

Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression[‡]

Ian R. White,^{a,*†} Jessica K. Barrett,^a Dan Jackson^a and Julian P. T. Higgins^{a,b}

Network meta-analysis (multiple treatments meta-analysis, mixed treatment comparisons) attempts to make the best use of a set of studies comparing more than two treatments. However, it is important to assess whether a body of evidence is consistent or inconsistent. Previous work on models for network meta-analysis that allow for heterogeneity between studies has either been restricted to two-arm trials or followed a Bayesian framework. We propose two new frequentist ways to estimate consistency and inconsistency models by expressing them as multivariate random-effects meta-regressions, which can be implemented in some standard software packages. We illustrate the approach using the `mvmeta` package in Stata. Copyright © 2012 John Wiley & Sons, Ltd.

Introduction

Network meta-analysis (NMA) in its standard form makes an assumption of 'consistency' (Lu and Ades, 2006), also called 'coherence' (Lumley, 2002), which means that estimates of treatment effects from direct and indirect evidence are in agreement, subject to the usual variation under the random-effects model for meta-analysis. Consider the case with three treatments A, B and C. If trials comparing B with A, C with A, and C with B estimate parameters δ^{AB} , δ^{AC} and δ^{BC} , respectively, then consistency means that $\delta^{AB} + \delta^{BC} = \delta^{AC}$: the effect of B relative to A, plus the effect of C relative to B, equals the effect of C relative to A.

The consistency assumption is often questionable. For example, the trials that do not include treatment A may include a population for whom A is inappropriate and hence, their results may differ systematically from trials that do include A (Song *et al.*, 2009). Alternatively, trials that include A may be older and therefore use different implementations of other treatments. Wrongly assuming consistency can lead to misleading conclusions, so procedures to investigate the possibility of inconsistency are important.

Some authors have proposed ways to assess consistency directly from fitting consistency models (Dias *et al.*, 2010, Lu *et al.*, 2011). This paper instead assesses consistency by fitting both consistency and inconsistency models. The companion paper (Higgins *et al.*, 2012) reviews the meaning of inconsistency and methods for assessing it (Salanti *et al.*, 2008; Lu and Ades, 2006; Lumley, 2002) and argues that inconsistency is best modelled by a design-by-treatment interaction. We follow that paper in using *design* to refer only to the set of treatments compared in a trial.

Models for consistency and especially inconsistency are complex. Lumley (2002) proposed a frequentist approach that was valid when all trials had only two arms. Lu and Ades (2006) proposed a Bayesian approach that accommodated multi-arm trials and could be implemented in WinBUGS software (Lunn *et al.*, 2000). This paper proposes two new computational strategies for estimating consistency and inconsistency models when the data include multi-arm trials. The new strategies differ from the Bayesian approach in two ways. First, they are frequentist

^aMRC Biostatistics Unit, Cambridge, UK

^bCentre for Reviews and Dissemination, University of York, UK

*Correspondence to: Ian R. White, MRC Biostatistics Unit, Cambridge, UK.

†E-mail: ian.white@mrc-bsu.cam.ac.uk

‡This article is published in *Research Synthesis Methods* as a special issue on *Network Meta-analysis*, edited by Georgia Salanti, University of Ioannina, Greece.

methods, so they aim to speed up computation, avoid sensitivity to the choice of priors and avoid Monte Carlo error. Second, they are two-stage estimation procedures, unlike the one-stage Bayesian procedure. Both strategies formulate the consistency and inconsistency models as multivariate random-effects meta-regressions which can now be easily fitted in standard software (White, 2009, 2011).

The paper is set out as follows. We first introduce data from an NMA of thrombolytic drugs and introduce multivariate meta-analysis and meta-regression in general terms. We then formulate a model for NMA allowing for heterogeneity and inconsistency and describe our new proposals for two simple estimation methods and the alternative Bayesian analysis. We compare the three methods using the thrombolytic drugs data and end with a discussion.

Example: thrombolytic drugs

As an example, we use a dataset consisting of 28 trials comparing eight thrombolytic treatments after acute myocardial infarction: streptokinase (A), accelerated alteplase (B), alteplase (C), streptokinase plus alteplase (D), tenecteplase (E), reteplase (F), urokinase (G) and anti-streptilase (H) (Lu and Ades, 2006). The data are shown in Table 1. Two of the studies are three-arm trials and the rest are two-arm trials. The usefulness of NMA is emphasised by examples such as this, where many pairs of treatments are not compared head to head in any trial. For example, the only way to assess the effect of reteplase (F) relative to urokinase (G) is by utilising the indirect evidence from trials in designs 5 and 6 which compare F and G with A and from designs 9 and 10 which compare F and G with B, although design 12, which compares G with C, also contributes to the indirect evidence. We aim to re-analyse these data to explore and test for consistency, and if consistency is not rejected, to rank the treatments.

Multivariate random-effects meta-regression

Suppose each study yields estimates of p different quantities of interest: in this paper, these will be different between-treatment contrasts, but in other applications of multivariate random-effects meta-regression, they may represent different study outcomes (Jackson *et al.*, 2011). We write these p estimates for study i as the $(1 \times p)$ vector \mathbf{y}_i . We assume we also know \mathbf{S}_i , the 'within-study' $(p \times p)$ variance-covariance matrix of \mathbf{y}_i . The meta-regression model is

$$\mathbf{y}_i \sim N(\boldsymbol{\alpha}\mathbf{X}_i, \boldsymbol{\Sigma} + \mathbf{S}_i) \quad (1)$$

where \mathbf{X}_i is a $(q \times p)$ design matrix containing the covariates (White, 2011). In this model, the unknown parameters are $\boldsymbol{\alpha}$, a $(1 \times q)$ vector of regression coefficients and $\boldsymbol{\Sigma}$, a $(p \times p)$ between-studies variance-covariance matrix, assumed the same for all studies.

Usually, the different quantities in \mathbf{y}_i have separate regressions, with the j th quantity regressed on q_j covariates. Then, the mean of the j th quantity is modelled as $\boldsymbol{\alpha}_j\mathbf{x}_{ij}$, where $\boldsymbol{\alpha}_j$ is a $(1 \times q_j)$ vector and \mathbf{x}_{ij} is a $(q_j \times 1)$ vector. In this case, $q = \sum_{j=1}^p q_j$ and (1) holds with

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_p) \quad (2)$$

and

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1} & 0 & \dots & 0 \\ 0 & \mathbf{x}_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{x}_{ip} \end{pmatrix}. \quad (3)$$

The multivariate meta-analysis model is the special case of (1) with \mathbf{X}_i equal to the $(p \times p)$ identity matrix, so that

$$\mathbf{y}_i \sim N(\boldsymbol{\alpha}, \boldsymbol{\Sigma} + \mathbf{S}_i) \quad (4)$$

and $\boldsymbol{\alpha}$ is a $(1 \times p)$ vector of the overall means of the estimates (White, 2009; Jackson *et al.*, 2011). The standard meta-regression model (Thompson and Sharp, 1999) is the special case of (1) with $p=1$, so that \mathbf{y}_i , \mathbf{S}_i and $\boldsymbol{\Sigma}$ are scalars.

