NICE DSU TECHNICAL SUPPORT DOCUMENT 3: HETEROGENEITY: SUBGROUPS, META-REGRESSION, BIAS AND BIAS-ADJUSTMENT

REPORT BY THE DECISION SUPPORT UNIT

September 2011

(last updated April 2012)

Sofia Dias[1], Alex J Sutton[2], Nicky J Welton[1], AE Ades[1]

[1]School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK
[2]Department of Health Sciences, University of Leicester, 2nd Floor Adrian Building, University Road, Leicester LE1 7RH, UK

Decision Support Unit, ScHARR, University of Sheffield, Regent Court, 30 Regent Street Sheffield, S1 4DA

Tel (+44) (0)114 222 0734
E-mail dsuadmin@sheffield.ac.uk

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University. The DSU is commissioned by The National Institute for Health and Clinical Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme.

Please see our website for further information www.nicedsu.org.uk

ABOUT THE TECHNICAL SUPPORT DOCUMENT SERIES

The NICE Guide to the Methods of Technology Appraisal[i] is a regularly updated document that provides an overview of the key principles and methods of health technology assessment and appraisal for use in NICE appraisals. The Methods Guide does not provide detailed advice on how to implement and apply the methods it describes. This DSU series of Technical Support Documents (TSDs) is intended to complement the Methods Guide by providing detailed information on how to implement specific methods.

The TSDs provide a review of the current state of the art in each topic area, and make clear recommendations on the implementation of methods and reporting standards where it is appropriate to do so. They aim to provide assistance to all those involved in submitting or critiquing evidence as part of NICE Technology Appraisals, whether manufacturers, assessment groups or any other stakeholder type.

We recognise that there are areas of uncertainty, controversy and rapid development. It is our intention that such areas are indicated in the TSDs. All TSDs are extensively peer reviewed prior to publication (the names of peer reviewers appear in the acknowledgements for each document). Nevertheless, the responsibility for each TSD lies with the authors and we welcome any constructive feedback on the content or suggestions for further guides.

Please be aware that whilst the DSU is funded by NICE, these documents do not constitute formal NICE guidance or policy.

Dr Allan Wailoo

Director of DSU and TSD series editor.

[i] National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal, 2008 (updated June 2008), London.

Acknowledgements

The DSU thanks Julian Higgins, Jeremy Oakley and Catrin Tudor Smith for reviewing this document. The editor for the TSD series is Allan Wailoo.

The production of this document was funded by the National Institute for Health and Clinical Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

This report should be referenced as follows:

Dias, S., Sutton, A.J., Welton, N.J., Ades, A.E. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. 2011; last updated April 2012; available from http://www.nicedsu.org.uk

EXECUTIVE SUMMARY

This Technical Support Document focuses on heterogeneity in relative treatment effects. Heterogeneity indicates the presence of effect-modifiers. A distinction is usually made between true variability in treatment effects due to variation between patient populations or settings, and biases related to the way in which trials were conducted. Variability in relative treatment effects threatens the *external validity* of trial evidence, and limits the ability to generalise from the results, imperfections in trial conduct represent threats to *internal validity*. In either case it is emphasised that, although we continue to focus attention on evidence from trials, the study of effect-modifying covariates is in every way a form of observational study, because patients cannot be randomised to covariate values. This document provides guidance on methods for outlier detection, meta-regression and bias adjustment, in pair-wise meta-analysis, indirect comparisons and network meta-analysis, using illustrative examples.

Guidance is given on the implications of heterogeneity in cost-effectiveness analysis. We argue that the predictive distribution of a treatment effect in a "new" trial may, in many cases, be more relevant to decision making than the distribution of the mean effect. Investigators should consider the relative contribution of true variability and random variation due to biases, when considering their response to heterogeneity.

Where subgroup effects are suspected, it is suggested that a single analysis including an interaction term is superior to running separate analyses for each subgroup.

Three types of meta-regression models are discussed for use in network meta-analysis where trial-level effect-modifying covariates are present or suspected: (1) Separate unrelated interaction terms for each treatment; (2) Exchangeable and related interaction terms; (3) A single common interaction term. We argue that the single interaction term is the one most likely to be useful in a decision making context. Illustrative examples of Bayesian meta-regression against a continuous covariate and meta-regression against "baseline" risk are provided and the results are interpreted. Annotated WinBUGS code is set out in an Appendix.

Meta-regression with individual patient data is capable of estimating effect modifiers with far greater precision, because of the much greater spread of covariate values. Methods for combining IPD in some trials with aggregate data from other trials are explained.

Finally, four methods for bias adjustment are discussed: meta-regression; use of external priors to adjust for bias associated with markers of lower study quality; use of network

synthesis to estimate and adjust for quality-related bias internally; and use of expert elicitation of priors for bias.

CONTENTS

TABLES, BOXES AND FIGURES

Abbreviations and Definitions

CEA	cost-effectiveness analysis
CI	Confidence interval
CrI	Credible interval
CZP	Certolizumab Pegol
DIC	Deviance information criterion
DMARD	Disease Modifying Anti-Rheumatic Drug
FE	Fixed effects
FN	Febrile Neutropenia
IC	Indirect comparisons
IPD	Individual patient data
MCMC	Markov chain Monte Carlo
MTX	Methotrexate
OR	Odds Ratio
RA	rheumatoid arthritis
RCT	Randomised controlled trial
RE	Random effects
TB	tuberculosis
TSD	Technical Support Document

1. INTRODUCTION

This Technical Support Document (TSD) is concerned with heterogeneity, and specifically with between-trials variation in relative treatment effects. It aims to provide guidance on techniques that can be used to explore the reasons for heterogeneity, as recommended in the NICE Guide to Methods of Technology Appraisal.[1] Variation in "baseline" natural history is dealt with in TSD5.[2] In common with other documents in this series, we focus particularly on the implications of different forms of heterogeneity in a decision making context, on the technical specification of models that can estimate or adjust for potential causes of heterogeneity, and on the interpretation of such models in a decision context. There is a considerable literature on the origins and implications of heterogeneity and the reader is referred to the Cochrane Handbook[3] for an introduction to the issues and further references.

Heterogeneity in treatment effects is an indication of the presence of effect-modifying mechanisms, in other words of *interactions* between the treatment effect and the trial or trial-level variable. A distinction is usually made between two kinds of interaction effect. The first results from variation between treatment effects due to different patient populations, settings, or variation in protocols across trials. We will refer to this as clinical variation in treatment effects. This variation is said to represent a threat to the *external validity* of trials, and it limits the extent to which one can generalise trial results from one situation to another. The trial may deliver an unbiased estimate of the treatment effect in a certain setting, but it may be "biased" with respect to the target population in a specific decision problem. Careful consideration of inclusion and exclusion criteria can help to minimise this type of bias, at the expense of having little or no evidence to base decisions on.

The second type of interaction effect is due to deficiencies in the way the trial was conducted, which threaten its *internal validity*. Here, the trial delivers a biased estimate of the treatment effect in its target population, which may or may not be the same as the target population for decision. Typically, these biases are considered to vary randomly in size over trials, and do not necessarily have a zero mean. The clearest examples are the biases associated with markers of poor trial quality such as lack of allocation concealment or lack of double blinding: these have been shown to be associated with larger treatment effects.[4,5] A general model for heterogeneity that encompasses both types can be found in Higgins et al.,[6] but it is seldom possible to determine what the causes of heterogeneity are, or how much is due to true variation in clinical factors and how much is due to other unknown causes of biases.

This document provides guidance on methods for meta-regression and bias adjustment that can address the presence of heterogeneity. In a network meta-analysis context, variability in relative treatment effects can also induce inconsistency (see TSD4[7]) across pair-wise comparisons. The methods introduced here are therefore also appropriate for dealing with inconsistency. Unless otherwise stated, when we refer to heterogeneity this can be interpreted as heterogeneity and / or inconsistency.

The document should be seen as an adjunct to TSD2,[8] which sets out a generalised linear modelling framework for network meta-analysis, indirect comparisons (IC) and pair-wise meta-analysis. TSD2[8] explains how the same core model can be applied with different likelihoods and linking functions. It should be understood that this carries over entirely to the Bayesian models developed for cross-validation (Section 3) sub-groups or meta-regression (Section 4) and bias-adjustment (Section 5) presented below.

1.1. AN OVERVIEW OF META-REGRESSION

Meta-regression is used to relate the size of a treatment effect obtained from a meta-analysis, to certain numerical characteristics of the included trials, with the aim of explaining some, or all, of the observed between-trial heterogeneity. These characteristics can be due to specific features of the individual participants in the trial, or they can be directly due to the trial setting or conduct. In common with other forms of meta-analysis, meta-regression can be based on aggregate (trial-level) outcomes and covariates, or Individual Patient Data (IPD) may be available. Textbooks[3,9] correctly emphasise that, even if we restrict attention to randomised controlled trial (RCT) data, the study of effect-modifiers is inherently observational. This is because it is not possible to randomise patients to one covariate value or another. As a consequence, the meta-regression techniques described in this document inherit all the difficulties of interpretation and inference that attach to non-randomised studies: confounding, correlation between covariates, and, most important, *the inability to infer causality from association*. However, although this restriction on the confidence we can have in inference based on meta-regression is applied across the board, there are major differences in the quality of evidence from meta-regression that depend on the nature of the covariate in question, and the structure of the data, as described below.

1.1.1. Within-trial and between-trial covariates

We will define trial-level covariates as covariates that relate to trial-characteristics or to trial participant characteristics which have been aggregated at trial-level and for which IPD, or a suitable breakdown of results by characteristic, are not available. Patient-level covariates are defined as covariates which relate to patient attributes and can be attributed to specific patients in each trial, either because IPD are available, or because a sufficient breakdown of results has been provided.

If we begin with categorical covariates, we can distinguish between the following scenarios:

A1. Trial-level covariates which relate to trial characteristics. For example, trials which have been conducted on primary and secondary prevention patient populations. This covariate relates to a *between-trial* treatment-covariate interaction. Methods for analysis are discussed under the heading Sub-Group effects (Section 4.1).

A2. Trial-level covariates which relate to patient characteristics. Examples include

(a) Separate trials on men and women: sex as a *between-trial* covariate. This is equivalent to A1 and methods are discussed in Section 4.1.

(b) Trials that include both men and women and report the proportions of men and women in the trial, but do not provide a separate breakdown of estimates (including uncertainty) by sex. The proportion is sometimes taken as a *between-trial* continuous covariate. Methods for this type of meta-regression are discussed in Section 4.3.

(c) Trials that include both men and women and do not report proportions or a breakdown of outcomes by sex. No meta-regression can be carried out unless further assumptions are made.

A3. Patient-level covariates

(a) Trials which have IPD available for the outcome and covariate of interest. In this case the covariate can be used to explore *within-trial* covariate effects, which can then be explored further in the meta-regression.

(b) Trials that include, for example, both men and women, but report the treatment effect with a measure of precision separately for each group. This is a *within-trial* effect, and for the purpose of meta-regression, is equivalent to having IPD on sex. This is true whether binary or continuous outcomes are reported, but only applies to categorical covariates.

A similar set of distinctions can be drawn for continuous covariates:

B1. Trial-level covariates which relate to trial characteristics. For example, the dose of a drug. Methods are discussed in Sections 4.3 and 4.4.

B2. Trial-level covariates which relate to patient characteristics. For example, the mean age of the patients in the trial. This is equivalent to B1 and methods are discussed in Sections 4.3 and 4.4.

B3. Patient-level covariates. With binary outcomes, if mean age and a measure of uncertainty are reported separately for events and non-events then, for the purpose of meta-regression, this is as good as having IPD with each patient's exact age recorded. If the mean covariate values are not reported separately, then IPD would be needed to perform meta-regression. For continuous outcomes with continuous covariates, IPD is always required for meta-regression. This is discussed in Section 4.5.

