Should we Abandon Acetaminophen? Network Meta-Analysis of NSAIDs and Acetaminophen for OA

Innovation and Clarity Conference

November 24, 2018

Objectives

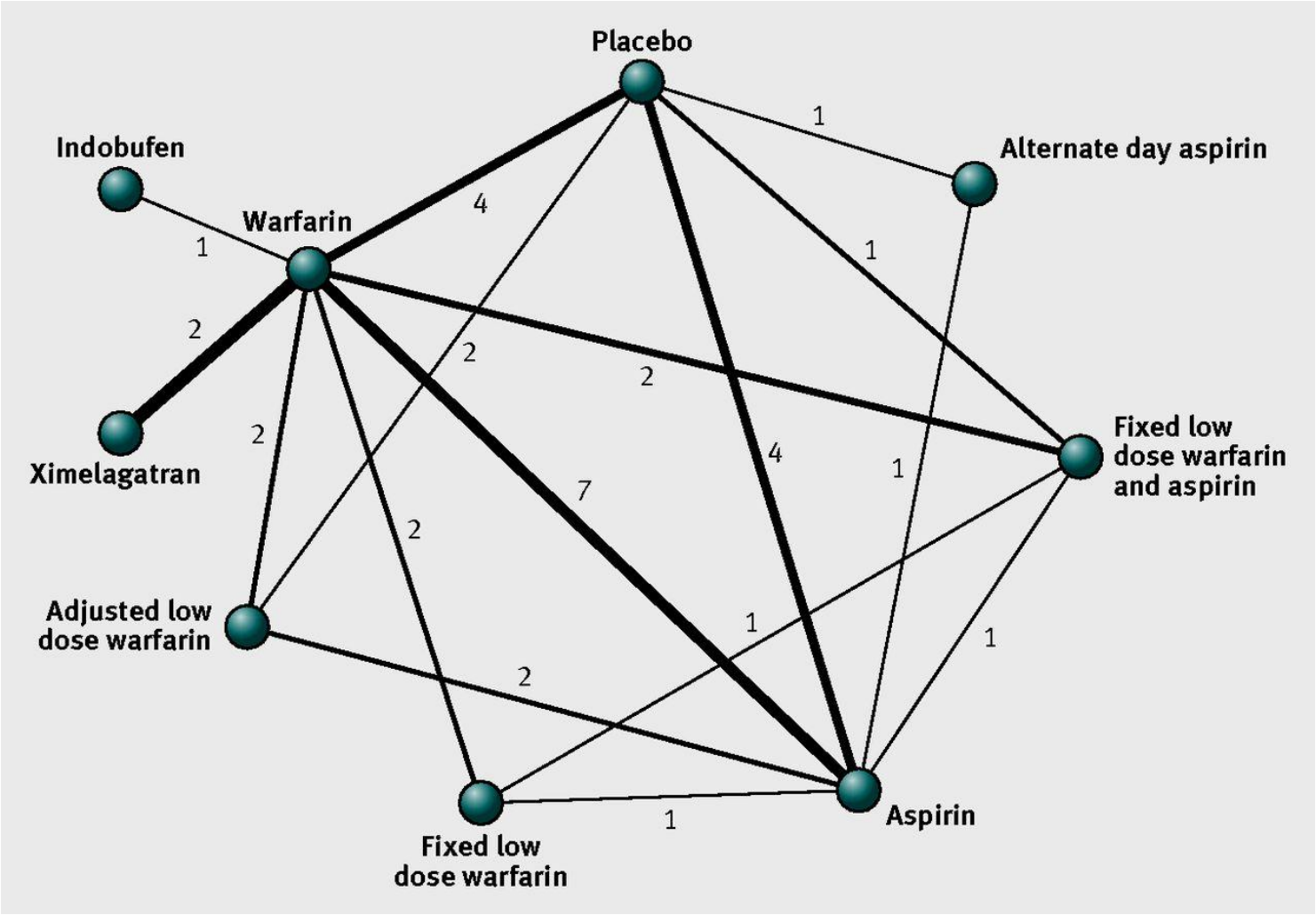
- Understand what is a (good) network meta-analysis
- Review evidence of benefit of NSAIDs, acetaminophen for knee of hip OA ***Lancet* 2016; 387: 2093–105**
- Consider other factors
- Decide whether to keep prescribing acetaminophen

Network Meta-Analysis (NMA)

- AKA multiple treatment comparison meta-analysis or mixed treatment meta-analysis
- Set of methods to visualize and interpret the wider picture of the evidence
- allows for estimation of comparative effects that have not been investigated head to head in randomized clinical trials

BMJ 2013;346:f2914

Fig 1 Network geometry of well connected network of randomized controlled trials (RCTs) evaluating stroke prevention among populations with atrial fibrillation.