Multivariate random-effects meta-analysis is implemented in Stata as `mvmeta` (White, 2009), which has recently been extended to implement multivariate random-effects meta-regression (White, 2011). Restricted maximum likelihood (REML) estimation is usually used for $\boldsymbol{\Sigma}$ to avoid the negative bias associated with maximum likelihood estimation of variance components. Likelihood-based estimation can straightforwardly accommodate studies in which some elements of \mathbf{y}_i are missing.

NMA model formulation

In this section, we state a model assuming that one of the treatments, which we call treatment A, is included in every trial. This is rarely true in practise and is not true for the thrombolytic drugs data. For the alternative case where the trials have no common treatment, we later describe different modifications and show how they can be expressed by the multivariate meta-regression model (1).

We consider a network including a total of T treatments A, B, C, and so on. Let $d = 1, \dots, D$ index the designs (the sets of treatments compared in a trial). Let there be n_d trials of the d th design, each comparing T_d treatments. Thus, in Table 1, $d = 1$ corresponds to the design comparing A, B and D; this is a three-arm design, so $T_1 = 3$, and only one such trial is present, so $n_1 = 1$.

Salanti *et al.* (2008) compared 'arm-based' models describing arm-specific parameters, such as the log odds, and 'contrast-based' models describing contrasts of arm-specific parameters, such as log odds ratios. We initially use a contrast-based approach.

Let y_{di}^{AJ} be the observed contrast of treatment J ($J = B, C, \dots$) with treatment A in the i th trial ($i = 1$ to n_d) in the d th design ($d = 1$ to D). y_{di}^{AJ} may represent any measure, such as a mean difference, a standardised mean difference, a log risk ratio or a log odds ratio. Our model for the observed data is

$$y_{di}^{AJ} = \delta^{AJ} + \beta_{di}^{AJ} + \omega_d^{AJ} + \varepsilon_{di}^{AJ}, J = B, C, \dots, \quad (5)$$

where δ^{AJ} represents a contrast (a summary effect) between J and A, β_{di}^{AJ} represents heterogeneity in the J–A contrast between studies within designs, ω_d^{AJ} represents inconsistency in the J–A contrast (heterogeneity between designs), and ε_{di}^{AJ} is a within-study error term. Equivalently, in vector notation,

$$\mathbf{y}_{di} = \boldsymbol{\delta} + \boldsymbol{\beta}_{di} + \boldsymbol{\omega}_d + \boldsymbol{\varepsilon}_{di} \quad (6)$$

where $\mathbf{y}_{di} = (y_{di}^{AB}, y_{di}^{AC}, \dots)'$, $\boldsymbol{\delta} = (\delta^{AB}, \delta^{AC}, \dots)'$, $\boldsymbol{\beta}_{di} = (\beta_{di}^{AB}, \beta_{di}^{AC}, \dots)'$, $\boldsymbol{\omega}_d = (\omega_d^{AB}, \omega_d^{AC}, \dots)'$ and $\boldsymbol{\varepsilon}_{di} = (\varepsilon_{di}^{AB}, \varepsilon_{di}^{AC}, \dots)'$. For the within-study error terms, we assume $\boldsymbol{\varepsilon}_{di} \sim N(0, \mathbf{S}_{di})$ where \mathbf{S}_{di} is assumed to be known. The treatment contrasts $\boldsymbol{\delta}$ are regarded as fixed parameters. We consider the specification of the heterogeneity terms $\boldsymbol{\beta}_{di}$ and the inconsistency terms $\boldsymbol{\omega}_d$ below.

For treatments J not used in design d , y_{di}^{AJ} is missing. This is not a problem; we simply use the model implied by (6) for the observed subvector of \mathbf{y}_{di} .

Modelling heterogeneity

Heterogeneity, represented by the β_{di}^{AJ} terms, refers to variation between true treatment effects in trials of the same design. We regard it as a random effect

$$\boldsymbol{\beta}_{di} \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \quad (7)$$

as in the conventional random-effects model for meta-analysis. We refer to model (7) without constraint on $\boldsymbol{\Sigma}$ as the unstructured model. In this model, the between-studies variance is $\text{var}(\beta_{di}^{AJ}) = \Sigma^{JJ}$ for the J–A contrast and $\text{var}(\beta_{di}^{AK} - \beta_{di}^{AJ}) = \Sigma^{JJ} - 2\Sigma^{JK} + \Sigma^{KK}$ for the K–J contrast ($J, K \neq A$); between-studies variances of different contrasts are distinct. $\boldsymbol{\Sigma}$ is identified under consistency if every pair of treatments has been compared in at least two studies, but estimation is typically very imprecise with small numbers of studies.

We now describe two possible structured heterogeneity models for $\boldsymbol{\Sigma}$, which are more readily identified (Salanti *et al.*, 2008; Lu and Ades, 2009).

In a fully structured heterogeneity model, we assume that all treatment contrasts have the same between-studies variance τ^2 . By considering the between-studies variance of the J–A contrasts, it follows that $\Sigma^{JJ} = \tau^2$ for all J. By considering the between-studies variance of the K–J contrasts, it also follows that $\Sigma^{JK} = \tau^2/2$ for $J \neq K$. Thus,

$$\boldsymbol{\Sigma} = \tau^2 \mathbf{P}(0.5) \quad (8)$$

where $\mathbf{P}(\rho)$ is a matrix with all diagonal entries equal to 1 and all off-diagonal entries equal to ρ .

Partially structured heterogeneity models are also possible. For example, the choice $\boldsymbol{\Sigma} = \tau^2 \mathbf{P}(\rho)$ for unknown ρ might be useful, because it implies that all treatment contrasts involving treatment A have variance τ^2 and all treatment contrasts not involving treatment A have a different variance $2(1 - \rho)\tau^2$.

Modelling inconsistency

Inconsistency is represented by the ω_d^{AJ} terms. These could be treated as random effects with mean zero (Lumley, 2002; Lu and Ades, 2006) or as fixed effects (Lu and Ades, 2006). In the companion paper, we argue that they are best treated as fixed effects (Higgins *et al.*, 2012). Here, we have an additional practical reason for treating them as fixed effects; if we treated them as random effects, then model (6) would have three variance components (within-study, between-study and inconsistency) so could not be expressed in the meta-regression framework (1).

We model inconsistency using the design-by-treatment interaction model described in the companion paper. This specifies no structure for the ω_d^{AJ} terms. Thus, each treatment contrast is allowed to differ freely across designs; for example, the B–A contrast in trials with design AB can differ from that in trials with design ABD. Inconsistency models with fewer parameters are possible, such as the model of Lu and Ades (2006), which depends on a particular ordering of the treatments (Higgins *et al.*, 2012).