When investigating an interaction between treatment and covariate, one is comparing the treatment efficacy at one covariate value with the efficacy at another. There are two key differences between within- and between-trial comparisons. Firstly, with a categorical covariate, like sex, the difference between the within-trial comparison and the between-trial comparison is very similar to the difference between a paired and an unpaired t-test. With between-trial comparisons, a given covariate effect (i.e. interaction) will be harder to detect as it has to be distinguishable from the "random noise" created by the between-trial variation. However, for within-trial comparisons the between-trial variation is controlled for, and the interaction effect needs only to be distinguishable from sampling error. With between-trial comparisons, because the number of observations (trials) may be very low while the precision of each trial may be relatively high, it is quite possible to observe a highly statistically significant relation between the treatment effect and the covariate that is entirely spurious.[10]

1.1.2. Ecologic Fallacy

A second difference is that between-trial comparisons are vulnerable to ecologic bias or ecologic fallacy.[11] This is a phenomenon in which for example, a linear regression coefficient of treatment effect against the covariate in the between-trial case can be entirely different to the coefficient for the within-trial data. It is perfectly possible, of course, to have both within-trial, A3(b), and between-trial information, A2(a), in the same evidence synthesis. With continuous covariates, if all the data are IPD (B3), it is possible to fit a model that estimates *both* a between-trial coefficient based on the mean covariate value, and a within-trial

coefficient based on individual variation of the covariate around the mean. Methods for IPD analysis are discussed in Section 4.5.

1.1.3. Greater power of IPD with continuous covariates

With continuous covariates and IPD, not only does the within-trial comparison avoid ecological bias, but it also has far greater statistical power to detect a true covariate effect. This is because the variation in patient covariate values will be many times greater than the variation between the trial means, and the precision in any estimated regression coefficient depends directly on the variance in covariate values.

1.1.4. Use of collapsed category data

The situation in A2(b) has been referred to as "collapsed category" data,[12,13] where the data have been pooled and the treatment effect statistic has been computed from the pooled data, as if the covariate had not been reported. In these cases there is a within-trial comparison, but the data has been degraded. A data structure that is quite commonly found is a mixture of trials: some on men, some on women, and a third category that report the proportion of men and women. It is possible to combine these trials into a single analysis with the proportion of men as a covariate in a between-trial comparison. The covariate would take the value one for trials on men, zero for trials on women. Such data can be analysed using the methodology for Sub-Groups (Section 4.1). However it is essential to note that this model is only strictly correct for linear models, in other words models with an identity link (see TSD2[8]). It is not valid for logit, log or other commonly used models.[11] There are collapsed category methods for incorporating all these forms of data, using non-linear models, without introducing bias. This is beyond the scope of this document, but readers are referred to published papers whose ideas can be adapted to solve this problem.[12-14] These methods can be extended still further to incorporate data from trials of type A2(c) in which information on the covariate is entirely "missing". This has not been attempted for treatment effects, but again ideas and programming code from similar applications[12,13] can be adapted.

1.1.5. Aggregation bias

Finally, it needs to be appreciated that in cases where the covariate does not interact with the treatment effect, but modifies the baseline risk, the effect of pooling data over the covariate is to bias the estimated treatment effect towards the null effect. This is a form of ecologic bias known as *aggregation bias*[11] but it does not affect strictly linear models, where pooling data across such covariates will not create bias. Usually it is significant only when both the

covariate effect on baseline risk and the treatment effect are quite strong. It is a particular danger in survival analysis because the effect of covariates like age on cancer risk can be particularly marked, and because the log-linear models routinely used are highly non-linear. When covariates that affect risk are present, even if they do not modify the treatment effect, the analysis must be based on pooled estimates of treatment effects from a stratified analysis for group covariates and regression for continuous covariates, and not on treatment effects estimated from pooled data.

1.2. OVERVIEW OF BIAS ADJUSTMENT

The aim of bias adjustment is in effect to transform estimates of treatment effect that are biased relative to the desired effect in the target population, into unbiased estimates. It is necessary in all cases to take into account the uncertainty in external data or prior opinions that are used. In Section 5 we discuss four methods, of which two are types of meta-regression. These are: covariate adjustment for external validity biases (Section 5.1); adjustment and down-weighting of evidence at risk of bias, based on external data, for internal biases (Section 5.2); estimation of bias associated with markers of risk of internal bias within a network meta-analysis (Section 5.3); and adjustment for internal and/or external biases based on expert opinion or other evidence (Section 5.4).

1.3. NETWORK META-ANALYSIS AS A FORM OF META-REGRESSION

It should be emphasised that although network meta-analysis can be understood as a form of meta-regression, it *is* based on randomised comparisons.[15] Indeed, it can be shown that the coherent estimates of treatment effects assuming consistency (see TSD2[8]) are weighted averages of the estimates from the individual trials,[16] just as is the case in pair-wise meta-analysis. It is also misleading to state that network meta-analyses or indirect comparisons suffer from the biases of observational studies.[3] They suffer from problems of unobserved effect modifiers, in the same way as pairwise meta-analysis. Both give unbiased estimates of the treatment effects in the target population, as long as their constituent trials are unbiased for that target population. Both are superior to observational studies as they are based on randomised comparisons.

2. MEASURES OF HETEROGENEITY

A number of standard methods for measuring between-trials heterogeneity have been proposed, and readers can be referred to standard texts.[3,9,17] In the literature, tests of the null hypothesis of homogeneity in Fixed Effect (FE) models, e.g. Cochran's Q, are often used to justify the choice of a Random Effects (RE) model. The I^2 statistic has the advantage of being scale-free, but it is dependent on the number and size of the included studies, making it hard to interpret in a typical meta-analysis.[18] The approach taken in TSD2,[8] in keeping with the Bayesian framework, has been to compare the Fixed and Random Effects models' residual deviance and DIC statistics.[19] An advantage of the Bayesian approach is that it provides a posterior distribution of the between-trials variance – or, perhaps easier to interpret – the between trial standard deviation, which gives investigators some insight into the range of values that are compatible with the data. It is also possible to obtain a measure of uncertainty for the between-trials variance using classical approaches,[20] but this is not often done.

We must, however, repeat the important warning given in TSD2[8] (Section 6.2) that the posterior for the between trial standard deviation is likely to be extremely sensitive to the prior, and in particular that our "default" practice of using vague priors is likely to result in posteriors which allow for unrealistically high levels of heterogeneity. This will inevitably occur whenever the number of trials is small, or when the majority of trials are small. The solution is to use informative priors, based on expert opinion or on meta-epidemiological data. The easiest approach might be to identify a large meta-analysis of other treatments for the same condition and using the same outcome measures, and use the posterior distribution for the between-trial heterogeneity from this meta-analysis to inform the current analysis.[21]

2.1. IMPLICATIONS OF HETEROGENEITY IN DECISION MAKING

The critical issue, which has received comparatively little attention, is how to respond to high levels of heterogeneity in a decision making context. It is essential that investigators compare the size of the treatment effect to the extent of between trials variation. Figure 1 portrays a situation where a RE model has been fitted. The posterior mean of the *mean* treatment effect is 0.70 with posterior standard deviation (sd)=0.2, making the mean effect clearly different from zero with 95% CrI (0.31, 1.09). However, the posterior mean of the between-trials standard deviation is σ=0.68, comparable in size to the mean effect. Now consider the question: what is a reasonable confidence interval for our prediction of the outcome of a

future trial of infinite size? An approximate answer in classical statistics is found by adding the variance of the mean to the between-trials variance, which gives $\mathrm{sd}^2 + \sigma^2 = 0.50$ giving a predictive standard deviation of 0.71. Note that the 95% predictive interval is now (-0.69, 2.09) easily spanning zero effect, including a range of harmful effects. If we interpret these distributions in a Bayesian way, we would find that the probability that the mean effect is less than zero only is 0.0002, while the probability that a new trial would show a negative effect is much higher: 0.162 (Figure 1).

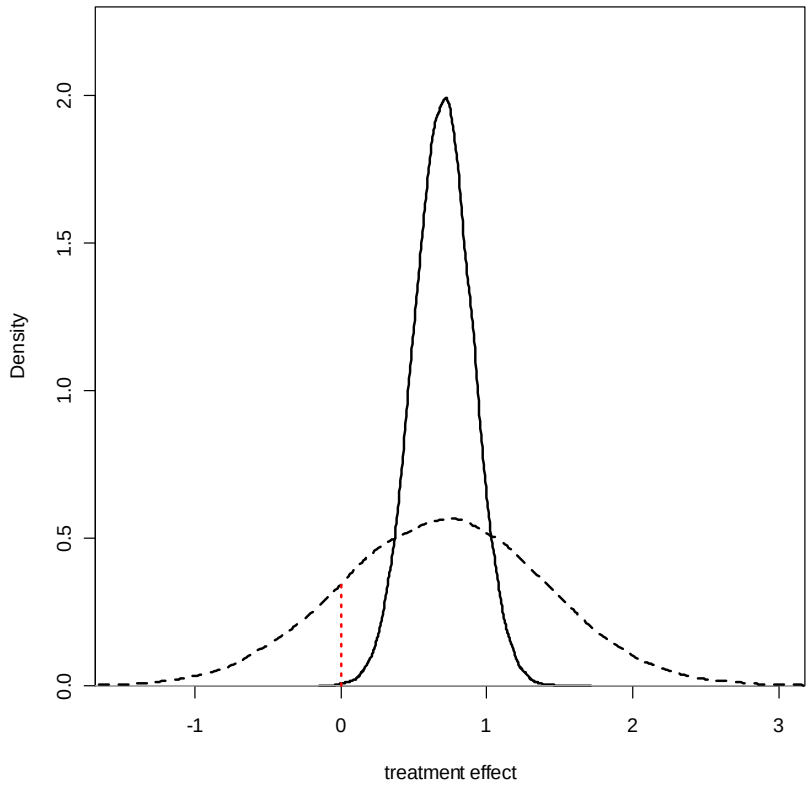

Figure 1 Posterior (solid) and predictive (dashed) densities for a treatment effect with mean=0.7, standard deviation=0.2 and heterogeneity (standard deviation)=0.68. The area under the curve to the left of the vertical dotted line is the probability of a negative value for the treatment effect.

This issue has been discussed before,[6,22-24] and it has been proposed that, in the presence of heterogeneity, the predictive distribution, rather than the distribution of the mean treatment effect, better represents our uncertainty about the comparative effectiveness of treatments in a future "roll out" of a particular intervention. In a Bayesian Markov chain Monte Carlo (MCMC) setting, a predictive distribution is easily obtained by drawing further samples from the distribution of effects:

$$\delta_{new} \sim N(d, \sigma^2)$$

where d is the estimated (common) mean treatment effect and σ^2, the estimated between-trial heterogeneity variance (see TSD2[8]).

The mean of the predictive distribution, on its linear scale, will be the same as the mean of the distribution of the mean effect. But the implications of this recommendation on the uncertainty in a decision, in cases where there are high levels of unexplained heterogeneity, could be quite profound, and it is therefore important that the degree of heterogeneity is not exaggerated. This immediately raises the question: what are the causes of the heterogeneity.[25,26] This is taken up in greater detail in subsequent sections, where we discuss methods that can reduce heterogeneity by *adjusting* trial results for factors that, putatively, cause it. For present purposes we can distinguish between true variability in the size of the treatment effect across patient populations, and apparent random variation due to biases caused by the way in which the trial was conducted.

Higgins et al.[6] make it clear that the variance term in the predictive distribution should consist *only* of true variation between trial populations. At the present time, however, there is no clear methodology, or source of information, that would allow one to distinguish the different sources of variation. Recent meta-epidemiological studies of very large numbers of meta-analysis are beginning to throw light on this, but all that can confidently be said at this time is that the observed heterogeneity is likely to be an over-estimate of the true variation in effect size.

This discussion has assumed exchangeability over all included trials. However, the target population for decision might be more similar to that of some trials than others. In this case adjustments for external validity should be considered – see Section 5.

3. OUTLIER DETECTION

Closely related to the question of heterogeneity is the matter of outlier detection. Here the focus is not on the overall level of variation in trial results, but on one or two trials that seem to have results that are particularly different from the others. The two issues are closely related, as a single outlying trial may impact greatly on the measure of heterogeneity. Conversely, a high level of heterogeneity makes it difficult to detect a true outlier.

