Sensitivity of treatment recommendations to bias in Network Meta-Analysis

David M Phillippo[1], Sofia Dias[1], AE Ades[1], Vanessa Didelez[2,3], Nicky J Welton[1]

[1]Population Health Sciences, Bristol Medical School, University of Bristol
[2]Leibniz Institute for Prevention Research and Epidemiology − BIPS
[3]Department of Mathematics, University of Bremen

A. APPENDICES

A.1 Lemma 1

Let $\tilde{k}^* = \underset{k=1,\dots,K}{\operatorname{argmax}} \widetilde{\mathbb{E}}(d_k)$ be the optimal treatment under the bias-adjusted data $\tilde{\boldsymbol{y}}(\boldsymbol{\beta})$. Then:

    a) $\tilde{k}^* \neq k^* \quad \Leftrightarrow \quad \widetilde{\mathbb{E}}(d_{bk^*}) < 0$ for at least one $b \in \{1,\dots,K\} \setminus k^*$.

    b)   For a value of $\boldsymbol{\beta}$ where $\widetilde{\mathbb{E}}(d_{bk^*}) < 0$ for only one $b \in \{1,\dots,K\} \setminus k^*$, the new optimal treatment is $\tilde{k}^* = b$.

Proof of Lemma 1a

Recall a general property of contrasts that $d_{ab} = -d_{ba}$. So without loss of generality, the set of contrasts associated with k^* is $\{d_{ak^*} : a \in \{1,\dots,K\} \setminus k^*\}$. Note that under the original biased data the set of contrasts associated with k^* have posterior expected values satisfying

$$\mathbb{E}_{d|y}(d_{ak^*}) > 0 \quad \forall a \in \{1,\dots,K\} \setminus k^*.$$

First show that

$$\tilde{k}^* \neq k^* \quad \Rightarrow \quad \widetilde{\mathbb{E}}(d_{bk^*}) < 0 \text{ for at least one } b \in \{1,\dots,K\} \setminus k^*.$$

We have that $\tilde{k}^*$ is the new optimal treatment under the bias-adjusted data, so by definition

$$\widetilde{\mathbb{E}}(d_{\tilde{k}^*}) > \widetilde{\mathbb{E}}(d_a) \quad \forall a \in \{1,\dots,K\} \setminus \tilde{k}^*,$$

and in particular $\widetilde{\mathbb{E}}(d_{\tilde{k}^*}) > \widetilde{\mathbb{E}}(d_{k^*})$. Thus

$$\widetilde{\mathbb{E}}\left(d_{\tilde{k}^*k^*}\right)=\widetilde{\mathbb{E}}\left(d_{k^*}\right)-\widetilde{\mathbb{E}}\left(d_{\tilde{k}^*}\right)<0.$$

That is, the posterior expectation of one of the contrasts associated with k^* has changed sign.

Now show that

$$\tilde{k}^* \neq k^* \quad \Longleftarrow \quad \widetilde{\mathbb{E}}\left(d_{bk^*}\right)<0 \text{ for at least one } b \in \{1,\ldots,K\}\setminus k^*.$$

As k^* is optimal under the original biased data, the posterior expectations of contrasts associated with k^* satisfy

$$\mathbb{E}_{d|y}\left(d_{ak^*}\right)=\mathbb{E}_{d|y}\left(d_{k^*}\right)-\mathbb{E}_{d|y}\left(d_{a}\right)>0 \quad \forall a \in \{1,\ldots,K\}\setminus k^*,$$

since $\mathbb{E}_{d|y}\left(d_{k^*}\right)=\max\limits_{a=1,\ldots,K}\mathbb{E}_{d|y}\left(d_{a}\right)$.

Now suppose we bias-adjust the data $\tilde{y}(\beta)$ enough that at least one of the posterior expectations of the contrasts associated with k^* changes sign; the set $B=\left\{b:\widetilde{\mathbb{E}}\left(d_{bk^*}\right)<0\right\}\subseteq\{1,\ldots,K-1\}$ is not empty.

We have then

$$\begin{aligned}\widetilde{\mathbb{E}}\left(d_{bk^*}\right)&=\widetilde{\mathbb{E}}\left(d_{k^*}\right)-\widetilde{\mathbb{E}}\left(d_{b}\right)<0 && \forall b \in B\\\widetilde{\mathbb{E}}\left(d_{bk^*}\right)&=\widetilde{\mathbb{E}}\left(d_{k^*}\right)-\widetilde{\mathbb{E}}\left(d_{b}\right)\geq 0 && \forall b \notin B\cup\{k^*\}\end{aligned}$$

Thus

$$\begin{aligned}\widetilde{\mathbb{E}}\left(d_{k^*}\right)&<\widetilde{\mathbb{E}}\left(d_{b}\right) && \forall b \in B\\\widetilde{\mathbb{E}}\left(d_{k^*}\right)&\geq\widetilde{\mathbb{E}}\left(d_{b}\right) && \forall b \notin B\cup\{k^*\}\end{aligned}$$

And so choosing $\tilde{k}^*=\arg\max\limits_{b\in B}\widetilde{\mathbb{E}}\left(d_{b}\right)$ we have that

$$\widetilde{\mathbb{E}}\left(d_{\tilde{k}^*}\right)\geq\widetilde{\mathbb{E}}\left(d_{b}\right)>\widetilde{\mathbb{E}}\left(d_{k^*}\right)\geq\widetilde{\mathbb{E}}\left(d_{b'}\right) \quad \forall \quad b \in B, \quad b' \notin B\cup\{k^*\}.$$

Therefore $\tilde{k}^*$ is the new optimal treatment and $\tilde{k}^* \neq k^*$ since $k^* \notin B$ and $\tilde{k}^* \in B$ by definition.