Only a limited number of ω_d^{AJ} parameters is usually needed in the model; for example, ω_d^{AC} is not needed in design AB. In the design-by-treatment interaction model, the number of identifiable ω_d^{AJ} parameters (the number of degrees of freedom (DOF) for inconsistency, df_{inc}) is the difference between the number of identified fixed parameters in the design-by-treatment interaction model, $\sum_d (T_d - 1)$, and the number of identified fixed parameters in the consistency model, $T - 1$, so $df_{inc} = \sum_d (T_d - 1) - (T - 1)$.

There are various equivalent ways to parameterise the model. A simple approach, illustrated below, considers the designs in order. It adds a parameter for inconsistency in any design which includes a pair of treatments whose contrast can be estimated either directly from a previous design or indirectly under the consistency model from two or more previous designs. For multi-arm designs, the number of inconsistency parameters added to a design depends on how many of its treatment contrasts can be estimated from previous designs.

Below we show two ways to estimate models (6) and (7) with fixed inconsistency parameters using a two-stage approach.

Testing consistency

Once an inconsistency model has been fitted, using the methods to be described below, it is possible to test the hypothesis of consistency. This is best performed by globally testing all the inconsistency parameters using the global Wald test statistic $\hat{\omega}' \hat{\text{var}}(\hat{\omega})^{-1} \hat{\omega}$ which under consistency follows a χ^2 distribution on df_{inc} DOF. Because we use REML estimation, likelihood ratio tests are not available for comparing models with different covariates. Different parameterisations for the models – different choices of the reference treatment or different ways to construct the inconsistency parameters – may be used, but the global Wald test should give the same results for all parameterisations.

It is also possible to examine and test individual ω parameters. However, it is important to remember that a different parameterisation – for example, a different choice of reference treatment – would lead to a different interpretation. This point is illustrated below.

Like any global test, the global Wald test may lack power. Thus, when the hypothesis of consistency is not rejected, inconsistency may nevertheless be present. It is important to use one's understanding of the nature and design of the studies in the NMA to decide how plausible is consistency. If large inconsistency is plausible, then the NMA should be broken down – by restricting the treatments or the trials – into a smaller problem where consistency is more plausible. If the hypothesis of consistency is not rejected and is plausible on subject-matter grounds, then it may be reasonable to base inferences on the consistency model.

If the significance test rejects the hypothesis of consistency, then treatment effects estimated from the consistency model are especially suspect. Instead, it is important to try to understand the source(s) of inconsistency, just as the sources of heterogeneity should be explored in standard meta-analysis (Thompson, 1994). Examination of the $\hat{\omega}$ parameters may help here. Possible explanations should be sought in terms of differences in study-level covariates such as the date of the study or the nature of the population (Salanti *et al.*, 2009). If successful, these covariates could be included in the NMA model or used to stratify the analysis. Another possible explanation is that particular studies are outliers. If outliers are also of inferior methodological quality, then they might be excluded from the analysis, but if they are of equal or superior quality, then there is no alternative to a careful understanding of the nature and reliability of the data.

Estimation by the standard approach

We now consider model estimation where no treatment is common to all trials. The standard contrast-based model (Salanti *et al.*, 2008) designates *trial-specific* reference treatments. For example, sensible choices of reference treatment in the data of Table 1 would be A for designs 1–7, B for designs 8–11 and C for designs 12–13. Contrasts of non-reference treatments with the reference treatment and their standard errors are then estimated.

First, consider the case when only two-arm trials are present. Define y_{di}^* as the single estimated treatment contrast in trial i of design d . If design d compares treatments J and K, then model (6) implies

$$y_{di}^* = (\delta^{AK} - \delta^{AJ}) + (\beta_{di}^{AK} - \beta_{di}^{AJ}) + (\omega_d^{AK} - \omega_d^{AJ}) + (\varepsilon_{di}^{AK} - \varepsilon_{di}^{AJ}). \quad (9)$$

We now describe how this model can be expressed as a standard univariate meta-regression (model (1) with $p = 1$), following Salanti *et al.* (2008).

The first term in (9) involves fixed parameters and enters the $\alpha\mathbf{X}_i$ term of model (1) by constructing covariates $x_{di}^B, x_{di}^C, \dots$ for each treatment except A. For the JK design, the covariate for K is $x_{di}^K = +1$, the covariate for J is $x_{di}^J = -1$, and all other covariates x_{di}^B, \dots are 0.

The second term in (9) is a random variable representing heterogeneity. Under the unstructured heterogeneity model (7), its variance is $\Sigma^{JJ} - 2\Sigma^{JK} + \Sigma^{KK}$, which varies between designs. Under the fully structured heterogeneity model (8), its variance is τ^2 in all designs. To use the meta-regression framework (1), which imposes a common between-studies variance for all studies, we must therefore assume the fully structured heterogeneity model (8).

The third term, representing inconsistency, involves fixed parameters and requires covariates constructed carefully as described above. Lumley (2002) proposed a similar model for NMAs comprising only two-arm trials with the inconsistency parameters as random effects.

The fourth term is a random variable, representing within-study error in the K–J contrast, and its variance can be estimated directly from the study-level data.

Now, consider the case with three-arm trials; each yield two estimated treatment contrasts with within-study covariances assumed known. Standard univariate meta-regression software is not now appropriate and to date, only Bayesian estimation has been described (Lu and Ades, 2006). We now express this case in the multivariate meta-regression framework (1).

Define $\mathbf{y}_{di}^* = (y_{di}^1, y_{di}^2)$ as the set of estimated contrasts in three-arm trial i of design d . For design d comparing treatments J, K and L, the expectation of \mathbf{y}_{di}^* under consistency is $\delta\mathbf{X}_{di}$, where $\delta = (\delta^{AJ}, \delta^{AK}, \delta^{AL})$ and $\mathbf{X}_{di} =$

$\begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$. In general, each outcome is regressed on $T-1$ covariates constructed to equal -1 for the reference treatment, $+1$ for the treatment compared with the reference and 0 for the other treatments. The coefficients δ are the same for all outcomes, so the structure differs from that in equations (2) and (3).

With this reformulation, the model can be fitted by a two-stage procedure using the Stata programme `mvmeta` (White, 2009, 2011) and the code in Appendix A.1. Imposing common coefficients across outcomes requires a previously unpublished option `commonparm` which was implemented for this paper. Extension to trials with four or more arms is straightforward.

Estimation by data augmentation

We now propose an alternative two-stage approach for data that include multi-arm trials; here, the same treatment is taken as the reference treatment in all designs. We accommodate designs which do not include the reference treatment by a data augmentation technique, which introduces an artificial reference treatment arm containing a very small amount of information. Data augmentation is a computational device to enable easy model estimation in a two-stage framework. The advantages of this approach over the standard approach are that model (6) can be estimated directly and intuitively; modelling and estimation are simplified; and more flexible modelling of heterogeneity, as in equation (7), is possible.

