Appendix 1

1 Elicitation method for continuous outcome measures

For a continuous outcome measure, let X_{ij} denote the sample mean in study i on treatment arm j, with $j = 1$ the control arm and $j = 2$ the experimental treatment arm. Suppose the sample means have the distributions $X_{i1} \sim N(\mu_i, \sigma^2/n_1)$ and $X_{i2} \sim N(\mu_i + \beta_i, \sigma^2/n_2)$, where $n_1$and n_2 are the sample size in group 1 and 2, respectively. Note that in the following, we assume the variances σ^2 are equal across arms and studies. The treatment effect, mean difference (MD) β_i, is on the original scale. A standardised mean difference (SMD), $\phi_i = \frac{\beta_i}{\sigma}$, may be used in meta-analysis if the included studies used different scales.

We further assume that the study-specific treatment effects are normally distributed: $\beta_1, \dots, \beta_S \sim N(d_{MD}, \tau^2_{MD})$, or $\phi_1, \dots, \phi_S \sim N(d_{SMD}, \tau^2_{SMD})$ depending on the scale used in each study. We suppose that the expert again prefers to consider variability in treatment effects via ratios of treatment effects, and we now consider a modification of the three-stage approach in Section 3.1.

If we can relate the treatment effects β_i or ϕ_i to an odds ratio (OR) δ_i, we could derive a distribution for τ_{MD} (the variability in mean differences (MDs) in a population of treatment effects) or τ_{SMD} (the variability in standardised mean differences (SMDs) in a population of treatment effects) via a distribution of τ (the variability in ORs in a population of treatment effects), elicited as before. We follow the approach by Chinn (2000) (38), where a continuous response is dichotomised, and a normal distribution is approximated by a logistic distribution.

A cut-off c of interest is chosen (a clinically meaningful threshold in the observed response), and the OR δ_i is defined as

$$\delta_i = \left(\frac{P(X_{i2} \geq c)}{P(X_{i2} < c)}\right) / \left(\frac{P(X_{i1} \geq c)}{P(X_{i1} < c)}\right). \qquad (1)$$

We can approximate a normal distribution $N(m, s^2)$ by a logistic distribution with same mean and variance, setting the location parameter in the logistic distribution to m and the scale parameter to $\frac{s\sqrt{3}}{\pi}$. Using the logistic distribution approximation, the OR (1) is

$$\delta_i = \exp\left(\frac{\phi_i \pi}{\sqrt{3}}\right) = \exp\left(\frac{\beta_i \pi}{\sigma\sqrt{3}}\right).$$

We now have

$$\tau_{SMD} = \frac{\sqrt{3}\tau}{\pi},$$

$$\tau_{MD} = \frac{\sqrt{3}\sigma\tau}{\pi},$$

where τ is the between-study standard deviation (SD) on the log OR scale. Hence, we can now use the method in Section 3.1 with the following modification.

1. Dichotomise the response using some appropriate cut-off c, to define a new treatment effect δ_i: the OR (1).

2. Considering ORs for the dichotomised response, use the three-stage procedure to elicit a prior distribution for τ, the variability in ORs in a population of treatment effects.

3. Given a prior distribution for τ, convert it to a prior distribution for the between-study SD τ_{MD} and τ_{SMD} on the continuous scale via $\tau_{MD} = \frac{\sqrt{3}\sigma\tau}{\pi}$ for MD, and $\tau_{SMD} = \frac{\sqrt{3}\tau}{\pi}$ for SMD, where σ is an estimate of an individual level standard deviation. The estimate could be a summary measure of the SDs in the included studies, pooled from included studies, or obtained from a single representative study.

2 Elicitation method for ordered categorical data

For ordered categorical data, the likelihood function for the data would be a multinomial distribution with either a logit link function (i.e. a proportional odds model) or a probit link function. Suppose that there are K outcome categories, denoted by $c_1, \ldots, c_K$. Define P_{ijk} to be the probability of an observation belonging to category k or above, on treatment $j = 1,2$, with $j = 1$ the control arm and $j = 2$ the experimental treatment arm, in study i. For a logit link function, the treatment effect in the ith study can be defined by a single OR δ_i, the OR

$$\frac{P_{i2k}}{1-P_{i2k}} \Big/ \frac{P_{i1k}}{1-P_{i1k}} \qquad (2)$$

which is constant for all k. Hence, the outcome can be dichotomised into the two category sets $c_1, \ldots, c_{k-1}$ and $c_k, \ldots, c_K$, and the elicitation can proceed as in Section 3.1.

If a probit link function is used, the treatment effect in study i may be described by a shift μ_i in the mean of the latent normal variable, and we again require a prior distribution for $\tilde{\tau}$, the variability in μ_i in a population of treatment effects. In this case, the OR (2) will change depending on the category k. However, an approximate prior for τ can be elicited using a similar approach to that in continuous outcome measures case: we dichotomise and approximate the latent normal variable by a latent logistic variable with scale parameter $\frac{\sqrt{3}}{\pi}$. We have the same modification as before:

1. Dichotomise the response using some appropriate category c_k, and define a new treatment effect δ_i : the OR (2).

2. Use the three-stage procedure to elicit a prior distribution for τ, the variability in ORs in a population of treatment effects.