34 RCTs of 45 possible comparisons

Edward J Mills et al. BMJ 2013;346:bmj.f2914

Network Geometry

  - Which treatments (nodes) have head-to-head trials?
  - Which connected via common comparators?
  - What is level of evidence for each?
- Nodes not well connected should be interpreted with caution
- Severe imbalance in amount of evidence for each intervention may affect power and reliability
- NMA only as good as evidence it includes

Heterogeneity and Incoherence

- Homo/heterogeneity describes different trials in one pairwise comparison
  - Measured by Cochrane's Q or I^2
- In/coherence describes different trials informing indirect vs direct comparisons
  - Whether we get different results from direct vs indirect comparisons
- Conceptual heterogeneity is differences in study characteristics that appear heterogeneous at face value

Heterogeneity and Incoherence

- Can have type I or II error when testing for it
- Combine statistical tests with conceptual reasoning
- If has clear heterogeneity and/or incoherence, may be unjustified to use NMA, but still may get ‘right answer’
- Random effects models may accommodate unexplained heterogeneity and make incoherence less prominent

Data synthesis

- Various models:
  - Fixed vs random effects
- Fixed effect:
  - Assumes little or no heterogeneity, gives artificially narrow CI if really is heterogeneous
- Random Effects:
  - Assumes and accounts for heterogeneity

Then Math happens

The weighted effect size $\hat{\beta}_{\text{Fixed}}$ under the fixed-effects model is

$$\hat{\beta}_{\text{Fixed}} = \frac{\sum_{i-1}^{k} w_i y_i}{\sum_{i-1}^{k} w_i} \tag{2}$$

where $w_i = 1/\sigma_i^2$ is the weight and k is the total number of studies. The sampling variance $\hat{s}^2_{\text{Fixed}}$ of $\hat{\beta}_{\text{Fixed}}$ is computed by

$$\hat{s}^2_{\text{Fixed}} = 1/\sum_{i-1}^{k} w_i. \tag{3}$$

If the population effect sizes are homogeneous, $\hat{\beta}_{\text{Fixed}}$ is an unbiased estimate of the population effect size. Moreover, $\hat{\beta}_{\text{Fixed}}$ has the smallest sampling variance of all possible weighted estimators when the sampling variances are truly known. Because σ_i^2 is usually estimated, the accuracy of $\hat{s}^2_{\text{Fixed}}$ in estimating the sampling variance of $\hat{\beta}_{\text{Fixed}}$ depends on how accurate the estimated value of σ_i^2 is (see Hedges, 2007; Hedges & Olkin, 1985). After the averaging, it is of interest to test whether the weighted effect size $\hat{\beta}_{\text{Fixed}}$ is statistically significant. We may compute a test statistic

$$Z_1 = \hat{\beta}_{\text{Fixed}}/\hat{s}_{\text{Fixed}}. \tag{4}$$

Under the null hypothesis $H_0 : \beta_{\text{Fixed}} = 0$, the test statistic Z_1 has an approximate standard normal distribution.

Combining estimates of effect sizes across studies with the fixed-effects model is appropriate only when the effect sizes are homogeneous (National Research Council, 1992); otherwise, the estimated standard error on the weighted mean under the fixed-effects models is smaller than its true value when the effect sizes are heterogeneous. To test the homogeneity of the effect sizes, we may compute a Q statistic (Cochran, 1954):

$$Q = \sum_{i-1}^{k} w_i (y_i - \hat{\beta}_{\text{Fixed}})^2. \tag{5}$$

Under the null hypothesis $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k$, the Q statistic has an approximate chi-square distribution with $(k - 1)$ degrees of freedom. However, this does not necessary mean that fixed-effects models should never be used whenever the effect sizes are heterogeneous. Hedges and Vevea (1998) pointed out that fixed-effects models are still appropriate even if the effect sizes are heterogeneous when the researchers are interested in only this collection of studies. This is what they call a conditional inference.

Models with covariates. Besides estimating a common effect size, study characteristics may be used as covariates to model the variability among the effect sizes. Study characteristics can be in the form of categorical covariates (Hedges, 1982a) and continuous covariates (Hedges, 1982b). A weighted least squares (WLS) approach is usually used to model the variability among the effect sizes with covariates (e.g., Hedges & Olkin, 1985). It would be more convenient to express the model in matrix notation

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{e}, \tag{6}$$

where $\mathbf{y}$ is a $k \times 1$ vector of effect sizes, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients including the intercept, $\mathbf{e}$ is a $k \times 1$ vector of residuals, and X is a $k \times p$ design matrix including ones in the first column. Because the effect sizes are assumed to be independent, the covariance matrix of the residuals V_e is a diagonal matrix, that is, $V_e = \text{diag}[\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2]$.

The vector of the estimated regression coefficients via WLS is

$$\hat{\boldsymbol{\beta}} = (X^T V_e^{-1} X)^{-1} X^T V_e^{-1} \mathbf{y}, \tag{7}$$

where $V_e^{-1} = \text{diag}[1/\sigma_1^2, 1/\sigma_2^2, \ldots, 1/\sigma_k^2]$ and the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\hat{V}_{\hat{\beta}} = (X^T V_e^{-1} X)^{-1}. \tag{8}$$

It is useful to test whether all of the p regression coefficients including the intercept are statistically significant. We may compute a large-sample test statistic

$$\tilde{Q} = \hat{\boldsymbol{\beta}}^T \hat{V}_{\hat{\beta}}^{-1} \hat{\boldsymbol{\beta}}. \tag{9}$$

Under the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$, the test statistic $\tilde{Q}$ is approximately distributed as a chi-square variate with p degrees of freedom. The above test can easily be modified to test hypotheses on some of the elements in $\boldsymbol{\beta}$, for instance, the regression coefficients excluding the intercept (Hedges & Olkin, 1985).