We use an arm-based approach to motivate the data augmentation approach in the contrast-based model. We take arm A as the reference category in all trials, irrespective of their design. For trials without arm A, we augment the data by introducing a very small amount of information in arm A. If the trial outcome is binary, the augmenting information is h individuals with success fraction m , for a small value of h . In the analyses below, we use $h=0.001$ and set m equal to the overall event fraction pooling all treatments and all trials (0.08 in the thrombolytics data). Sensitivity to choices of reference treatment, h and m is explored below. If the trial outcome is quantitative, the augmenting information could be a mean equal to the overall mean and a standard error $M \gg 1$ times the largest arm-specific standard error.

Table 2 shows why the augmentation method works, using studies 2, 6 and 27 from the thrombolytics data which compare three treatments A, C and H. Study 2 includes all three treatments and hence yields estimates of both contrasts, which are assembled into the vector \mathbf{y}_{di} with the variance–covariance matrix \mathbf{S}_{di} shown. Both variances in \mathbf{S}_{di} are small, because there is information about both contrasts. The \mathbf{X}_{di} matrix shows that \mathbf{y}_{di} estimates the two basic parameters. Study 6 includes A and C but not H, so it only yields an estimate for the first component of \mathbf{y}_{di} ; the second component is missing, and correspondingly, \mathbf{S}_{di} has just one non-missing element, and \mathbf{X}_{di} has just one non-missing column. Study 27 includes C and H but not the reference treatment A. For the standard approach, \mathbf{y}_{di} has just one non-missing component corresponding to the H–C contrast. The \mathbf{X}_{di} matrix shows that it estimates the difference of the two basic parameters. For the data augmentation approach, an arm A is introduced. The estimated \mathbf{S}_{di} contains very large variances, representing the lack of information in this trial about the C–A and H–A contrasts. However, the equally large covariance in \mathbf{S}_{di} means that the information about the H–C contrast is correctly conveyed; setting $\mathbf{c} = (-1, 1)^T$, the H–C contrast is $\mathbf{c}^T \mathbf{y}_{di} = -0.262$ with variance

Table 2. Subset of the thrombolytic drugs data (deaths/patients), showing coding for the standard and data augmentation approaches. Parameter vector is $\delta = (\delta^{AC}, \delta^{AH})$.

Study	A	C	H	\mathbf{y}_{di}	\mathbf{S}_{di}	\mathbf{X}_{di}
2	1455/13 780	1418/13 746	1448/13 773	(-0.026 -0.005)	$\begin{pmatrix} 0.0016 & 0.0015 \\ 0.0008 & 0.0008 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
6	887/10396	929/10372	.	(0.053 .)	$\begin{pmatrix} 0.0024 & . \\ . & . \end{pmatrix}$	$\begin{pmatrix} 1 & . \\ 0 & . \end{pmatrix}$
27, Standard	.	13/164	10/161	(-0.262 .)	$\begin{pmatrix} 0.19 & . \\ . & . \end{pmatrix}$	$\begin{pmatrix} -1 & . \\ 1 & . \end{pmatrix}$
27, Data augmentation	0.00008/0.001	13/164	10/161	(-0.010 -0.272)	$\begin{pmatrix} 13587.04 & 13587.06 \\ 13586.96 & 13586.96 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$\mathbf{c}^T \mathbf{S}_{di} \mathbf{c} = 0.19$, the same as with the standard approach. The \mathbf{X}_{di} matrix shows that \mathbf{y}_{di} estimates the two basic parameters as in study 2.

In the case of binary outcomes, zero counts in observed treatment arms require additional treatment; the standard procedure is to add 1/2 to all cells in observed treatment arms for any trial with a zero count (Sweeting *et al.*, 2004). A better but more computationally demanding, the way to handle binary data is to use the full binomial likelihood (Hamza *et al.*, 2008).

Fitting the models to the augmented dataset

The consistency model (equation (5) with all ω terms equal to zero) can be written as

$$\mathbf{y}_{di} \sim N(\boldsymbol{\delta}, \mathbf{S}_{di} + \boldsymbol{\Sigma}). \quad (10)$$

This is a multivariate random-effects meta-analysis as in equation (4). With the ω parameters as fixed effects, the inconsistency model can be written as a multivariate random-effects meta-regression

$$\mathbf{y}_{di} \sim N(\boldsymbol{\delta} + \boldsymbol{\omega}_d, \mathbf{S}_{di} + \boldsymbol{\Sigma}). \quad (11)$$

The random effects in such a model represent heterogeneity within designs. For trials missing one or more of arms B, C and so on, \mathbf{y}_{di} is incomplete, but likelihood-based procedures naturally accommodate this.

Ranking in the consistency model

If the consistency model is supported by the data, then it is useful to report the probability that each treatment is the best treatment (Caldwell *et al.*, 2005). This is most naturally performed in the Bayesian approach but can also be approximated in the frequentist approach using a parametric bootstrap procedure implemented in `mvmeta` (White, 2011). We describe the procedure in the case where positive elements of $\boldsymbol{\delta}$ identify treatments that are worse than the reference treatment. Parameter vectors $\boldsymbol{\delta}_b^*$, for $b = 1, \dots, B$, are drawn from the approximate posterior $N(\hat{\boldsymbol{\delta}}, \mathbf{V})$, where $\mathbf{V} = \hat{\text{var}}(\hat{\boldsymbol{\delta}})$. For each b , the best treatment is identified as treatment A if all elements of $\boldsymbol{\delta}_b^*$ are positive and otherwise as the treatment corresponding to the lowest element of $\boldsymbol{\delta}_b^*$. The probability that each treatment is best is estimated by the fraction of the B draws for which that treatment was best. Other summaries, such as rankograms or the probability that each treatment is the worst (Salanti *et al.*, 2011) can be estimated similarly but are usually of less clinical interest.

As described, the method ranks the overall treatment effects $\boldsymbol{\delta}$; the predicted treatment effects in a new trial, $\boldsymbol{\delta} + \boldsymbol{\beta}_{dir}$, can be compared by instead drawing $\boldsymbol{\delta}_b^*$ from $N(\boldsymbol{\delta}, \mathbf{V} + \hat{\boldsymbol{\Sigma}})$ (Higgins *et al.*, 2009; Riley *et al.*, 2011).

Estimation by the Bayesian approach

We also consider a Bayesian version of our model, which we implement for the thrombolytics data in WinBUGS. This model uses binomial within-study distributions in a one-stage analysis and does not require data augmentation or any approximation. We use an arm-based approach, where for each trial, we model a baseline treatment outcome μ_{di} and other treatment outcomes as comparisons to the baseline treatment. For convenience, we directly model the treatment comparison in each design, $\delta^{AJ} + \omega_d^{AJ}$, and reparameterise to obtain results for the inconsistency parameters ω_d^{AJ} . A Bayesian version of the Wald test statistic for inconsistency can be computed. Alternatively, consistency and inconsistency models can be compared using the deviance information criterion (Spiegelhalter *et al.*, 2002).