Proof of Lemma 1b

Let us choose a value of the bias adjustment $\boldsymbol{\beta}$ so that the posterior expectation of only one contrast changes sign, say $\widetilde{\mathbb{E}}\left(d_{bk^*}\right) < 0$.

Following on from the proof of part a) we see that $B = \{b\}$, a singleton set. Since $\tilde{k}^* \in B$ by definition, we therefore know that $\tilde{k}^* = b$.

A.2   Theorem: Posterior distribution for FE model

Given the following model with multivariate Normal likelihood and prior distribution

$$\boldsymbol{y} \mid \boldsymbol{d} \sim \mathrm{N}\left(\boldsymbol{Xd}, \boldsymbol{V}\right)$$
$$\boldsymbol{d} \sim \mathrm{N}\left(\boldsymbol{d}_0, \boldsymbol{\Sigma}_d\right)$$

Then the prior distribution and likelihood are conjugate and the posterior distribution is also multivariate Normal

$$\boldsymbol{d} \mid \boldsymbol{y} \sim \mathrm{N}\left(\boldsymbol{\Sigma}_n\left(\boldsymbol{\Sigma}_d^{-1}\boldsymbol{d}_0 + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}\right), \boldsymbol{\Sigma}_n\right),$$

where

$$\boldsymbol{\Sigma}_n^{-1} = \boldsymbol{\Sigma}_d^{-1} + \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}.$$

Proof

Using Bayes rule (Bayes 1763), the posterior distribution is

$$
\begin{aligned}
p\left(\boldsymbol{d} \mid \boldsymbol{y}\right) &\propto p\left(\boldsymbol{y} \mid \boldsymbol{d}\right) p\left(\boldsymbol{d}\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\left(\boldsymbol{y} - \boldsymbol{Xd}\right)^T \boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{Xd}\right) + \left(\boldsymbol{d} - \boldsymbol{d}_0\right)^T \boldsymbol{\Sigma}_d^{-1}\left(\boldsymbol{d} - \boldsymbol{d}_0\right)\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\boldsymbol{y}^T\boldsymbol{V}^{-1}\boldsymbol{y} - \boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{V}^{-1}\boldsymbol{Xd} + \boldsymbol{d}^T\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{Xd} \right. \right. \\
&\qquad\qquad \left. \left. + \boldsymbol{d}^T\boldsymbol{\Sigma}_d^{-1}\boldsymbol{d} - \boldsymbol{d}_0^T\boldsymbol{\Sigma}_d^{-1}\boldsymbol{d} - \boldsymbol{d}^T\boldsymbol{\Sigma}_d^{-1}\boldsymbol{d}_0 + \boldsymbol{d}_0^T\boldsymbol{\Sigma}_d^{-1}\boldsymbol{d}_0\right)\right)
\end{aligned}
$$

Then since V and Σ are symmetric, and noting proportionality with d,

$$\propto \exp\left(-\frac{1}{2}\left(d^T\left(\Sigma_d^{-1}+X^TV^{-1}X\right)d-2d^T\left(X^TV^{-1}y+\Sigma_d^{-1}d_0\right)\right)\right)$$

And completing the square,

$$\propto \exp\left(-\frac{1}{2}\left(d-d_n\right)^T\Sigma_n^{-1}\left(d-d_n\right)\right)$$

which we recognise as the multivariate Normal as required. See also Gelman *et al.* (2013, p. 71).

A.3   Derivation of bias adjustment thresholds for the basic FE model

We derive thresholds for the basic FE model described in section 2.3.2 by rearranging the expression for the posterior mean from equation (8). Under the bias adjusted data, this becomes

$$\begin{aligned}\widetilde{\mathbb{E}}(d) &= \Sigma_n\left(\Sigma_d^{-1}d_0+X^TV^{-1}\left(y+\beta_m\right)\right)\\ &= \mathbb{E}(d)+\Sigma_nX^TV^{-1}\beta_m\end{aligned} \tag{A.1}$$

Notice here that we can think of $\Sigma_nX^TV^{-1}$ as an influence matrix, determining the change in the posterior mean resulting from a change in the data vector by β_m. Let us denote this influence matrix as H, and rewrite equation (A.1) as

$$\widetilde{\mathbb{E}}(d) = \mathbb{E}(d)+H\beta_m. \tag{A.2}$$

Consequently from Lemma 1 (appendix A.1), we must solve $K-1$ equations of the form

$$\widetilde{\mathbb{E}}\left(d_{ak^*}\right) = 0, \quad \forall a \in \{1,\ldots,K\}\setminus k^* \tag{A.3}$$

to determine the point at which the expected value changes sign, and then take the smallest positive and negative solutions to be the positive and negative threshold values. Using (A.2) we can find solutions $u_{ak^*,m}$ to equation (A.3) expectation by rearranging the following equations:

$$\begin{aligned}0 = \widetilde{\mathbb{E}}\left(d_{ak^*}\right) &= \widetilde{\mathbb{E}}\left(d_{k^*}\right)-\widetilde{\mathbb{E}}\left(d_a\right)\\ &= \left[\widetilde{\mathbb{E}}(d)\right]_{k^*-1}-\left[\widetilde{\mathbb{E}}(d)\right]_{a-1}\\ &= \mathbb{E}\left(d_{ak^*}\right)+\left(\left[H\right]_{k^*-1,m}-\left[H\right]_{a-1,m}\right)u_{ak^*,m}\end{aligned} \tag{A.4}$$

for $a \in \{2,\ldots,K\}\setminus k^*$ when $k^* \neq 1$, and

4

$$0 = \widetilde{\mathbb{E}}\left(d_{1k^*}\right) = \widetilde{\mathbb{E}}\left(d_{k^*}\right)$$
$$= \mathbb{E}\left(d_{k^*}\right) + \left[\boldsymbol{H}\right]_{k^*-1,m} u_{1k^*,m} \tag{A.5}$$

for $a = 1$ when $k^* \neq 1$. If $k^* = 1$ then the equations are of the form

$$0 = \widetilde{\mathbb{E}}\left(d_{1a}\right) = \widetilde{\mathbb{E}}\left(d_a\right)$$
$$= \mathbb{E}\left(d_a\right) + \left[\boldsymbol{H}\right]_{a-1,m} u_{1a,m} \tag{A.6}$$

for $a \in \{2,\ldots,K\}$.

Now we simply re-arrange equations (A.4), (A.5), and (A.6) to arrive at the solutions:

For $k^* \neq 1$

$$u_{ak^*,m} = \frac{-\mathbb{E}\left(d_{ak^*}\right)}{\left[\boldsymbol{H}\right]_{k^*-1,m} - \left[\boldsymbol{H}\right]_{a-1,m}}, \qquad \text{for } a \in \{2,\ldots,K\} \setminus k^*$$

$$u_{1k^*,m} = \frac{-\mathbb{E}\left(d_{1k^*}\right)}{\left[\boldsymbol{H}\right]_{k^*-1,m}}, \qquad \text{for } a = 1$$

For $k^* = 1$

$$u_{1a,m} = \frac{-\mathbb{E}\left(d_{1a}\right)}{\left[\boldsymbol{H}\right]_{a-1,m}}, \qquad \text{for } a \in \{2,\ldots,K\}$$

The positive and negative threshold values $\beta_m^{+\text{thresh}}$ and $\beta_m^{-\text{thresh}}$ for a data point y_m are then

$$\beta_m^{+\text{thresh}} = u_{bk^*,m} \text{ where } b = \underset{a \in \{1,\ldots,K\} \setminus k^*}{\operatorname{argmin}} \left\{ u_{ak^*,m} : u_{ak^*,m} > 0 \right\}$$
$$\beta_m^{-\text{thresh}} = u_{bk^*,m} \text{ where } b = \underset{a \in \{1,\ldots,K\} \setminus k^*}{\operatorname{argmax}} \left\{ u_{ak^*,m} : u_{ak^*,m} < 0 \right\} \tag{A.7}$$

Furthermore, from Lemma 1 (appendix A.1) and equation (A.7) we know that the new optimal treatment at the threshold is $\tilde{k}^* = b$. Note that, although it is in theory possible for there to be two or more treatment effects that are both exactly maximal at the threshold value of β, this happens with probability zero.

A.4 Theorem: Posterior distribution for RE model

Given the RE model

Prior: $\quad\quad d \sim \mathrm{N}\left(d_0, \Sigma_d\right)$

Likelihood: $\quad y \mid \delta \sim \mathrm{N}\left(\delta, V\right)$

RE Model: $\quad \delta \mid d \sim \mathrm{N}\left(Xd, \Sigma_{\tau^2}\right)$

where the between studies covariance matrix Σ_{τ^2} is known and fixed, the joint posterior distribution for d and δ is

$$\binom{d}{\delta} \middle| y \sim \mathrm{N}\left(\Sigma_n \binom{\Sigma_d^{-1} d_0}{V^{-1} y}, \Sigma_n\right)$$

where

$$\Sigma_n = \begin{pmatrix} X^T \Sigma_{\tau^2}^{-1} X + \Sigma_d^{-1} & -X^T \Sigma_{\tau^2}^{-1} \\ -\Sigma_{\tau^2}^{-1} X & V^{-1} + \Sigma_{\tau^2}^{-1} \end{pmatrix}^{-1}.$$

Proof

Similarly to the proof for the posterior distribution of the FE model, the posterior distribution is

$$p(d, \delta \mid y) \propto p(y \mid \delta) p(\delta \mid d) p(d)$$

$$\propto \exp\left\{-\frac{1}{2}\left[(y-\delta)^T V^{-1}(y-\delta) + (\delta - Xd)^T \Sigma_{\tau^2}^{-1}(\delta - Xd) + (d-d_0)^T \Sigma_d^{-1}(d-d_0)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\binom{d}{\delta}^T \begin{pmatrix} X^T \Sigma_{\tau^2}^{-1} X + \Sigma_d^{-1} & -X^T \Sigma_{\tau^2}^{-1} \\ -\Sigma_{\tau^2}^{-1} X & V^{-1} + \Sigma_{\tau^2}^{-1} \end{pmatrix}\binom{d}{\delta} - 2\binom{d}{\delta}\binom{\Sigma_d^{-1} d_0}{V^{-1} y}\right]\right\}$$

which we recognise as the multivariate Normal as required. See also Gelman *et al.* (2013, p. 582).