3.1. PREDICTIVE CROSS-VALIDATION IN PAIR-WISE META-ANALYSIS

Figure 2 shows a forest plot with the crude log-odds ratios calculated from the data in Table 1, and the "shrunken" estimates from a RE model (i.e. the trial-specific treatment effects, assumed to be exchangeable), for a set of 16 trials of intravenous magnesium against placebo, for patients with acute myocardial infarction.[27] WinBUGS code for all analyses is presented in the Appendix (Example 1).

Table 1 Number of deaths out of the total number of patients for 16 trials of intravenous magnesium against placebo, for patients with acute myocardial infarction.[27]

			Placebo		Magnesium	
Trial ID	Trial Name	Year	Deaths	Total	Deaths	Total
1	Morton	1984	2	36	1	40
2	Rasmussen	1986	23	135	9	135
3	Smith	1986	7	200	2	200
4	Abraham	1987	1	46	1	48
5	Feldstedt	1988	8	148	10	150
6	Shechter	1989	9	56	1	59
7	Ceremuzynski	1989	3	23	1	25
8	Bertschat	1989	1	21	0	22
9	Singh	1990	11	75	6	76
10	Pereira	1990	7	27	1	27
11	Shechter1	1991	12	80	2	89
12	Golf	1991	13	33	5	23
13	Thorgersen	1991	8	122	4	130
14	LIMIT-2	1992	118	1157	90	1159
15	Shechter2	1995	17	108	4	107
16	ISIS-4	1995	2103	29039	2216	29011

The choice of a RE model for this data was based on a posterior mean of the residual deviance of 29.6 (which compares well to 32 data points) and DIC=54.2, compared to a posterior mean of the residual deviance of 77.5 and DIC=94.5 for a FE model (see TSD2[8] for more details). The posterior median of the standard deviation is 0.68 with 95% CrI (0.35, 1.30), which is comparable in size to the mean treatment effect of -0.89 with 95% CrI (-1.49, -0.41) on the log-odds ratio scale. This indicates that there is substantial heterogeneity.

Figure 2 shows that one particular trial, the ISIS-4 "mega-trial", has an estimated trial-specific log-odds ratio of 0.055 with 95% CrI (-0.007, 0.117) which is somewhat different from the other trials. In particular neither the crude 95% Confidence Interval (CI) nor the "shrunken" 95% CrI for this trial overlap with the 95% CrI for the mean treatment effect (Figure 2). Investigators might wonder whether this trial is an "outlier" in some sense. The appropriate tool for examination of single trials in a meta-analysis is cross-validation[28,29]

based on a "leave one out" approach. The procedure is to remove the trial from the synthesis, and compare the observed treatment effect to the predictive distribution of effects that we would expect *based on an analysis of the remaining trials*. So, the first step in predictive cross-validation is to fit the RE meta-analysis model to the data in Table 1, excluding trial 16, ISIS-4.

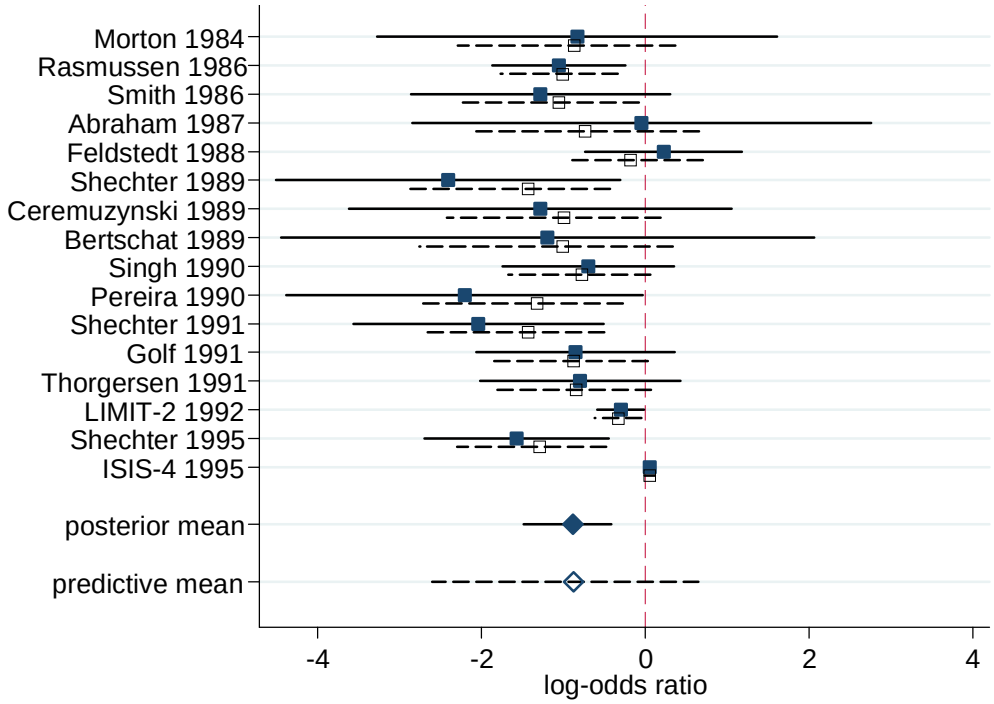

Figure 2 Magnesium Example: Crude log-odds ratios with 95% CI (filled squares, solid lines); posterior mean with 95% CrI of the trial-specific log-odds ratios, "shrunken" estimates, (open squares, dashed lines); posterior mean with 95% CrI of the posterior (filled diamond, solid line) and predictive distribution (open diamond, dashed line) of the pooled treatment effect, obtained from a RE model including all the trials.

Following the notation in TSD2[8], r_{ik} represents the number of events (deaths), out of the total number of patients in each arm, n_{ik}, for arm k of trial i, and is assumed to have a Binomial likelihood $r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$, where p_{ik} represents the probability of an event in arm k of trial i for i=1,…, 15 (excluding ISIS-4, trial 16); k=1,2. The RE model is

$$\text{logit}(p_{ik}) = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}}$$

where

$$I_{\{u\}} = \begin{cases} 1 & \text{if } u \text{ is true} \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

and the trial-specific log-odds ratios come from a common distribution: $\delta_{i,12} \sim N(d, \sigma^2)$. The next step is to draw the predicted treatment effect in a future trial, δ_{new}, from the predictive distribution

$$\delta_{new} \sim N(d, \sigma^2)$$

where d and σ are drawn from the posterior distributions. We now need to draw a replicate study of the same size and with the same baseline risk as ISIS-4, onto which we will apply the predictive treatment effect δ_{new}. In this example the baseline effect is the logit of the probability of mortality on Placebo, p_{base}, which could be estimated from the proportion of mortalities on the placebo arm of ISIS-4 as 2103/29039=0.072. However, this would not convey our uncertainty about this probability. Instead we can assume that the probability of mortality in a new study like ISIS-4 has a Beta distribution

$$p_{base} \sim \text{Beta}(a, b)$$

where a=$r_{16,1}$=2103, the number of events in the placebo arm of trial 16 (ISIS-4) and b=$n_{16,1}$-$r_{16,1}$=26936, the number of non-events in the control arm of trial 16. The predictive probability of mortality on Magnesium in a future study like ISIS-4, given the remaining 15 trials, p_{new}, is given by

$$\text{logit}(p_{new}) = \text{logit}(p_{base}) + \delta_{new}$$

and the predicted number of events, r_{new}, in the Magnesium arm of a future trial of the same size as trial 16 (ISIS-4) can be drawn from a binomial distribution with probability p_{new}

$$r_{new} \sim \text{Binomial}(p_{new}, n_{16,2})$$

and compared to the observed number of events in trial 16 (ISIS-4) to obtain a Bayesian p-value: the probability of obtaining a value as extreme as that observed in trial 16, i.e. Pr(r_{new}>$r_{16,2}$). Within a Bayesian MCMC framework, this is done by setting up a variable that, at each iteration, takes the value 1 if r_{new}>$r_{16,2}$ and is 0 otherwise. By averaging over a large number of iterations this variable gives the desired probability.

WinBUGS code to fit the original RE model is given in TSD2[8] (Programs 1(a) or (c)). Code for predictive cross-validation is provided in the Appendix (Program 1). The result is a p-value of 0.056, indicating that a trial with a result as extreme as ISIS-4 would be unlikely, but still possible, given our model for the remaining data (convergence was achieved after 20,000 burn-in iterations and results are based on 50,000 samples from three independent chains). In examining these results, however, one must take into account the effective number of tests that could be undertaken. In carrying out cross-validation for ISIS-4 we have picked the most

extreme of 16 trials, so there is an implication that n=16 tests could be performed and the test on ISIS-4 would give the most extreme result (i.e. have the smallest p-value). To correctly interpret the significance of the observed p-value we need to compare it to its expected value, which is $1/(n+1) = 0.059$, the value of the n-th Uniform order statistic. The observed p-value therefore suggests that ISIS-4 is not necessarily incompatible with a RE model fitted to the remaining data. This can also be seen in Figure 3 which now presents the "shrunken" estimates (δ_{i2}, i=1,…,15), mean and predictive treatment effects for a RE meta-analysis excluding the ISIS-4 trial, but includes the observed log-odds ratio and CI for this trial. It can be seen that the observed log-odds ratio from the ISIS-4 trial although well outside the CrI for the posterior mean still lies within the bounds of the CrI for the predictive mean treatment effect, which is the basis for predictive cross-validation.

This is a statistical result only: it is impossible to deduce whether ISIS-4 is a deviant result, or whether the other trials are. This particular meta-analysis has been discussed repeatedly[30,31] and current opinion is that ISIS-4 is in fact the "correct" result.[32]

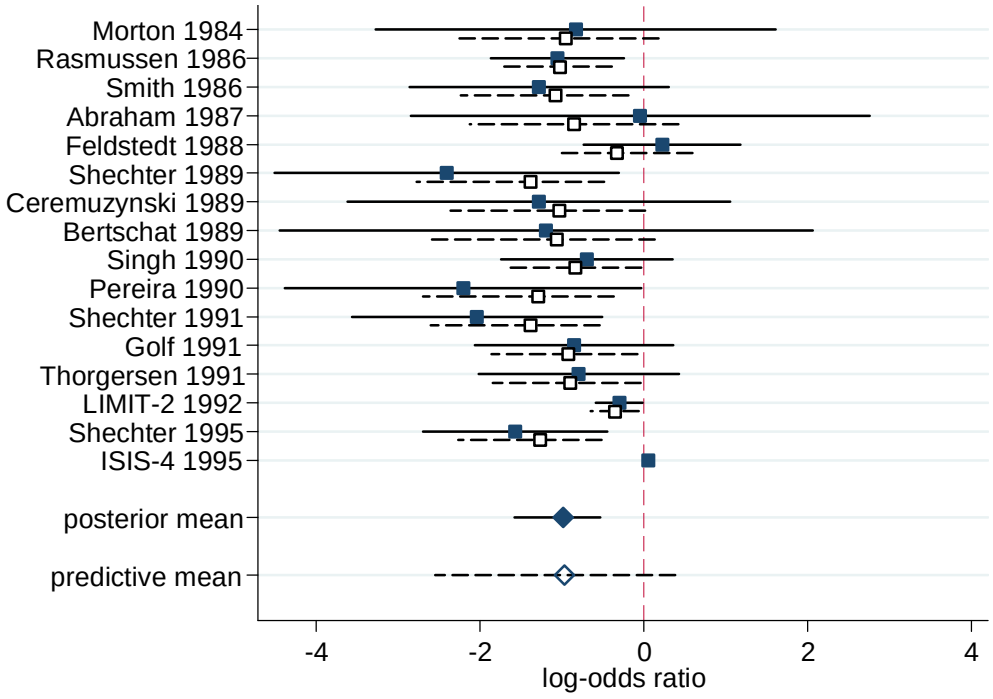

Figure 3 Magnesium Example: Crude log-odds ratios with 95% CI (filled squares, solid lines); posterior mean with 95% CrI of the trial-specific log-odds ratios, "shrunken" estimates, (open squares, dashed lines); posterior mean with 95% CrI of the posterior (filled diamond, solid line) and predictive distribution (open diamond, dashed line) of the pooled treatment effect, obtained from a RE model excluding the ISIS-4 trial.

