882). Size of symbol indicates the number of patients (inset scale).

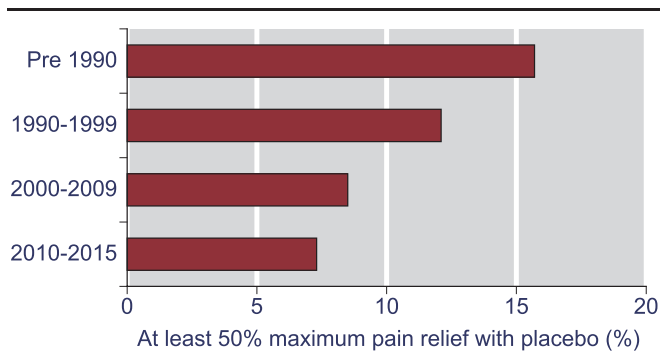


Figure 2. Response rate with placebo over time.

3.3. Results of the network meta-analysis

The network is shown in **Figure 3**, for all drugs, doses, and combinations, and for all surgery types. It was extensive, and although placebo comparisons dominated, there were many cross comparisons, with large data sets for ibuprofen acid 200 and 400 mg, aspirin 600/650 mg, paracetamol 600/650 mg and 975/1000 mg, and codeine 60 mg. Rofecoxib 50 mg also contributed a large amount of cross-comparison data.

The main results of the NMA using all trial data from all types of surgery for 28 common analgesics in comparisons with placebo are presented in **Figure 4**. Outputs are shown in 3 ways: as ORs, RD, and NNT, in order of highest OR. There was a consistent order using all 3 presentations, as expected. Supplementary file 1 (available online as supplemental digital content at <http://links.lww.com/PAIN/A611>) shows the same presentations by type of surgery, and for all drugs and doses.

The NMA of the 28 analgesics plus placebo showed moderate between-study heterogeneity ($I^2 = 0.6$, $Q = 1083.9$ on 441 degrees of freedom). The quasi-likelihood approach we used to analyse the data adjusts for this: the confidence intervals obtained from it are wider than ones that would have been obtained under an assumption of no between-study heterogeneity. The 95% confidence intervals for the NNT are wide if the number of patients is small or if effect sizes are small. Point estimates for NNTs ranged from around 2 to over 5 for paracetamol 500 mg and codeine 60 mg. Nineteen of the 28 comparisons in **Figure 4C** had point estimates of the NNT below 3.

Similar results were found for different surgery types. Notwithstanding that the bulk of data derived from dental studies, analyses were also performed for trials in episiotomy, gynaecological surgery, orthopaedic surgery, mixed surgery, and any other surgery. Supplementary file 1 (available online as supplemental digital content at <http://links.lww.com/PAIN/A611>) shows these results presented as ORs, except for dental surgery for which there was sufficient information to present risk ratio and NNT in addition.

Table 3
Calculation of NNT for ibuprofen acid 400 mg according to trial data available by decade.

Period	No. of		Percent with		NNT (95% CI)
	Trials	Patients	Active	Placebo	
Pre-1990	8	677	58	13	2.2 (2.0-2.6)
1990-1999	23	2157	53	13	2.5 (2.3-2.8)
2000-2009	13	2214	53	9	2.3 (2.1-2.5)
2010-2015	2	247	52	8	2.2 (1.8-2.9)

CI, confidence interval; NNT, number needed to treat.

Demonstrating significant differences between effective analgesics was difficult, even with the large amounts of information available for the NMA. For example, **Figure 5** shows the ORs and RD when the 28 common analgesics were compared with diclofenac K 100 mg rather than placebo (NNTs are obviously unhelpful here, as NNT values tend to infinity with only small RDs, and confidence intervals can be unhelpfully wide in that circumstance). As an example, the relative efficacy of diclofenac K 100 mg was “at par” with that of celecoxib 400 mg. In both cases, for only a few analgesics did the upper limit of the 95% confidence interval not include 1 (OR) or 0 (RD). These were placebo, all doses of paracetamol, paracetamol 600/650 mg + codeine 60 mg, ketoprofen 50 mg, etodolac 100 mg, and codeine 60 mg. For no comparison was the lower limit of the confidence interval above 1 or 0.

3.4. Contrast network meta-analysis and Cochrane indirect comparisons

The Cochrane indirect comparison uses NNT values for active drug compared with placebo. Number needed to treat is a measure now readily understood and widely used in both the pain literature and clinical decision-making. This comparison uses only trials where the active and placebo were tested.

Figure 6 shows the differences between the NMA indirect comparison estimate of NNT compared with placebo as a difference from the Cochrane indirect estimate, plotted against the Cochrane indirect estimate NNT. The size of symbol is proportional to the amount of data in the Cochrane indirect estimate. The difference would be zero if there were perfect concordance, irrespective of the actual NNT.

Only modest differences were actually found over NNT values from 2 to 5 for larger data sets. The sum of differences for all active drugs was 2 NNT units over 52 comparisons (an average of 0.04 NNT units). Larger differences (more than plus or minus 1 NNT unit) were found in 5 comparisons. Three of these (ibuprofen 100 mg + caffeine 100 mg, dexketoprofen 20/25 mg, and celecoxib 200 mg) had small amounts of data in indirect comparisons (200-416 patients), and one had a large data set but high NNT (codeine 60 mg, 2411 patients, NNT 7.3 NMA and 12 indirect). Only one, paracetamol 500 mg, had a reasonable amount of data (561 patients in six studies).

