Spondyloarthritis: NMA on pain outcome

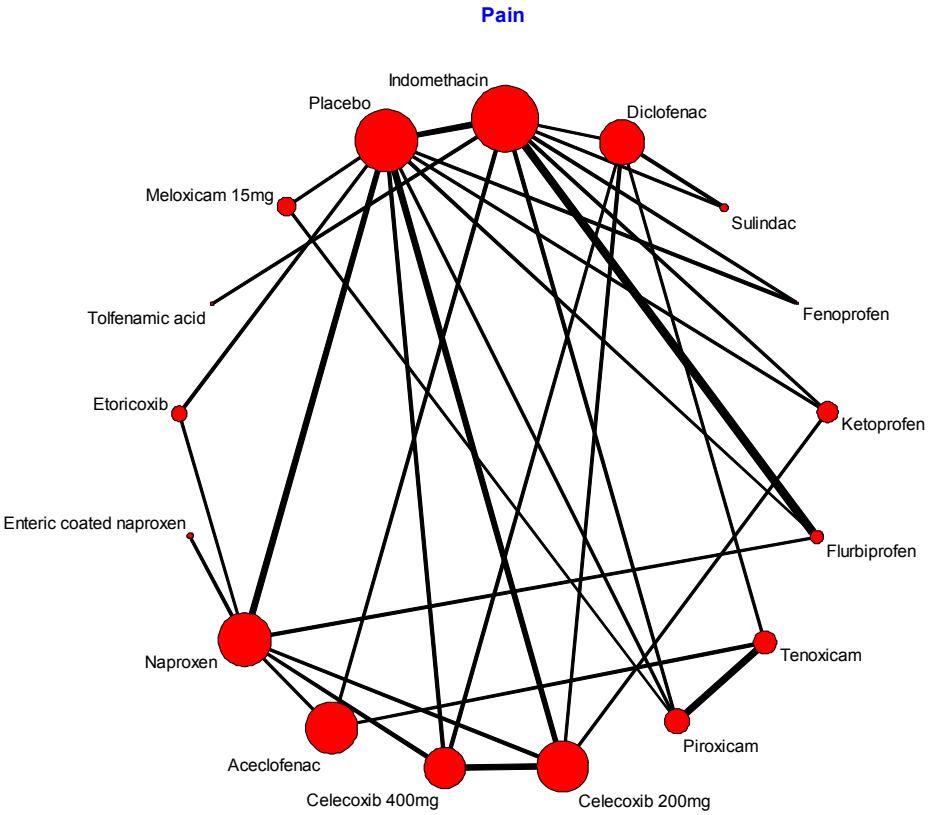
CGTSU, Bristol (Edna Keeney and Sofia Dias)

The purpose of this analysis was to estimate the comparative effectiveness of the following pharmacological interventions for management of pain associated with axial spondyloarthritis:

1. Indomethacin (Reference)
2. Diclofenac
3. Sulindac
4. Fenoprofen
5. Ketoprofen
6. Flurbiprofen
7. Tenoxicam
8. Piroxicam
9. Celecoxib 200mg
10. Celecoxib 400mg
11. Aceclofenac
12. Naproxen
13. Enteric coated Naproxen
14. Etoricoxib
15. Tolfenamic acid
16. Meloxicam 15mg
17. Placebo

23 studies were included in the analyses. The network diagram is shown in Figure 1.

Figure 1. Network Diagram for pain outcome



METHODS

In order to take all trial information into consideration, Mixed Treatment Comparison meta-analytic techniques, also termed Network meta-analysis (NMA), were employed. NMA is a generalization of standard pairwise meta-analysis for A versus B trials, to data structures that include, for example, A versus B, B versus C, and A versus C trials.[1-3] A basic assumption of NMA methods is that direct and indirect evidence estimate the same parameter, that is, the relative effect between A and B measured directly from a A versus B trial, is the same as the relative effect between A and B estimated indirectly from A versus C and B versus C trials. NMA techniques strengthen inference concerning the relative effect of two treatments by including both direct and indirect comparisons between treatments, and, at the same time, allow simultaneous inference on all treatments while respecting randomisation.[2,3] Simultaneous inference on the relative effects of all treatments is possible whenever treatments are part of a single "network of evidence", that is, every treatment is linked to at least one of the other treatments under assessment. The correlation between the random effects of multi-arm trials (i.e. those with more than 2 arms) in the network is taken into account in the analysis.[1]

A Bayesian framework is used to estimate all parameters, using Markov chain Monte Carlo simulation methods implemented in WinBUGS 1.4.3.[4,5] In order to test whether starting values have an impact on the results, two chains with different initial values were run simultaneously. Convergence was assessed by inspection of the Gelman–Rubin diagnostic plots and by examining the

history plots. Pre-convergence iterations were discarded, and further iterations on all chains were run on which results are based.

Sample WinBUGS code is provided in Appendix 1.

RE-SCALING OF OUTCOMES

The pain outcome was reported on various scales as shown in Table 1.

Table 1. Number of studies reporting on each scale

Scale	No. of studies
0-100 VAS	10
1-17 VAS	1
0-10 VAS	1
0-4	5
0-3	2
Absent, slight, moderate, severe (Number of patients in each category reported)	1
None, very mild, mild, moderate, severe, very severe (Number of patients in each category reported)	1
1 = no pain - 5=very severe (Assumed continuous 1-5, mean reported)	2

In order to carry out the analysis all data had to be reported as means and standard deviations (SDs) on the same scale. The VAS 0-100 scale was chosen as it was the most commonly reported and the easiest to interpret. Therefore, for each study not reporting on the 0-100 scale, a linear transformation of the original measurement scale was carried out assuming that the underlying shape of the distribution was the same for both scales.

In the studies which reported number of patients in each category, we assumed that these categories were continuous. For example 'absent, slight, moderate, severe' was assumed to be equivalent to 0-3 on a continuous scale and the data were reconstructed as such to calculate the mean and SD.

Then, for each study with mean and SD not on the 0-100 VAS scale, the following equations (1) were used to transform the reported values onto the 0-100 scale, which were used in the analyses.

$$Y = \frac{100}{u-l} X \quad \text{and} \quad \mathrm{SD}_Y = \frac{100}{u-l} \mathrm{SD}_X \qquad (1)$$

Where X and SD_X are the reported mean and SD on the old scale from the lower limit (l) to the upper limit (u), and Y and SD_Y are the transformed mean and SD on the 0-100 scale.

In doing this we made two assumptions. First, that the mean and SD are good summaries of the scale distribution for all scales i.e. that they all have approximately symmetric and unimodal distributions. Second, that outcomes on all scales have the same distributional shape i.e. that none have particular properties such as bimodal or asymmetric which differ from others. The second assumption is in part already covered by the first.

IMPUTING VARIANCES

Of the 23 studies included, 9 (39%) did not report the SD. Excluding these trials would have resulted in two treatments (tolfenamic acid and fenoprofen) dropping out of the network and a possible reduction in the precision of the estimates. In order to include all trials the SD had to be imputed. This can be done using one of three methods:

1. Obtain a value for the SDs from the literature or from the studies reporting SD (e.g. mean or median SD) and impute as if known.
2. Model the distribution of reported SDs and impute the mean or median of that, as if known.
3. Model the distribution of reported SDs and predict SD for each study from that distribution.[6]

We chose the 3rd method as it is the only one that allows for the uncertainty in the imputed SDs.