The Magnesium dataset holds several important messages about RE models in decision making. First, note that the RE models with and without the ISIS-4 trial fit equally well (the posterior means of the residual deviances for the two models are 29.7 and 27.9 to compare to 32 and 30 data points respectively). This is because a RE model can generally fit any random distribution of effects, it is not greatly affected by the spread. Second, it illustrates the weakness in basing inference on the mean effect. Within the entire ensemble of trials, whether including or even excluding itself, ISIS-4 is not particularly remarkable. It is, however, markedly different from the mean effect. To base the decision on the mean effect is, therefore, to base a decision on a model *in which the different sources of evidence are in an unexplained conflict.* A model based on the predictive distribution is compatible with all the data.

3.2. PREDICTIVE CROSS-VALIDATION FOR INDIRECT COMPARISONS AND NETWORK META-ANALYSIS

Cross-validation can be applied without modification to broader networks of evidence, including multiple treatments and multi-arms trials. However, it needs to be borne in mind that, when there are multiple treatments, the predictive distribution is multi-variate normal. So, for a network with *s* treatments, the predictive distribution for the *s*-1 treatment effects relative to treatment 1 (the basic parameters, see TSD2[8]) is given by

$$\boldsymbol{\delta}^{new} = \begin{pmatrix} \delta_{12}^{new} \\ \vdots \\ \delta_{1s}^{new} \end{pmatrix} \sim N_{s-1}\left(\begin{pmatrix} d_{12} \\ \vdots \\ d_{1s} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \dots & \sigma^2/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \cdots & \sigma^2 \end{pmatrix} \right) \tag{2}$$

where d and σ are sampled from the posterior distributions (given the data). This can be re-written as a series of conditional univariate normal distributions[33]

$$\delta_{1k}^{new} \Big| \begin{pmatrix} \delta_{12}^{new} \\ \vdots \\ \delta_{1(k-1)}^{new} \end{pmatrix} \sim N\left(d_{1k} + \frac{1}{(k-1)} \sum_{j=1}^{k-1} \left(\delta_{1j}^{new} - d_{1j} \right), \frac{k}{2(k-1)} \sigma^2 \right) \tag{3}$$

Either the multivariate distribution in equation (2) or the conditional distributions in equation (3) must be used to estimate the predictive random effects of each treatment relative to treatment 1 (the reference treatment). The code presented in the Appendix (Program 2) follows the code in TSD2[8] and uses the formulation in equation (3) as it allows for a more generic code which works for networks with any number of treatments.

To ensure that the correlations between the predictive treatment effects are carried through correctly to all treatment contrasts, the predictive distributions for the other treatment comparisons are obtained from the consistency equations (TSD2[8]):

$$\delta_{XY}^{new} = \delta_{1Y}^{new} - \delta_{1X}^{new}$$

See Example 2 in the Appendix for an illustration and WinBUGS code.

In the context of network meta-analysis, cross-validation for outlier detection is closely related to methods of inconsistency checking such as the node-split,[34] where "direct" evidence from trials on a specific contrast is separated from the rest of the network to produce an estimate of the relative treatment effect, which is then compared to the relative effect predicted from the rest of the network. In effect, the node-split method is analogous to a cross-validation where a subset of trials, rather than just one trial, is removed from the original analysis.

However, one crucial difference between these methods is that although cross-validation is essentially a method for detecting "outliers", the concept of inconsistency between "direct" and "indirect" evidence refers to inconsistency in expected (i.e. mean) effects. It is for this reason that node-splitting for inconsistency checking, as presented in TSD4,[7] is based on the posterior distributions of the mean effects. This will frequently result in a situation where, with a triangular network in which one edge consists of a singleton trial, node splitting might show inconsistency in the expected effects, while cross-validation fails to show that the singleton trial is an outlier. Such an outcome is by no means paradoxical: the ISIS-4 trial is not an outlier when the predictive distribution is considered, although it departs very markedly from the expected effect based on the remaining evidence. However, this example indicates that investigators need to be clear about whether they are looking for evidence that, for example, the mean AB and AC effects are inconsistent with the mean BC effect (based albeit on a single trial), or whether they are concerned that the single BC trial is an "outlier" in the context of an evidence synthesis.

Of course, technically there is no reason why inconsistency checks cannot be made on the predictive distributions of the treatment effects, and this may be desirable if inference is to be based on the predictive treatment effects from a network meta-analysis.

4. SUBGROUPS, META-REGRESSION AND ADJUSTING FOR BASELINE RISK

4.1. AN INTRODUCTION TO META-REGRESSION: SUB-GROUP EFFECTS

In the context of treatment effects in RCTs, a sub-group effect can be understood as a categorical trial level covariate that interacts with the treatment, and this corresponds to scenario A1 in Section 1.1.1. The hypothesis would be that the size of treatment effect is different in, for example, male and female patients, or that it depends on age group, previous treatment, etc. The simplest way of analysing such data is to carry out *separate analyses* for each group and then examine the estimates of the relative treatment effects. However, this approach has two disadvantages. First, if the models have random treatment effects, having separate analyses means having different estimates of between-trial variation. As there is seldom enough data to estimate the between-trial variation, it may make more sense to assume that it is the same for all subgroups. A second problem is that running separate analyses does not immediately produce the *test of interaction* that is required to reject the null hypothesis of equal effects. The alternative to running separate analyses for each subgroup is a single integrated analysis with a shared between-trial heterogeneity parameter, and an interaction term, β, introduced on the treatment effect.

The RE model for separate pairwise meta-analyses, introduced in TSD2,[8] is

$$\theta_{ik} = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}}$$

where θ_{ik} is the linear predictor (for example the log-odds) in arm k of trial i, μ_i are the trial-specific baseline effects in a trial i, treated as unrelated nuisance parameters and $\delta_{i,1k}$ are the trial-specific treatment effects of the treatment in arm k relative to the control treatment in arm 1 in that trial, with k=1,2 and I defined in equation (1).

The meta-regression model with random treatment effects is

$$\theta_{ik} = \mu_i + \left(\delta_{i,1k} + \beta x_i\right) I_{\{k \neq 1\}} \qquad (4)$$

where x_i is the trial-level covariate for trial i, which can represent a subgroup, a continuous covariate or baseline risk. We can re-write equation (4) as

$$\theta_{i1} = \mu_i$$
$$\theta_{i2} = \mu_i + \delta_{i,12} + \beta x_i$$

and note that the treatment and covariate interaction effects (δ and β) only act in the treatment arm, not in the control. For a RE model the trial-specific log-odds ratios come from a

common distribution: $\delta_{i,12} \sim N(d, \sigma^2)$. For a FE model we replace equation (4) with $\theta_{ik} = \mu_i + (d + \beta x_i) I_{\{k \neq 1\}}$. In the Bayesian framework *d*, β and σ will be given independent (non-informative) priors: for example *d*, $\beta \sim$ N(0, 100^2) and $\sigma \sim$ Uniform(0,5).

Section 4.1.1 provides a worked example contrasting the results obtained with separate analyses and those from a sub-group interaction analysis.

Ideally, we would want to include subgroup terms whether they were "statistically significant" or not, possibly using informative priors elicited from clinical experts. However, the NICE Methods Guide[1] suggests that subgroup effects should be statistically robust if they are to be considered in a cost-effectiveness model, as well as having some *a priori* justification. In practice, it would be difficult to sustain an argument that a treatment should be accepted or rejected based on a statistically weak interaction.

4.1.1. Subgroups in a pair-wise meta-analysis: Statins Example

A meta-analysis of 19 trials of Statins for cholesterol lowering, against placebo or usual care[35] included some trials on which the aim was primary prevention (patients included had no previous heart disease), and others on which the aim was secondary prevention (patients had previous heart disease). Note that the subgroup indicator is a trial-level covariate. The outcome of interest was all-cause mortality and the data are presented in Table 2. The potential effect-modifier, primary vs secondary prevention, can be considered a subgroup in a pair-wise meta-analysis of all the data, or two separate meta-analyses can be conducted on the two types of study.

Table 2 Meta-analysis of Statins against Placebo for cholesterol lowering in patients with and without previous heart disease:[35] number of deaths due to all-cause mortality in the control and Statin arms of 19 RCTs.

	Placebo/Usual care		Statin		
Trial ID	number of deaths r_{i1}	number of patients n_{i1}	number of deaths r_{i2}	number of patients n_{i2}	Type of prevention x_i
1	256	2223	182	2221	Secondary
2	4	125	1	129	Secondary
3	0	52	1	94	Secondary
4	2	166	2	165	Secondary
5	77	3301	80	3304	Primary
6	3	1663	33	6582	Primary
7	8	459	1	460	Secondary
8	3	155	3	145	Secondary
9	0	42	1	83	Secondary
10	4	223	3	224	Primary
11	633	4520	498	4512	Secondary
12	1	124	2	123	Secondary
13	11	188	4	193	Secondary
14	5	78	4	79	Secondary
15	6	202	4	206	Secondary
16	3	532	0	530	Primary
17	4	178	2	187	Secondary
18	1	201	3	203	Secondary
19	135	3293	106	3305	Primary

The number of deaths in arm k of trial i, r_{ik}, is assumed to have a Binomial likelihood $r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$, i=1,…,19; k=1,2. Defining x_i as the trial-level subgroup indicator such that

$$x_i = \begin{cases} 0 & \text{if study } i \text{ is a primary prevention study} \\ 1 & \text{if study } i \text{ is a secondary prevention study} \end{cases}$$

our interaction model is given in equation (4) where θ_{ik}=logit(p_{ik}) is the linear predictor (see TSD2[8]). In this setup, μ_i represent the log-odds of the outcome in the 'control' treatment (i.e. the treatment indexed 1) and $\delta_{i,12}$ are the trial-specific log-odds ratios of success on the treatment group compared to control for primary prevention studies.

WinBUGS code to fit two separate fixed or random effects models is given in TSD2[8] (programs 1(a) to 1(d)). Code for a single analysis with an interaction term for subgroup is given in the Appendix to this document (Example 3, Programs 3(a) and 3(b)).

The results (including the model fit statistics introduced in TSD2[8]) of the two separate analyses and the single analysis using the interaction model for fixed and random treatment

effects models are shown in Table 3. For the FE models, convergence was achieved after 10,000 burn-in iterations for separate analyses (20,000 iterations for the joint analysis) and results are based on 50,000 samples from three independent chains. For the RE models 40,000 burn-in iterations were used for the separate analyses, 50,000 burn-in iterations were used for the joint analysis and results are based on 100,000 samples from three independent chains. Note that in a FE context the two analyses deliver exactly the same results for the treatment effects in the two groups, while in the RE analysis, due to the shared variance, treatment effects are not quite the same: they are more precise in the single analysis, particularly for the primary prevention subgroup where there was less evidence available to inform the variance parameter, leading to very wide Credible Intervals (CrI) for all estimates in the separate RE meta-analysis. However, within the Bayesian framework, only the joint analysis offers a direct test of the interaction term β, which, in both cases has a 95% Credible Interval (CrI) which includes the possibility of no interaction, although the point estimate is negative, suggesting that Statins might be more effective in secondary prevention patients.

Table 3 Posterior summaries, mean, standard deviation (sd) and 95% Credible Interval (CrI) of the log-odds ratio (LOR), odds ratio (OR) and posterior median, sd and 95% CrI between-trial heterogeneity (σ) of all-cause mortality when using Statins (LOR<0 and OR<1 favour Statins) for primary and secondary prevention groups for both fixed and random effects models; and measures of model fit: posterior mean of the residual deviance (resdev), number of parameters (pD) and DIC.