Table 4 shows the NNTs for 28 common analgesic drugs and doses compared with placebo for both the NMA indirect comparison and those from the Cochrane indirect analyses. In 17 cases, the NMA point estimate NNT was numerically smaller; in 9, the Cochrane point estimate NNT was numerically smaller; and in 2 cases, they were identical. In all cases, there were overlapping confidence intervals, with only modest differences between the point estimates.

4. Discussion

This NMA used 261 unique clinical trials in postoperative pain, with 39,753 patients providing data. Data came from a Cochrane overview of 39 separate Cochrane reviews, which, together with other minor sources of data, provided over 50 comparisons of analgesic drugs, doses, and combinations with placebo. We are unaware of any NMA in any topic presenting a comparison between NMA and Cochrane indirect analyses. The only other NMA in acute postoperative pain (135 trials, 13,287 patients) addressed largely nonsteroidal anti-inflammatory drugs and cyclooxygenase inhibitors compared with placebo, with little cross comparison.²⁶

5. Strengths and weaknesses

The study had considerable strength. The network itself was extensive, with many individual trials using a design that included

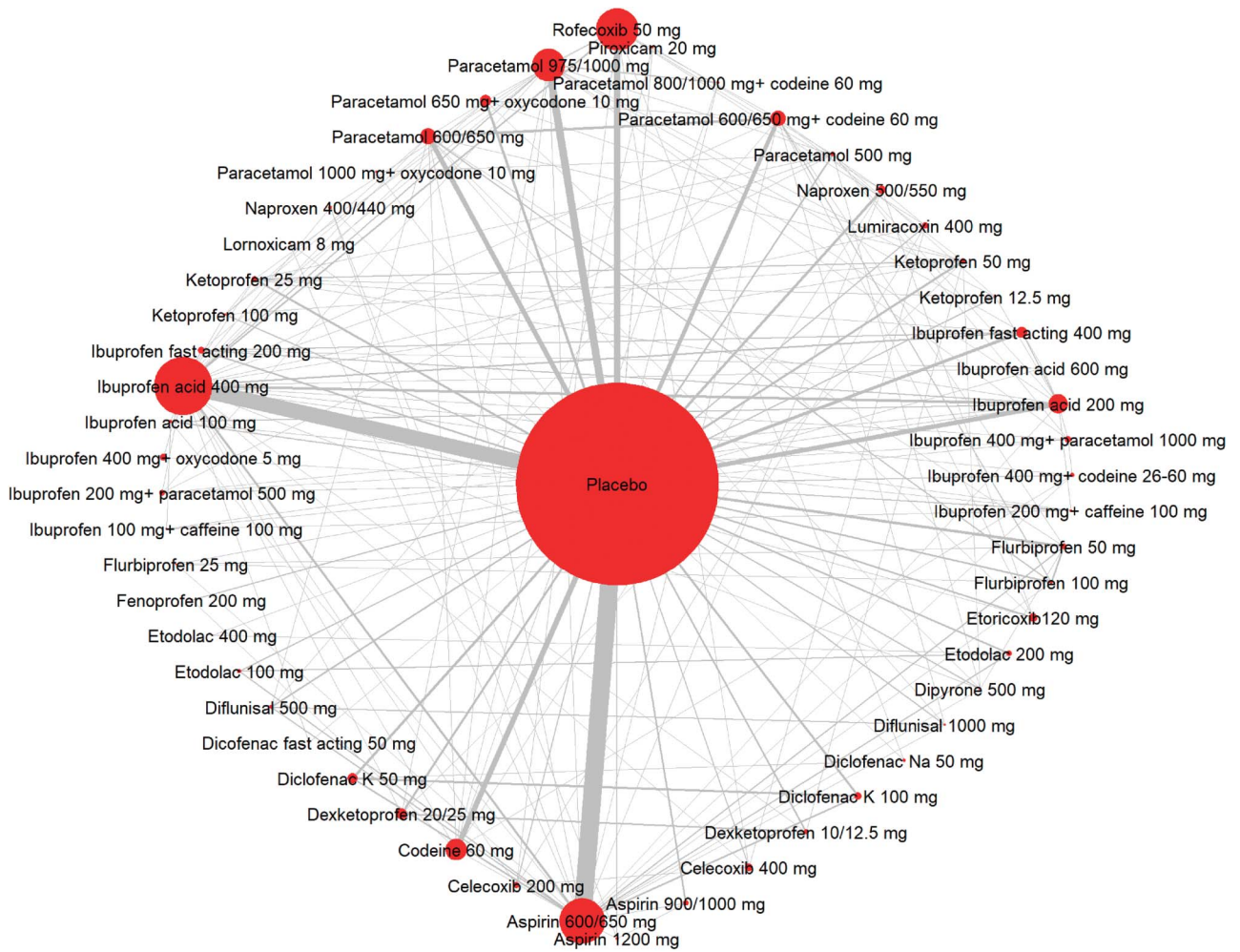


Figure 3. All data for all drugs, doses, and combinations, and for all surgery types for Network Meta Analysis of analgesic response. The size of the symbols refers to the number of patients treated with particular drugs, and the thickness of lines is proportional to the numbers in each possible comparison.

an active comparator as well as placebo and the test drug, providing considerable cross-comparisons and a richness of data for analysis. Results were presented in 3 ways (OR, RD, and NNT); users of evidence might use these outputs differently.

