National Institute for Health and Care Excellence

version 2.0

Eating disorders: recognition and treatment

Appendix N - TSU report on NMA for bulimia

NICE Guideline

Methods, evidence and recommendations

May 2017

Final

Developed by the National Guideline Alliance, hosted by the Royal College of Obstetricians and Gynaecologists

Contents

1

2

Appendices

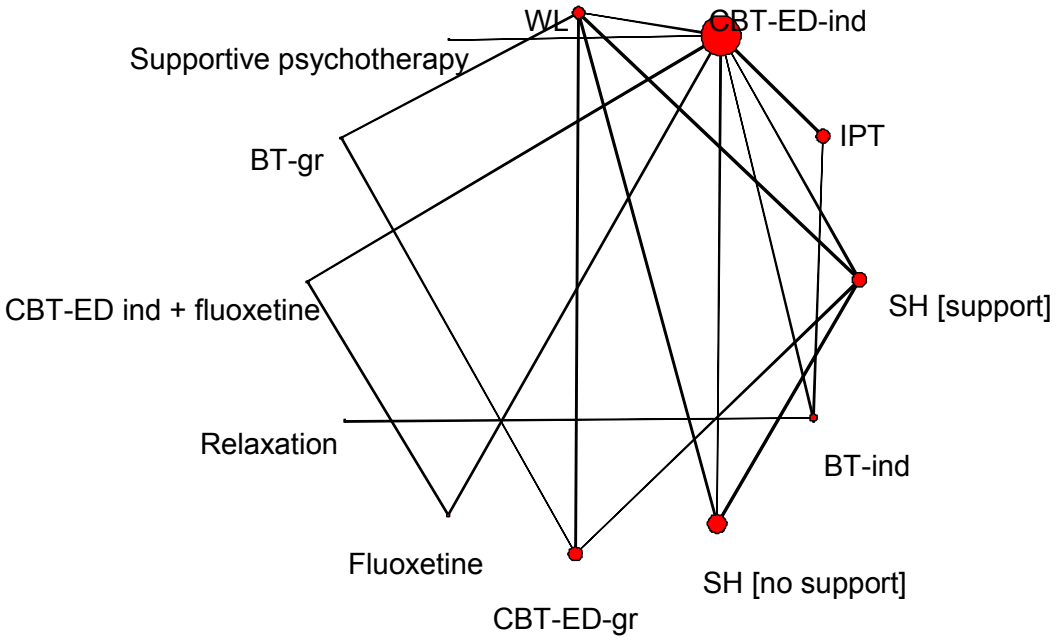
Appendix N: BULIMIA NERVOSA: NMA ON REMISSION OUTCOME, INCONSISTENCY CHECKS AND BIAS ADJUSTMENT

The purpose of this analysis was to test the robustness of the estimates of comparative effectiveness identified from the base case network meta-analysis (NMA) of the following interventions for remission for bulimia nervosa:

1. WL
2. CBT-ED-ind
3. IPT
4. SH [support]
5. BT-ind
6. SH [no support]
7. CBT-ED-gr
8. Fluoxetine
9. Relaxation
10. CBT-ED ind + fluoxetine
11. BT-gr
12. Supportive psychotherapy

22 studies were included in the analyses. The network diagram is shown in **Error! eference source not found.**.

Figure 1: Network Diagram for remission outcome



1

2

3

1 Methods

Inconsistency checks

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 an 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. Inconsistency can be thought of as a conflict between direct evidence on a comparison between treatments A and B, and indirect evidence gained from AC and BC trials.

We tested for inconsistency firstly by comparing the standard network consistency model to an "inconsistency", or unrelated mean effects, model (Dias, 2013). The latter is equivalent to having separate, unrelated, meta-analyses for every pair-wise contrast but with a common variance parameter in random effects (RE) models. The WinBUGS code for the inconsistency model is provided in Appendix 1.

The goodness-of-fit of each model to the data was measured by comparing the posterior mean of the summed deviance contributions to the number of data points (Dempster, 1997). The Deviance Information Criterion (DIC), which is equal to the sum of the posterior mean of the residual deviance and the effective number of parameters was used as a basis for model comparison (Spiegelhalter, 2002). Model selection was also based on the posterior mean between study heterogeneity (SD).