	Fixed effects						**Random effects**					
	Primary Prevention			**Secondary Prevention**			**Primary Prevention**			**Secondary Prevention**		
	Separate analyses						**Separate analyses**					
	mean	sd	CrI	mean	sd	CrI	mean/median	sd	CrI	mean/median	sd	CrI
LOR	-0.11	0.10	(-0.30,0.09)	-0.31	0.05	(-0.42,-0.21)	-0.18	0.74	(-2.01,1.12)	-0.36	0.16	(-0.72,-0.06)
OR	0.90	0.09	(0.74,1.09)	0.73	0.04	(0.66,0.81)	1.12	3.65	(0.13,3.07)	0.71	0.11	(0.49,0.94)
σ	-	-	-	-	-	-	0.79	0.98	(0.06,3.90)	0.16	0.23	(0.01,0.86)
resdev	16.9†			29.0‡			11.9†			28.3‡		
pD	6.0			15.0			9.3			16.8		
DIC	22.9			44.0			21.1			45.1		
	Single analysis with interaction term, β, for subgroup						**Single analysis with interaction term, β, for subgroup**					
	mean	sd	CrI	mean	sd	CrI	mean/median	sd	CrI	mean/median	sd	CrI
β	-0.21	0.11	(-0.42,0.01)				-0.29	0.26	(-0.86,0.20)			
LOR	-0.11	0.10	(-0.30,0.09)	-0.31	0.05	(-0.42,-0.21)	-0.07	0.20	(-0.48,0.36)	-0.36	0.16	(-0.72,-0.07)
OR	0.90	0.09	(0.74,1.09)	0.73	0.04	(0.66,0.81)	0.95	0.21	(0.62,1.43)	0.70	0.11	(0.49,0.94)
σ	-	-	-	-	-	-	0.19	0.20	(0.01,0.76)			
resdev*	45.9						42.6					
pD	21.0						24.2					
DIC	66.9						66.8					

† compare to 10 data points
‡ compare to 28 data points
* compare to 38 data points

These ideas extend naturally, but not necessarily easily, from binary effect modifiers to multiple categories. For example, for trials on patients categorised as mild, moderate and severe, two interaction terms can be introduced one for moderate compared to mild, the second for severe compared to mild. Alternatively, disease severity can be examined as a continuous covariate (see Section 4.3) or as regression on baseline risk (see Section 4.4). A further variant is to introduce random interaction terms. Applications in decision making are probably rare, but such a model could be valuable in the analysis of variation in treatment effects between countries or regions, assuming that a sufficiently large number of trials within regions are available for synthesis. In this case, a different interaction term is proposed for each region, and these are sampled randomly from a common distribution with a mean and between-region variance. For a meta-analysis of S studies, the random interaction model is then

$$\theta_{ik} = \mu_i + \left(\delta_{i,1k} + \beta_i x_i\right) I_{\{k \neq 1\}}$$

with $\beta_i = B_j$ if trial i was conducted in region j, $i=1,\ldots,S$, $k=1,2$ and

$$B_j \sim N(b, \tau_b^2)$$

where B_j represent the region-specific interaction effects, b represents the mean interaction effect across regions and τ_b^2 is the between-region variability.

4.2. THE RANGE OF INTERACTION MODELS AND THEIR INTERPRETATION IN NETWORK META-ANALYSIS

In principle the same ideas apply to a network synthesis with multiple treatments. However, there are a very large number of models that can be proposed, each with very different implications. Below we set out the range of models available, and discuss their interpretation. Note that although we develop the range of models in the context of sub-group effects, sub-group interaction models are structurally the same as meta-regression with continuous covariates (Section 4.3) or meta-regression on baseline risk (Section 4.4), and exactly the same range of models can be developed in these cases, too. We argue that only a restricted class of interaction models have interpretations that are likely to be useful in a practical decision making context. This conclusion is then applied not only to sub-group interactions, but to continuous covariates and to baseline risk as a covariate.

We set out three general approaches to meta-regression models in a multiple treatment context: separate and unrelated interaction terms for each treatment; exchangeable and related interaction terms; and one single interaction effect for all treatments.

Consider a binary between-trial covariate, for example primary versus secondary prevention, in a case where multiple treatments $T_1,T_2,\ldots,T_s$ are being compared. Following the approach to consistency models adopted in TSD2,[8] we have (s-1) basic parameters for the relative treatment effects $d_{12},d_{13},\ldots,d_{1s}$ of each treatment relative to treatment 1. As before, we shall assume that treatment 1 is a placebo or standard treatment, which will be taken as the reference treatment in the network meta-analysis (see TSD2[8]). The remaining (s-1)(s-2)/2 treatment contrasts are expressed in terms of these parameters using the consistency equations: for example the effect of treatment 4 compared to treatment 3 is written as $d_{34}=d_{14}-d_{13}$ (see TSD2[8] for details). We can now set out a range of fixed treatment effect interaction models as detailed in Box 1. These models can be easily extended to allow for between-trial variation in treatment effects. Examples are given in Sections 4.3.2 and 4.4.1.

Box 1

1. Independent, treatment-specific interactions.

In this case there is an interaction effect between say, primary/secondary prevention and treatment, but these interactions are different for every treatment. To model this, we introduce as many interaction terms as there are basic treatment effects, for example $\beta_{12},\beta_{13},\ldots,\beta_{1s}$.

Each of these added terms represents the *additional* (interaction) treatment effect in secondary prevention (compared to primary) in comparisons of treatments 2,3,…,s to treatment 1. These terms are exactly parallel to the main effects $d_{12},d_{13},\ldots,d_{1s}$, which now represent the treatment effects in primary prevention populations. As with the main effects for trials comparing say, treatments 3 and 4, the interaction term would be the difference between the interaction terms on the effects relative to treatment 1, so that $\beta_{34}=\beta_{14}-\beta_{13}$. Following the notation in TSD2[8], the fixed treatment effects model for the linear predictor would be

$$\theta_{ik} = \mu_i + \left(d_{t_{i1},t_{ik}} + \beta_{t_{i1},t_{ik}} x_i\right) I_{\{k\neq 1\}} = \mu_i + \left(d_{1t_{ik}} - d_{1t_{i1}} + (\beta_{1t_{ik}} - \beta_{1t_{i1}})x_i\right) I_{\{k\neq 1\}} \qquad (5)$$

with t_{ik} representing the treatment in arm k of trial i, x_i the covariate/subgroup indicator and I defined in equation (1). In this model we set $d_{11}=\beta_{11}=0$. The remaining interaction terms are all unrelated, and would be given unrelated vague prior distributions in a Bayesian analysis.

Thus, the relative treatment effects in secondary prevention are $d_{12}+\beta_{12}$, $d_{13}+\beta_{13},\ldots,d_{1s}+\beta_{1s}$. The interpretation of this model would be that, in effect, the relative efficacy of each of the s treatments in primary prevention populations is *entirely unrelated* to their relative efficacy in secondary prevention populations. One might, indeed, carry out two separate analyses, except that this would make it harder to test the interaction terms, and would also prevent the use of shared variance terms in random treatment effect models, as noted in Section 4.1.1.

2. Exchangeable, related, treatment-specific interactions

This model has the same structure, and the same number of parameters as the model above. The *only* difference is that the (*s*-1) 'basic' interaction terms would not be given unrelated vague priors, but would be drawn from a random distribution with a common mean and between-treatment variance: $\beta_{1k} \sim N(b, \tau^2)$, for treatment k=2,…,s.

The mean interaction effect and its variance would be estimated from the data, although informative priors, that limit how similar or different the interaction terms could also be used.

3. The same interaction effect for all treatments

In this final variant there is a single interaction b term that applies to relative effects of all the treatments relative to treatment 1, so we have $\beta_{1k} = b$ for all treatments k=2,…s. Thus the treatment effects *relative to treatment 1*, d_{12}, d_{13}…d_{1s} in primary prevention, are all higher or lower by the same amount, b in secondary: $d_{12}+b$, $d_{13}+b$ …$d_{1s}+b$. However, the effects of treatments 2,3,...,*s relative to each other* in primary and secondary prevention populations are exactly the same, because the interaction terms now cancel out. This means that the choice of reference treatment 1 becomes important and results for models with covariates are sensitive to this choice. Readers should be aware of the interpretation of parameters when coding models. For example, consider the effect of treatment 4 relative to treatment 3 in secondary prevention. This will be $d_{14} + b - (d_{13} + b) = d_{14} - d_{13}$, which is the same as in primary prevention.

Box 1 A Range of interaction models

When considering models that allow for effect-modification, we come to a series of choice points in model construction. One of the factors that can influence choice of model is the amount of data available. If a fixed treatment effect model is being considered, the unrelated interactions model (model 1, in Box 1) requires *two* connected networks (one for each subgroup) including all the treatments, i.e. with at least (*s*-1) trials in each. With random treatment effects even more data is required to estimate the common between-trials variance.

It may be possible to estimate the exchangeable interaction model (model 2, in Box 1) with less data. However, to use this model we need to have a clear rationale for exchangeability. One rationale could be to allow for different covariate effects for different treatments within the same *class*. Thus, treatment 1 is a standard or placebo treatment while some of the treatments 2,…,*s* belong to a "class". For example, one might imagine one set of

exchangeable interaction terms for aspirin-based treatments for atrial fibrillation relative to placebo, and another set of interactions for warfarin-based treatments relative to placebo.[36]

Although related and exchangeable interactions might seem at first sight to offer an attractive approach, the difficulty is that, even with ample data, their use in clinical practice and in decision making could lead to recommendations that are counter-intuitive and difficult to defend. The claim made by the related and exchangeable interactions model is that there are *real* differences between the relative efficacies of the treatments within the class. If models 1 or 2 were used as a basis for treatment recommendation, a strict application of incremental cost-effectiveness analysis (CEA) could lead to different treatments being recommended for different sub-groups. This might be considered perverse, unless the hypothesis of different interaction effects was shown to be statistically robust.

For these reasons, this document explores only the last of the three general models described in Box 1, which assumes an *identical* interaction effect across all treatments with respect to treatment 1, the reference treatment. An example is given in Section 4.3.2. We do not rule out alternative models for unrelated or exchangeable interaction effects: they certainly have a role in exploratory analyses, or hypothesis-forming exercises, and readers may consult literature for examples and approaches to coding.[36,37]

There are situations where it is reasonable to propose an even more restricted model. Rather than a single interaction term for all active treatments within a class, we could simply have a single interaction term for *all* active treatments, regardless of class. For example, some treatments are so effective that they can virtually eliminate adverse symptoms: here it is almost inevitable that there will be an "interaction" between severity and treatment efficacy, because the extent of improvement is inevitably greater in more severely affected patients. Potential examples might be different classes of biologic therapy for inflammatory arthritis, or perhaps certain treatments for pain relief. In these cases the "interaction" may reflect a property of the scale of measurement, rather than the pharmacological effects of the treatment. Informed clinical and scientific input to model formulation is, as ever, critical.

4.3. META-REGRESSION WITH A CONTINUOUS COVARIATE

When dealing with a continuous covariate, the analysis should use centred covariate values to improve the mixing of the MCMC chains. This is achieved by subtracting the mean covariate value, $\bar{x}$, from each x_i. For the simple pairwise meta-analysis case, the model in equation (4) becomes

$$\theta_{ik} = \mu_i + \left(\delta_{ik} + \beta(x_i - \bar{x})\right) I_{\{k \neq 1\}} \quad (6)$$

The treatment effects are estimated at the mean covariate value and can be un-centred and transformed to produce treatment effect estimates at any covariate value. So the mean treatment effect at covariate value *z*, is $d - \beta(\bar{x} - z)$.

For network meta-analysis, the model in equation (5) can be centred in the same way.

4.3.1. Pair-wise meta-regression with continuous covariate: BCG vaccine Example

A meta-analysis of trials evaluating the efficacy of a BCG vaccine for preventing tuberculosis (TB) suggested that the absolute latitude, or distance from the equator, at which the trials was conducted might influence vaccine efficacy.[38] This corresponds to scenario B1 in Section 1.1.1. Data were available on the number of vaccinated and unvaccinated patients and the number of patients diagnosed with TB during the study follow-up period for each group as well as the absolute latitude at which the trial was conducted (Table 4).

Table 4 BCG Example: number of patients diagnosed with TB, *r*, out of the total number of patients, *n*, in the vaccinated and unvaccinated groups, and absolute latitude at which the trial was conducted, *x*.

	Not vaccinated		Vaccinated		
Trial number	number diagnosed with TB r_{i1}	total number of patients n_{i1}	number diagnosed with TB r_{i2}	total number of patients n_{i2}	Absolute degrees latitude x_i
1	11	139	4	123	44
2	29	303	6	306	55
3	11	220	3	231	42
4	248	12867	62	13598	52
5	47	5808	33	5069	13
6	372	1451	180	1541	44
7	10	629	8	2545	19
8	499	88391	505	88391	13
9	45	7277	29	7499	27
10	65	1665	17	1716	42
11	141	27338	186	50634	18
12	3	2341	5	2498	33
13	29	17854	27	16913	33

The crude odds ratios obtained from Table 4, are plotted (on a log-scale) against distance from the equator in Figure 4 where, for each study, the size of the plotted bubble is proportional to its precision so that larger, more precise studies have larger bubble diameters. It seems plausible that the effect of the vaccine may differ at varying latitudes according to a linear relationship (on the log-odds ratio scale).