A.5   Bias adjustment thresholds for the basic RE model

The joint posterior distribution of d and δ (conditional on Σ_{τ^2}) is given in equation (9)(with proof in appendix A.4). We partition the joint covariance matrix as

$$\Sigma_n = \begin{pmatrix} X^T \Sigma_{\tau^2}^{-1} X + \Sigma_d^{-1} & -X^T \Sigma_{\tau^2}^{-1} \\ -\Sigma_{\tau^2}^{-1} X & V^{-1} + \Sigma_{\tau^2}^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} A_* & B_* \\ B_*^T & C_* \end{pmatrix}.$$

Under bias-adjusted data $\tilde{y} = y + \beta_m$, the joint posterior mean of d and δ becomes

$$\widetilde{\mathbb{E}}(d) = \left[\Sigma_n \begin{pmatrix} \Sigma_d^{-1} d_0 \\ V^{-1}(y + \beta_m) \end{pmatrix} \right]_{1:K-1}$$
$$= A_* \Sigma_d^{-1} d_0 + B_* V^{-1} y + B_* V^{-1} \beta_m$$
$$= \mathbb{E}(d) + B_* V^{-1} \beta_m$$

where we make use of a subscript colon notation to indicate subvectors (i.e. for a general vector x of length l, we define $x_{p:q} = (x_p, \ldots, x_q)^T$ for $1 \leq p < q \leq l$). Then following the same arguments as the basic FE case (given in appendix A.3), the influence matrix is $H = B_* V^{-1}$ and the thresholds are given by equations (5) and (6).

A.6 Bias adjustment thresholds for the extended FE model

The extended FE model with parameter vector $\gamma = \begin{pmatrix} d \\ \mu \end{pmatrix}$ is written

$$y \mid \gamma \sim \mathrm{N}(X\gamma, V)$$
$$\gamma \sim \mathrm{N}(\gamma_0, \Sigma_\gamma)$$

with posterior distribution

$$\gamma \mid y \sim \mathrm{N}\left(\Sigma_n \left(\Sigma_\gamma^{-1} \gamma_0 + X^T V^{-1} y \right), \Sigma_n \right),$$

where the posterior covariance matrix is $\Sigma_n = \left(\Sigma_\gamma^{-1} + X^T V^{-1} X \right)^{-1}$.

Then the joint posterior expectation of the basic treatment effect parameters d under the bias-adjusted data is

$$\widetilde{\mathbb{E}}(d) = \left[\Sigma_n \left(\Sigma_\gamma^{-1} \gamma_0 + X^T V^{-1} (y + \beta_m) \right) \right]_{1:K-1}$$
$$= \mathbb{E}(d) + \left[\Sigma_n X^T V^{-1} \beta_m \right]_{1:K-1}$$
$$= \mathbb{E}(d) + \left[\Sigma_n X^T V^{-1} \right]_{\text{rows } 1:K-1} \beta_m$$

where we make use of a subscript colon notation to denote subvectors (as in appendix A.5) and similarly to subset rows of a matrix.

7

From here we continue exactly as in the basic case (appendix A.3) to derive bias adjustment thresholds using the threshold equations (5) and (6), where the influence matrix is now

$$H = \left[\boldsymbol{\Sigma}_n \boldsymbol{X}^T \boldsymbol{V}^{-1} \right]_{\text{rows } 1:K-1}.$$

A.7   Bias adjustment thresholds for the extended RE model

We extend the RE model to include additional parameters $\boldsymbol{\mu}$, which are given a Normal prior distribution, and have an associated design matrix $\boldsymbol{M}$. We also allow for random effects terms to only be included for certain data points, using a design matrix $\boldsymbol{L}$, for example in the case of absolute effect measure data where only non-reference arms have a random effect term (Dias *et al.* 2013a); in most other cases the matrix $\boldsymbol{L}$ is the identity matrix. The extended RE model is written as

$$\begin{aligned} \text{Priors:} \qquad & \boldsymbol{d} \sim \mathrm{N}\left(\boldsymbol{d}_0, \boldsymbol{\Sigma}_d\right), \qquad \boldsymbol{\mu} \sim \mathrm{N}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_\mu\right) \\ \text{Likelihood:} \quad & \boldsymbol{y} \mid \boldsymbol{\delta}, \boldsymbol{\mu} \sim \mathrm{N}\left(\boldsymbol{L}\boldsymbol{\delta} + \boldsymbol{M}\boldsymbol{\mu}, \boldsymbol{V}\right) \\ \text{RE Model:} \quad & \boldsymbol{\delta} \mid \boldsymbol{d}, \tau^2 \sim \mathrm{N}\left(\boldsymbol{X}\boldsymbol{d}, \boldsymbol{\Sigma}_{\tau^2}\right) \end{aligned}$$