Individual reviews from which data were taken had used the highest quality of evidence from clinical trial designs performed consistently over 60 years²⁸; individual studies with high risk of bias were excluded, and raw data were usually freely available in

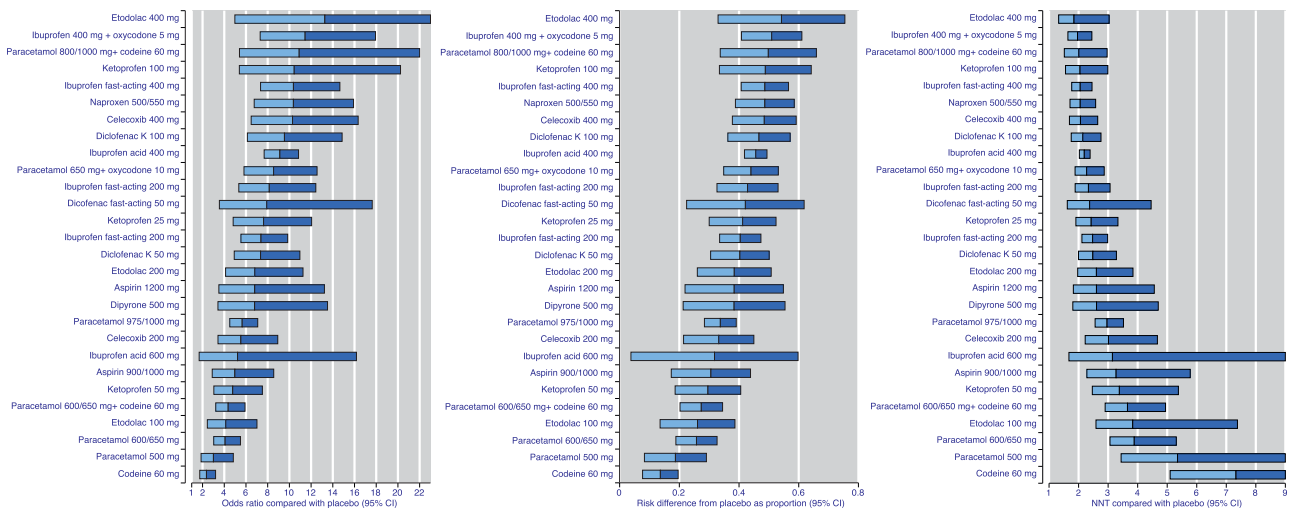


Figure 4. Results using all available data for the comparison of active drug with placebo, presented as odds ratio, risk difference, and NNT. Horizontal bars represent the width of the 95% confidence interval, and colour change the point estimate. NNT, number needed to treat.