Figure 2 shows the distribution of the observed SDs. We attempted to fit a gamma distribution to this but the fit was poor. We achieved a better fit by fitting a normal distribution to the variances (SD^2) for the studies reporting it (Figure 3).

Figure 2. Distribution of observed SDs

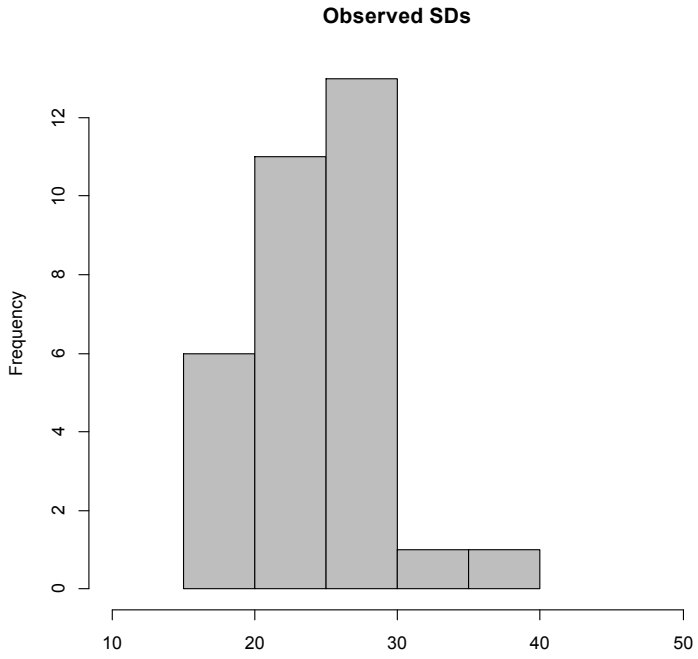
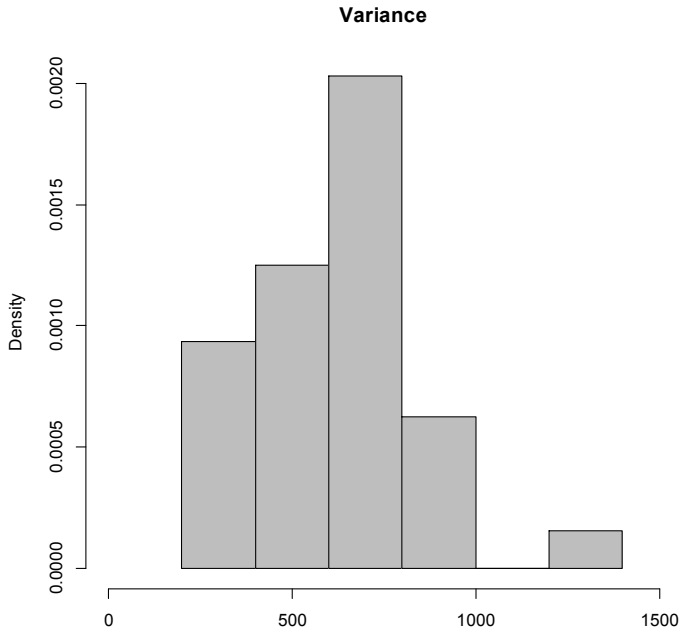
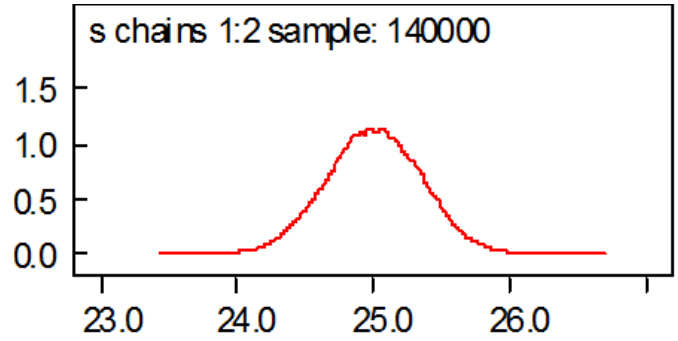


Figure 3. Distribution of observed variances (SD^2)



The overall SD was then estimated as shown in Figure 4 with an approximately normal distribution and a mean/median/mode of around 25. Only one study (Tannenbaum 1984) had a noticeably different SD with 36.4 on the piroxicam arm. We therefore proceeded with the imputation of SD from this distribution, for studies where it was missing.

Figure 4. Overall SD from WinBUGS



Imputed values are shown in Table 2. All values were similar and very close to 25 on the VAS 0-100 scale. These values were then converted to standard errors using the reported sample size for each study.

Table 2. Imputed Standard Deviations for studies where it was missing

Study	arm	Median	95% CrI
Shipley 1980	1	25.01	(20.69,28.68)
Shipley 1980	2	25.00	(20.66,28.68)
Shipley 1980	3	25.02	(20.67,28.70)
Juvakoski 1982	1	25.00	(20.69,28.70)
Juvakoski 1982	2	24.99	(20.70,28.69)
Khan 1987	1	25.01	(20.65,28.68)
Khan 1987	2	25.00	(20.68,28.70)
Good 1977	1	25.00	(20.66,28.66)
Good 1977	2	25.01	(20.71,28.69)
Lomen 1986	1	25.00	(20.69,28.71)
Lomen 1986	2	25.02	(20.67,28.68)
Burry 1980	1	25.01	(20.69,28.70)
Burry 1980	2	25.00	(20.69,28.68)
Villa Alacazar 1996	1	25.01	(20.69,28.67)
Villa Alacazar 1996	2	25.01	(20.69,28.69)
Pasero 1994	1	24.99	(20.69,28.68)
Pasero 1994	2	25.00	(20.65,28.70)
Rejholec 1981	1	25.00	(20.69,28.68)
Rejholec 1981	2	25.01	(20.68,28.70)

We also carried out sensitivity analyses to test the robustness of the results to the imputed SDs. These are reported further on in the document.

NMA MODEL FOR CONTINUOUS DATA (PAIN)

A random effects NMA model was used to estimate the relative effects of each treatment compared to the reference treatment. For each arm *k* of a trial *i*, the observed mean difference in pain on the VAS 0-100 scale, y_{ik} , has a normal likelihood

$$y_{ik} \sim \text{Normal}\left(\theta_{ik}, s_{ik}^2\right) \quad (2)$$

where θ_{ik} is the underlying (true) mean pain and s_{ik} is the standard error of the mean pain in arm *k* of trial *i* (which could be imputed with uncertainty – see Table 2).

The mean pain is modelled using a NMA model such that

$$\theta_{ik} = \mu_i + \delta_{ik} \quad (3)$$

with μ_i being given non-informative normal priors, Normal(0,100^2), and $\delta_{i1} = 0$, since there is no relative treatment effect estimated for arm 1 of each trial.