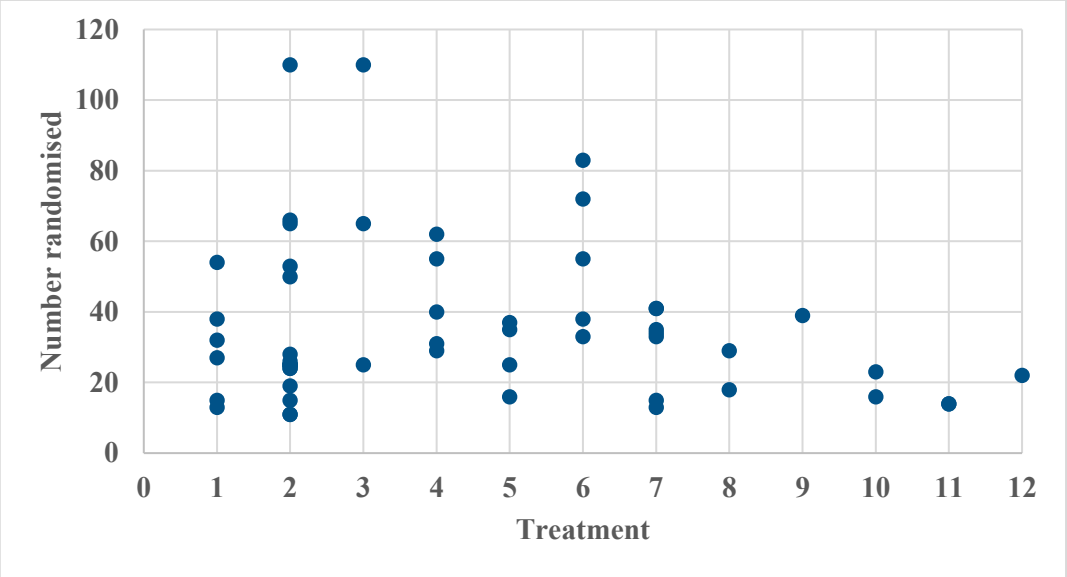
Another approach we used to test for inconsistency was node-splitting (Dias, 2010). This involves splitting the information contributing to estimates of a parameter (AB), into two distinct components: the "direct" based on all the AB data (which may come from AB, ABC, DAB etc. trials) and the "indirect" based on all the remaining evidence. This was done using the GeMTC package in R (van Valkenhoef, 2012).

1.1 Bias adjustment

It is commonly known that small studies are more likely to be published if they show a significant effect. Smaller size is also often associated with less rigorous conduct of trials. Figure 2 shows the number of patients randomised to each treatment arm in the analysis, ordered by treatment. It is clear that some treatments such as treatments 10, 11 and 12 (CBT-ED-individual + fluoxetine, BT-group, and supportive psychotherapy) have only been tested on quite low numbers of patients.

As there were a number of small studies included in the base case analysis we carried out a sensitivity analysis adjusting for bias associated with such small study effects.

Figure 2 NUMBER OF PATIENTS RANDOMISED TO EACH TREATMENT ARM



Starting with the assumption that the smaller the study the greater the bias, the analysis attempted to estimate the "true" treatment effect which is that which would be obtained in a study of infinite size. This was taken to be the intercept in a regression of the treatment effect against the study variance. Both random and fixed effect bias adjustment models were run. The WinBUGS code for this analysis is given in Appendix 2. The effect that this adjustment had on relative effects and between trial heterogeneity is explored below.

2 Results

2.1 Inconsistency checks

Both consistency (i.e. standard NMA) and inconsistency models were run using the full dataset. Convergence was satisfactory by at least 70,000 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.

Some differences were observed in posterior mean residual deviance and DIC values suggesting that, for the full network, there was evidence of inconsistency (Table 1). The addition of a continuity correction of 0.5 for studies with zero events (on either arm) did not improve model fit.