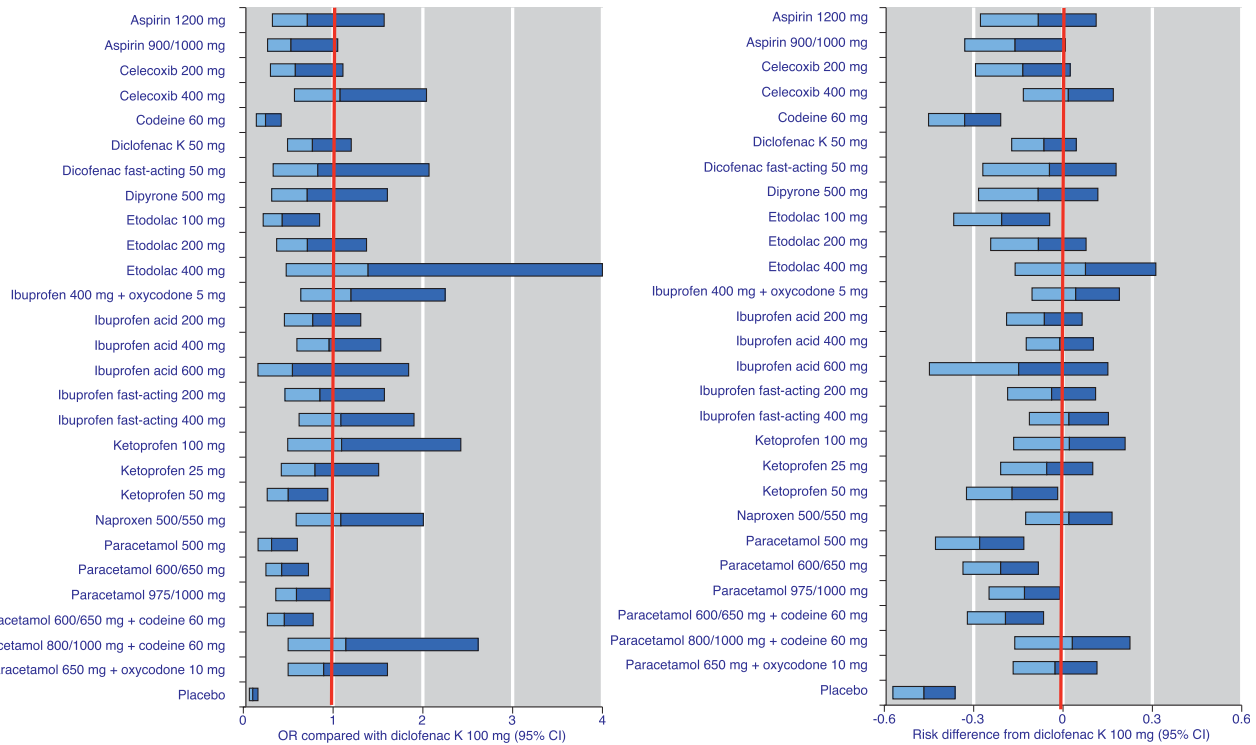


Figure 5. Results using all available data for the comparison of active drug with diclofenac K 100 mg, presented as odds ratio and risk difference. Horizontal bars represent the width of the 95% confidence interval, and colour change the point estimate. Horizontal bars crossing the red vertical line indicate no significant difference. OR, odds ratio.

appendices for others to replicate analyses. Reviews were based on extensive literature searching, including hand searching for older trials.²⁰ Only comparisons with a minimum of available data were used in the network, minimising potential problems with random error and small size bias.^{1,9,10,19,39,43,50} The outcome used was calculated from outcomes reported in individual trials, was uniform, and used validated methods.^{38,42} Moreover, it is an outcome of relevance to patients and of clinical importance.⁴¹

Other strengths come from the use of all surgery types, with dental surgery, mixed surgery, episiotomy, orthopaedic, and gynaecological surgery providing 97% of trials and 98% of

patients. Men and women were represented (64% women). Although there was a mix of younger and older adults, only 7 trials had a mean age above 50 years. The effects of temporal changes were examined and shown not to make a difference in terms of effect size, although response to placebo fell over time.

A major potential weakness was that children and older adults were clearly not represented, limiting applicability to important demographics, especially as much elective surgery involves older people.¹⁷ Much information derived from drugs, or doses of drugs, that were of limited clinical relevance, but even so they provided useful information for comparing methods.

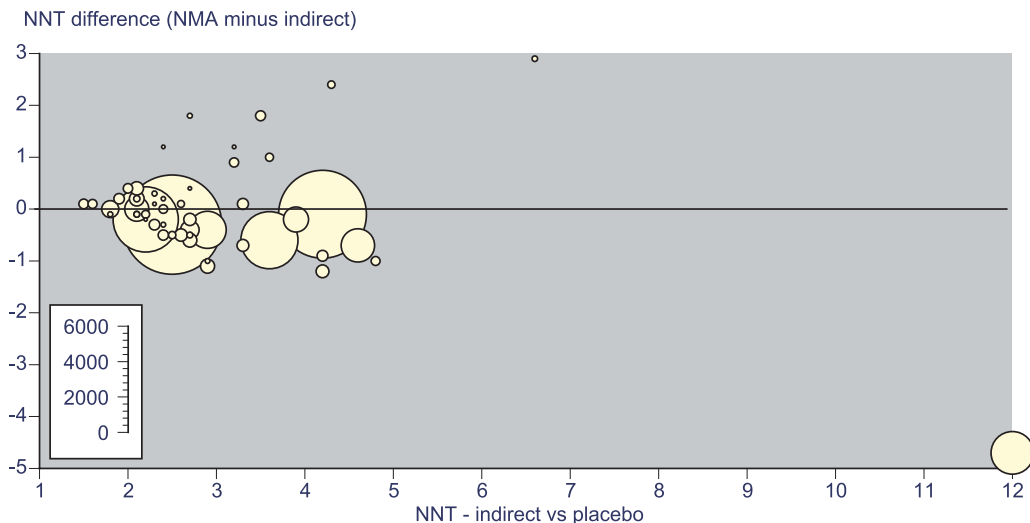


Figure 6. Difference between the NMA indirect and Cochrane indirect estimates of NNT, plotted against the Cochrane indirect estimate NNT. Size of symbol indicates the number of patients (inset scale). NMA, network meta-analysis; NNT, number needed to treat.

Table 4**Numbers needed to treat from NMA and indirect comparisons with placebo for 28 analgesic drugs and doses.**