3. Given a prior for τ, convert this to a prior for $\tilde{\tau}$ via

$$\tilde{\tau} = \frac{\sqrt{3}}{\pi}\tau.$$

The interpretation of the heterogeneity parameter can be found in Table 1.

Heterogeneity	'range' of treatment effect, R, for scale-free outcome measure	τ for scale-free outcome measure	τ for outcome measure using probit or standardised mean difference scale	τ for outcome measure using mean difference scale
No heterogeneity	1	0	0	0
Low	1.21	0.05	0.028	0.028σ
Moderate	1.48	0.1	0.06	0.06σ
	2.19	0.2	0.11	0.1σ
	3.24	0.3	0.17	0.17σ
	4.80	0.4	0.22	0.22σ
	7.10	0.5	0.28	0.28σ
High	10.51	0.6	0.33	0.33σ
	15.55	0.7	0.39	0.39σ
	23.01	0.8	0.44	0.44σ
	34.06	0.9	0.50	0.50σ
	50.40	1.0	0.55	0.55σ
Extremely high	357.81	1.5	0.83	0.83σ
	2540.20	2	1.10	1.10σ

Table 1: Suggested interpretation of the between-study standard deviation. The scale-free outcome measure refers to odds ratio, relative risk, hazard ratio and ratio of means. The estimate of σ could be a summary measure of the standard deviations in the included studies, pooled from included studies, or obtained from a single representative study.

Appendix 2

1 BUGS code for example TA163

```
model{                    # *** PROGRAM STARTS
  for(i in 1:ns){         # LOOP THROUGH STUDIES
    w[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(0,0.0001)       # vague priors for all trial baselines
    for (k in 1:na[i]) {          # LOOP THROUGH ARMS
      r[i,k] ~ dbin(p[i,k],n[i,k])   # binomial likelihood
      logit(p[i,k]) <- mu[i] + delta[i,k]  # model for linear predictor
      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]))
                  +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))       #Deviance contribution  }
       resdev[i] <- sum(dev[i,1:na[i]])      #  summed residual deviance contribution for this trial
    for (k in 2:na[i]) {          # LOOP THROUGH ARMS
      # trial-specific LOR distributions
      delta[i,k] ~ dnorm(md[i,k],precisiond[i,k])
      # mean of LOR distributions (with multi-arm trial correction)
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
      # precision of LOR distributions (with multi-arm trial correction)
      precisiond[i,k] <- precision *2*(k-1)/k
      # adjustment for multi-arm RCTs
      w[i,k] <- (delta[i,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[])          # Total Residual Deviance
 d[1]<-0       # treatment effect is zero for reference treatment
 # vague priors for treatment effects
 for (k in 2:nt){  d[k] ~ dnorm(0,0.0001)
  OR[k] <- exp(d[k])
  d.new[k] ~ dnorm(d[k],precision)
  OR.new[k] <- exp(d.new[k])
 }
 # vague prior U[0,5] for between-trial SD
 tau ~ dunif(0,5)
 precision <- pow(tau,-2)   # between-trial precision = (1/between-trial variance)
 # informative prior using Turner et al (2012)
 #tau2~dlnorm(-2.57,0.33)   # prior for between study variance from lognormal (-2.57, 1.74^2)
 #tau<-sqrt(tau2)
 #precision<-1/tau2       # between-trial precision = (1/between-trial variance)
 # informative prior using Turner et al (2012) truncated so that the ratio of ORs can't exceed 10
 #tau2~dlnorm(-2.57,0.33)I(,0.345)   # R=exp(3.92tau)=> tau^2=(log(10)/3.92)^2=0.345
 #tau<-sqrt(tau2)
 #precision<-1/tau2       # between-trial precision = (1/between-trial variance)
 # informative prior using elicitation
 #R~dgamma(2.68,0.721) #elicited prior for the 'range' of OR
 #tau<-log(R+1)/3.92    #minimum of R is 1; convert the 'range' of OR to the between-study standard
deviation
 #precision<-pow(tau,-2)       # between-trial precision = (1/between-trial variance)
}                              # *** PROGRAM ENDS
```

#Data (1=placebo, 2=infliximab, 3=ciclosporin)

list(ns=4,nt=3)

t[,1]	t[,2]	n[,1]	r[,1]	n[,2]	r[,2]	na[]
1	2	21	14	24	7	2
1	2	3	3	3	0	2
1	3	9	4	11	3	2
1	3	15	3	14	3	2