After testing the significance of all (or some) regression coefficients, we may want to construct an approximate Z test to test the significance of an individual regression coefficient under $H_0 : \beta_i = 0$,

$$Z_2 = \hat{\beta}_i / \sqrt{(\hat{V}_{\hat{\beta}})_{ii}}, \tag{10}$$

where $(\hat{V}_{\hat{\beta}})_{ii}$ is the sampling variance of $\hat{\beta}_i$.

Random-Effects Models

Fixed-effects models assume that the population effect sizes share a common value. Many researchers have argued that studies are not direct replications of each other. It is expected that there will be differences in the population

Ways to display probabilities

- Probabilities can be fragile when network sparse, may change dramatically with new data
- Safer to focus on treatment effects and their uncertainty

Table 1 Treatment effect estimates from example network

Network comparator treatments	Rate ratio (95% credible interval)	Treatment	Probability of being best treatment (%)
—	—	Placebo	0
Adjusted standard dose warfarin *v* placebo	0.37 (0.26 to 0.53)	Adjusted standard dose warfarin	3
Adjusted low dose warfarin *v* placebo	0.32 (0.18 to 0.56)	Adjusted low dose warfarin	16
Fixed low dose warfarin *v* placebo	0.76 (0.30 to 1.76)	Fixed low dose warfarin	1
Aspirin *v* placebo	0.62 (0.43 to 0.86)	Aspirin	0
Fixed low dose warfarin and aspirin *v* placebo	0.98 (0.60 to 1.67)	Fixed low dose warfarin and aspirin	0
Ximelagatran *v* placebo	0.35 (0.19 to 0.65)	Ximelagatran	11
Alternate day aspirin *v* placebo	0.17 (0.01 to 1.15)	Alternate day aspirin	66
Indobufen *v* placebo	0.46 (0.19 to 1.14)	Indobufen	5

ASA EOD comes out as ‘likely best treatment’ due to its best point estimate of benefit, even though its estimate of treatment benefit is imprecise (may have been chance finding, may not differ from placebo)

Summary

- A strong NMA has:
  - A robust network geometry with many studies comparing multiple treatments
  - Homogeneity and coherence
  - Likely uses random effects models
  - Focuses on estimates of treatment effects

NSAIDs for OA

- OA is really common
- Pain causes disability, affects all-cause mortality
- NSAIDs 'mainstay' of treatment
  - 65% of US people with OA receive NSAIDs
  - Many choices, often discontinue initial treatment or switch drugs
  - Inadequate pain control and/or side effects

Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis *B R da Costa*, S Reichenbach*, N Keller, L Nartey, S Wandel, P Jüni, S Trelle* Lancet 2016; 387: 2093–105

Selection criteria

- Large (>100/group) RCTs in knee or hip OA
  - OA reported separately, or >80% of subjects OA
- Any NSAID, acetaminophen or placebo
- Excluded lone abstracts, any language
- Cochrane CENTRAL registry 1980 to 2015, checked Medline and Embase, their own database, screened reference lists, checked ClinicalTrials.gov and other trial registries

Study Selection

- 2 investigators independently screened all trials, translator if non-English
- “reached consensus using a standardised, piloted web-based data management tool for systematic reviews, accompanied by a codebook”
- Used ITT if possible

Endpoints

- Primary outcome pain
  - Highest available of:
    - (1) global pain score; (2) pain on walking;(3)WOMAC osteoarthritis index pain subscore; (4) composite pain scores other than WOMAC; (5) pain on activities other than walking (such as stair climbing); (6) WOMAC global score; (7) Lequesne osteoarthritis index global score; (8) other algofunctional composite scores; (9) patient's global assessment; (10) physician's global assessment
  - 1, 2, 4, 6, weeks, 6, 12 months if reported
- Secondary outcome, physical function, same deal

Statistics

- Methodologic quality by ‘adapted’ version of Cochrane
- Multivariable Bayesian random effects model for mixed multiple treatment comparisons
  - “fully preserves the direct comparisons”, “allows comparison of all treatments….trials, accounts for multiple comparisons….. within trials”
  - “random walk” assumes that outcomes at adjacent time points are more similar than at remote time points

Sensitivity Analysis

- Explored assumptions about about relation between time and treatment effect
- Adjusted results for trial characteristics (blinding, data completeness, LOCF, site of OA) by regression coefficient
- Dose-response by drug-specific covariates
- Separate analyses per time point

More Stats

- Medians with 95% CI
- Effect size = difference between median divided by pooled SD (or SE)
- Goodness of fit by “number of means of standardised node-based residuals within 1·96 of the standard normal distribution”
- Visually inspecting the distribution of residuals on Q–Q plots

Heterogeneity…

- Estimated from the posterior median between trial variance τ^2
- Consistency of the network determined by the difference in effect sizes derived from direct and indirect comparisons
  - This conceptual reasoning, not statistically tested
- To examine the data for small study effects, we constructed comparison-adjusted funnel plots