Drug and dose	Network meta-analysis NNT (95% CI)	Cochrane review NNT (95% CI)
Etodolac 400 mg	1.8 (1.3-3.0)	2.9 (2.3-4.0)
Ibuprofen + oxycodone 400 + 5 mg	2.0 (1.6-2.4)	2.3 (2.0-2.8)
Ketoprofen 100 mg	2.0 (1.6-3.0)	2.1 (1.7-2.6)
Paracetamol + codeine 800/1000 + 60 mg	2.0 (1.5-3.0)	2.2 (1.8-2.9)
Ibuprofen fast-acting 400 mg	2.1 (1.4-3.4)	2.1 (1.9-2.4)
Celecoxib 400 mg	2.1 (1.7-2.6)	2.6 (2.3-3.0)
Naproxen 500/550 mg	2.1 (1.7-2.6)	2.7 (2.3-3.3)
Diclofenac K 100 mg	2.1 (1.8-2.8)	1.9 (1.7-2.3)
Ibuprofen acid 400 mg	2.2 (1.7-3.1)	2.5 (2.4-2.6)
Paracetamol + oxycodone 650 + 10 mg	2.3 (1.9-2.9)	2.7 (2.4-3.1)
Ibuprofen fast-acting 200 mg	2.3 (1.9-3.1)	2.1 (1.9-2.4)
Dicofenac fast-acting 50 mg	2.4 (1.6-4.5)	2.4 (2.0-3.0)
Ketoprofen 25 mg	2.4 (1.9-3.3)	2.0 (1.8-2.3)
Diclofenac K 50 mg	2.5 (2.0-3.3)	2.1 (1.9-2.5)
Ibuprofen acid 200 mg	2.5 (2.1-3.0)	2.9 (2.7-3.2)
Aspirin 1200 mg	2.6 (1.8-4.6)	2.4 (1.9-3.2)
Dipyron 500 mg	2.6 (1.8-4.7)	2.3 (1.9-3.1)
Etodolac 200 mg	2.6 (2.0-3.8)	3.3 (2.7-4.2)
Celecoxib 200 mg	3.0 (2.2-4.7)	4.2 (3.4-5.6)
Paracetamol 975/1000 mg	3.0 (2.6-3.5)	3.6 (3.2-4.1)
Ibuprofen acid 600 mg	3.1 (1.7-26)	2.7 (2.0-4.2)
Aspirin 900-1000 mg	3.3 (2.3-5.8)	4.2 (3.8-4.6)
Ketoprofen 50 mg	3.4 (2.5-5.4)	3.3 (2.7-4.3)
Paracetamol + codeine 600/650 + 60 mg	3.7 (2.9-4.9)	3.9 (3.3-4.7)
Etodolac 100 mg	3.8 (2.6-7.4)	4.8 (3.5-7.8)
Paracetamol 600/650 mg	3.9 (3.1-5.3)	4.6 (3.9-5.5)
Paracetamol 500 mg	5.3 (3.4-12)	3.5 (2.7-4.8)
Codeine 60 mg	7.3 (5.1-13)	12 (8.4-18)

CI, confidence interval; NMA, network meta-analysis; NNT, number needed to treat.

Another potential weakness is the uncertain comparability of the single doses used to achieve analgesia between analgesics, together with the fact that dosing schedules of analgesics in clinical practice are often quite different. For example, the recommended dose for acute pain of celecoxib is 400 mg initially, followed by an additional 200-mg dose if needed on the first day, whereas for diclofenac K the recommended dosage for acute pain is 50 mg 3 times daily or initial dose of 100 mg followed by a 50-mg dose if needed.

Any interpretation of the efficacy beyond single doses has to be done with caution and requires quite different trial designs.

6. Network meta-analysis

A broad range of relative efficacy estimates was produced by the NMA, consistent across the different outputs. Number needed to treat values ranged from under 2 (good) to over 10 (poor). There was broad agreement between efficacy results from the NMA and Cochrane indirect analyses. Not unexpectedly some

disagreements occurred when data sets were small, or effect sizes small, or both.^{19,30,39,50}

This consistent agreement provides great confidence in our results, and the trials on which they are based. Avoidance of trials with high risk of bias, methodological competence, an important outcome, and avoidance of small patient numbers are probably the key factors underpinning our confidence. Those same factors probably also underpin the consistent estimates—high quality means that simple direct comparisons with placebo provide as accurate an estimate of efficacy as one from an extensive and rich network.

This NMA is one of a small but growing number of such analyses in acute and chronic pain. Most have involved small numbers of patients and trials, although several in musculoskeletal conditions included larger numbers of patients (146,524⁵³; 58,451⁶) than this NMA (39,753 patients). The larger studies have tended to support what is already known. For example, the NMA analyses of musculoskeletal conditions both showed paracetamol to have little efficacy, but diclofenac Na 150 mg/day

and etoricoxib 60 mg/day to have the highest probability of being the best intervention.

What this means, of course, is that complicated NMA methods are not necessarily superior, despite theoretical considerations that imply that they might or should be. Neither this NMA nor other NMAs in pain have shown any important insights beyond what was already known. That is despite this NMA having considerably more cross-linking data; in pain and other areas it is common for comparisons with placebo to dominate an NMA, although that is not always the case.^{8,44,51,53}

One reason that the NMA and Cochrane indirect analyses were so similar is likely to arise from the source of data. Trials for both analyses had to meet stringent quality criteria, with consequent low risk of bias. The additional consistency in methods and outcomes would make major discrepancies unlikely.