In a random effects (RE) model the trial-specific treatment effects of the treatment in arm *k*, relative to the treatment in arm 1, are drawn from a common random effects distribution, under the assumption of consistency:

$$\delta_{ik} \sim N(d_{1t_{ik}} - d_{1t_{i1}}, \tau^2) \quad (4)$$

where $d_{t_{ik}}$ represents the mean effect of the treatment in arm *k* in trial *i*, t_{ik}, relative to the reference treatment, and τ^2 represents the between-trial variability in treatment effects (heterogeneity). The between-trials standard deviation was given a Uniform (0,100) prior.

In the FE model we replace equation (3) with

$$\theta_{ik} = \mu_i + d_{1t_{ik}} - d_{1t_{i1}}$$

Convergence was satisfactory by at least 36,000 iterations in all cases. Models were then run for a further 70,000 iterations on two separate chains, and all results are based on this further sample.

Model comparison using the DIC showed the random effects consistency model as the preferred model with no evidence of inconsistency (Table ~~2~~3).

Table 3. Model fit statistics – Base case analysis

Model	No. of data points	Residual Deviance over studies with complete data	Residual Deviance over all studies	Between-trials SD (posterior median) and 95% credible intervals	DIC
RE consistency	53	30.96	52.26	5.58 (2.63, 9.84)	847.67
RE inconsistency	53	32.5	54.16	10.08 (0.59, 26.96)	855.21
FE consistency	53	41.03	84.64	-	872.49

RESULTS

Figure 4 shows the mean differences on the VAS 0-100 scale compared to Indomethacin. Negative values mean that the treatment reduces pain compared to Indomethacin. The figure shows that there is not much difference between drugs in the reduction of pain. The one exception is placebo which increases pain by 15.0 points (95% CrI 8.2 to 23.0) on the VAS 0-100 scale, compared to Indomethacin. The comparison of all treatments against each other is given in Appendix 2.

Figure 4. Mean differences in pain compared to Indomethacin - base case analysis

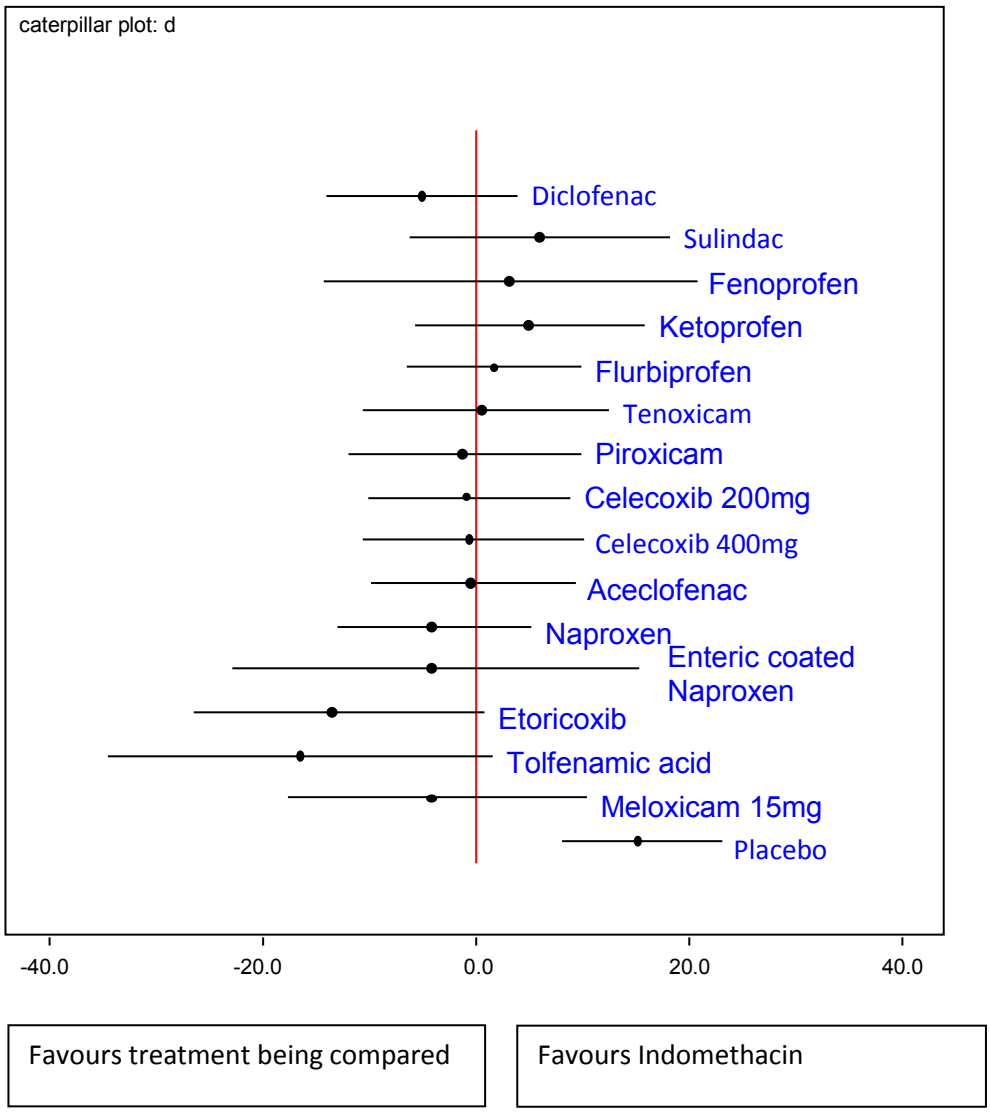


Table 3 lists the drugs in terms of mean and median rank. Tolfenamic acid is ranked highest followed closely by etoricoxib however the credible intervals around these estimates indicate the considerable uncertainty in the rankings, mainly due to the fact that there are no differences in effect between these drugs. However placebo ranks low, as expected.

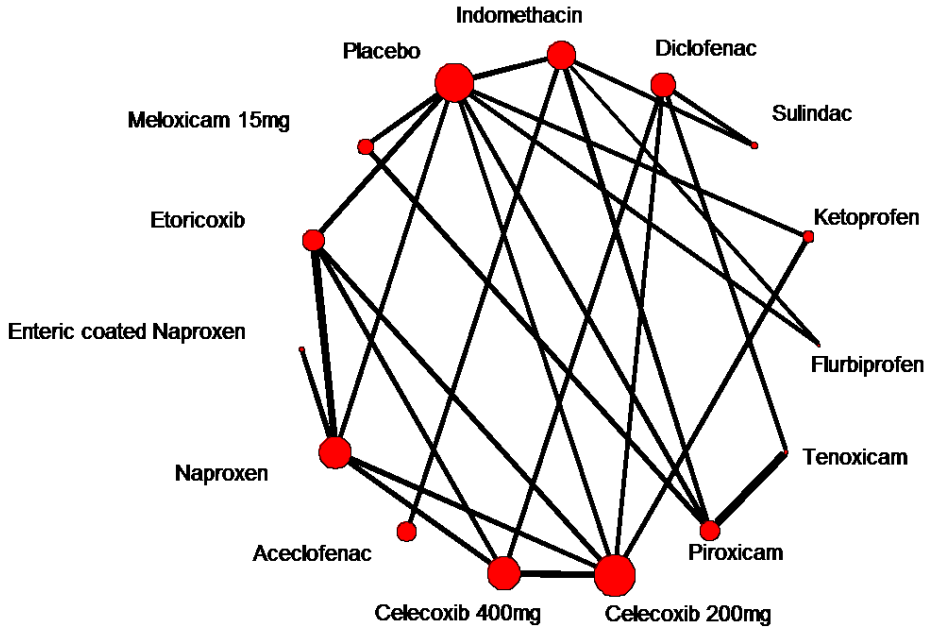
Table 3. Ranking on efficacy – base case analysis