The posterior distribution is then

$$\begin{pmatrix} \boldsymbol{d} \\ \boldsymbol{\delta} \\ \boldsymbol{\mu} \end{pmatrix} \Bigg\| \, \boldsymbol{y}, \tau^2 \sim \mathrm{N}\left(\boldsymbol{\Sigma}_n \begin{pmatrix} \boldsymbol{\Sigma}_d^{-1} \boldsymbol{d}_0 \\ \boldsymbol{L}^T \boldsymbol{V}^{-1} \boldsymbol{y} \\ \boldsymbol{\Sigma}_\mu^{-1} \boldsymbol{\mu}_0 + \boldsymbol{M}^T \boldsymbol{V}^{-1} \boldsymbol{y} \end{pmatrix}, \boldsymbol{\Sigma}_n \right)$$

where

$$\boldsymbol{\Sigma}_n = \begin{pmatrix} \boldsymbol{X}^T \boldsymbol{\Sigma}_{\tau^2}^{-1} \boldsymbol{X} + \boldsymbol{\Sigma}_d^{-1} & -\boldsymbol{X}^T \boldsymbol{\Sigma}_{\tau^2}^{-1} & \boldsymbol{0} \\ -\boldsymbol{\Sigma}_{\tau^2}^{-1} \boldsymbol{X} & \boldsymbol{\Sigma}_{\tau^2}^{-1} + \boldsymbol{L}^T \boldsymbol{V}^{-1} \boldsymbol{L} & -\boldsymbol{L}^T \boldsymbol{V}^{-1} \boldsymbol{M} \\ \boldsymbol{0} & -\boldsymbol{M}^T \boldsymbol{V}^{-1} \boldsymbol{L} & \boldsymbol{\Sigma}_\mu^{-1} + \boldsymbol{M}^T \boldsymbol{V}^{-1} \boldsymbol{M} \end{pmatrix}^{-1} =: \begin{pmatrix} \boldsymbol{A}_* & \boldsymbol{B}_* & \boldsymbol{D}_* \\ \boldsymbol{B}_*^T & \boldsymbol{C}_* & \boldsymbol{E}_* \\ \boldsymbol{D}_*^T & \boldsymbol{E}_*^T & \boldsymbol{F}_* \end{pmatrix} \quad \text{(A.8)}$$

noting that, as in the basic RE case, we can partition the posterior covariance matrix into blocks. Proof of the posterior distribution follows closely that of the basic RE model in appendix A.4.

The joint posterior expectation of the basic treatment effect parameters $\boldsymbol{d}$ under the bias-adjusted data $\tilde{\boldsymbol{y}}$ is then

8

$$\widetilde{\mathbb{E}}(d) = \left[\Sigma_n \begin{pmatrix} \Sigma_d^{-1} d_0 \\ L^T V^{-1} (y + \beta_m) \\ \Sigma_\mu^{-1} \mu_0 + M^T V^{-1} (y + \beta_m) \end{pmatrix} \right]_{1:K-1}$$

$$= \mathbb{E}(d) + \left(B_* L^T + D_* M^T \right) V^{-1} \beta_m$$

Following the same arguments as before (appendix A.2), we see that the thresholds are given by equations (5) and (6) where the influence matrix is now $H = \left(B_* L^T + D_* M^T \right) V^{-1}$.

A.8 Bias adjustment thresholds for RE models including class effects

Further extending the RE model of section 2.3.4 to include class effects, we have

Priors: $\qquad z \sim N(z_0, \Sigma_z), \qquad \mu \sim N(\mu_0, \Sigma_\mu)$

Likelihood: $\qquad y \mid \delta, \mu \sim N(L\delta + M\mu, V)$

RE Model: $\qquad \delta \mid d, \tau^2 \sim N(Xd, \Sigma_{\tau^2})$

Class Effects: $\qquad d \mid z \sim N(Zz, \Sigma_d)$

where z is the vector of class effect parameters and the matrix Z is a design matrix for the treatment clasfses. Each row of Z corresponds to a treatment, which is assigned a class by a 1 in the corresponding column and zeros elsewhere in the row. As before, we must work with the posterior distribution assuming that Σ_{τ^2} is known, fixed, and invariant to bias adjustments; furthermore, we now assume that the between-treatment covariance matrix Σ_d is also known, fixed, and invariant to bias adjustments. Both of these assumptions can be tested using sensitivity analyses.

The posterior distribution is then

$$\begin{pmatrix} d \\ \delta \\ \mu \\ z \end{pmatrix} \Bigg| y, \tau^2, \Sigma_d \sim N\left(\Sigma_n \begin{pmatrix} 0 \\ L^T V^{-1} y \\ \Sigma_\mu^{-1} \mu_0 + M^T V^{-1} y \\ \Sigma_z^{-1} z_0 \end{pmatrix}, \Sigma_n \right)$$

where

9

$$\Sigma_n = \begin{pmatrix} X^T \Sigma_{\tau^2}^{-1} X + \Sigma_d^{-1} & -X^T \Sigma_{\tau^2}^{-1} & 0 & -\Sigma_d^{-1} Z \\ -\Sigma_{\tau^2}^{-1} X & \Sigma_{\tau^2}^{-1} + L^T V^{-1} L & -L^T V^{-1} M & 0 \\ 0 & -M^T V^{-1} L & \Sigma_\mu^{-1} + M^T V^{-1} M & 0 \\ -Z^T \Sigma_d^{-1} & 0 & 0 & \Sigma_d^{-1} + Z^T \Sigma_z^{-1} Z \end{pmatrix}^{-1} =: \begin{pmatrix} A_* & B_* & D_* & G_* \\ B_*^T & C_* & E_* & H_* \\ D_*^T & E_*^T & F_* & I_* \\ G_*^T & H_*^T & I_*^T & J_* \end{pmatrix}$$

noting that, as in the basic RE case, we can partition the posterior covariance matrix into blocks. Proof of the posterior distribution follows closely that of the basic RE model in appendix A.4.