We use prior distributions for modelling the baseline treatment outcomes, $\mu_{di} \sim N(0, 100)$, and treatment comparisons, $(\delta^{AJ} + \omega_d^{AJ}) \sim N(0, 100)$. If $\boldsymbol{\Sigma}$ is unconstrained, a Wishart prior could be used for $\boldsymbol{\Sigma}^{-1}$. For the constrained model (8), an appropriate prior distribution is $\tau \sim U(0, 2)$. Results may be sensitive to the prior distribution for the heterogeneity parameters (Lambert *et al.*, 2005).

For each WinBUGS run, we used a burn-in of 30 000 updates. Convergence was checked using the Gelman–Rubin statistic as modified by Brooks and Gelman (1998), calculated for three chains with different initial parameter values. Every 20th update was sampled to reduce auto-correlation. Results were calculated from 150 000 sampled updates, ensuring that the estimated Monte Carlo error for all parameters was less than 0.005.

Results for thrombolytics data

Stata code for the frequentist analyses of the thrombolytics data is given in Appendix A, and WinBUGS code for the Bayesian analysis is given in Appendix B.

Choice of heterogeneity model

Table 1 shows that only eight contrasts are found in more than one trial, and only seven contrasts are found in more than one trial of the same design. The arguments above show that we cannot estimate an unstructured heterogeneity model. Instead, we fit the fully structured heterogeneity model (8), which assumes the same heterogeneity variance between each pair of treatments. The partially structured heterogeneity model $\Sigma = \tau^2 P(\rho)$ gives restricted log likelihoods that vary by no more than 0.02 as ρ varies from 0 to 1 (results not shown).

Consistency model

The consistency model (Table 3) shows that treatments B, E and G are most likely to be the best, with similar results for REML and Bayesian estimation. Point estimates were similar between the two estimation methods, except that the estimated heterogeneity was very small for REML but larger for Bayesian estimation. Standard errors were larger for the Bayesian analysis because of the larger estimated heterogeneity and because Bayesian analysis better allows for uncertainty in the heterogeneity parameter.

Inconsistency model

For these data, the design-by-treatment interaction model has 15 fixed parameters (two for each of the two three-arm designs and one for each of the 11 two-arm designs), and the consistency model has seven fixed parameters. There are thus $15 - 7 = 8$ DOF for inconsistency.

We parameterise the design-by-treatment inconsistency model by working down the designs in Table 1. Bold font in Table 1 indicates designs to which we attach inconsistency parameters. Designs 1 and 2 have no potential for inconsistency. Design 3 introduces potential for inconsistency because the C–A contrast can differ between designs 2 (ACH) and 3 (AC); we call this ‘design inconsistency’ in the companion paper (Higgins *et al.*, 2012). We therefore attach an inconsistency parameter ω_3^{AC} to arm C in design 3. We similarly attach inconsistency parameters to D in design 4 and H in design 7. Design 9 introduces potential for inconsistency in a different way, because it directly estimates the F–B contrast which is indirectly estimated under consistency by combining designs 1 (ABD) and 5 (BF); we call this ‘loop inconsistency’ in the companion paper (Higgins *et al.*, 2012). We therefore attach an inconsistency parameter ω_9^{AF} to arm F in design 9. We similarly attach inconsistency parameters to G and H in designs 10–13 because the contrasts in these designs (BG, BH, CG and CH) are all estimable under consistency from earlier designs.

Table 3 reports results for REML estimation via data augmentation and for Bayesian estimation. The REML Wald test for inconsistency was 8.60 on 8 DOF, showing no overall evidence for inconsistency. The Wald-like statistic for

Table 3. Thrombolytic drugs data: results from consistency and inconsistency models. ‘REML’ is the data augmentation approach using $h = 0.001$, $m = 0.08$ and with 10 000 parametric bootstrap samples to compute P(best). ‘Bayes’ is the Bayesian approach and estimates are posterior means.

Treatment	Parameter	Consistency model				Inconsistency model	
		Estimate (standard error)		P(best)		Estimate (standard error)	
		REML	Bayes	REML	Bayes	REML	Bayes
A	-			0.00	0.00		
B	δ^{AB}	-0.16 (0.05)	-0.23 (0.14)	0.19	0.17	-0.16 (0.22)	-0.16 (0.31)
C	δ^{AC}	0.00 (0.03)	-0.02 (0.10)	0.00	0.01	-0.03 (0.22)	-0.03 (0.31)
D	ω_3^{AC}					-0.16 (0.32)	-0.18 (0.38)
	δ^{AD}	-0.04 (0.05)	-0.06 (0.14)	0.00	0.02	-0.04 (0.22)	-0.04 (0.31)
E	ω_4^{AD}					0.45 (0.73)	0.48 (0.82)
	δ^{AE}	-0.16 (0.08)	-0.22 (0.22)	0.23	0.28	-0.15 (0.32)	-0.15 (0.45)
F	δ^{AF}	-0.11 (0.06)	-0.18 (0.16)	0.07	0.11	-0.06 (0.23)	-0.06 (0.32)
	ω_9^{AF}					-0.18 (0.40)	-0.21 (0.52)
G	δ^{AG}	-0.20 (0.22)	-0.23 (0.24)	0.51	0.41	-0.35 (0.55)	-0.37 (0.60)
	ω_{10}^{AG}					0.33 (0.71)	0.38 (0.80)
	ω_{12}^{AG}					0.05 (0.69)	0.05 (0.77)
H	δ^{AH}	0.01 (0.04)	0.04 (0.11)	0.00	0.04	-0.00 (0.22)	-0.00 (0.30)
	ω_7^{AH}					-0.06 (0.41)	-0.06 (0.47)
	ω_{11}^{AH}					1.20 (0.53)	1.25 (0.64)
	ω_{13}^{AH}					-0.31 (0.45)	-0.32 (0.52)
Heterogeneity	τ	0.02 (0.08)	0.12 (0.10)			0.22 (0.14)	0.26 (0.15)
Wald test of consistency (χ_8^2)						8.61	7.91
Deviance information criterion			95.92				97.96

the Bayesian analysis was similar at 7.80. The deviance information criterion was lower for, and hence shows greater support for, the consistency model.

In the design-by-treatment inconsistency model (Table 3), seven of the eight inconsistency parameters are smaller than their standard errors, but $\hat{\omega}_{11}^{AH}$ is slightly more than twice its standard error. Given the multiple testing and the overall lack of evidence for inconsistency, this is likely to be a chance finding. However, to illustrate interpretation as in Table 5 of the companion paper (Higgins *et al.*, 2012), we now take $\hat{\omega}_{11}^{AH}$ at face value; it represents a discrepancy between the direct estimate of the H–B contrast from design 11 and the indirect estimate from the other designs, especially designs 1 (AB) and 2 and 7 (AH), suggesting inconsistency around loop ABH. Indeed, design 11 seems to show large benefit of B over H, whereas designs 1, 2 and 7 show little difference between A, B and H.