7. Clinical implications

This NMA, together with individual Cochrane reviews and a Cochrane overview, provides a solid basis of knowledge about drug efficacy in acute pain and increases our confidence in these results. The range of drugs and doses covered should make these results of importance globally. They should help decision-making and medicines management at the local, national, and international level. It is unlikely that any large set of new data will become available to change the results. The results justify a GRADE evaluation of high-quality evidence, meaning that this research provides a very good indication of the likely effects, so the likelihood that the effects will be substantially different is low.¹³

Adverse events were not addressed. An overview of adverse events in acute pain trials is available,³⁶ but inconsistencies in reporting adverse events make it unlikely that an NMA would provide any more insight. Adverse event rates are usually low in these short duration studies, which, combined with low-quality outcome reporting, limits what is possible.

It is much more difficult to distinguish between drugs of high efficacy, which we arbitrarily consider to be an NNT of 3 or below in this circumstance, as **Figure 5** demonstrates. Neither NMA nor Cochrane indirect comparisons provide any more insight. Experience is that the determination of dose or drug differences is likely to require either very large single studies or meta-analysis of head-to-head comparisons.²⁹ In making decisions about appropriate use, results of efficacy analyses will be combined with patient characteristics, experience, availability, and cost, among other considerations.

8. Future research

There are further issues highlighted by the NMA. Perhaps, the most interesting is that of temporal changes for the number responding with placebo, although with no apparent impact on analgesic efficacy estimates. In some ways that is similar to different surgery types, which have been found to have different placebo responses with no difference in efficacy estimate.³

Temporal changes in placebo have been noticed in trials of neuropathic pain, but in that case increasing in time, and then in only U.S. trials.⁵² Other chronic pain studies report increases in fibromyalgia,¹⁵ decreases in irritable bowel syndrome,⁴⁵ or no change in neuropathic or chronic pain.^{22,46,54} As far as we know, this is the first observation of changes in placebo response with time in acute pain studies. This is clearly of potential importance and requires better understanding.

To highlight the absence of information in the extremes of age is not new. Reviews providing data omitted children, but the

inadequacy of knowledge about interventions for paediatric pain is frequently attested.^{7,12,48,55} There was, however, no restriction on maximum age. The fact that only 7 trials reported a mean age above 50 years is a concern when the older population is growing rapidly. A more focussed examination of drug efficacy for acute pain in older adults is probably in order.

And finally, there may be an opportunity to examine our ability to determine whether there is a meaningful difference between the efficacies of analgesics with NNTs below 3. The network was rich and is likely to contain a small number of trials with the same 2 active drugs. Methods previously demonstrated to be useful for dose-response²⁹ might be used in that circumstance also, allowing for a comparison of head-to-head analysis with the overall result.

9. Conclusion

The concordance between NMA and Cochrane indirect analyses gives further validity for the use of indirect analyses. When there is low heterogeneity in methods and measurement, and the trials are of an adequate size, NMA adds little to the overall conclusion. That said, NMA is a useful analytical tool. Here it increases our confidence in the estimates of efficacy of analgesics in acute postoperative pain and provokes novel insight into the changes in trial design over time.

Conflict of interest statement

The research was led by Oxford Medical Knowledge (OMK). R.A. Moore is the owner of Oxford Medical Knowledge (OMK), which has received funding from Novartis Pharma AG for this research. S. Banerjee is an employee of Novartis Healthcare Pvt Ltd, and R. Karan, E. Glimm, and A. Wiksten are employees of Novartis Pharma AG. S. Derry, P.J. Wiffen, D. Aldington, and C. Eccleston have no interests to declare.

The Oxford Pain Relief Trust provided institutional support.

Acknowledgements

Author contributions: R.A. Moore had the original idea for the study, and R.A. Moore, S. Derry, P.J. Wiffen, and C. Eccleston determined the initial structure. R.A. Moore and S. Derry performed the data extraction for the NMA. R.A. Moore, R. Karan, S. Banerjee, A. Wiksten, and E. Glimm determined NMA analysis methods and ran the analyses. D. Aldington provided clinical insight. R.A. Moore wrote the original draft of the paper, and all authors collaborated on its development.

Appendix A. Supplemental digital content

Supplemental digital content associated with this article can be found online at <http://links.lww.com/PAIN/A611>.

Article history:

Received 22 March 2018

Received in revised form 1 June 2018

Accepted 20 June 2018

Available online 10 July 2018

References

- [1] AlBalawi Z, McAlister FA, Thorlund K, Wong M, Wetterslev J. Random error in cardiovascular meta-analyses: how common are false positive and false negative results? *Int J Cardiol* 2013;168:1102–7.