	Treatments	Mean rank	Median rank	95% CrI
15	Tolfenamic acid	2.4	1	(1,12)
14	Etoricoxib	2.5	2	(1,9)
2	Diclofenac	5.7	5	(2,12)
12	Naproxen	6.2	6	(2,12)
16	Meloxicam 15mg	6.8	6	(1,15)
13	Enteric coated Naproxen	7.2	6	(1,16)
8	Piroxicam	8.6	8	(3,15)
9	Celecoxib 200mg	9.1	9	(3,15)
10	Celecoxib 400mg	9.3	9	(3,15)
11	Aceclofenac	9.4	9	(3,15)
1	Indomethacin	9.8	10	(5,14)
7	Tenoxicam	10.2	11	(3,16)
6	Flurbiprofen	11.2	12	(4,16)
4	Fenoprofen	11.3	13	(2,17)
5	Ketoprofen	13.1	14	(5,16)
3	Sulindac	13.4	15	(5,17)
17	Placebo	16.8	17	(15,17)

SENSITIVITY ANALYSIS (i) NO IMPUTED SDs

A sensitivity analysis was carried out testing the robustness of the results to the imputed SDs. In this analysis any study which did not report an SD was removed (9 studies were removed). This resulted in two interventions (tolfenamic acid and fenoprofen) dropping out of the network. The revised network plot is shown in Figure 5.

Figure 5. Network plot for sensitivity analysis – studies not reporting SD removed



A random effects NMA model was again used to estimate the relative effects of each treatment compared to the reference treatment. Convergence was satisfactory by at least 35,000 iterations in all cases. Models were then run for a further 70,000 iterations on two separate chains, and all results are based on the further samples.

Model comparison using the DIC showed the random effects consistency model as the preferred model with no evidence of inconsistency (Table 4). However, there was a considerable increase in heterogeneity as compared to the base case analysis.

Table 4. Model fit statistics – Sensitivity analysis (i)

Model	Number of data points	Residual Deviance	Between-trials SD (posterior median) and 95% credible intervals	DIC
RE consistency	32	33.78	14.13 (7.4, 26.7)	209.32
RE inconsistency	32	34.26	18.42 (1.2, 50.97)	210.62
FE consistency	32	101.1	-	270.61

Figure 6 shows the mean differences on the VAS 0-100 scale compared to Indomethacin. The figure shows that there is very little difference between drugs in the reduction of pain. The estimates are a lot more uncertain than in the base case analysis.

Figure 6. Mean differences in pain compared to Indomethacin - sensitivity analysis (i)

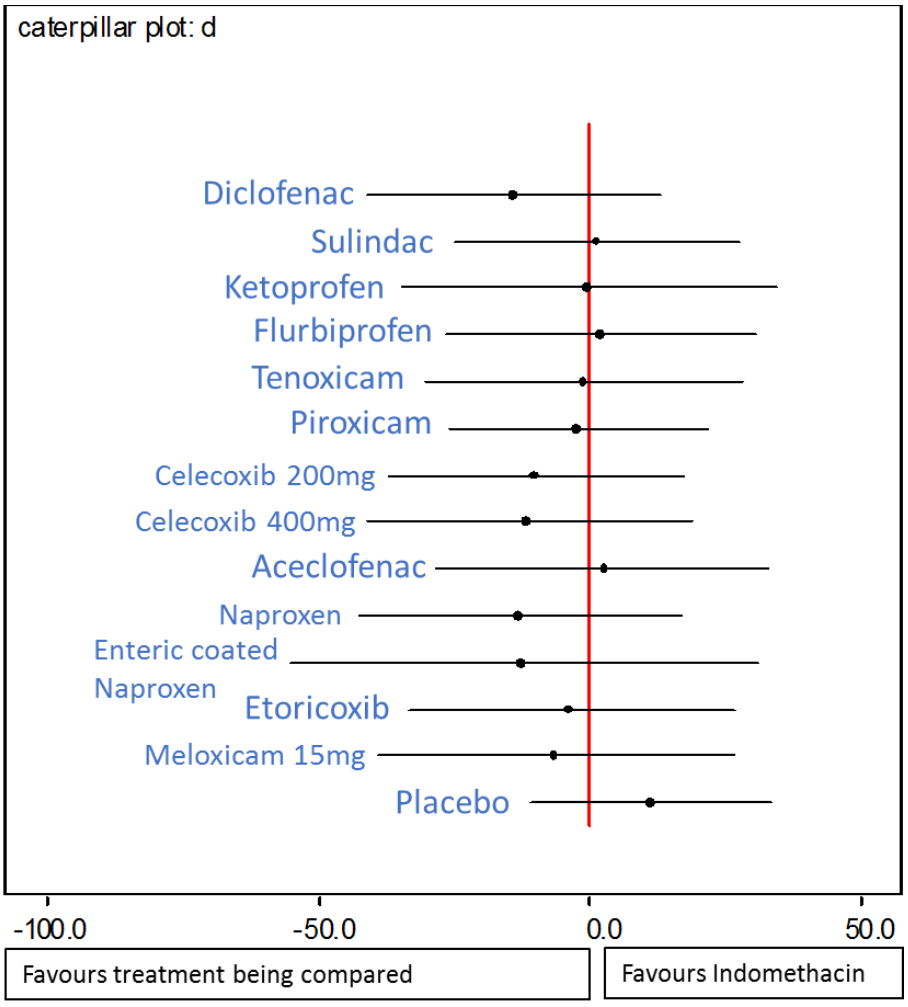


Table 5 lists the drugs in terms of mean and median rank. Diclofenac is ranked highest followed closely by Naproxen however the credible intervals around these estimates indicate the considerable uncertainty in the rankings, ranging from 1st to 12th out of 15 treatments.

Table 5. Ranking on efficacy – sensitivity analysis (i)

	Treatments	Mean rank	Median rank	95% CrI
2	Diclofenac	4.3	4	(1,12)
11	Naproxen	4.8	4	(1,12)
9	Celecoxib 400mg	5.3	5	(1,13)
8	Celecoxib 200mg	5.7	5	(1,12)
12	Enteric coated Naproxen	5.7	4	(1,15)
14	Meloxicam 15mg	7.1	7	(1,15)
13	Etoricoxib	8.2	8	(2,14)
7	Piroxicam	8.5	9	(2,14)
6	Tenoxicam	9.0	9	(1,15)
4	Ketoprofen	9.2	10	(1,15)
1	Indomethacin	9.5	10	(3,14)
3	Sulindac	9.8	10	(2,15)
5	Flurbiprofen	9.9	11	(1,15)
10	Aceclofenac	9.9	11	(1,15)
15	Placebo	13.2	14	(8,15)

SENSITIVITY ANALYSIS (ii) HIGHEST SD ASSUMED

As an alternative sensitivity analysis, rather than remove studies which did not report an SD or predicting the missing SDs, we assumed that the missing SDs would be equal to the highest reported SD (36.4 from Tannenbaum 1984).