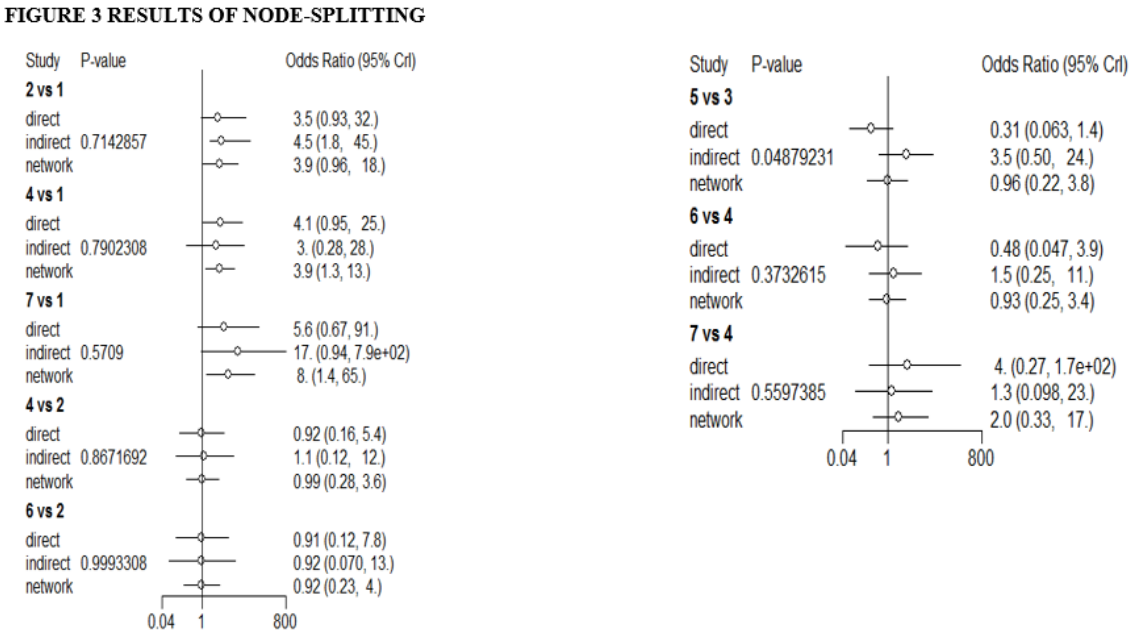
We examined the effect of removing one study (Mitchell 1993) in which all treatment arms were classified as CBT-ED-gr but numbers achieving remission varied substantially. This treatment classification assumes that the effect of each of the options compared to each of the others is zero as they are the same intervention. However, the data suggested that not all intensities of this intervention have the same effectiveness and this was translated as high heterogeneity in model. As the study did not contribute to the estimates of the relative effects of CBT-ED-gr compared to any of the other treatments it was removed. The random-effects model, continuity corrected and excluding this trial, provided an adequate fit to the data (Table 1). The final data file used is shown in Appendix 3.

Table 1 model fit statistics - base case analysis

Error! Reference ource not found.	**No. of data points**	**Residual Deviance over all studies**	**Between-trials SD (posterior median) and 95% credible intervals**	**DIC**
RE consistency	55	54.98	0.77 (0.39 – 1.25)	276.76
RE inconsistency	55	53.91	0.71 (0.37 – 1.17)	277.74
RE consistency – Continuity Corrected	55	54.15	0.74 (0.37 – 1.20)	279.78
RE consistency – Mitchell removed, Continuity Corrected	51	49.42	0.42 (0.04 – 0.93)	254.01
RE inconsistency – Mitchell removed, Continuity Corrected	51	50.13	0.48 (0.04 – 0.96)	265.77

Error! Reference source not found.3 shows the results from the node-slitting exercise, lotting the direct, indirect, and combined information on each comparison where both are available. A p-value less than 0.05 indicates a discrepancy between the direct and indirect information and it is shown on the plots that this only occurs in the comparison between treatments 5 and 3 (BT individual and IPT). Here the direct information favours BT individual but the indirect information favours IPT.

Figure 3: Results of node-splitting



Source: NICE TSU

Four different bias scenarios were tested after consultation with the Guideline Committee:

1. In trials of active treatments versus waitlist, active treatments are favoured.
2. i) Active treatments are favoured against waitlist and
   ii) In trials of supportive psychotherapy or relaxation versus other active treatments, the other active treatments are favoured.
3. i) Active treatments are favoured against waitlist and
   ii) other active treatments are favoured against supportive psychotherapy and relaxation and
   iii) in trials of CBT versus other treatments, CBT is favoured.
4. i) All active treatments are favoured against waitlist and
   ii) CBT is favoured against other treatments.