Rankings

- Calculated median rank
- Probability of treatment reaching MCID (-0.37SD units~ 9mm on VAS)
- SUCRA (surface under the cumulative ranking line)
  - SUCRA of 100 = certain to be the best

Conflicts of Interest

- This study was funded by the Swiss National Science Foundation and by a grant from the Arco Foundation, Switzerland
  - Not sure where Arco Fdn gets its money
- The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report
- 2 principal authors declare no COI, others have some

Results

- 8973 reports yielded 74 RCTs
- 23 nodes in NMA
  - Most studies for cox-II drugs, highest for celecoxib
  - 4 drugs had only 1 trial each
- Mean age ranged 58 to 71, 49 to 90% female
- Median f/u 12 weeks (1-52)
- 58,556 subjects in primary analysis

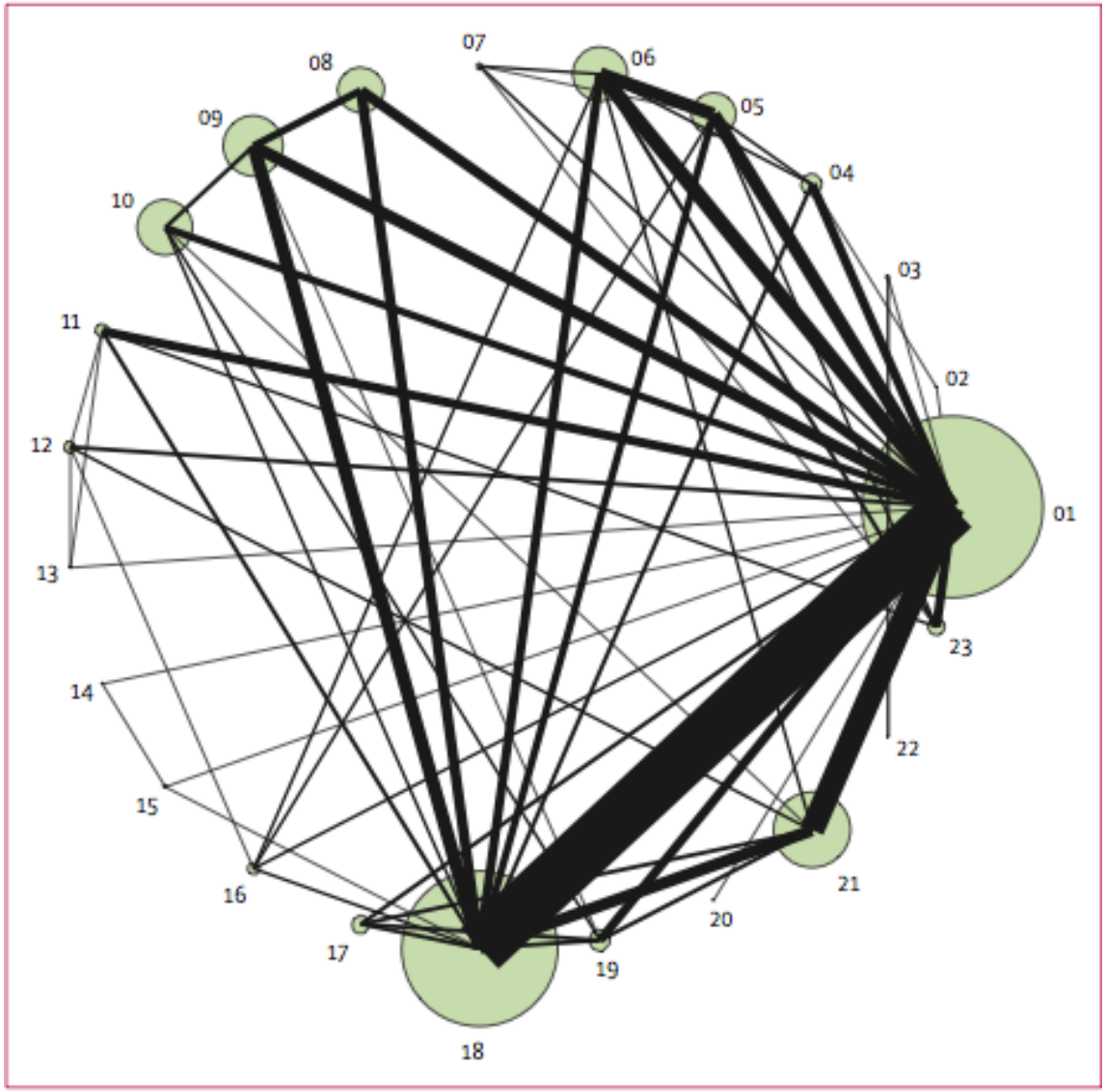


***Figure 1*: Network of comparisons included in the analyses**

The size of every circle is proportional to the number of randomly assigned patients and indicates the sample size. The width of the lines corresponds to the number of trials. 01=placebo. 02=paracetamol <2000 mg. 03=paracetamol 3000 mg. 04=paracetamol 3900–4000 mg. 05=rofecoxib 12·5 mg. 06=rofecoxib 25 mg. 07=rofecoxib 50 mg. 08=lumiracoxib 100 mg. 09=lumiracoxib 200 mg. 10=lumiracoxib 400 mg. 11=etoricoxib 30 mg. 12=etoricoxib 60 mg. 13=etoricoxib 90 mg. 14=diclofenac 70 mg. 15=diclofenac 100 mg. 16=diclofenac 150 mg. 17=celecoxib 100 mg. 18=celecoxib 200 mg. 19=celecoxib 400 mg. 20=naproxen 750 mg. 21=naproxen 1000 mg. 22=ibuprofen 1200 mg. 23=ibuprofen 2400 mg.