A random effects NMA model was again used to estimate the relative effects of each treatment compared to the reference treatment. Convergence was satisfactory by at least 100,000 iterations in all cases. Models were then run for a further 50,000 iterations on two separate chains, and all results are based on this further sample.

Model comparison using the DIC showed the random effects consistency model as the preferred model with no evidence of inconsistency (Table 6).

Table 6. Model fit statistics – Sensitivity analysis (ii)

Model	Number of data points	Residual Deviance	Between-study SD (posterior median) and 95% credible intervals	DIC
RE consistency	53	53.83	3.32 (0.18, 8.29)	342.75
RE inconsistency	53	53.36	9.57(0.45, 26.49)	351.05
FE consistency	53	57.17	-	340.59

Figure 7 shows the mean differences on the VAS 0-100 scale compared to Indomethacin. The figure shows that again there is very little difference between drugs in the reduction of pain. However, placebo increases pain by 16.95 points (95% CrI 9.8 to 25.6) on the VAS 0-100 scale, compared to Indomethacin, and etoricoxib decreases pain by 13.1 points (95% CrI -23.7 to -0.3). These estimates are similar to the base case analysis but slightly more certain as the SD is now assumed to be known.

Figure 7. Mean differences in pain compared to Indomethacin - sensitivity analysis (ii)

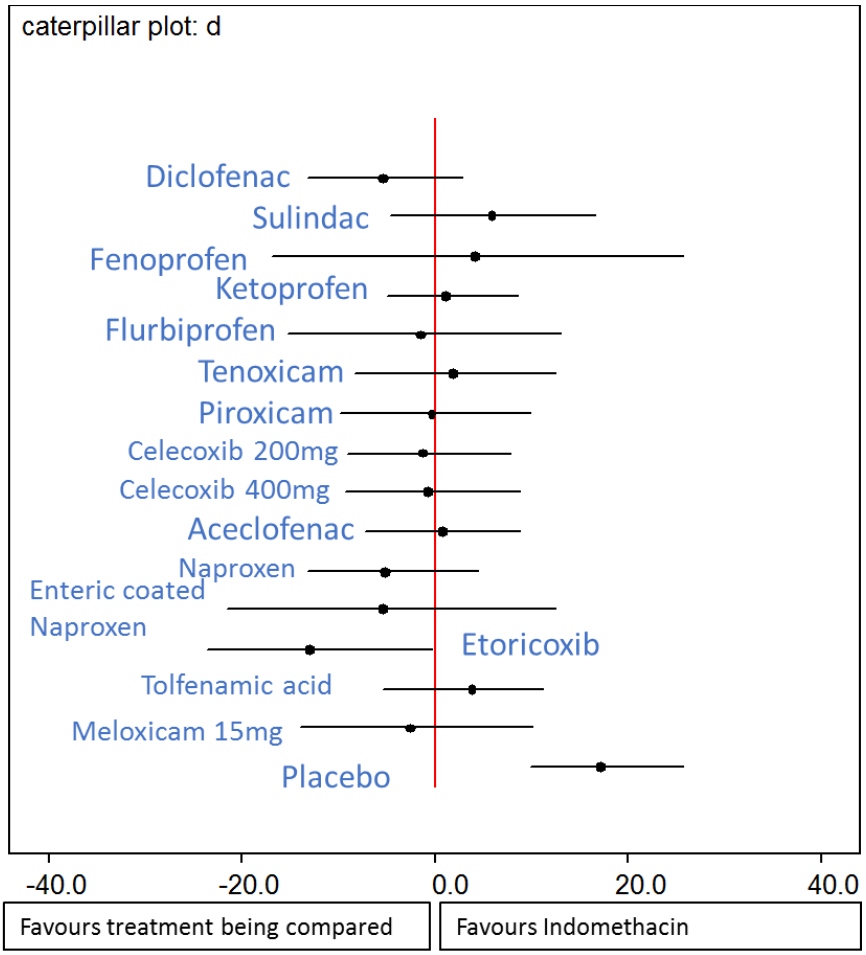


Table 7 lists the drugs in terms of mean and median rank. Etoricoxib is ranked highest followed closely by diclofenac however the credible intervals around these estimates indicate the considerable uncertainty in the rankings. Tolfenamic is no longer ranked first, as in the base case analysis, but has a median rank of 14th (95% CrIs 4th – 16th).

Table 7. Ranking on efficacy – sensitivity analysis (ii)

	Treatments	Mean rank	Median rank	95% CrI
14	Etoricoxib	1.7	1	(1,7)
2	Diclofenac	4.6	4	(1,11)
12	Naproxen	4.7	4	(2,11)
13	Enteric coated Naproxen	5.8	4	(1,16)
16	Meloxicam 15mg	7.1	6	(1,15)
6	Flurbiprofen	8.1	8	(1,16)
9	Celecoxib 200mg	8.2	8	(3,14)
10	Celecoxib 400mg	8.8	9	(3,15)
8	Piroxicam	9.0	9	(3,15)
1	Indomethacin	9.4	10	(4,14)
11	Aceclofenac	10.1	11	(3,15)
5	Ketoprofen	10.4	11	(4,15)
7	Tenoxicam	10.9	12	(3,16)
4	Fenoprofen	11.2	14	(1,17)
15	Tolfenamic acid	12.6	14	(4,16)
3	Sulindac	13.5	15	(5,16)
17	Placebo	16.8	17	(16,17)