It was not possible to obtain results from scenarios 2 and 3 due to sparsity of the data. Table 2 looks at the bias coefficients (B) from scenarios 1 and 4. These are the change in the log odds ratio of the favoured intervention for a one unit increase in the study variance. If bias is present in this network we would expect the coefficient to be positive as remission is a positive outcome and the log odds of remission will be increased due to bias.

In nearly all cases the bias coefficient is positive suggesting that the treatment effect is exaggerated in smaller studies, although the 95% credible intervals (CrI) include the possibility of no bias. The effect seems to be more exaggerated in comparisons of active treatments against waitlist than in comparisons of CBT with other treatments.

Table 2 also compares the model fit statistics from the random and fixed effects bias adjustment models. In all cases the random effects model is a better fit to the data with the residual deviance closer to the number of data points and a lower between-trials SD and DIC. Comparing these statistics to the model fit statistics from the base case analysis (Table 1) shows that adjusting for the bias does not reduce the between-trials heterogeneity as would be expected if the bias was the cause of the heterogeneity.

Table 2: Bias coefficients and model fit statistics for each scenario

	B	95% CrIs	Data points	Residual Deviance	Between trials SD	DIC
Scenario 1 (RE)						
Active treatments favoured v WL	0.44	(-0.80, 2.01)	52	50.4	0.45 (0.04, 0.99)	256.94
Scenario 1 (FE)						
Active treatments favoured v WL	0.44	(-0.67, 1.74)	52	56.81	-	257.52
Scenario 4 (RE)						
Active treatments favoured v WL	0.44	(-1.11, 2.05)	52	51.09	0.46 (0.03, 1.01)	258.65
CBT favoured v other treatments	-0.16	(-3.8, 2.97)	52	51.09	0.46 (0.03, 1.01)	258.65
Scenario 4 (FE)						
Active treatments favoured v WL	0.37	(-0.97, 2.02)	52	57.24	-	259.83
CBT favoured v other treatments	-0.69	(-3.87, 2.44)	52	57.24	-	259.83

Error! Reference source not found.4 shows the log odds ratio of remission of each reatment compared to waitlist in the base case analysis (not adjusting for bias). Positive values mean that the treatment is more likely to lead to remission compared to waitlist. The figure shows that compared to waitlist, 5 interventions resulted in a significant increase in remission: CBT-ED-individual, SH (support), SH (no support), CBT-ED-group, and BT-group.

Error! Reference source not found.5 and **Error! Reference source not found.**6 show the g odds ratios of remission under bias scenarios 1 and 4. These show that adjusting for the presence of small study effect bias by extrapolating to an infinitely sized study increases the overall uncertainty in the network.

This indicates the presence of small study effects as when the trials are adjusted to account for the bias no treatments result in a significant increase in remission.

Figure 4: MEAN DIFFERENCES IN REMISSION COMPARED TO WAITLIST - BASE CASE ANALYSIS

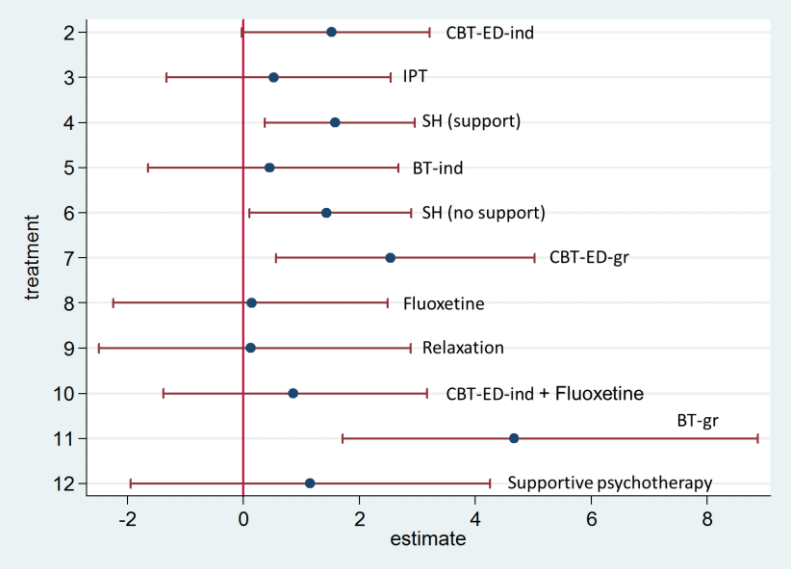
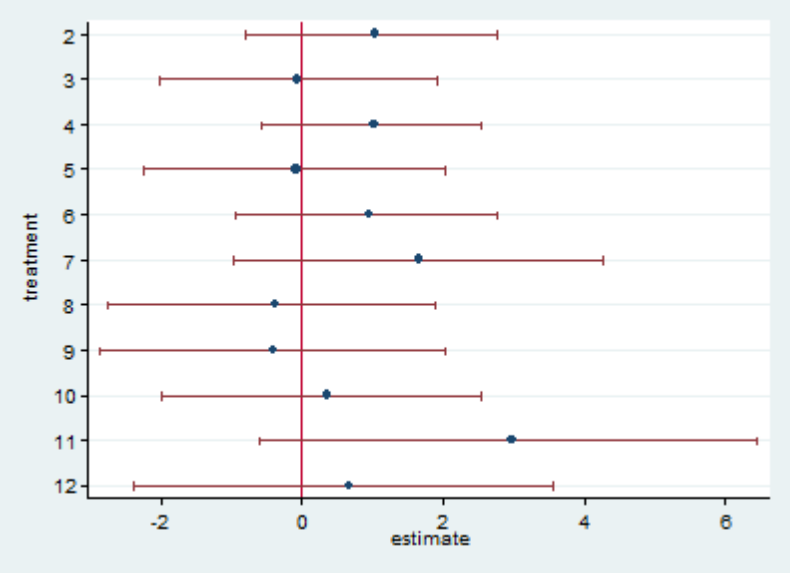