The joint posterior expectation of the basic treatment effect parameters d under the bias-adjusted data $\tilde{y}$ is then

$$\widetilde{\mathbb{E}}(d) = \left[\Sigma_n \begin{pmatrix} 0 \\ L^T V^{-1} y \\ \Sigma_\mu^{-1} \mu_0 + M^T V^{-1} y \\ \Sigma_z^{-1} z_0 \end{pmatrix} \right]_{1:K-1}$$

$$= \mathbb{E}(d) + \left(B_* L^T + D_* M^T \right) V^{-1} \beta_m$$

Following the same arguments as before (appendix A.3), we see that the thresholds are given by equations (5) and (6) where the influence matrix is now $H = \left(B_* L^T + D_* M^T \right) V^{-1}$. Note that the influence matrix is identical to that in the extended RE case; we need do nothing different to the extended RE case despite the presence of class effects.

A.9 Forming an approximate dataset for contrast-level analysis

Suppose that we have the posterior summaries available from a one-stage Bayesian analysis for basic treatment effect parameters $d = (d_2, \ldots, d_K)^T$, with joint posterior distribution given by $d \mid y \sim \mathrm{N}(\eta, \Sigma)$. We wish to obtain a bias adjustment threshold for the combined data on each contrast d_{ab} for which there is direct evidence, $d_{ab} \in \mathcal{D}_{\mathrm{dir}}$, where $\mathcal{D}_{\mathrm{dir}}$ is the set of contrasts with direct evidence. We may think of the posterior distribution as arising from a NMA on $|\mathcal{D}_{\mathrm{dir}}| \le K(K-1)/2$ data points y_{ab} with variances v_{ab}. These data points are hypothetical and are treated as independent of one another, each representing the combined evidence on a single contrast.

Assume a multivariate Normal likelihood for these hypothetical data points: $\boldsymbol{y}\,|\,\boldsymbol{d}\sim\mathrm{N}\left(\boldsymbol{Xd},\boldsymbol{V}\right)$ where $\boldsymbol{y}=\left(y_{ab}:d_{ab}\in\mathcal{D}_{\mathrm{dir}}\right)^{T}$, $\boldsymbol{V}=\mathrm{diag}\left(v_{ab}:d_{ab}\in\mathcal{D}_{\mathrm{dir}}\right)$, and $\boldsymbol{X}$ is a design matrix. Giving $\boldsymbol{d}$ the conjugate prior distribution $\boldsymbol{d}\sim\mathrm{N}\left(\boldsymbol{d}_{0},\boldsymbol{\Sigma}_{d}\right)$ we have that the reconstructed posterior distribution based on the hypothetical data points is

$$\boldsymbol{d}\,|\,\boldsymbol{y}\sim\mathrm{N}\left(\left(\boldsymbol{\Sigma}_{d}^{-1}+\boldsymbol{X}^{T}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{\Sigma}_{d}^{-1}\boldsymbol{d}_{0}+\boldsymbol{X}^{T}\boldsymbol{V}^{-1}\boldsymbol{y}\right),\left(\boldsymbol{\Sigma}_{d}^{-1}+\boldsymbol{X}^{T}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\right).$$

Equating this with the true joint posterior distribution $\mathrm{N}\left(\boldsymbol{\eta},\boldsymbol{\Sigma}\right)$ reported by the original NMA, we see that $\boldsymbol{\eta}=\left(\boldsymbol{\Sigma}_{d}^{-1}+\boldsymbol{X}^{T}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{\Sigma}_{d}^{-1}\boldsymbol{d}_{0}+\boldsymbol{X}^{T}\boldsymbol{V}^{-1}\boldsymbol{y}\right)^{-1}$ and

$$\boldsymbol{\Sigma}=\left(\boldsymbol{\Sigma}_{d}^{-1}+\boldsymbol{X}^{T}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}. \tag{A.9}$$

In order to calculate bias adjustment thresholds using the results of section 2.3.2, we require the influence matrix $\boldsymbol{H}=\boldsymbol{\Sigma}\boldsymbol{X}^{T}\boldsymbol{V}^{-1}$. However, we cannot immediately evaluate $\boldsymbol{H}$ as the hypothetical likelihood precision matrix $\boldsymbol{V}^{-1}$ is required. Instead, exploiting the structure of the design matrix $\boldsymbol{X}$, we obtain $\boldsymbol{V}^{-1}$ (or an approximation) using equation (A.9) which we then use to evaluate the influence matrix and derive thresholds.

To obtain $\boldsymbol{V}^{-1}$, firstly, note that there are $\left|\mathcal{D}_{\mathrm{dir}}\right|$ unknowns in $\boldsymbol{V}^{-1}$, as this matrix is diagonal with elements $p_{ab}=v_{ab}^{-1}$ by assumption. Secondly, note that $\boldsymbol{X}$ is structured with a row for each data point y_{ab}, with a 1 in column $b-1$ and, provided $k>1$, with a -1 in column $a-1$; this structure results in the symmetric matrix