Quality

- Low risk of bias for blinding
- 74% had incomplete outcome data of varying degree
- 68% used LOCF
- 27% didn’t do ITT
- 92% financed by ‘a commercial body’, 8% were unclear

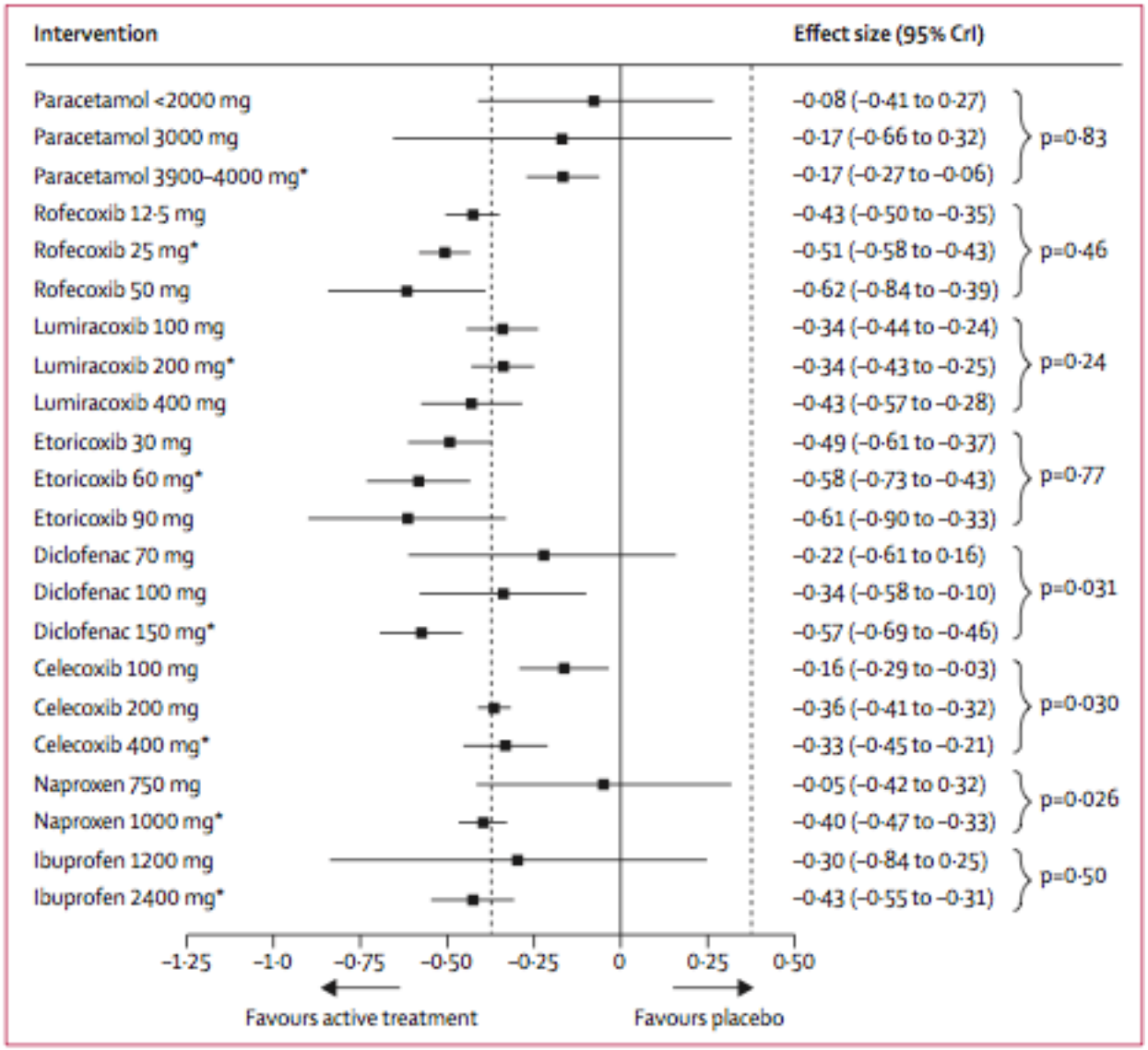


***Figure 2*: Estimates of the treatment effects on pain for different daily doses of NSAIDs and paracetamol compared with placebo**

Analysis considers data from all timepoints as available. Area between dashed lines shows the treatment effect estimates below the minimum clinically important difference. Two-sided p values are derived from tests of linear dose–effect. NSAID=non-steroidal anti-inflammatory drug. CrI=credibility interval. *Maximum approved daily dose.