Appendix 1. WinBUGS code for pain outcome

```
model{
#Impute missing SDs
for (i in 1:complete){          #loop through studies that contain SDs
  for(k in 1:na.a[i]){            # loop through arms
    var[i,k] <- pow(sd.a[i,k], 2)
    var[i,k] ~ dnorm(v, b)I(0,) #likelihood for observed variances
# residual deviance for normal likelihood, mean unknown
# predicted value is v in this case
    dev.n[i,k] <- ((var[i,k]-v)*(var[i,k]-v))*b
   }
  resdev.n[i] <- sum(dev.n[i,1:na.a[i]])
 }
v <- pow(s,2)     # s is "true" population SD, v is pop variance
for (i in complete+1:ns.a){       #loop through studies that do not report
SDs
  for (k in 1:na.a[i]){
    var[i,k] ~ dnorm(v, b)I(0,) #predict unknown variances
    sd.p[i,k] <- sqrt(var[i,k]) # predicted SDs for studies where missing
  }
}
log(s) <- lns
lns ~ dunif(-10,10)     # prior for log of true population SD
#s ~ dunif(0,100)
sd.b ~ dunif(0,100) # prior for SD of distr of variances
b <- pow(sd.b,-2)   # precision of distr of variances
# Total Residual Deviance for normal (complete data)
totresdev.n <- sum(resdev.n[])
#
for (i in 1:ns.a){       #loop through all studies converting var to prec
  for (k in 1:na.a[i]){
# use cut to avoid updating of var[i,k]
    var.final[i,k] <- cut(var[i,k])
# convert to observed and imputed precisions (1/se^2)
    prec[i,k] <- n.a[i,k]/var.final[i,k]
  }
}
#TSD code
for(i in 1:ns.a){        # LOOP THROUGH STUDIES WITH ARM DATA
  w.a[i,1] <- 0 # adjustment for multi-arm trials is zero for control arm
  delta[i,1] <- 0 # treatment effect is zero for control arm
  mu[i] ~ dnorm(0,.0001) # vague priors for all trial baselines
  for (k in 1:na.a[i]) { # LOOP THROUGH ARMS
    y.a[i,k] ~ dnorm(theta[i,k],prec[i,k]) # normal likelihood
    theta[i,k] <- mu[i] + delta[i,k] # model for linear predictor
#Deviance contribution
    dev[i,k] <- (y.a[i,k]-theta[i,k])*(y.a[i,k]-theta[i,k])*prec[i,k]
  }
# summed residual deviance contribution for this trial
  resdev[i] <- sum(dev[i,1:na.a[i]])
  for (k in 2:na.a[i]) { # LOOP THROUGH ARMS
    delta[i,k] ~ dnorm(md[i,k],taud.a[i,k]) # trial-specific LOR
distributions
# mean of treat effects distributions (with multi-arm trial correction)
    md[i,k] <- d[t.a[i,k]] - d[t.a[i,1]] + sw.a[i,k]
# precision of treat effects distributions (with multi-arm trial
correction)
    taud.a[i,k] <- tau *2*(k-1)/k
# adjustment for multi-arm RCTs
    w.a[i,k] <- (delta[i,k] - d[t.a[i,k]] + d[t.a[i,1]])
```

```
# cumulative adjustment for multi-arm trials
    sw.a[i,k] <- sum(w.a[i,1:k-1])/(k-1)
  }
}
for(i in 1:ns.t){    # LOOP THROUGH STUDIES WITH TRIAL DATA
  w[i,1] <- 0 # adjustment for multi-arm trials is is zero for control arm
  delta[i+ns.a,1] <- 0 # treatment effect is zero for control arm
  for (k in 1:(na[i]-1)) { # LOOP THROUGH ARMS
    for (j in 1:(na[i]-1))  {Sigma[i,j,k]  <- V[i]*(1-equals(j,k)) +
        pow(se[i,k+1],2)*equals(j,k)  }
  }
  Omega[i,1:(na[i]-1),1:(na[i]-1)] <- inverse(Sigma[i,,]) #Precision matrix
# multivariate normal likelihood for 3-arm trials
  y[i,2:na[i]] ~ dmnorm(delta[i+ns.a,2:na[i]],Omega[i,1:(na[i]-1),1:(na[i]-
1)])
#Deviance contribution for trial i
  for (k in 1:na[i]-1){ # multiply vector & matrix
    ydiff[i,k]<- y[i,(k+1)] - delta[i+ns.a,(k+1)]
    z[i,k]<- inprod2(Omega[i,k,1:(na[i]-1)], ydiff[i,1:(na[i]-1)])
  }
  resdev[i+ns.a]<- inprod2(ydiff[i,1:(na[i]-1)], z[i,1:(na[i]-1)])
  for (k in 2:na[i]) { # LOOP THROUGH ARMS
# trial-specific LOR distributions
    delta[i+ns.a,k] ~ dnorm(md[i+ns.a,k],taud[i,k])
# mean of LOR distributions, with multi-arm trial correction
    md[i+ns.a,k] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]
# precision of LOR distributions (with multi-arm trial correction)
   taud[i,k] <- tau *2*(k-1)/k
# adjustment, multi-arm RCTs
   w[i,k] <- (delta[i+ns.a,k] - d[t[i,k]] + d[t[i,1]])
   sw[i,k] <- sum(w[i,1:k-1])/(k-1) # cumulative adjustment for multi-arm
trials
  }
}
totresdev <- sum(resdev[1:complete]) #Total Residual Deviance complete data
totresdev2 <- sum(resdev[]) # Total Residual Deviance ALL data
d[1]<-0 # treatment effect is zero for reference treatment
for (k in 2:nt){  # LOOP THROUGH TREATMENTS
  d[k] ~ dnorm(0,.0001) # vague priors for treatment effects
}
sdev ~ dunif(0,100) # vague prior for between-trial SD.
tau <- pow(sdev,-2) # between-trial precision = (1/between-trial variance)
# pairwise mean differences for all possible pair-wise comparisons
for (c in 1:(nt-1)){
  for (k in (c+1):nt){ diff[c,k] <- d[k] - d[c]  }
}
# Ranking and probabilities
for(k in 1:nt){
  #rk[k]<-nt+1-rank(d[],k)
  rk[k] <- rank(d[],k)  # smaller values are better
  best[k]<-equals(rk[k],1)
  for (h in 1:nt){  prob[h,k]<-equals(rk[k],h)}
}
}                    # **PROGRAM ENDS



list(ns.a=22, complete=13, ns.t = 1, nt=17)

#Arm-level data
```

```
na.a[]      t.a[,1]     y.a[,1]     n.a[,1]     sd.a[,1]    t.a[,2]
      y.a[,2]     n.a[,2]     sd.a[,2]    t.a[,3]     y.a[,3]     n.a[,3]
      sd.a[,3]    t.a[,4]     y.a[,4]     n.a[,4]     sd.a[,4]
4     9     -29.5 100   28    10    -30   118   24    12    -36   118
      22.81 17    -10   72    18.67
3     2     33.8  154   27    9     37.4  152   25.6  10    38.7  148   25
      NA    NA    NA    NA
3     12    -35.4 97    23    14    -42.9 100   22    17    -12.6 93    22
      NA    NA    NA    NA
3     5     -21   90    26    9     -27   80    30    17    -13   76    29
      NA    NA    NA    NA
3     8     -29   108   28    16    -31   120   27    17    -11   121   28
      NA    NA    NA    NA
2     2     25    30    19    3     39    30    21    NA    NA    NA    NA
      NA    NA    NA    NA
2     1     37.5  23    15    3     40    23    25    NA    NA    NA    NA
      NA    NA    NA    NA
2     2     30    12    20    7     43.33 12    26.67 NA    NA    NA    NA
      NA    NA    NA    NA
2     7     20.83 8     17.25 8     33.33 9     29    NA    NA    NA    NA
      NA    NA    NA    NA
2     7     36.92 13    16.01 8     25.71 14    24    NA    NA    NA    NA
      NA    NA    NA    NA
2     1     -42.5 27    26    8     -39.38      28    36.38 NA    NA    NA
      NA    NA    NA    NA    NA
2     1     -25   153   27.21 11    -22.4 155   24.9  NA    NA    NA    NA
      NA    NA    NA    NA
2     12    37.5  39    30.85 13    37.5  39    23    NA    NA    NA    NA
      NA    NA    NA    NA
3     1     24.67 19    NA    4     32.78 19    NA    17    49.78 19    NA
      NA    NA    NA    NA
2     1     13.64 44    NA    5     16.48 44    NA    NA    NA    NA    NA
      NA    NA    NA    NA
2     1     23.25 81    NA    2     23.25 93    NA    NA    NA    NA    NA
      NA    NA    NA    NA
2     1     -22.5 13    NA    6     -17.5 13    NA    NA    NA    NA    NA
      NA    NA    NA    NA
2     1     37.5  21    NA    6     37.5  23    NA    NA    NA    NA    NA
      NA    NA    NA    NA
2     6     51.91 29    NA    12    52.39 29    NA    NA    NA    NA    NA
      NA    NA    NA    NA
2     7     -27.5 115   NA    11    -25.7 120   NA    NA    NA    NA    NA
      NA    NA    NA    NA
2     11    25    60    NA    12    29    60    NA    NA    NA    NA    NA
      NA    NA    NA    NA
2     1     40    25    NA    15    23.33 25    NA    NA    NA    NA    NA
      NA    NA    NA    NA
END

#Trial-level data
t[,1] t[,2] t[,3]     y[,2]    y[,3]    se[,2]    se[,3]    V[]   na[]
1           6           17       -1.4       4.7        1.7         2.3        2.09
3
END
```

```
#chain 1

list(d=c( NA, 0,0,0,0,      0,0,0,0,0,      0,0,0,0,0,      0,0),

mu=c(0,0,0,0,0,    0,0,0,0,0,    0,0,0,0,0,    0,0,0,0,0,    0,0),
sdev=70)


#chain 2

list(d=c( NA, -3,-3,-3,-3,      -3,-3,-3,-3,-3,      -3,-3,-3,-3,-3,
-3,-3),

mu=c(-3,-3,-3,-3,-3,    -3,-3,-3,-3,-3,    -3,-3,-3,-3,-3,    -3,-
3,-3,-3,-3,    -3,-3), sdev=30)
```