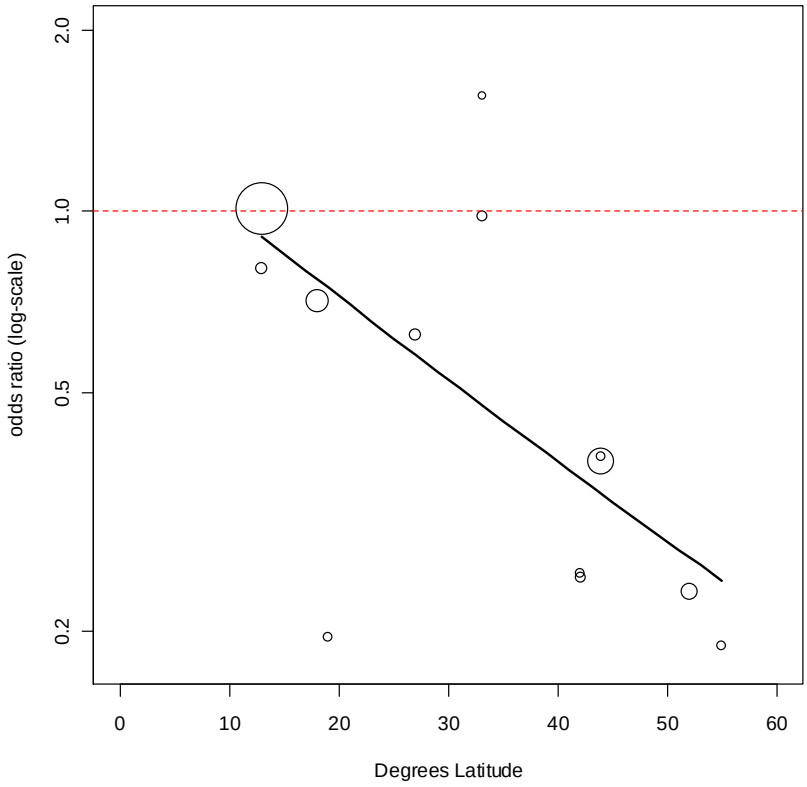

Figure 4 BCG Vaccine for prevention of TB: Plot of the crude odds ratios against absolute distance from the equator in degrees latitude on a log-scale. The size of the bubbles is proportional to the studies' precisions, the horizontal line (dashed) represents no treatment effect and the solid line is the regression line estimated by the RE interaction model. An odds ratio below 1 favours the vaccine.

Assuming a binomial distribution for the number of cases of diagnosed TB in arm k of trial i, $r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$, and letting x_i be the continuous covariate representing absolute degrees latitude, the meta-regression model in equation (6) was fitted to the data with both fixed and random treatment effects and mean covariate value $\bar{x} = 33.46°$ latitude. The treatment effects obtained are log-odds ratios at the mean covariate value. WinBUGS code is presented in the Appendix (Example 4, Programs 4(a) and 4(b)).

The fixed effects model had a very poor fit to the data (posterior mean of the residual deviance of 40 compared to 26 data points) so we present only the results for the RE model (based on 50,000 iterations from 3 independent chains after a burn-in of 20,000). The results obtained for a RE model with and without the covariate 'absolute degrees latitude' are presented in Table 5. Note that, the treatment effect for the model with covariate adjustment is interpreted as the effect at the mean value of the covariate (33.46° latitude). The estimated log-odds ratios at different degrees latitude are represented by the solid line in Figure 4.

Table 5 Posterior mean, standard deviation (sd) and 95% Credible Interval (CrI) of the log-odds ratio (LOR), odds ratio (OR) and the interaction estimate (*b*), and posterior median, sd and 95% CrI of the between-trial heterogeneity (σ) for the number of patients diagnosed with TB (LOR<0 and OR<1 favour Vaccination) for the RE models without covariate and measures of model fit: posterior mean of the residual deviance (resdev), number of parameters (pD) and DIC.

	No covariate			Model with Covariate†		
	mean/median	sd	CrI	mean/median	sd	CrI
b	-	-	-	-0.032	0.009	(-0.05,-0.01)
LOR	-0.762	0.220	(-1.21,-0.34)	-0.763	0.126	(-1.04,-0.52)
OR	0.478	0.107	(0.30,0.71)	0.470	0.059	(0.35,0.59)
σ	0.649	0.202	(0.39,1.17)	0.272	0.188	(0.03,0.75)
resdev*	26.1			30.4		
pD	23.5			21.1		
DIC	49.6			51.5		

* Compare to 26 data points

† treatment effects are at the mean value of the covariate Latitude=33.46°

Comparing the values of the DIC, it would appear that the models with and without the covariate are not very different, differences of less than 3 or 5 are not considered important – although the model without covariates has a smaller posterior mean of the residual deviance, the model with the covariate allows for more shrinkage of the random treatment effects, resulting in a smaller effective number of parameters (pD). We can however see that the heterogeneity is considerably reduced in the model with the covariate: the posterior medians are 0.649 for the model with no covariate and 0.270 for the model with covariate, and the CrI for the interaction term *b* does not include zero (Table 5). In deciding whether a covariate should be included, the posterior mean of the regression coefficient should be compared to the posterior standard deviation. The DIC is not a reliable criterion for deciding whether to include a covariate in RE models. This is because RE models can fit the data equally well, whatever the between-trial variation.

4.3.2. Network meta-regression with continuous covariate: Certolizumab Example

A review of trials of Certolizumab Pegol (CZP) for the treatment of rheumatoid arthritis (RA) in patients who had failed on disease-modifying anti-rheumatic drugs (DMARDs), including Methotrexate (MTX), was conducted for a recent single technology appraisal at NICE.[39] Twelve MTX controlled trials were identified, comparing seven different treatments: Placebo plus MTX (coded 1), CZP plus MTX (coded 2), Adalimumab plus MTX (coded 3), Etanercept plus MTX (coded 4), Infliximab plus MTX (coded 5), Rituximab plus MTX (coded 6) and Tocilizumab plus MTX (coded 7); forming the network presented in Figure 5. This type of network, where comparisons are all relative to the same treatment, is often called a “star network”.

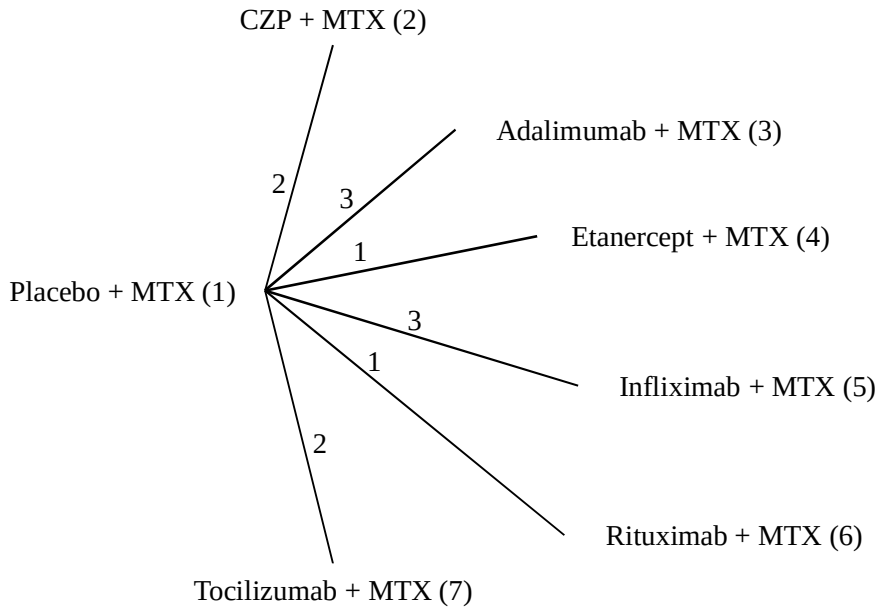

Figure 5 Certolizumab example: Treatment network. Lines connecting two treatments indicate that a comparison between these treatments has been made. The numbers on the lines indicate how many RCTs compare the two connected treatments.

Table 6 shows the number of patients achieving ARC50 at 6 months (ARC50 at 3 months was used when this was not available), r_{ik}, out of all included patients, n_{ik}, for each arm of the included trials, along with the mean disease duration in years for patients in each trial, x_i (i=1,…,12; k=1,2). It is thought that mean disease duration can affect relative treatment efficacy, and this corresponds to scenario B2 in Section 1.1.1. The crude odds ratios (OR) from Table 6, are plotted (on a log-scale) against mean disease duration in Figure 6, with the numbers 2 to 7 representing the OR of that treatment relative to Placebo plus MTX (chosen as the reference treatment). The crude OR for the Abe 2006 study was calculated by adding 0.5 to each cell.

Table 6 Certolizumab Example: number of patients achieving ACR50 at 6 months, *r*, out of the total number of patients, *n*, in the arms 1 and 2 of the 12 trials, and mean disease duration (in years) for patients in trial i, x_i. All trial arms had MTX in addition to the placebo or active treatment.

			Arm 1		Arm 2		
Study name	Treatment in arm 1 t_{i1}	Treatment in arm 2 t_{i2}	number achieving ACR50 r_{i1}	total number of patients n_{i1}	number achieving ACR50 r_{i2}	total number of patients n_{i2}	Mean disease duration (years) x_i
RAPID 1	Placebo	CZP	15	199	146	393	6.15
RAPID 2	Placebo	CZP	4	127	80	246	5.85
Kim 2007	Placebo	Adalimumab	9	63	28	65	6.85
DE019	Placebo	Adalimumab	19	200	81	207	10.95
ARMADA	Placebo	Adalimumab	5	62	37	67	11.65
Weinblatt 1999	Placebo	Etanercept	1	30	23	59	13
START	Placebo	Infliximab	33	363	110	360	8.1
ATTEST	Placebo	Infliximab	22	110	61	165	7.85
Abe 2006*	Placebo	Infliximab	0	47	15	49	8.3
Strand 2006	Placebo	Rituximab	5	40	5	40	11.25
CHARISMA*	Placebo	Tocilizumab	14	49	26	50	0.915
OPTION	Placebo	Tocilizumab	22	204	90	205	7.65