Different parameterisations can lead to different conclusions about individual parameters but not overall. For example, if we parameterise the design-by-treatment inconsistency model by working up, instead of down, the designs in Table 1, then the overall test for inconsistency changes by less than 0.001 (code in Appendix A.2, results in Table 4). However, now, two parameters ($\hat{\omega}_1^{AB}$ and $\hat{\omega}_5^{AF}$) are larger than twice their standard errors. Large $\hat{\omega}_1^{AB}$ suggests inconsistency around loop ABH, as with the first parameterisation, whereas $\hat{\omega}_5^{AF}$ suggests a second form of inconsistency which was not seen with the first parameterisation.

Surprisingly, the estimated heterogeneity is much larger in the inconsistency model than in the consistency model. Because the contrasts between designs contribute to the heterogeneity in the consistency model and not in the inconsistency model, this may arise because one or more contrasts between designs are smaller than expected by chance. Univariate analysis shows that the C–A contrast, which probably contributes most information about heterogeneity, is very similar between designs 2 (AC) and 3(ACH).

It is possible to fit other inconsistency models using our framework, as described in the companion paper (Higgins *et al.*, 2012).

Choice of reference categories, h and m

Table 5 explores sensitivity of the data augmentation approach to different choices of the parameters *h* and *m* and different choices of reference category. The results of the Wald test of consistency are very similar in all cases except that mild differences are seen with *h* = 0.1 and *m* = 0.5. Estimates of treatment contrasts AB and AG showed excellent agreement in all cases. Similar comparisons for the other treatment contrasts showed even closer agreement (results not shown). In the data augmentation approach, we calculated matrices S_{di} using double precision computation (that is, with accuracy to about 16 significant figures); by contrast, using single precision

Table 4. Thrombolytic drugs data: results from two different parameterisations of the inconsistency model, using the data augmentation approach with *h* = 0.001, *m* = 0.08.

Treatment	Parameter	Parameterisation 1		Parameterisation 2	
		Estimate	(Standard error)	Estimate	(Standard error)
A	-				
B	δ^{AB}	-0.16	(0.22)	-1.42	(0.56)
	ω_1^{AB}			1.26	(0.60)
C	δ^{AC}	-0.03	(0.22)	0.22	(0.52)
	ω_2^{AC}			-0.25	(0.56)
	ω_3^{AC}	-0.16	(0.32)	-0.41	(0.57)
D	δ^{AD}	-0.04	(0.22)	0.41	(0.69)
	ω_1^{AD}			-0.45	(0.73)
	ω_4^{AD}	0.45	(0.73)		
E	δ^{AE}	-0.15	(0.32)	-1.41	(0.60)
F	δ^{AF}	-0.06	(0.23)	-1.51	(0.60)
	ω_5^{AF}			1.45	(0.64)
	ω_9^{AF}	-0.18	(0.40)		
G	δ^{AG}	-0.35	(0.55)	-0.05	(0.63)
	ω_6^{AG}			-0.30	(0.84)
	ω_{10}^{AG}	0.33	(0.71)	-1.23	(0.79)
	ω_{12}^{AG}	0.05	(0.69)		
H	δ^{AH}	-0.00	(0.22)	-0.06	(0.35)
	ω_2^{AH}			0.06	(0.41)
	ω_7^{AH}	-0.06	(0.41)		
	ω_{11}^{AH}	1.20	(0.53)		
	ω_{13}^{AH}	-0.31	(0.45)		
Heterogeneity	τ	0.2156	(0.1445)	0.2156	(0.1445)
Wald test of consistency (χ^2_8)		8.6053		8.6050	

Table 5. REML estimation of thrombolytic drugs data: Wald χ^2 statistics testing consistency, selected estimated treatment effects and their standard errors, comparing standard approach with data augmentation approach for various choices of h and m and the reference treatment.

	Standard approach	Reference treatment	Data augmentation approach						
			m	0.08			0.5		
				h	0.001	0.01	0.1	0.001	0.01
Wald test of consistency (χ^2_8)	8.61	A	8.61	8.61	8.60	8.60	8.60	8.59	8.42
		B	8.61	8.61	8.62	8.61	8.62	8.62	8.64
		C	8.61	8.61	8.60	8.60	8.59	8.35	
Treatment effect δ^{AB}	-0.161	A	-0.161	-0.161	-0.161	-0.161	-0.161	-0.161	-0.162
		B	-0.161	-0.161	-0.161	-0.161	-0.161	-0.161	-0.158
		C	-0.161	-0.161	-0.161	-0.161	-0.161	-0.161	-0.162
Treatment effect δ^{AG}	-0.197	A	-0.197	-0.197	-0.197	-0.197	-0.197	-0.198	-0.205
		B	-0.197	-0.197	-0.198	-0.197	-0.198	-0.202	
		C	-0.197	-0.197	-0.197	-0.197	-0.198	-0.202	
Standard error $se(\delta^{AB})$	0.046	A	0.046	0.046	0.046	0.046	0.046	0.046	0.046
		B	0.046	0.046	0.046	0.046	0.046	0.046	0.046
		C	0.046	0.046	0.046	0.046	0.046	0.047	0.047
Standard error $se(\delta^{AG})$	0.222	A	0.222	0.222	0.222	0.222	0.222	0.222	0.221
		B	0.222	0.222	0.222	0.222	0.222	0.222	0.221
		C	0.222	0.222	0.222	0.222	0.222	0.222	0.221

(that is, with accuracy to about eight significant figures), agreement between approaches was poorer and computation failed for $h < 0.001$.

These results confirm that the choice of reference category is unimportant, support the argument for taking m to be the observed mean of the data and show that $h = 0.001$ was a good choice in these data.

Discussion

We have demonstrated two frequentist estimation procedures for consistency and inconsistency models for NMA. Our procedures are approximate in two ways. First, they are two-stage procedures which summarise the data from each study as a point estimate and variance, so they implicitly approximate the within-study log-likelihood by a quadratic function of the parameters. This is typically a good approximation except with very sparse data, that is, where many study arms have no events (or no non-events). Second, the data augmentation approach approximates by adding a small amount of information in the reference category. The comparisons in Table 5 shows that this is an excellent approximation.

The alternative Bayesian procedure avoids both of these approximations, but it has its own difficulties. First, it is computationally intensive; fitting the inconsistency model took about 2 h for the Bayesian procedure and 8–14 s for each frequentist procedure on a Windows Server running on a multi-user Sun computer with 2.59-GHz processors. Second, results can be very sensitive to the prior distribution chosen for the heterogeneity parameters.