$$\boldsymbol{X}^{T}\boldsymbol{V}^{-1}\boldsymbol{X}=\begin{pmatrix} p_{12}+\sum_{b=3}^{K}p_{2b} & -p_{23} & -p_{24} & \cdots & -p_{2K} \\ & p_{13}+p_{23}+\sum_{b=4}^{K}p_{3b} & -p_{34} & \cdots & -p_{3K} \\ & & p_{14}+\sum_{a=2}^{3}p_{a4}+\sum_{b=5}^{K}p_{4b} & \ddots & \vdots \\ & & & \ddots & -p_{(K-1)K} \\ & & & & p_{1K}+\sum_{a=2}^{K-1}p_{aK} \end{pmatrix}$$

where $p_{ab} = 0$ if $d_{ab} \notin \mathcal{D}_{\text{dir}}$. We then rearrange equation (A.9) to equate $X^T V^{-1} X = \Sigma^{-1} - \Sigma_d^{-1}$, resulting in $K(K-1)/2$ equations to solve in $|\mathcal{D}_{\text{dir}}|$ unknowns as the matrices are symmetric.

When the evidence network is complete $|\mathcal{D}_{\text{dir}}| = K(K-1)/2$ and the number of equations equals the number of unknowns. We therefore easily find a unique solution for each p_{ab}.

When there are some treatments with no direct evidence comparing them $|\mathcal{D}_{\text{dir}}| < K(K-1)/2$ and we have fewer unknowns than equations; the system of equations is overdetermined. We proceed to approximate V^{-1} using non-negative least squares (Lawson and Hanson 1995).

To examine how well the hypothetical likelihood recreates the posterior distribution we suggest evaluating the Kullback-Leibler (KL) divergence (Kullback and Leibler 1951) of the reconstructed posterior distribution with covariance matrix $\hat{\Sigma} := \left(\Sigma_d^{-1} + X^T V^{-1} X \right)^{-1}$ from the true posterior distribution with covariance matrix Σ; the KL divergence between two multivariate Normal distributions with the same mean is given by

$$\frac{1}{2} \left(\log \frac{|\hat{\Sigma}|}{|\Sigma|} + \text{tr}\left(\hat{\Sigma}^{-1} \Sigma \right) - (K-1) \right) .$$

The KL divergence is always non-negative and smaller values are desirable, indicating that the approximation reconstructs the posterior distribution well. Noting that the KL divergence is equivalent to the expected value of a log Bayes factor, we refer to Kass and Raftery (1995) for interpretation: for example, a KL divergence less than 1 is negligible, and values greater than 3 may be considered large. If evaluation of the KL divergence suggests a bad approximation, the contrast-level threshold analysis may give inaccurate results.

Once the hypothetical likelihood covariance matrix has been reconstructed, the thresholds are then evaluated as before using equations (5) and (6) with the influence matrix $H = \Sigma X^T V^{-1}$. Note that we cannot re-evaluate the posterior means under the bias-adjusted data to obtain $\tilde{k}^*$ as we do not have the hypothetical data, but we can use the result of Lemma 1 (appendix A.1) to efficiently obtain the new optimal treatment.

12

A.10 Computation

The threshold methods described in sections 2.3, 2.4, and 2.5 are implemented in R (version 3.0.1 or later) (R Core Team 2016), where we take advantage of the computational efficiency of vector and matrix operations to derive every threshold at the same time instead of looping over all data points and all treatments. In section A.10.1 we briefly demonstrate how such an approach is formulated mathematically. The resulting computation times are almost instantaneous; the examples presented in section 3 of this paper each took less than 0.01 seconds to calculate thresholds on a standard desktop PC. An R package *nmathresh* is provided in the supplementary material, containing all of the functions described below plus the data required for the examples in section 3. The package also contains a vignette which details the exact commands used for the examples.

The R function *nma_thresh* implements the study-level threshold method for FE models (sections 2.3.2 and 2.3.4) and RE models (sections 2.3.3, 2.3.4, and 2.3.5), depending on the value of the *nmatype* parameter (either *"fixed"* or *"random"* respectively). This function takes the posterior mean of the relative treatment effects, the likelihood and posterior covariance matrices, and the design matrix/matrices as inputs. The R function *nma_thresh* is also used to perform contrast-level analysis (section 2.4). In this scenario the hypothetical likelihood covariance matrix is constructed from the prior and posterior covariance matrices and the design matrix using the R function *recon_vcov*, either exactly or using non-negative least squares (NNLS) (Lawson and Hanson 1995) via the function *nnls* from the package *nnls* (Mullen and van Stokkum 2012). The R function *thresh_forest* takes the results from the NMA and threshold analysis and presents them graphically on a forest plot.

A.10.1 Mathematical derivation

Here we briefly describe mathematically the derivation of bias adjustment thresholds in a vectorised manner, allowing for highly efficient computation that does not rely on looping.

First, define a matrix U which contains the elements $u_{ak^*,m}$:

$$
U = \begin{pmatrix}
u_{1k^*,1} & \cdots & u_{1k^*,N} \\
\vdots & & \vdots \\
u_{(k^*-1)k^*,1} & \cdots & u_{(k^*-1)k^*,N} \\
u_{(k^*+1)k^*,1} & \cdots & u_{(k^*+1)k^*,N} \\
\vdots & & \vdots \\
u_{Kk^*,1} & \cdots & u_{Kk^*,N}
\end{pmatrix}
$$

where each row corresponds to a contrast d_{ak^*} and each column to a data point y_m.