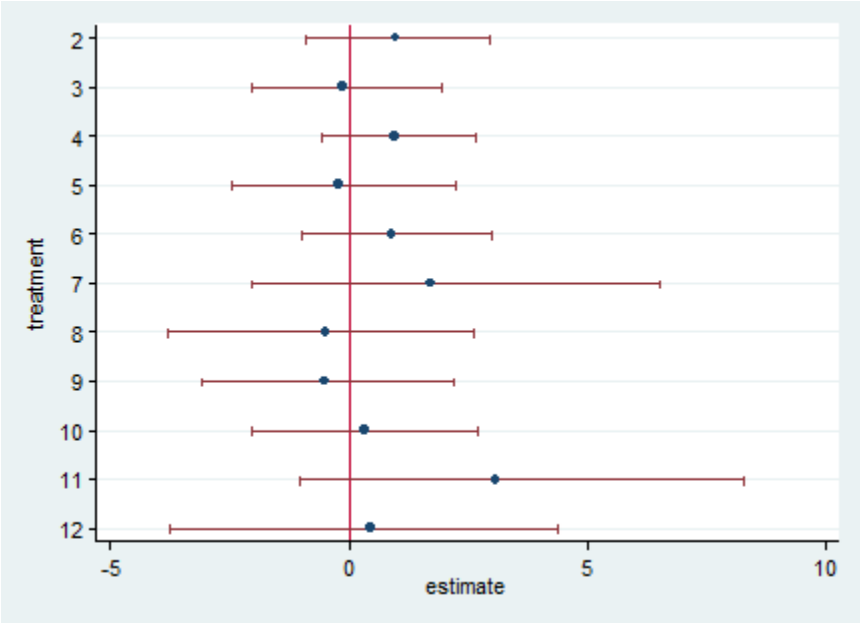
Figure 5: MEAN DIFFERENCES IN REMISSION COMPARED TO WAITLIST - SCENARIO 1

Treatment numbering	
2	CBT-ED-ind
3	IPT
4	SH [support]
5	BT-ind
6	SH [no support]
7	CBT-ED-gr
8	Fluoxetine
9	Relaxation
10	CBT-ED ind + fluoxetine
11	BT-gr
12	Supportive psychotherapy



1

Figure 6: MEAN DIFFERENCES IN REMISSION COMPARED TO WAITLIST - SCENARIO 4



Conclusion

The inconsistency checks did not identify any significant inconsistency in the direct and indirect evidence included in the network meta-analysis. This strengthens the conclusions from the base case analysis.

The bias adjustment sensitivity analysis suggested that bias due to small study effects may be exaggerating the treatment effects in this network. However, as the bias coefficient included zero in all scenarios and there was no reduction in heterogeneity as a result of the bias adjustment, no strong conclusions about the presence of bias can be made.