- [2] Barden J, Derry S, McQuay HJ, Moore RA. Bias from industry trial funding? A framework, a suggested approach, and a negative result. *PAIN* 2006;121:207–18.
- [3] Barden J, Edwards JE, McQuay HJ, Andrew Moore R. Pain and analgesic response after third molar extraction and other postsurgical pain. *PAIN* 2004;107:86–90.
- [4] Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331:897–900.
- [5] Collins SL, Edwards J, Moore RA, Smith LA, McQuay HJ. Seeking a simple measure of analgesia for mega-trials: is a single global assessment good enough? *PAIN* 2001;91:189–94.
- [6] Cooper SA. Single-dose analgesic studies: the upside and downside of assay sensitivity. In: Max MB, Portenoy RK, Laska EM, editors. *The design of analgesic clinical trials. Advances in pain research and therapy. Vol. 18.* New York: Raven Press, 1991. p. 117–24.
- [7] Cooper TE, Heathcote LC, Clinch J, Gold JI, Howard R, Lord SM, Schechter N, Wood C, Wiffen PJ. Antidepressants for chronic non-cancer pain in children and adolescents. *Cochrane Database Syst Rev* 2017;8:CD012535.
- [8] da Costa BR, Reichenbach S, Keller N, Nartey L, Wandel S, Jüni P, Trelle S. Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. *Lancet* 2017;390:e21–e33.
- [9] Dechartres A, Altman DG, Trinquart L, Boutron I, Ravaud P. Association between analytic strategy and estimates of treatment outcomes in meta-analyses. *JAMA* 2014;312:623–30.
- [10] Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ* 2013;346:f2304.
- [11] Derry S, Cooper TE, Phillips T. Single fixed-dose oral dextropropofol plus tramadol for acute postoperative pain in adults. *Cochrane Database Syst Rev* 2016;9:CD012232.
- [12] Eccleston C, Cooper TE, Fisher E, Anderson B, Wilkinson NM. Non-steroidal anti-inflammatory drugs (NSAIDs) for chronic non-cancer pain in children and adolescents. *Cochrane Database Syst Rev* 2017;8:CD012537.
- [13] Effective Practice and Organisation of Care (EPOC). 23. Worksheets for preparing a summary of findings using GRADE. Resources for review authors. Oslo: Norwegian Knowledge Centre for the Health Services, 2015. Available at: epoc.cochrane.org/epoc-specific-resources-review-authors. Accessed January 19, 2018.
- [14] Gaskell H, Derry S, Wiffen PJ, Moore RA. Single dose oral ketoprofen or dextropropofol for acute postoperative pain in adults. *Cochrane Database Syst Rev* 2017;5:CD007355.
- [15] Häuser W, Bartram-Wunn E, Bartram C, Reinecke H, Tölle T. Systematic review: placebo response in drug trials of fibromyalgia syndrome and painful peripheral diabetic neuropathy—magnitude and patient-related predictors. *PAIN* 2011;152:1709–17.
- [16] Hearn L, Derry S, Moore RA. Single dose dipyrone (metamizole) for acute postoperative pain in adults. *Cochrane Database Syst Rev* 2016;4:CD011421.
- [17] Hospital episode statistics, admitted patient care—England, 2014–15. NHS Digital, 2015. Available at: <https://digital.nhs.uk/catalogue/PUB19124>. Accessed March 1, 2018.
- [18] Houde RW, Wallenstein SL, Rogers A. Clinical pharmacology of analgesics. 1. A method of assaying analgesic effect. *Clin Pharmacol Ther* 1960;1:163–74.
- [19] Int'Hout J, Ioannidis JP, Borm GF, Goeman JJ. Small studies are more heterogeneous than large ones: a meta-meta-analysis. *J Clin Epidemiol* 2015;68:860–9.
- [20] Jadad AR, McQuay HJ. A high-yield strategy to identify randomized controlled trials for systematic reviews. *Online J Curr Clin Trials* 1993;Doc No 33. PMID: 8306000.
- [21] Jansen JP, Fleurence R, Devine B, Iltzler R, Barrett A, Hawkins N, Lee K, Boersma C, Anjemans L, Cappelleri JC. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health* 2011;14:417–28.
- [22] Katz J, Finnerup NB, Dworkin RH. Clinical trial outcome in neuropathic pain: relationship to study characteristics. *Neurology* 2008;70:263–72.
- [23] Li T, Puhon MA, Vedula SS, Singh S, Dickersin K. Ad Hoc Network Meta-analysis Methods Meeting Working Group. Network meta-analysis—highly attractive but more methodological research is needed. *BMC Med* 2011;9:79.
- [24] Li Wan Po A, Petersen B. How high should total pain-relief score be to obviate the need for analgesic re-medication in acute pain? Estimation using signal detection theory and individual-patient meta-analysis. *J Clin Pharm Ther* 2006;31:161–5.
- [25] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;23:3105–24.
- [26] Martinez V, Beloeil H, Marret E, Fletcher D, Ravaud P, Trinquart L. Non-opioid analgesics in adults after major surgery: systematic review with network meta-analysis of randomized trials. *Br J Anaesth* 2017;118:22–31.
- [27] McCullagh P, Nelder JA. Quasi-likelihood functions. In: McCullagh P, Nelder JA, editors. *Generalized linear models.* London: Chapman and Hall, 1989. p. 323–56.
- [28] McQuay HJ, Derry S, Eccleston C, Wiffen PJ, Moore RA. Evidence for analgesic effect in acute pain—50 years on. *PAIN* 2012;153:1364–7.
- [29] McQuay HJ, Moore RA. Dose-response in direct comparisons of different doses of aspirin, ibuprofen and paracetamol (acetaminophen) in analgesic studies. *Br J Clin Pharmacol* 2007;63:271–8.
- [30] Mhuircheartaigh RJ, Moore RA, McQuay HJ. Analysis of individual patient data from clinical trials: epidural morphine for postoperative pain. *Br J Anaesth* 2009;103:874–81.
- [31] Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. *BMJ* 2013;346:f2914.
- [32] Moore A, McQuay H, Gavaghan D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics. *PAIN* 1996;66:229–37.
- [33] Moore A, McQuay H, Gavaghan D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics: verification from independent data. *PAIN* 1997;69:127–30.
- [34] Moore A, Moore O, McQuay H, Gavaghan D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics: use of pain intensity and visual analogue scales. *PAIN* 1997;69:311–5.
- [35] Moore RA, Barden J, Derry S, McQuay HJ. Managing Potential publication bias. In: McQuay HJ, Kalso E, Moore RA, editors. *Systematic reviews in pain research: methodology refined.* Seattle: IASP Press, 2008. p. 15–23.
- [36] Moore RA, Derry S, Aldington D, Wiffen PJ. Adverse events associated with single dose oral analgesics for acute postoperative pain in adults—an overview of Cochrane reviews. *Cochrane Database Syst Rev* 2015;10:CD011407.
- [37] Moore RA, Derry S, Aldington D, Wiffen PJ. Single dose oral analgesics for acute postoperative pain in adults—an overview of Cochrane reviews. *Cochrane Database Syst Rev* 2015;9:CD008659.
- [38] Moore RA, Edwards JE, McQuay HJ. Acute pain: individual patient meta-analysis shows the impact of different ways of analysing and presenting results. *PAIN* 2005;116:322–31.
- [39] Moore RA, Gavaghan D, Tramèr MR, Collins SL, McQuay HJ. Size is everything—large amounts of information are needed to overcome random effects in estimating direction and magnitude of treatment effects. *PAIN* 1998;78:209–16.
- [40] Moore RA, McQuay HJ. Single-patient data meta-analysis of 3453 postoperative patients: oral tramadol versus placebo, codeine and combination analgesics. *PAIN* 1997;69:287–94.
- [41] Moore RA, Straube S, Aldington D. Pain measures and cut-offs—“no worse than mild pain” as a simple, universal outcome. *Anaesthesia* 2013;68:400–12.
- [42] Moore RA, Straube S, Paine J, Derry S, McQuay HJ. Minimum efficacy criteria for comparisons between treatments using individual patient meta-analysis of acute pain trials: examples of etoricoxib, paracetamol, ibuprofen, and ibuprofen/paracetamol combinations after third molar extraction. *PAIN* 2011;152:982–9.
- [43] Nguyen TL, Collins GS, Lamy A, Devereaux PJ, Daurès JP, Landais P, Le Manach Y, Le Manach Y. Simple randomization did not protect against bias in smaller trials. *J Clin Epidemiol* 2017;84:105–13.
- [44] Nüesch E, Häuser W, Bernardy K, Barth J, Jüni P. Comparative efficacy of pharmacological and non-pharmacological interventions in fibromyalgia syndrome: network meta-analysis. *Ann Rheum Dis* 2013;72:955–62.
- [45] Pitz M, Cheang M, Bernstein CN. Defining the predictors of the placebo response in irritable bowel syndrome. *Clin Gastroenterol Hepatol* 2005;3:237–47.
- [46] Quessy SN, Rowbotham MC. Placebo response in neuropathic pain trials. *PAIN* 2008;138:479–83.
- [47] Salanti G, Higgins J, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;17:279–301.
- [48] Schnabel A, Thyssen NM, Goeters C, Zheng H, Zahn PK, Van Aken H, Pogatzki-Zahn EM. Age- and procedure-specific differences of epidural analgesia in children—a database analysis. *Pain Med* 2015;16:544–53.
- [49] Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326:472.