* ACR50 at 3 months

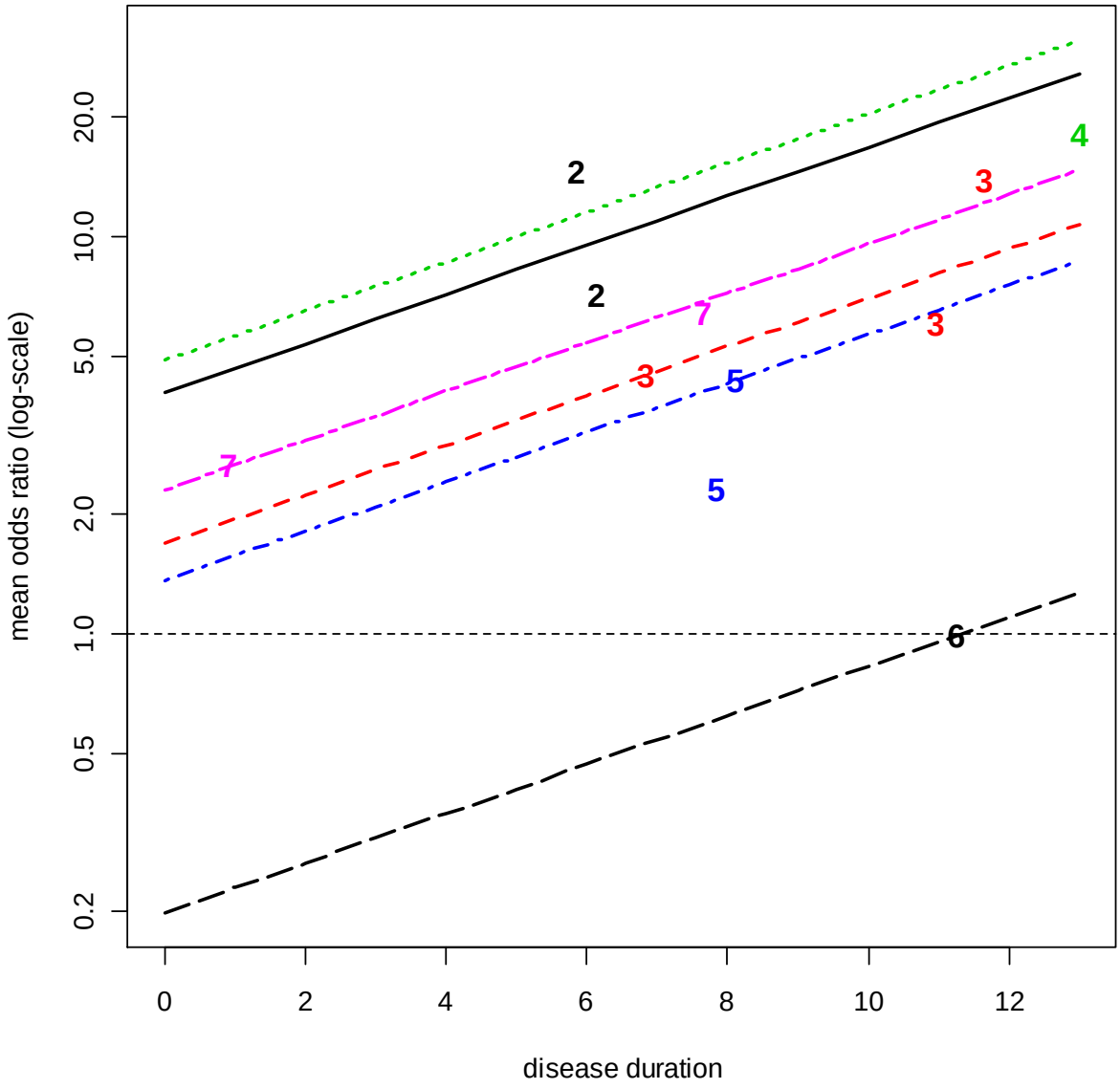

Figure 6 Certolizumab Example: Plot of the crude odds ratios (on a log-scale) of the six active treatments relative to Placebo plus MTX, against mean disease duration (in years). The plotted numbers refer to the treatment being compared to Placebo plus MTX and the lines represent the relative effects of the following treatments (from top to bottom) compared to Placebo plus MTX based on a RE meta-regression model: Etanercept plus MTX (treatment 4, dotted green line), CZP plus MTX (treatment 2, solid black line), Tocilizumab plus MTX (treatment 7, short-long dash purple line), Adalimumab plus MTX (treatment 3, dashed red line), Infliximab plus MTX (treatment 5, dot-dashed dark blue line) and Rituximab plus MTX (treatment 6, long-dashed black line). Odds ratios above 1 favour the plotted treatment and the horizontal line (thin dashed) represents no treatment effect.

We will fit a model which assumes a common interaction effect for all treatments. The FE model with common interaction term is described in Box 1. To fit the equivalent random treatment effects model with covariate centring, we re-write equation (5) as

$$\theta_{ik} = \text{logit}(p_{ik}) = \mu_i + \left(\delta_{i,1k} + (\beta_{1t_{ik}} - \beta_{1t_{i1}})(x_i - \bar{x})\right) I_{\{k \neq 1\}} \quad (7)$$

where $\bar{x} = 8.21$, β_{11}=0, β_{1k}=b (k=2,…,7) and $\delta_{i,1k} \sim N(d_{1t_{ik}} - d_{1t_{i1}}, \sigma^2)$.

The model can be expressed, and coded for computer implementation, in many ways. In this formulation we retain the treatment-specific interaction effects, but set them all equal to *b*. This guarantees that the terms cancel out in active vs active comparisons. This formulation mirrors the code provided in the Appendix (Example 5).

Finally, note that since pairwise meta-analysis is a special case of network meta-analysis (TSD2[8]), in the case of only two treatments, the model in equation (7) simplifies to the model in equation (6).

The basic parameters d_{1k} and b are given non-informative normal priors. See Example 5 in the Appendix for details on the prior for the between-trials standard deviation and corresponding WinBUGS code.

Since the analysis used centred covariate values, the treatment effects obtained are the estimated log-odds ratios at the mean covariate value (8.21 years in this case), which can be un-centred and transformed to produce the estimate at covariate value z from $d_{1k} - b(\bar{x} - z)$, k=2,…,7.

Table 7 shows the results of fitting fixed and random treatment effects network meta-analyses (see TSD2[8]) and interaction models with disease duration as the covariate (results are based on 100,000 iterations from 3 independent chains after a burn-in of 40,000). The estimated odds ratios for different durations of disease are represented by the parallel lines in Figure 6.

Table 7 Certolizumab Example: Posterior mean, standard deviation (sd) and 95% Credible Interval (CrI) for the interaction estimate (b), and log-odds ratios d_{XY} of treatment Y relative to treatment X, and posterior median, sd and 95% CrI of the between-trial heterogeneity (σ) for the number of patients achieving ACR50 for the fixed and random effects models with and without covariate 'disease duration' and measures of model fit: posterior mean of the residual deviance (resdev), effective number of parameters (pD) and DIC. Treatment codes are given in Figure 5.

	No covariate						Covariate 'disease duration'					
	FE			RE†			FE			RE†		
	mean	sd	CrI	mean/ median	sd	CrI	mean	sd	CrI	mean/ median	sd	CrI
b	-	-	-	-	-	-	0.14	0.06	(0.01,0.26)	0.14	0.09	(-0.03,0.32)
d_{12}	2.21	0.25	(1.73,2.72)	2.27	0.39	(1.53,3.10)	2.50	0.29	(1.96,3.08)	2.57	0.42	(1.79,3.44)
d_{13}	1.93	0.22	(1.52,2.37)	1.97	0.33	(1.33,2.64)	1.66	0.25	(1.19,2.16)	1.71	0.34	(1.04,2.41)
d_{14}	3.47	1.34	(1.45,6.74)	3.46	1.41	(1.26,6.63)	2.82	1.34	(0.71,5.96)	2.77	1.42	(0.42,6.01)
d_{15}	1.38	0.17	(1.06,1.72)	1.48	0.33	(0.90,2.21)	1.40	0.17	(1.08,1.74)	1.48	0.30	(0.95,2.15)
d_{16}	0.00	0.71	(-1.40,1.39)	0.01	0.82	(-1.61,1.63)	-0.42	0.73	(-1.86,1.04)	-0.44	0.84	(-2.08,1.21)
d_{17}	1.65	0.22	(1.22,2.10)	1.56	0.38	(0.77,2.28)	1.98	0.28	(1.45,2.53)	2.00	0.45	(1.12,2.93)
σ	-	-	-	0.34	0.20	(0.03,0.77)	-	-	-	0.28	0.19	(0.02,0.73)
resdev*	37.6			30.9			33.8			30.2		
pD	18.0			21.2			19.0			21.3		
DIC	55.6			52.1			52.8			51.4		

* compare to 24 data points

† Using informative prior for σ

The DIC and posterior means of the residual deviances for the 4 models in Table 7 do not decisively favour a single model. Comparing only the FE models we can see that the fit is improved by including the covariate interaction term *b* which also has a CrI which does not include zero. Although the RE model with covariate reduces the heterogeneity (from a posterior median of 0.34 in the RE model with no covariate to 0.28 for the RE model with covariate) the CrI for the interaction parameter *b* includes zero. The meta-regression models are all reasonable but not strongly supported by the evidence. Nevertheless the finding of smaller treatment effects with a shorter disease duration has been reported with larger sets of studies,[37] and the implications for the decision model need to be considered. The issue is whether or not the use of biologics should be confined to patients whose disease duration was above a certain threshold. This is not an unreasonable idea but it would be difficult to determine this threshold on the basis of the regression in Figure 6 alone. The slope is largely determined by treatments 3 and 7 (Adalimumab and Tocilizumab) which are the only treatments trialled at more than one disease duration, and which appear to have different effects at each duration. The linearity of relationships is highly questionable and the prediction of negative effects for treatment 6 (Rituximab) is not plausible. This suggests that the meta-regression model used is not plausible and other explorations of the causes of heterogeneity should be undertaken (see Section 4.4.1).

4.4. META-REGRESSION ON BASELINE RISK

The meta-regression model on baseline risk is the same as in equation (7), but now $x_i=\mu_i$, the trial-specific baseline for the control arm in each trial. An important property of this Bayesian formulation is that it takes the "true" baseline (as estimated by the model) as the covariate and automatically takes the uncertainty in each μ_i into account.[40,41] Naïve approaches which regress against the observed baseline risk fail to take into account the correlation between the treatment effect and baseline risk, and the consequent regression to the mean phenomenon.

It is important to note that the covariate value μ_i is on the same scale as the linear predictor (e.g. the logit, log or identity scales – see TSD2[8]) and therefore the mean covariate value for centring needs to be on this scale too. For example, when using a logit link function, the covariate should be centred by subtracting the mean of the log-odds in the baseline arm (k=1) of each trial which compares treatment 1 from μ_i. In a network meta-analysis context, the treatment in arm 1 will not always be treatment 1 (the reference treatment). However, for the

model in equation (7), which assumes the same interaction effect for all treatments compared to treatment 1, the regression terms will cancel for all other comparisons, so no baseline risk adjustment is performed for trials which do not include treatment 1. If fitting one of the other models in Box 1, care should be taken to ensure that the risk being adjusted for refers to the estimated risk for the reference treatment (treatment 1) which may not have been compared in every trial.

4.4.1. Network Meta-regression on baseline risk: Certolizumab Example

Figure 7 shows the crude OR obtained from Table 6 plotted against the baseline odds of ACR50 (on a log-scale), for the Certolizumab example. Numbers 2 to 7 represent the OR of that treatment relative to Placebo plus MTX (chosen as the reference treatment). For plotting purposes, the crude OR for the Abe 2006 study was calculated by adding 0.5 to each cell and the baseline log-odds was assumed to be 0.01. Figure 7 seems to suggest a strong linear relationship between the treatment effect and the baseline risk (on the log-scale). As discussed in Section 4.3.2 the model in equation (7) assumes that parallel regression lines are fitted to the points in Figure 7, where the differences between the lines represent the true mean treatment effects adjusted for baseline risk.

Both fixed and random treatment effects models with a common interaction term were fitted. The basic parameters d_{1k} and b are given non-informative normal priors N(0,100^2) and $\sigma \sim$ Uniform(0,5). WinBUGS code for meta-regression on baseline risk is given in Example 6 in the Appendix.

The analysis used centred covariate values, achieved by subtracting the mean covariate value (mean of the observed log-odds on treatment 1, $\bar{x} = -2.421$) from each of the estimated μ_i. The treatment effects obtained are then the estimated log-odds ratios at the mean covariate value, which can be un-centred and transformed to produce the estimate at baseline risk z from $d_{1k} - b(\bar{x} - z)$, k=2,…,7.

Table 8 shows the results of the interaction models with fixed and random treatment effects, with baseline risk as the covariate (results are based on 100,000 iterations from 3 independent chains after a burn-in of 60,000). The treatment effects for the models with covariate adjustment are interpreted as the effects for patients with a baseline logit probability of ACR50 of -2.421 which can be converted to a baseline probability of ACR50 of 0.082, using the inverse logit function (TSD2[8], Table 3).