Appendix 1. WinBUGS code for inconsistency model

```
# Random effects inconsistency model
model{
for(i in 1:ns){
    delta[i,1]<-0           # treatment effect is zero in control arm
    mu[i] ~ dnorm(0,.001)  # vague priors for trial baselines

for (k in 1:na[i])  {
        r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
        logit(p[i,k]) <- mu[i] + delta[i,k]

#Deviance contribution
        rhat[i,k] <- p[i,k] * n[i,k] # expected value of the numerators
#Deviance contribution
    p0[i,k]<-0.5+.999999*(p[i,k]-0.5)
      r0[i,k]<-r[i,k]+0.01*equals(r[i,k],0) -0.01*equals(r[i,k],n[i,k])
      r.hat[i,k]<- p0[i,k]*n[i,k]                                    #
expected value of the numerators

#Deviance calculation for binomial data with adjustments
dev[i,k]<-      2*(r0[i,k]*log(r0[i,k]/r.hat[i,k])      +      (n[i,k]      -
r0[i,k])*log((n[i,k] - r0[i,k])/(n[i,k] - r.hat[i,k])))
}

# summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
   for (k in 2:na[i]) {

# trial-specific LOR distributions
        delta[i,k] ~ dnorm(d[t[i,1],t[i,k]] ,tau)    }

  }

totresdev <- sum(resdev[])    # Total Residual Deviance
for (c in 1:(nt-1)) {  d[c,c]<-0                      # priors for all mean
trt effects
    for (k in (c+1):nt)  { d[c,k] ~ dnorm(0,.001)
        or[c,k] <- exp(d[c,k])}
        }

sd ~ dunif(0,4)  # vague prior for between-trial standard deviation
var <- pow(sd,2) # between-trial variance
tau <- 1/var     # between-trial precision
}
```

Appendix 2 WinBUGS code for bias adjustment model (Scenario 4)

```
model{
for(i in 1:ns){
     w[i,1] <- 0
     beta[i,1] <- 0      # no bias term in baseline arm
     delta[i,1] <- 0
     mu[i] ~ dnorm(0,.0001)   # vague priors for trial baselines
     for (k in 1:na[i])  {
         r[i,k] ~ dbin(p[i,k],n[i,k])

 rhat[i,k] <- p[i,k] * n[i,k]                                          #
expected value of the numerators
     dev[i,k]    <-    2    *    (r[i,k]    *    (log(r[i,k])-log(rhat[i,k]))
#Deviance contribution
         + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))}

  resdev[i] <- sum(dev[i, 1:na[i]])     # residual deviance for study i

#Parameterization#
logit(p[i,1])<- mu[i]



    for (k in 2:na[i])  {
lamda.1[i,k]<- equals(r[i,k],0)
lamda.2[i,k]<- equals(n[i,k],r[i,k])
lamda.3[i,k]<- equals(r[i,1],0)
lamda.4[i,k]<- equals(n[i,1],r[i,1])
lamda.a[i,k]<- max(lamda.1[i,k],lamda.2[i,k])
lamda.b[i,k]<- max(lamda.3[i,k],lamda.4[i,k])
lamda[i,k]<-max(lamda.a[i,k],lamda.b[i,k])

var[i,k]<-1/(r[i,k]+(0.5*lamda[i,k]))+1/(r[i,1]+(0.5*lamda[i,k]))
+1/(n[i,k]-r[i,k]+(0.5*lamda[i,k]))+1/(n[i,1]-r[i,1]+(0.5*lamda[i,k]))

logit(p[i,k]) <- mu[i] + delta[i,k]+ beta[i,k]*var[i,k]*I[i,k]          #
model for linear predictor
I[i,k]<-0.5*(Z[i,1]-Z[i,k])

# model for bias parameter beta
         beta[i,k] ~ dnorm(A[C[i,(k-1)]], Pkappa)
# distributions for trial-specific logHR
         delta[i,k] ~ dnorm(md[i,k], taud[i,k])
         md[i,k] <- (d[t[i,k]] - d[t[i,1]]) + sw[i,k]
#precision of diff in means distributions
         taud[i,k] <- tau *2*(k-1)/k
#adjustment, multi-arm RCTs
         w[i,k] <- delta[i,k] - d[t[i,k]] + d[t[i,1]]
# cumulative adjustment for multi-arm trials
         sw[i,k] <-sum(w[i,1:k-1])/(k-1)
        }

}
totresdev <- sum(resdev[])           # Total Residual Deviance
d[1]<-0
for (k in 2:nt){d[k] ~ dnorm(0,.0001) }    # vague priors for basic parameters
sd.d ~ dunif(0,5)                          # vague prior for RE st dev
tau <- pow(sd.d,-2)
```

```
# mean bias: assumptions
A[1] <- 0              # WL v WL
A[2] <- b[1]              # WL v A
A[3] <- b[1]              # WL v CBT
A[4] <- b[2]              # A v CBT
A[5] <- 0              # A v A
A[6] <- 0              # CBT v CBT
# bias model prior for variance
kappa ~ dunif(0,10)
kappa.sq <- pow(kappa,2)
Pkappa <- 1/kappa.sq
# bias model prior for mean
for (k in 1:2) {b[k] ~ dnorm(0,.0001)}
# all pairwise differences
for (c in 1:(nt-1)) { for (k in (c+1):nt) { or[c,k] <- exp(d[k] - d[c])
lor[c,k] <- (d[k]-d[c])
} }
}
```