Appendix 2. Comparison of all treatments against each other for main analysis (RE model using imputed SDs). Blank spaces indicate that the treatments were not directly compared.

		Network meta-analysis		Pairwise analysis (from inconsistency model)	
Control Treatment	**Active Treatment**	**Mean difference**	**95% CrIs**	**Mean difference**	**95% CrIs**
Indomethacin	Diclofenac	-5.02	(-13.92,3.87)	0.04	(-26.22,26.08)
Indomethacin	Sulindac	6.02	(-6.26,18.27)	2.63	(-24.60,29.68)
Indomethacin	Fenoprofen	3.12	(-14.22,20.69)	4.43	(-22.44,34.98)
Indomethacin	Ketoprofen	4.92	(-5.59,15.84)	2.74	(-23.82,29.59)
Indomethacin	Flurbiprofen	1.71	(-6.43,9.94)	0.42	(-15.47,17.18)
Indomethacin	Tenoxicam	0.60	(-10.67,12.52)		
Indomethacin	Piroxicam	-1.37	(-11.88,9.92)	3.01	(-26.52,32.21)
Indomethacin	Celecoxib 200mg	-0.86	(-10.03,8.92)		
Indomethacin	Celecoxib 400mg	-0.65	(-10.73,10.06)		
Indomethacin	Aceclofenac	-0.47	(-9.83,9.25)	2.52	(-23.24,28.12)
Indomethacin	Naproxen	-4.22	(-12.88,5.31)		
Indomethacin	Enteric coated Naproxen	-4.16	(-22.82,15.28)		
Indomethacin	Etoricoxib	-13.42	(-26.52,0.77)		
Indomethacin	Tolfenamic acid	-16.48	(-34.53,1.63)	-16.31	(-44.04,11.99)
Indomethacin	Meloxicam 15mg	-4.13	(-17.72,10.36)		
Indomethacin	Placebo	15.18	(8.18,23.03)	11.52	(-5.81,32.68)
Diclofenac	Sulindac	11.04	(-1.01,22.99)	13.82	(-13.55,40.52)
Diclofenac	Fenoprofen	8.14	(-10.85,27.35)		
Diclofenac	Ketoprofen	9.94	(-2.54,22.76)		
Diclofenac	Flurbiprofen	6.73	(-4.58,18.11)		
Diclofenac	Tenoxicam	5.63	(-6.67,18.49)	13.30	(-17.34,43.40)
Diclofenac	Piroxicam	3.66	(-8.69,16.89)		
Diclofenac	Celecoxib 200mg	4.16	(-5.40,14.33)	3.50	(-22.56,29.29)
Diclofenac	Celecoxib 400mg	4.37	(-5.71,15.01)	4.80	(-21.24,30.75)
Diclofenac	Aceclofenac	4.55	(-7.03,16.43)		
Diclofenac	Naproxen	0.81	(-9.68,12.12)		
Diclofenac	Enteric coated Naproxen	0.86	(-18.75,21.21)		
Diclofenac	Etoricoxib	-8.39	(-22.88,7.12)		
Diclofenac	Tolfenamic acid	-11.45	(-31.62,8.69)		
Diclofenac	Meloxicam 15mg	0.89	(-14.19,16.80)		

Diclofenac	Placebo	20.20	(10.76,30.50)		
Sulindac	Fenoprofen	-2.90	(-23.71,18.27)		
Sulindac	Ketoprofen	-1.10	(-16.57,14.71)		
Sulindac	Flurbiprofen	-4.31	(-18.66,10.22)		
Sulindac	Tenoxicam	-5.41	(-21.13,10.88)		
Sulindac	Piroxicam	-7.38	(-22.80,8.80)		
Sulindac	Celecoxib 200mg	-6.88	(-20.71,7.61)		
Sulindac	Celecoxib 400mg	-6.67	(-21.00,8.23)		
Sulindac	Aceclofenac	-6.49	(-21.31,8.57)		
Sulindac	Naproxen	-10.23	(-24.34,4.73)		
Sulindac	Enteric coated Naproxen	-10.18	(-31.84,12.24)		
Sulindac	Etoricoxib	-19.43	(-36.80,-1.19)		
Sulindac	Tolfenamic acid	-22.49	(-44.23,-0.60)		
Sulindac	Meloxicam 15mg	-10.15	(-27.78,8.35)		
Sulindac	Placebo	9.17	(-4.06,23.14)		
Fenoprofen	Ketoprofen	1.80	(-17.96,21.60)		
Fenoprofen	Flurbiprofen	-1.41	(-20.22,17.13)		
Fenoprofen	Tenoxicam	-2.52	(-22.79,17.90)		
Fenoprofen	Piroxicam	-4.49	(-24.15,15.50)		
Fenoprofen	Celecoxib 200mg	-3.98	(-22.80,14.84)		
Fenoprofen	Celecoxib 400mg	-3.77	(-23.20,15.69)		
Fenoprofen	Aceclofenac	-3.59	(-23.07,15.77)		
Fenoprofen	Naproxen	-7.34	(-25.83,11.40)		
Fenoprofen	Enteric coated Naproxen	-7.28	(-32.24,17.80)		
Fenoprofen	Etoricoxib	-16.54	(-37.36,4.50)		
Fenoprofen	Tolfenamic acid	-19.60	(-44.85,5.45)		
Fenoprofen	Meloxicam 15mg	-7.25	(-28.40,14.44)		
Fenoprofen	Placebo	12.06	(-5.32,29.52)		
Ketoprofen	Flurbiprofen	-3.21	(-15.91,9.10)		
Ketoprofen	Tenoxicam	-4.32	(-18.88,10.61)		
Ketoprofen	Piroxicam	-6.29	(-20.08,8.14)		
Ketoprofen	Celecoxib 200mg	-5.78	(-16.98,5.54)	-5.96	(-32.18,20.73)
Ketoprofen	Celecoxib 400mg	-5.57	(-18.37,7.40)		
Ketoprofen	Aceclofenac	-5.39	(-18.86,8.05)		
Ketoprofen	Naproxen	-9.14	(-21.13,3.36)		