Effect Sizes

- All interventions have some effect, maybe not above placebo for lower dose acetaminophen and older NSAIDs
- Higher dose diclofenac and the Cox II inhibitors above the MCID
- Some dose effect, linear only for celecoxib, diclofenac and naproxen

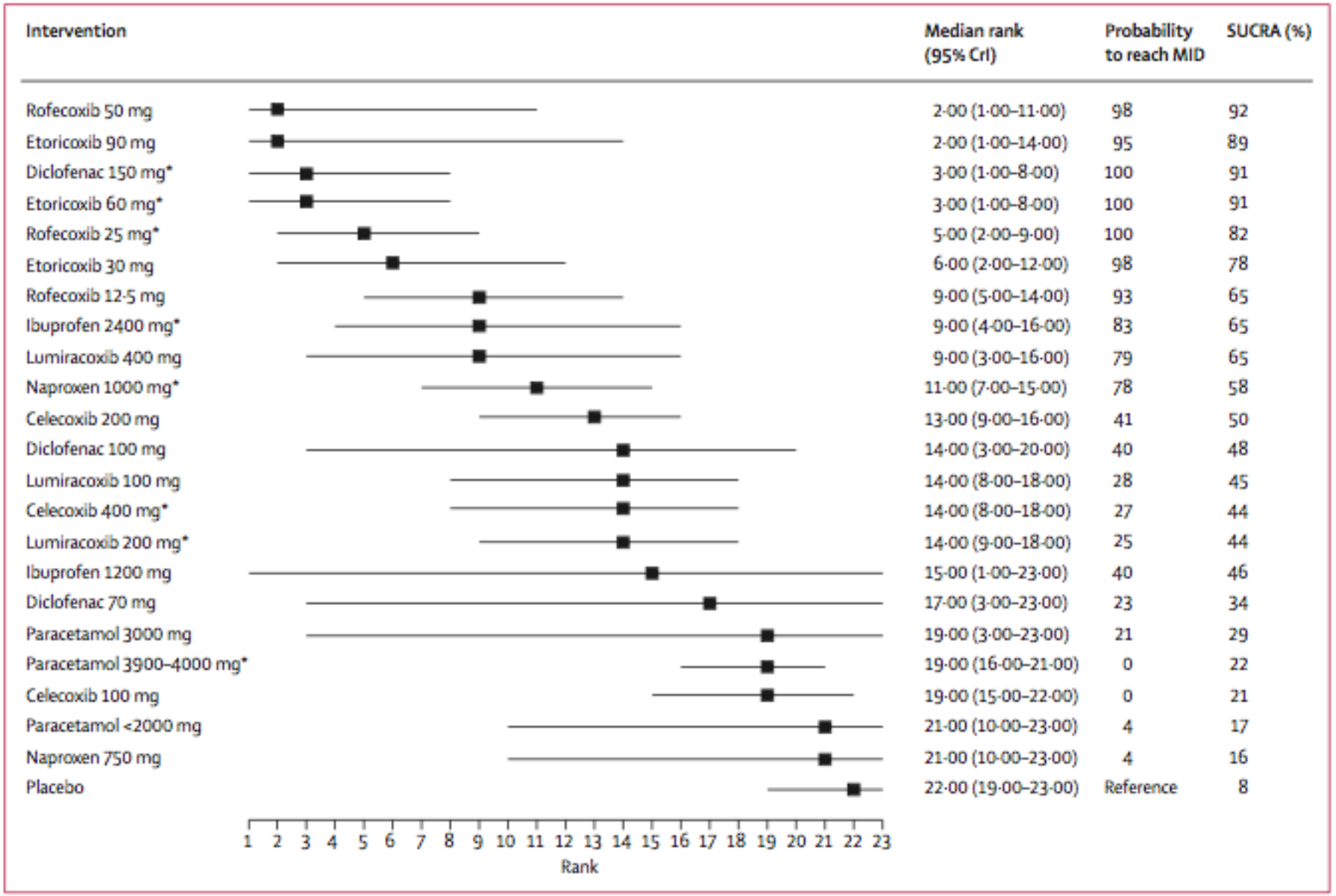


***Figure 3*: Median rank, probability of reaching MID, and SUCRA values of competing interventions and daily doses**

MID=minumum clinically important difference. SUCRA=surface under the cumulative ranking curve. CrI=credibility interval. *Maximum daily dose