Appendix 3. Data file for Bulimia Nervosa

t[,1]	t[,2]	t[,3]	t[,4]	r[,1]	n[,1]	r[,2]	n[,2]	r[,3]	n[,3]	r[,4]	n[,4]	na[]	Study name
1	2	6	NA	2	27	5	28	9	55	NA	NA	3	Treasure 1994
1	4	NA	NA	6	54	14	55	NA	NA	NA	NA	2	Banasiak 2005
1	4	4	6	0.5	32	4.5	29	3.5	31	2.5	33	4	Palmer 2002
1	6	NA	NA	1	38	7	38	NA	NA	NA	NA	2	Sanchez-Ortiz 2011
1	7	NA	NA	1	15	4	15	NA	NA	NA	NA	2	Lee 1986
1	7	13	13	0.5	13	1.5	13	4.5	14	4.5	14	4	Leitenberg 1988
2	2	NA	NA	26	53	29	50	NA	NA	NA	NA	2	Fairburn 2009
2	2	NA	NA	22	24	18	26	NA	NA	NA	NA	2	Ghaderi 2006
2	2	NA	NA	7	11	7	11	NA	NA	NA	NA	2	Wilson 1991
2	2	NA	NA	11	25	10	25	NA	NA	NA	NA	2	Thomson-Brenner 2016
2	3	NA	NA	35	110	8	110	NA	NA	NA	NA	2	Agras 2000
2	3	NA	NA	22	65	7	65	NA	NA	NA	NA	2	Fairburn 2015
2	3	5	NA	9	25	11	25	5	25	NA	NA	3	Fairburn 1993
2	4	NA	NA	19	66	17	62	NA	NA	NA	NA	2	Mitchell 2008
2	5	NA	NA	7	15	6	16	NA	NA	NA	NA	2	Cooper 1995
2	8	10	NA	6	24	2	29	3	23	NA	NA	3	Goldbloom 1997
2	8	10	NA	5	19	2	18	3	16	NA	NA	3	Jacobi 2002
2	14	NA	NA	3	25	2	22	NA	NA	NA	NA	2	Walsh 1997
4	7	NA	NA	1	40	3	41	NA	NA	NA	NA	2	Bailer 2004
5	5	9	NA	24	37	15	35	18	39	NA	NA	3	Bulik 1998
6	6	NA	NA	12	83	11	72	NA	NA	NA	NA	2	Wagner 2013b

References

Dempster AP. 1997. The direct use of likelihood for significance testing. *Statistics and Computing* 7(4): 247-252

Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. (2013). Evidence Synthesis for Decision Making 4: Inconsistency in Networks of Evidence Based on Randomized Controlled Trials. *Medical Decision Making*. 33(5):641-656.

Dias, S, Welton, N, Caldwell, D & Ades, A, 2010, 'Checking consistency in mixed treatment comparison meta-analysis'.*Statistics in Medicine*, vol 29., pp. 932 – 944

van Valkenhoef, G., Lu, G., de Brock, B., Hillege, H., Ades, A. E., & Welton, N. J. (2012). Automating network meta-analysis. *Research Synthesis Methods*, *3*(4), 285–299

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society (B)* 64(4): 583-616