The two frequentist procedures fit the same basic model but have different strengths and weaknesses. The disadvantage of the data augmentation approach is the extra approximation involved. Its advantages are (i) the between-studies variance matrix Σ need not follow the fully structured model (8), unlike in the standard approach, and (ii) the modelling framework is more intuitive, with treatment effects appearing as overall means rather than as coefficients of design variables. To allow treatment effects to vary with a study-level covariate W , the data augmentation approach enters W as a covariate, whereas the standard approach requires the interaction between W and the design variables to be entered in the model.

If the data augmentation approach is used for binary outcomes, m should be the observed mean and h should take a small value such as 0.001. Although the choice of reference category should not affect the model fit, it may be convenient in terms of interpretation to use an untreated control or placebo group as reference. Alternatively, computational speed might be improved by using a treatment found in many designs as reference category. Finally, when substantial data augmentation has been performed, it may be wise to check that results are similar between different choices of the reference treatments, h and m .

Although we have illustrated the methods using Stata, other software can be used. In R, the data augmentation approach can be implemented using the `mvmeta` package (Gasparrini, 2011). In SAS, both frequentist approaches can be implemented using `proc mixed` (van Houwelingen *et al.*, 2002), and a referee has pointed out that the models can also be fitted in a one-stage procedure using conditional logistic regression (Stijnen *et al.*, 2010).

Future research should explore the interpretation of individual parameters in the design-by-treatment interaction model; the power to test for inconsistency in this model; possible ways to assess inconsistency by

posterior predictive checks; possible reduced inconsistency models with fewer DOF and greater power under specific forms of inconsistency; and partially structured models for heterogeneity.

Network meta-analysis is increasingly popular. This paper has contributed a frequentist analysis approach based on multivariate meta-regression, which should make analysis easier and faster in the presence of multi-arm trials, gives a global test for consistency, and is able to rank the treatments in a way that was formerly only possible with a Bayesian analysis.

Appendix A Code for frequentist analysis in Stata

The code below uses data with one record per study, with identifiers `design` and `study`. Numbers of events/individuals are `r1/n1` for treatment 1 (A), and so on, up to `r8/n8` for treatment 8 (H). `mvmeta` may be installed by typing `net install mvmeta, from(http://www.mrc-bsu.cam.ac.uk/IW_Stata/meta)` in Stata.

A.1 Standard approach

```
// FIRST STAGE OF ANALYSIS
// Organise the data
reshape long r n, i(study) j(trt)
drop if missing(n)
gen logodds=log(r/(n-r))
gen varlogodds=1/r+1/(n-r)
qui by study: gen arm=_n-1
drop r n
reshape wide trt logodds varlogodds, i(study) j(arm)
// Create trial-level summaries
gen y1=logodds1 - logodds0
gen y2=logodds2 - logodds0
gen S11=varlogodds1+varlogodds0
gen S22=varlogodds2+varlogodds0
gen S12=varlogodds0 if !missing(y2)

// SECOND STAGE OF ANALYSIS
// Create covariates
forvalues j=2/8{
    local jj=substr("ABCDEFGH", `j', 1)
    gen treat1`jj'=(trt1==`j') - (trt0==`j')
    gen treat2`jj'=(trt2==`j') - (trt0==`j')
}
drop trt*
// Create interactions for inconsistency models
tab design, gen(des)
forvalues k=1/2{
    gen inter`k' Cdes3=treat`k'C * des3
    gen inter`k' Fdes9=treat`k'F * des9
    gen inter`k' Gdes10=treat`k'G * des10
    gen inter`k' Gdes12=treat`k'G * des12
    gen inter`k' Hdes7=treat`k'H * des7
    gen inter`k' Hdes13=treat`k'H * des13
}
// Fit consistency model
mat P2=I(2)+J(2,2,1)
mvmeta y S, eq(y1:treat1*, y2:treat2*) commonparm bscov(prop P2) nocons
// Fit inconsistency model
mvmeta y S, eq(y1:treat1* inter1*, y2:treat2D treat2H) bscov(prop P2) nocons
* Wald test of consistency
testparm inter1* //ignore this test
test[y1] treat1D=[ y2] treat2D, accum //ignore this test
test[y1] treat1H=[ y2] treat2H, accum //this is the correct test
```

A.2. Data augmentation approach

```
// FIRST STAGE OF ANALYSIS
// Augment data in reference arm
local h 0.001
local m 0.08
replace r1 = 'm' * 'h' if missing(n1)
replace n1 = 'h' if missing(n1)
// Form log odds ratios and variance-covariance matrix
forvalues j = 1/8 {
    gen logodds `j' = log(r`j' / (n`j' - r`j' ))
    gen var `j' = 1/r`j' + 1/(n`j' - r`j' )
}
forvalues j = 2/8 // NB need double precision for (co)variances S*
    gen y`j' = logodds `j' - logodds1
    gen double S`j' `j' = var `j' + var1
    local j1 = `j' + 1
    forvalues k = `j1' / 8 {
        gen double S`j' `k' = var1 if !mi(logodds `j' , logodds `k' )
    }
}
drop logodds* var*
// SECOND STAGE OF ANALYSIS
// Fit consistency model with fully structured Sigma
mat P7 = 0.5 * (I(7) + J(7,7,1))
mvmeta y S, bscov(prop P7) longparm
* Estimate P(best)
mvmeta, pbest(min in 1, zero reps(10000) seed(4187))
// Fit inconsistency model
tab design, gen(des)
mvmeta y S, bscov(prop P7) eq(y3:des3, y4:des4, ///
    y6:des9, y7: des10 des12, y8:des7 des11 des13)
* Wald test of consistency
test[ y3] des3[ y4] des4[ y6] des9[ y7] des10 des12 ///
    [ y8] des7 des11 des13
// Alternative parameterisation of inconsistency model
mvmeta y S, bscov(prop P7) eq(y2: des1, y3: des2 des3, ///
    y4: des1, y6: des5, y7: des10 des6, y8: des2)
test[ y2] des1[ y3] des2 des3[ y4] des1[ y6] des5 ///
    [ y7] des10 des6[ y8] des2
```

Appendix B Code for Bayesian analysis in WinBUGS

This code fits the inconsistency model to the thrombolytics data and should be easily adaptable to other examples. The main data are arranged with one record per arm: *d* and *study* indicate which design and study that arm belongs to, *t* indicates its treatment, and *b* indicates the first treatment in that design. *r* and *n* are the numbers of successes and individuals in the arm. The supplementary data *offset* and *offset.design* list the rows in which the first arm of each trial and of each design is found. *S*=28 is the number of studies, *T*=8 is the number of treatments, *A*=58 is the total number of treatment arms in all trials, and *D*=13 is the number of designs.

To simplify the WinBUGS code, we use a reparameterised version of the model (6). We re-express δ and ω_d as a $D \times T$ matrix *eff.des*: for each design, $\text{eff.des}[d, t] = \delta^{At} + \omega_d^{At}$ is the summary treatment effect of treatment *t* relative to the baseline treatment of that design, allowing for inconsistency; for treatments not in the design, corresponding elements of *eff.des* are undefined. For each study *i*, $\text{eff.study}[i, t, b] = \delta^{At} + \omega_d^{At} + \beta_{di}^{At} - \beta_{di}^{Ab}$ is the true effect comparing treatment *t* with *b* for that study, allowing for heterogeneity. Because the data are expressed in this way, the inconsistency parameters *w* and one of the treatment effects *dAE* (for which no direct comparison is available) must be computed indirectly.