Ketoprofen	Enteric coated Naproxen	-9.08	(-29.47,11.84)		
Ketoprofen	Etoricoxib	-18.34	(-33.88,-2.34)		
Ketoprofen	Tolfenamic acid	-21.39	(-42.56,-0.46)		
Ketoprofen	Meloxicam 15mg	-9.05	(-25.08,7.48)		
Ketoprofen	Placebo	10.26	(-0.20,21.09)	7.75	(-18.79,34.15)
Flurbiprofen	Tenoxicam	-1.10	(-14.10,12.66)		
Flurbiprofen	Piroxicam	-3.07	(-15.30,9.97)		
Flurbiprofen	Celecoxib 200mg	-2.57	(-13.37,8.81)		
Flurbiprofen	Celecoxib 400mg	-2.36	(-13.99,9.79)		
Flurbiprofen	Aceclofenac	-2.18	(-13.65,9.57)		
Flurbiprofen	Naproxen	-5.92	(-15.49,4.42)	0.66	(-27.07,28.40)
Flurbiprofen	Enteric coated Naproxen	-5.87	(-25.09,13.85)		
Flurbiprofen	Etoricoxib	-15.12	(-29.02,-0.23)		
Flurbiprofen	Tolfenamic acid	-18.18	(-38.10,1.66)		
Flurbiprofen	Meloxicam 15mg	-5.84	(-20.61,9.66)		
Flurbiprofen	Placebo	13.48	(4.90,22.86)		
Tenoxicam	Piroxicam	-1.97	(-13.51,9.65)	-1.75	(-22.45,21.05)
Tenoxicam	Celecoxib 200mg	-1.47	(-14.77,11.61)		
Tenoxicam	Celecoxib 400mg	-1.25	(-15.10,12.42)		
Tenoxicam	Aceclofenac	-1.07	(-12.07,9.46)	1.75	(-24.36,27.67)
Tenoxicam	Naproxen	-4.82	(-17.32,7.71)		
Tenoxicam	Enteric coated Naproxen	-4.77	(-25.51,16.08)		
Tenoxicam	Etoricoxib	-14.02	(-30.26,2.26)		
Tenoxicam	Tolfenamic acid	-17.08	(-38.69,4.17)		
Tenoxicam	Meloxicam 15mg	-4.73	(-20.42,10.96)		
Tenoxicam	Placebo	14.58	(2.76,26.38)		
Piroxicam	Celecoxib 200mg	0.50	(-12.27,12.90)		
Piroxicam	Celecoxib 400mg	0.72	(-12.92,13.80)		
Piroxicam	Aceclofenac	0.90	(-11.84,12.97)		
Piroxicam	Naproxen	-2.85	(-14.92,9.15)		
Piroxicam	Enteric coated Naproxen	-2.80	(-23.36,17.83)		
Piroxicam	Etoricoxib	-12.05	(-27.72,3.57)		

Piroxicam	Tolfenamic acid	-15.11	(-36.49,5.65)		
Piroxicam	Meloxicam 15mg	-2.76	(-15.75,10.03)	-2.03	(-28.01,24.10)
Piroxicam	Placebo	16.55	(6.06,26.85)	17.76	(-8.62,43.55)
Celecoxib 200mg	Celecoxib 400mg	0.21	(-8.90,9.22)	-0.51	(-26.66,25.42)
Celecoxib 200mg	Aceclofenac	0.39	(-11.72,12.13)		
Celecoxib 200mg	Naproxen	-3.35	(-12.83,6.51)	-6.38	(-32.37,19.73)
Celecoxib 200mg	Enteric coated Naproxen	-3.30	(-22.43,16.18)		
Celecoxib 200mg	Etoricoxib	-12.55	(-26.56,1.78)		
Celecoxib 200mg	Tolfenamic acid	-15.61	(-36.13,4.60)		
Celecoxib 200mg	Meloxicam 15mg	-3.27	(-18.17,11.86)		
Celecoxib 200mg	Placebo	16.04	(7.72,24.66)	19.24	(-7.21,45.09)
Celecoxib 400mg	Aceclofenac	0.18	(-12.46,12.60)		
Celecoxib 400mg	Naproxen	-3.57	(-13.62,6.96)		
Celecoxib 400mg	Enteric coated Naproxen	-3.51	(-23.00,16.18)		
Celecoxib 400mg	Etoricoxib	-12.77	(-27.35,2.23)		
Celecoxib 400mg	Tolfenamic acid	-15.83	(-36.82,4.83)		
Celecoxib 400mg	Meloxicam 15mg	-3.48	(-18.94,12.33)		
Celecoxib 400mg	Placebo	15.83	(6.40,25.61)		
Aceclofenac	Naproxen	-3.75	(-13.73,6.85)	3.93	(-22.94,30.14)
Aceclofenac	Enteric coated Naproxen	-3.69	(-23.10,16.23)		
Aceclofenac	Etoricoxib	-12.95	(-27.59,2.37)		
Aceclofenac	Tolfenamic acid	-16.01	(-36.51,4.40)		
Aceclofenac	Meloxicam 15mg	-3.66	(-18.78,12.19)		
Aceclofenac	Placebo	15.65	(5.64,26.29)		
Naproxen	Enteric coated Naproxen	0.05	(-16.75,16.79)	0.21	(-27.03,27.68)
Naproxen	Etoricoxib	-9.20	(-21.39,2.99)	-7.48	(-33.37,18.68)
Naproxen	Tolfenamic acid	-12.26	(-32.78,7.62)		
Naproxen	Meloxicam 15mg	0.09	(-14.50,14.69)		
Naproxen	Placebo	19.40	(11.54,27.12)	22.42	(-3.99,48.19)

Enteric coated Naproxen	Etoricoxib	-9.25	(-29.99,11.34)		
Enteric coated Naproxen	Tolfenamic acid	-12.31	(-38.97,13.70)		
Enteric coated Naproxen	Meloxicam 15mg	0.03	(-22.17,22.23)		
Enteric coated Naproxen	Placebo	19.34	(0.87,37.76)		
Etoricoxib	Tolfenamic acid	-3.06	(-26.02,19.19)		
Etoricoxib	Meloxicam 15mg	9.29	(-8.12,26.69)		
Etoricoxib	Placebo	28.60	(16.41,40.66)		
Tolfenamic acid	Meloxicam 15mg	12.35	(-10.19,35.40)		
Tolfenamic acid	Placebo	31.66	(12.43,51.43)		
Meloxicam 15mg	Placebo	19.31	(6.51,32.13)		

REFERENCES

1. Dias S, Welton NJ, Sutton AJ, Ades AE. *NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials.* 2011.
2. Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med.* 2004 2004;23:3105-3124.
3. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ.* 2005 2005;331:897-900.
4. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing.* 2000 2000;10:325-337.
5. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. *The BUGS book.* Boca Raton, FL: CRC Press; 2013.
6. Stevens, J.W., A note on dealing with missing standard errors in meta-analyses of continuous outcome measures in WinBUGS. *Pharmaceutical Statistics*. 2011; 10:4:374-378