- [50] Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, Guyatt G, Devereaux PJ, Thabane L. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis—a simulation study. *PLoS One* 2011;6:e25491.
- [51] Thorlund K, Toor K, Wu P, Chan K, Druyts E, Ramos E, Bhambri R, Donnet A, Stark R, Goadsby PJ. Comparative tolerability of treatments for acute migraine: a network meta-analysis. *Cephalalgia* 2017;37:965–78.
- [52] Tuttle AH, Tohyama S, Ramsay T, Kimmelman J, Schweinhardt P, Bennett GJ, Mogil JS. Increasing placebo responses over time in U.S. clinical trials of neuropathic pain. *PAIN* 2015;156:2616–26.
- [53] van Walsem A, Pandhi S, Nixon RM, Guyot P, Karabis A, Moore RA. Relative benefit-risk comparing diclofenac to other traditional non-steroidal anti-inflammatory drugs and cyclooxygenase-2 inhibitors in patients with osteoarthritis or rheumatoid arthritis: a network meta-analysis. *Arthritis Res Ther* 2015;17:66.
- [54] Vase L, Vollert J, Finnerup NB, Miao X, Atkinson G, Marshall S, Nemeth R, Lange B, Liss C, Price DD, Maier C, Jensen TS, Segerdahl M. Predictors of the placebo analgesia response in randomized controlled trials of chronic pain: a meta-analysis of the individual data from nine industrially sponsored trials. *PAIN* 2015;156:1795–802.
- [55] Wiffen PJ, Cooper TE, Anderson AK, Gray AL, Grégoire MC, Ljungman G, Zernikow B. Opioids for cancer-related pain in children and adolescents. *Cochrane Database Syst Rev* 2017;7:CD012564.