The consistency model can be fitted by replacing $\text{eff.des}[d, k]$ with $\text{eff.des}[k]$ and $\text{eff.study}[i, t, b[k]]$ with $\text{eff.study}[i, k]$.

```

model {
  for(i in 1:S) {
    eff.study[ i, b[ offset[ i ] ], b[ offset[ i ] ] ] <- 0
    for(k in (offset[ i ] + 1) : (offset[ i + 1 ] - 1)) {
      eff.study[ i, t[ k ], b[ k ] ] <- eff.des[ d[ k ], t[ k ] ] + RE[ i, t[ k ] ] - RE[ i, b[ k ] ]
    }
  }
  # Random effects for heterogeneity
  for(i in 1:S) {
    RE[ i, 1 ] <- 0
    RE[ i, 2:T ] ~ dnorm(zero[ ], Prec[ , ] )
  }
  # Prec is the inverse of the structured heterogeneity matrix
  for(i in 1:(T-1)) {
    for(j in 1:(T-1)){
      Prec[ i, j ] <- 2* (equals(i, j)-1/T) / (tau*tau)
    }
  }
  for(i in 1:A) {
    logit(p[ i ] ) <- mu[ study[ i ] ] + eff.study[ study[ i ], t[ i ], b[ i ] ]
    r[ i ] ~ dbin(p[ i ], n[ i ] )
  }
  # For computing DIC
  for(i in 1:A) {
    rhat[ i ] <- p[ i ] * n[ i ]
    dev[ i ] <- 2* ( r[ i ] * (log(r[ i ] ) - log(rhat[ i ] )) + (n[ i ] - r[ i ] ) * (log(n[ i ] - r[ i ] ) - log(n[ i ] - rhat[ i ] )) )
  }
  devs <- sum(dev[ ])
  # Priors
  for(i in 1:S) {
    mu[ i ] ~ dnorm(0, 0.01)
  }
  tau ~ dunif(0, 2)
  for(i in 1:D) {
    for(k in (offset.design[ i ] + 1) : (offset.design[ i ] + num.ests[ i ] )) {
      eff.des[ i, t[ k ] ] ~ dnorm(0, 0.01)
    }
  }
  # Define inconsistency parameters and dAE indirectly
  w[ 1 ] <- eff.des[ 3, 3 ] - eff.des[ 2, 3 ] # wAC3
  w[ 2 ] <- eff.des[ 4, 4 ] - eff.des[ 1, 4 ] # wAD4
  w[ 3 ] <- eff.des[ 9, 6 ] - eff.des[ 5, 6 ] + eff.des[ 1, 2 ] # wAF9
  w[ 4 ] <- eff.des[ 10, 7 ] - eff.des[ 6, 7 ] + eff.des[ 1, 2 ] # wAG10
  w[ 5 ] <- eff.des[ 12, 7 ] - eff.des[ 6, 7 ] + eff.des[ 2, 3 ] # wAG12
  w[ 6 ] <- eff.des[ 7, 8 ] - eff.des[ 2, 8 ] # wAH7
  w[ 7 ] <- eff.des[ 11, 8 ] - eff.des[ 2, 8 ] + eff.des[ 1, 2 ] # wAH11
  w[ 8 ] <- eff.des[ 13, 8 ] - eff.des[ 2, 8 ] + eff.des[ 2, 3 ] # wAH13
  dAE <- eff.des[ 8, 5 ] + eff.des[ 1, 2 ]
}

```

Acknowledgements

This work was supported by the Medical Research Council (Unit Programme numbers U105260558, U105285807; Grant numbers G0902100, G0600650) and the National Science Foundation under Grant DMS-0635449 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank Guobing Lu for providing WinBUGS code from which the code in Appendix B was derived.

References

Brooks SP, Gelman A. 1998. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 7: 434–455.

- Caldwell DM, Ades AE, Higgins JPT. 2005. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Br Med J* **331**: 897–900.
- Dias S, Welton NJ, Caldwell DM, Ades AE. 2010. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* **29**: 932–944.
- Gasparrini A. 2011. mvmeta: multivariate meta-analysis and meta-regression. R package version 0.2.3.
- Hamza TH, van Houwelingen HC, Stijnen T. 2008. The binomial distribution of meta-analysis was preferred to model within-study variability. *J Clin Epidemiol* **61**: 41–51.
- Higgins JPT, Thompson SG, Spiegelhalter DJ. 2009. A re-evaluation of random-effects meta-analysis. *J Roy Stat Soc A* **172**: 137–159.
- Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. 2012. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Meth* **3**: 98–110.
- van Houwelingen HC, Arends LR, Stijnen T. 2002. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* **21**: 589–624.
- Jackson D, Riley R, White IR. 2011. Multivariate meta-analysis: potential and promise. *Stat Med* **30**: 2481–2498.
- Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. 2005. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med* **15**: 2401–2428.
- Lu G, Ades AE. 2006. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc* **101**: 447–459.
- Lu G, Ades AE. 2009. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* **10**: 792–805.
- Lu G, Welton NJ, Higgins JPT, White IR, Ades AE. 2011. Linear inference for mixed treatment comparison meta-analysis: a two-stage approach. *Res Synth Meth* **2**: 43–60.
- Lumley T. 2002. Network meta-analysis for indirect treatment comparisons. *Stat Med* **21**: 2313–2324.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D. 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* **10**: 325–337.
- Riley RD, Higgins JPT, Deeks JJ. 2011. Interpretation of random effects meta-analyses. *Br Med J* **342**: d549.
- Salanti G, Higgins JP, Ades A, Ioannidis JP. 2008. Evaluation of networks of randomized trials. *Stat Methods Med Res* **17**: 279–301.
- Salanti G, Marinho V, Higgins J. 2009. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol* **62**: 857–864.
- Salanti G, Ades A, Ioannidis J. 2011. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* **64**: 163–171.
- Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. 2009. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *Br Med J* **338**: b1147.
- Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. 2002. Bayesian measures of model complexity and fit. *J Roy Stat Soc B Stat Meth* **64**: 583–639.
- Stijnen T, Hamza TH, Özdemir P. 2010. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med* **29**: 3046–3067.
- Sweeting MJ, Sutton AJ, Lambert PC. 2004. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* **23**: 1351–1375.
- Thompson SG. 1994. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* **309**: 1351–1355.
- Thompson SG, Sharp S. 1999. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* **18**: 2693–2708.
- White IR. 2009. Multivariate random-effects meta-analysis. *Stata J* **9**: 40–56.
- White IR. 2011. Multivariate random-effects meta-regression. *Stata J* **11**: 255–270.