Recall that each $u_{ak^*,m}$ is of the form $u_{ak^*,m} = -\mathbb{E}(d_{ak^*})\big/\big([H]_{k^*-1,m} - [H]_{a-1,m}\big)$ where H is the influence matrix of the data y on the joint posterior mean of d, and so each $u_{ak^*,m}$ consists of a numerator and denominator. Next construct the matrix D which describes the linear combinations of elements of the influence matrix found in the denominator of each $u_{ak^*,m}$ element of B:

$$
D = \begin{pmatrix}
0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \\
-1 & \ddots & \vdots & 1 & \vdots & & & \vdots \\
0 & \ddots & 0 & \vdots & \vdots & & & \vdots \\
\vdots & \ddots & -1 & 1 & 0 & & & \vdots \\
\vdots & & 0 & 1 & -1 & \ddots & & \vdots \\
\vdots & & \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\
\vdots & & \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & 1 & 0 & \cdots & 0 & -1
\end{pmatrix}
$$

where the (k^*-1)-th column of D is filled with 1s. We can now write the matrix U using the formula

$$
U = \Big[-\operatorname{diag}\big(\mathbb{E}(d_{ak^*}) : a \in \{1,\ldots,K\} \setminus k^*\big)\Big]\big[1/(DH)\big] \tag{A.10}
$$

where $\operatorname{diag}\big(\mathbb{E}(d_{ak^*}) : a \in \{1,\ldots,K\} \setminus k^*\big)$ denotes the diagonal matrix with elements being the posterior expected values of contrasts associated with k^*, and $1/(DH)$ denotes division performed elementwise, that is $\big[1/(DH)\big]_{i,j} = 1/[DH]_{i,j}$. Threshold values and invariant intervals for each data point y_m are then simply found by examining the m-th column of U from equation (A.10).

14

A.11 Thrombolytics Example: Formulating the NNLS problem for the contrast-level case

The design matrix for the hypothetical data is

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix}.$$

We need now to reconstruct the hypothetical likelihood covariance matrix V. Letting the likelihood precision matrix V^{-1} have diagonal elements $p_{12}, p_{13}, p_{14}, p_{15}, p_{34}, p_{35}, p_{36}$, we see that

$$X^T V^{-1} X = \begin{pmatrix} p_{12} & 0 & 0 & 0 & 0 \\ & p_{13} + p_{34} + p_{35} + p_{36} & -p_{34} & -p_{35} & -p_{36} \\ & & p_{14} + p_{34} & 0 & 0 \\ & & & p_{15} + p_{35} & 0 \\ & & & & p_{36} \end{pmatrix} \qquad \text{(A.11)}$$

then using the fact that $\Sigma_n^{-1} = \Sigma_d^{-1} + X^T V^{-1} X$ we equate (A.11) with

$$\Sigma_n^{-1} - \Sigma_d^{-1} = \begin{pmatrix} q_{22} & q_{23} & q_{24} & q_{25} & q_{26} \\ & q_{33} & q_{34} & q_{35} & q_{36} \\ & & q_{44} & q_{45} & q_{46} \\ & & & q_{55} & q_{56} \\ & & & & q_{66} \end{pmatrix} \qquad \text{(A.12)}$$

Thus we have a system of $K(K-1)/2 = 15$ equations with $\text{length}(y) = 7$ unknowns to solve; an overdetermined system. We can write the system of equations (A.11) and (A.12) in the form $Rp = q$, where

$$p = \left(p_{12}, p_{13}, p_{14}, p_{15}, p_{34}, p_{35}, p_{36} \right)^T$$
$$q = \left(q_{12}, q_{13}, q_{14}, q_{15}, q_{16}, q_{23}, q_{24}, q_{25}, q_{26}, q_{34}, q_{35}, q_{36}, q_{45}, q_{46}, q_{56} \right)^T$$

and

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The NNLS problem is then

$$\text{Minimise} \qquad \|Rp - q\|^2$$
$$\text{Subject to} \qquad p \geq 0$$

A.12 Social Anxiety Example: Treatment Codes

Table A1: Social Anxiety treatment codes and classes.

Code	Treatment	Class
1	Wait List	Wait List
2	Pill Placebo	Pill Placebo
3	Psychological Placebo	Psychological Placebo
4	Exercise Promotion	Exercise
5	Book	Self-help without support
6	Internet	
7	Book	Self-help with support
8	Internet	
9	Pregabalin	Anticonvulsants
10	Levetiracetam	
11	Gabapentin	
12	Mirtazapine	Noradrenaline and selective serotonin antagonists
13	Sertraline	SSRIs
14	Citalopram	
15	Escitalopram	
16	Fluoxetine	
17	Fluvoxamine	
18	Paroxetine	

19	Venlafaxine	
20	Alprazolam	Benzodiazepines
21	Clonazepam	
22	Moclobemide	MAOIs
23	Phenelzine	
24	Exposure in vivo	Exposure in vivo and social skills training
25	Social skills training	
26	Supportive therapy	
27	Mindfullness	Other psychological interventions
28	Interpersonal	
29	Short-term psychodynamic	Short-term psychodynamic
30	CBT	
31	CBT (Heimberg)	CBT - Group
32	CBT (Enhanced)	
33	CT (Shortened)	
34	CBT (Heimberg)	CBT - Individual
35	CBT	
36	CT	
37	Group CBT with fluoxetine	
38	Psychodynamic with clonazepam	
39	Paroxetine with clonazepam	Combined psychological and drug
40	Group CBT with moclobemide	
41	Group CBT with phenelzine	