Similar deal for physical function

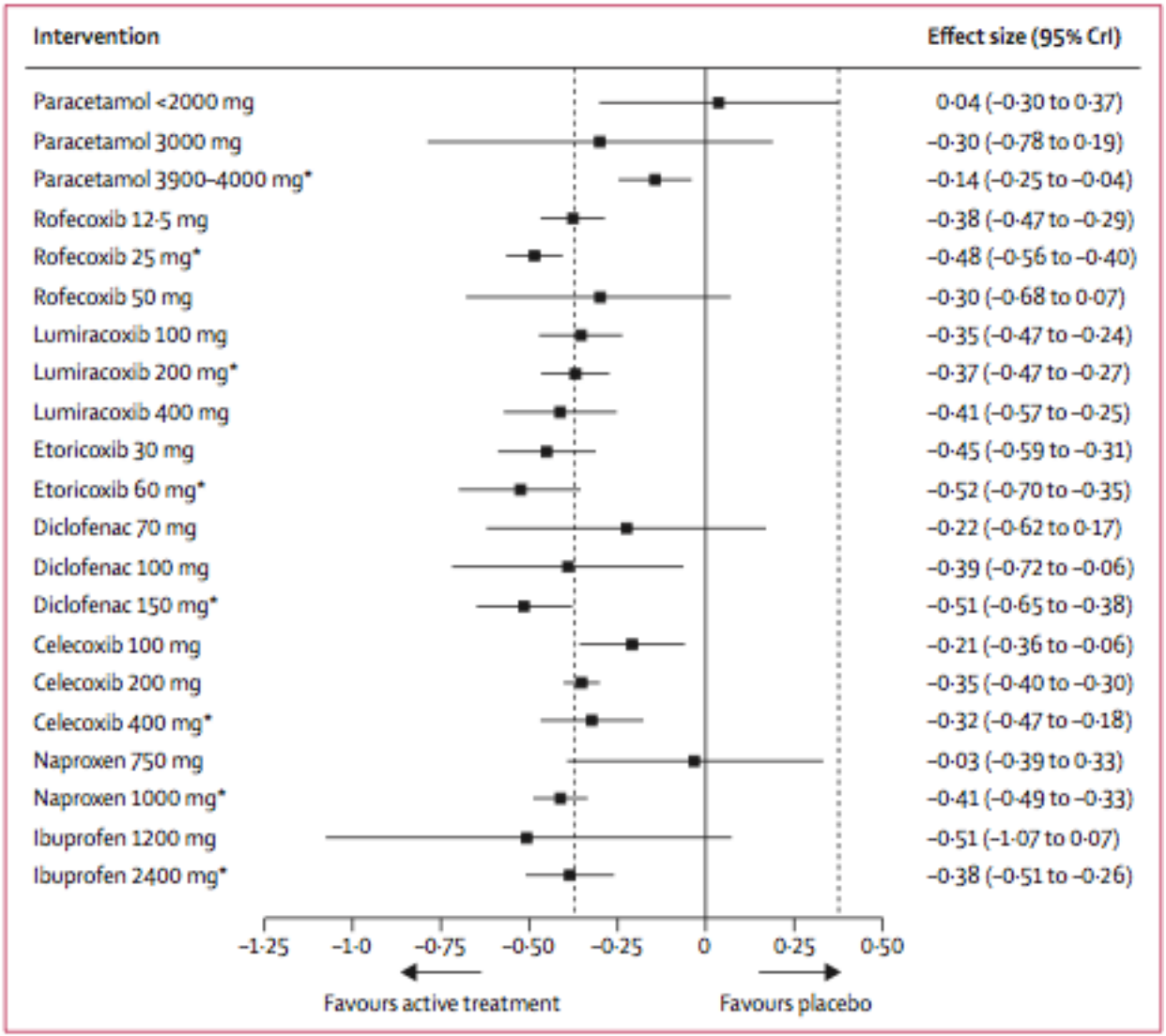


***Figure 4*: Estimates of the treatment effects on physical function for different daily doses of NSAIDs and paracetamol compared with placebo**

Analysis considers data from all timepoints as available. Area between dashed lines shows treatment effect estimates below the minimum clinically important difference. NSAID=non-steroidal anti-inflammatory drug. CrI=credibility interval. *Maximum approved daily dose.

Heterogeneity…

- “model fit was good for both pain and physical function outcomes”
- τ^2 suggests low heterogeneity (0.011, 0.007 - 0.017)
- “no relevant inconsistency”
- Analyses with alternate models similar results
- No interaction of trial characteristics with treatment effect
- Funnel plot no asymmetry

Inconsistency

- **Web-appendix 7. Inconsistency**
- Inconsistency was assessed based on the main model which used timepoint 6 as reference timepoint
- **Pain outcome**
- The median inconsistency factor (ICF) for pain was 0.06 and it ranged from 0.00 to 0.42. Four out of 69 ICFs were statistically significant. None of the ICFs for pain were clinically relevant.
- Web-appendix Figure 2. Inconsistency factors for pain outcome analysis. CI: confidence interval. RoM: Ratio of means. 01 = Placebo; 02 =Paracetamol <2000mg; 03 = Paracetamol 3000mg; 04 = Paracetamol 3900-4000mg; 05 = Rofecoxib 12.5mg; 06 = Rofecoxib 25mg; 07 =Rofecoxib 50mg; 08 = Lumiracoxib 100mg; 09 = Lumiracoxib 200mg; 10 = Lumiracoxib 400mg; 11 = Etoricoxib 30mg; 12 = Etoricoxib 60mg; 13 = Etoricoxib 90mg; 14 = Diclofenac 70mg; 15 = Diclofenac 100mg; 16= Diclofenac 150mg; 17 = Celecoxib 100mg; 18 = Celecoxib 200mg; 19 = Celecoxib 400mg; 20 = Naproxen 750mg; 21 = Naproxen 1000mg; 22 = Ibuprofen 1200mg; 23 = Ibuprofen 2400mg