END

2 BUGS code for example TA336

```
model{                          # *** PROGRAM STARTS
 for(i in 1:ns){                  #   LOOP THROUGH STUDIES
   w[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(0,0.0001)          # vague priors for all trial baselines
   for (k in 1:na[i]) {            #  LOOP THROUGH ARMS
     var[i,k] <- pow(se[i,k],2)   # calculate variances
     prec[i,k] <- 1/var[i,k]      # set precisions
     y[i,k] ~ dnorm(theta[i,k],prec[i,k]) # binomial likelihood
     theta[i,k] <- mu[i] + delta[i,k]  # model for linear predictor
     dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k]  #Deviance contribution
   }
   #  summed residual deviance contribution for this trial
   resdev[i] <- sum(dev[i,1:na[i]])
   for (k in 2:na[i]) {            # LOOP THROUGH ARMS
     # trial-specific distributions
```

```
    delta[i,k] ~ dnorm(md[i,k],precisiond[i,k])
    # mean of distributions, with multi-arm trial correction
    md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k]
    # precision of distributions (with multi-arm trial correction)
    precisiond[i,k] <- precision *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       # treatment effect is zero for control arm
d.new[1]<-0
 # vague priors for treatment effects
for (k in 2:nt){  d[k] ~ dnorm(0,0.0001)
  d.new[k] ~ dnorm(d[k],precision) }
# vague prior U[0,5] for between-trial SD
tau ~ dunif(0,5)
precision <- pow(tau,-2)   # between-trial precision = (1/between-trial variance)
 # informative prior using Turner et al (2012) on odds ratio scale
#tau2~dlnorm(-2.56,0.33)  #odds ratio scale
#tau<-sqrt(tau2)/1.81*2.61   #mean difference scale; 2.61 is the mean of individual level standard deviation
#precision <- pow(tau,-2)   # between-trial precision = (1/between-trial variance)
```

```
  # informative prior using Turner et al (2012) truncated so that the ratio of ORs can't exceed 10 on the odds ratio scale, R=exp(3.92tau)=>tau^2=(log(10)/3.92)^2=0.345

  #tau2~dlnorm(-2.56,0.33)I(,0.345)

  #tau<-sqrt(tau2)/1.81*2.61  #mean difference scale

  #precision <- pow(tau,-2)   # between-trial precision = (1/between-trial variance)

  # informative prior using elicitation

  #R~dgamma(1.94,0.823)        #odds ratio scale

  #tau<-log(R+1)/3.92/1.81*2.61  #mean difference scale, #minimum of R is 1

  #precision<-pow(tau,-2)       # between-trial precision = (1/between-trial variance)
}                              # *** PROGRAM ENDS


#Data (1=Placebo+Met+SU, 2=Sita+Met+SU, 3=Empa 10mg+Met+SU, 4=Lina+Met+SU,
5=Saxa+Met+SU, 6=Can 300mg+Met, 7=Can 100mg+Met, 8=Empa 25mg+Met+SU)

list(ns=6,nt=8)
t[,1]   t[,2]   t[,3]   y[,1]   y[,2]   y[,3]   se[,1]  se[,2]  se[,3]  na[]

1       3       8       -0.39   -2.16   -2.39   0.15    0.15    0.16    3

1       4       NA      -0.06   0.27    NA      0.16    0.09    NA      2

1       2       NA      -0.70   0.40    NA      0.3316  0.2551  NA      2

1       5       NA      -0.60   0.20    NA      0.1849  0.1945  NA      2

2       6       NA      0.2649  -2.384  NA      0.1325  0.1325  NA      2

1       7       6       -0.648  -1.945  -2.408  0.2362  0.2362  0.2362  3

END
```

Appendix 3

1 R code instructions

A function `elicitHeterogen()` is available in the R package `SHELF` (30). The elicitation tool can be run using the commands

```
library(SHELF)

elicitHeterogen()
```

Type `?elicitHeterogen` for further instructions.

2 Elicited prior distribution for the re-analysis of TA163 and TA336

Table 1 shows the number of bins used and the number of probs/chips allocated in each bin for the re-analysis of TA163 and TA336. The elicited prior for $R-1$ was gamma (2.62, 0.721). It presented the beliefs that the probability of heterogeneity being low, moderate and high as 0.01, 0.85, and 0.14, respectively. The R function used was `elicitHeterogen(lower=1,upper=10,nbins=9)`.

The elicited prior for $R-1$ was gamma (1.94, 0.741). It presented the beliefs that the probability of heterogeneity being low, moderate and high as 0.06, 0.88, and 0.06, respectively. The R function used was `elicitHeterogen(lower=1,upper=10,nbins=9,sigma=2.61,scale.free=FALSE)`.

Bin boundary	**[1, 2)**	**[2, 3)**	**[3, 4)**	**[4, 5)**	**[5, 6)**	**[6, 7)**	**[7, 8)**	**[8,9)**	**[9, 10)**
Number of probs allocated (TA136)	4	5	6	6	5	4	2	1	1
Number of probs allocated (TA336)	4	5	4	3	2	1	1	0	0

Table 1: The number of bins and the number of probs allocated in each bin for the re-analysis of TA163 and TA336.