Authors' Discussion

- Diclofenac won
- Consider adverse effects:
  - Diclofenac increases CV events, GI similar to Cox II
  - Naproxen no effect on CV events but more GI
  - Suggest intermittent or short-term use
- Short to intermediate follow-up in most trials
- Mixed quality of studies, limited by omitting small studies
- Results comparable to other meta-analyses

My thoughts

- Evidence of effect size for diclofenac comes from only a few trials
  - CIs for its median rank are wide
- Only coxib in Canada is celecoxib and meloxicam
- Studies on younger adults, few frail elderly
- Does raise questions about efficacy of acetaminophen

Editorial

- Effect size appears proportional to affinity to the COX-2 enzyme
  - Diclofenac 17 - 87 x more potent than celecoxib
- Patients enrolled in NSAID trials don't represent real life
- NSAIDs commonly used intermittently or short term
  - Should study efficacy and safety of intermittent use

Lancet 2016; 387:2056-7

Editorial

- “Not surprised” that acetaminophen ineffective
- Should it be first line?
- Does that make people suffer by delaying their use of NSAIDs?
- No better options:
  - opioids side effects
  - glucosamine etc ineffective
- Invent new analgesics!

Other MA of Acetaminophen

- 12 RCTs
- High quality evidence that acetaminophen in ineffective for reducing back pain or its disability or QOL
- High quality evidence that acetaminophen provides a significant though not clinically important effect on pain (3.7 mm) and disability (2.9 mm)
- No AE except liver enzymes

BMJ. 2015; 350: h1225.

Other RCT of Acetaminophen

- Cluster RCT, 80 clusters in 18 NH in Norway
- 352 residents with moderate to severe BPSD
- R to stepwise pain Rx vs placebo
- 120 got acetaminophen 3 g, 30 got buprenorphine patch, few got other Rx
- Reduced pain in active Rx group
- Also reduced agitation, effect size comparable to effect of antipsychotic in other trials

Husebo, Ballard et al BMJ. 2011; 343: d4065.

Acetaminophen Harms

- Systematic review of observational studies
- 8 cohort studies
- One of two studies reporting mortality showed a dose–response and reported an increased relative rate of mortality from 0.95 (0.92 to 0.98) to 1.63 (1.58 to 1.68).
- Four studies reporting cardiovascular AEs showed a dose–response with one reporting an increased risk ratio of all cardiovascular AEs from 1.19 (0.81 to 1.75) to 1.68 (1.10 to 2.57)

Ann Rheum Dis. 2016 Mar; 75(3): 552–559.

Acetaminophen Harms

- One study reporting GI AEs reported a dose–response with increased relative rate of GI AEs or bleeds from 1.11 (1.04 to 1.18) to 1.49 (1.34 to 1.66).
- Three of four studies reporting renal AEs, reported a dose–response with one reporting an increasing OR of ≥30% decrease in estimated glomerular filtration rate from 1.40 (0.79 to 2.48) to 2.19 (1.4 to 3.43).

NSAID Harms

- Meta-analysis of CV and GI effects of NSAIDs
- 280 + 474 randomised trials
- Major vascular (mostly coronary) events increased by coxibs (RR 1.37, 1.14-1.66) and diclofenac (1.41, 1.12-1.78)
- NNH_{1year} for coxibs = 333 for major vascular
- Ibuprofen increased coronary events, naproxen did not

Lancet. 2013 Aug 31;382(9894):769-79.

NSAID Harms

- CHF doubled by all NSAIDs
- All NSAIDs increased GI complications:
  - Coxib or diclofenac RR ~ 1.8
  - Rate for coxibs 0.38%/year vs 0.19% pbo, NNH ~526, 2% of GI Bleeds fatal
  - Ibuprofen or naproxen RR ~4
- Coxibs increase all-cause mortality (1.22, 1.04-1.44), other NSAIDs show similar trend not statistically significant

What else is there?

- Most guidelines for treating pain especially in older adults still include acetaminophen as first line
- Oral NSAIDs ‘use for shortest time possible’
- Topical NSAIDs do reduce pain and improve physical function, useful for OA knee but not hip
- Topical lidocaine has effect in OA, low AE

Guidelines

- Opioids risk of falls, fractures, constipation, cognitive changes, N&V….
  - ‘consider for moderate to severe pain with impairments in ADL or QOL that have failed other treatments’
- Adjuvants (TCAs, Gaba, SNRIs) proven for neuropathic pain
- One study of duloxetine showed benefit for knee OA

Intraarticular steroids

- Cochrane review found low quality evidence with imprecision and heterogeneity
- Couldn’t rule out either a significant benefit or no benefit
- Best quality study showed no benefit
- Low side effects

Non-Pharmacologic/Herbs

- Evidence for CBT, acupuncture, massage, exercise, Tai Chi, yoga
  - Issues with engagement for people with cognitive impairment
- One trial each (Cochrane) suggest arnica extract gel and comfrey extract gel have benefit, but capsicum does not
- Oral boswellia serrata in 2 small trials reduced OA pain and improved function

Should we still use Acetaminophen first line?

- Is it more effective in our frail elderly where same dose of drugs tends to have larger effect?
  - If so is it worse for their livers?
- Is it useful as a quasi-placebo to keep them away from more dangerous drugs like NSAIDs and opiates?
- Vote