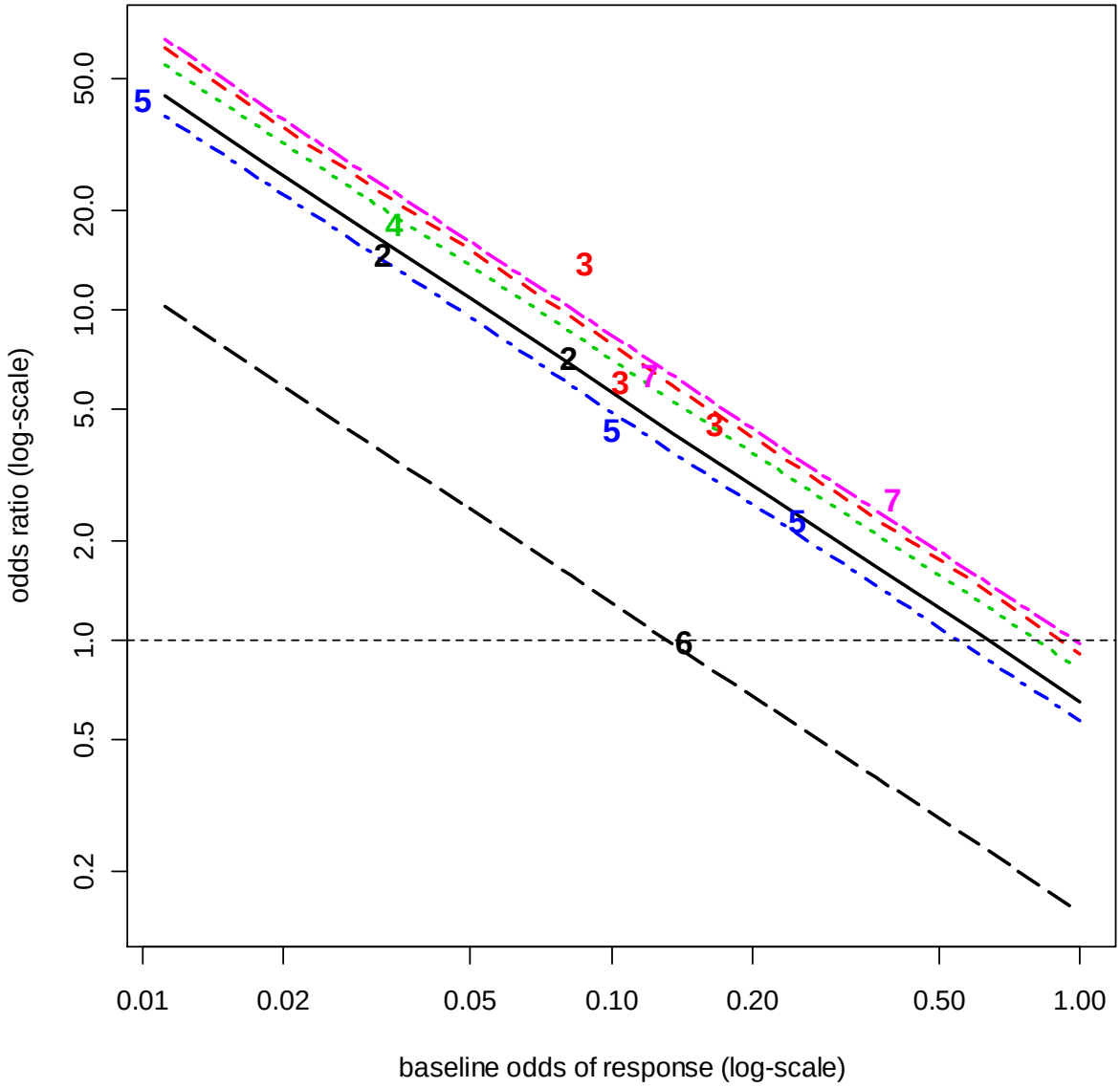

Figure 7 Certolizumab Example: Plot of the crude odds ratios of the six active treatments relative to Placebo plus MTX, against odds of baseline response on a log-scale. The plotted numbers refer to the treatment being compared to Placebo plus MTX and the lines represent the relative effects of the following treatments (from top to bottom) compared to Placebo plus MTX based on a RE meta-regression model: Tocilizumab plus MTX (7, short-long dash purple line), Adalimumab plus MTX (3, dashed red line), Etanercept plus MTX (4, dotted green), CZP plus MTX (2, solid black line), Infliximab plus MTX (5, dot-dashed dark blue line), Rituximab plus MTX (6, long-dashed black line). Odds ratios above 1 favour the plotted treatment and the horizontal line (dashed) represents no treatment effect.

Table 8 Certolizumab Example: Posterior mean, standard deviation (sd) and 95% Credible Interval (CrI) for the interaction estimate (*b*) and log-odds ratios d_{XY} of treatment *Y* relative to treatment *X*. Posterior median, sd and 95% CrI of the between-trial heterogeneity (σ) for the number of patients achieving ACR50 for the fixed and random effects models with covariate 'baseline risk' with measures of model fit: posterior mean of the residual deviance (resdev), number of parameters (pD) and DIC. Treatment codes are given in Figure 5.

	FE			RE		
	mean	sd	CrI	mean/median	sd	CrI
b	-0.93	0.09	(-1.03,-0.69)	-0.95	0.10	(-1.10,-0.70)
d_{12}	1.85	0.10	(1.67,2.06)	1.83	0.24	(1.35,2.29)
d_{13}	2.13	0.11	(1.90,2.35)	2.18	0.22	(1.79,2.63)
d_{14}	2.08	0.34	(1.47,2.80)	2.04	0.46	(1.19,2.94)
d_{15}	1.68	0.10	(1.49,1.86)	1.71	0.22	(1.30,2.16)
d_{16}	0.36	0.50	(-0.72,1.27)	0.37	0.59	(-0.86,1.45)
d_{17}	2.20	0.14	(1.93,2.46)	2.25	0.27	(1.75,2.79)
σ	-	-	-	0.19	0.19	(0.01,0.70)
resdev*	27.3			24.2		
pD	19.0			19.4		
DIC	46.3			43.6		

* compare to 24 data points

Both the fixed and random effects models with covariate have a credible region for the interaction term which is far from zero, suggesting a strong interaction effect between the baseline risk and the treatment effects. The estimated odds ratios for different durations for the RE model with baseline risk interaction are represented by the different parallel lines in Figure 7. The DIC statistics and the posterior means of the residual deviance also marginally favour the RE model with the covariate.

An important point to note is that the assumption of a common regression term *b* allows the interaction parameter to be estimated even for comparisons which only have one trial. It also allows estimation of treatment effects at values of the baseline risk outside the ranges measured in trials involving certain comparisons. For example, there is only one trial comparing Rituximab plus MTX (treatment 6) with Placebo plus MTX. The model assumptions imply that a line parallel to the others is drawn through this point (Figure 7). This analysis also suggests that adding Rituximab to MTX may be of much less benefit to patients than the other treatments and predicts, perhaps implausibly, that it can be harmful if baseline risk is above 0.15.

The striking support in Figure 7 for a single interaction term for all treatments, except maybe treatment 6, has several implications for decision making and for synthesis in practice. Firstly it clearly suggests a relation between efficacy and baseline risk that needs to be incorporated into CEA models. Secondly, Figure 7 illustrates how variation in effect size due to a

covariate will, if not controlled for, introduce severe heterogeneity in pairwise meta-analysis and potential inconsistency in network synthesis. It is clear that *both* the differences between trials (within treatments) *and* the differences between the anti TNF-α drugs are minimal once baseline risk is accounted for.

4.5. INDIVIDUAL PATIENT DATA IN META-REGRESSION

Individual Patient Data meta-analyses have been described as the gold standard[42] and they enjoy certain advantages over syntheses conducted on summary data, including the possibility of standardising analysis methods.[43] Further, when *patient level* covariates are of interest, using the IPD to regress individual patient characteristics on individual patient outcomes will produce a more powerful and reliable analysis[44,45] compared to the use of aggregate outcome and covariate data considered in Sections 4.1-4.4. Not only is such an analysis usually much more powerful than one based on aggregate data, it can avoid the potential ecological biases. An IPD meta-regression analysis is essential when dealing with a continuous covariate and a continuous outcome.

Below we distinguish the situation where IPD is available on all trials and where it is only available on a subset of trials.

4.5.1. How to use Individual Patient Data on patient level covariates to explore heterogeneity

In meta-analysis of IPD, historically, two broad approaches have been considered, the one- and two-step approaches.[46] In a two-step approach, the analyst first estimates the effect size(s) of interest from each study, together with a measure of uncertainty (e.g. standard error), and then conducts a meta-analysis in the standard way using this summary data. In the context of exploring heterogeneity, the effect size could relate to a treatment by covariate interaction.[47] In some circumstances, it may be possible to carry out an IPD analysis such as this even if the analyst does not have access to all the IPD, i.e. owners of the data may be willing to calculate and supply such interaction effects when they are not willing to supply the whole IPD dataset. However, such an approach becomes cumbersome/infeasible if multiple covariates are to be considered simultaneously.

The two-stage approach can be useful for inference about the existence of an interaction, but it is unhelpful for decision making where the main effects and interactions need to be estimated simultaneously so that parameter correlations can be propagated through the model.

In a one-step approach all the IPD is compiled into a single dataset and analysed simultaneously while preserving the within study comparisons within the data. IPD random-effects pairwise meta-analysis models have been developed for continuous,[48,49] binary,[50] survival[51] and ordinal[52] variables and all can allow the inclusion of patient level covariates. Although most of the models are presented in the single pair-wise comparison context, it is possible to extend them to a network meta-analysis context.[48,53] A recent paper[47] considers simple criteria for determining the potential benefits of IPD to assess patient level covariates and this is recommended reading.

Thus, treatment by covariate interactions can be estimated exclusively using between-study information when only summary data are available (meta-regression) and exclusively using within-study (variability) information if IPD are available. However, a subtlety when using IPD is that both between- *and* within-study coefficients can be estimated.[48] This can be achieved by including two covariates: the mean covariate value in that study (i.e. each individual in a study gets the same value – which is the value that would be used if an aggregate meta-regression analysis were being conducted), and a second covariate which is the individual patient response minus the mean value in that study. Specific modelling details are available elsewhere.[54] Note that this applies most naturally to continuous covariates, but it can also be applied to binary covariates (for example if the binary covariate is sex, the between-study covariate would be the proportion of women).

There are a number of ways in which these dual effect (within and between) models can be used. The most appealing option is to use the estimate derived exclusively from the within-trial variability, since this is free from ecological/aggregation biases and other potential sources of confounding between studies. Potentially, power could be gained by including the information in the between trial variability by having the same parameter for within and between covariates. This, of course, comes at the cost of potentially inducing bias. It has been suggested[54] that a statistical test of the difference between the two estimates could be carried out and the decision of whether to have the same interaction effect for within and between covariates could be based on this test. However, we suspect this test will have low power in many situations, and further investigation of this approach is required before it can be recommended.

4.5.2. Using a combination of Individual Patient Data and Aggregate Data

The situation may exist where IPD is available from a number of, but not all, relevant studies. When this is the case, in a pair wise meta-analysis context, there are three potential options

available for exploring heterogeneity. The first is to exclude all trials for which IPD is not available. This keeps the analysis simple, and can be based exclusively on within study comparisons (using the methods described in the previous section), but has the disadvantage of not including all of the relevant trials. Furthermore, the analysis could potentially be biased if the reason for not providing IPD is related to the treatment effect. The second is to carry out a meta-regression on the aggregate data. This would potentially mean all trials could be included, but the benefits of having some IPD would be forgone. Finally, models have been developed which allow the incorporation if IPD, where available, and aggregate data where not.[55] This approach allows all the data to be included at the most detailed level available from all the studies, but as for an IPD only analysis, a decision has to be made on whether between study variability is to be included in the estimation of effects. Again, a test of the difference between the effect using between and within study variability can be constructed and used to decide which approach to take (but again noting its probable low power in many contexts may make this a problematic approach). Models which allow the incorporation of IPD and aggregate data have been described for binary[55,56] and continuous[57] outcomes.

Little work has been done to date on the simultaneous use of IPD and aggregate data in a network meta-analysis context. It is quite conceivable that IPD may be available for all trials of some comparisons, while none may be available for others. This may be particularly true for Single Technology Appraisals done by industry where a company may have complete access to trial data for their own products, but only aggregate data on competitors' products. As described in Section 4.2, a decision has to be made on whether interaction effects with placebo/usual care are assumed to be the same, exchangeable, or different across treatments. Although we have suggested a single interaction parameter for all treatments within the same class, models for all these possibilities can be constructed. Extensions to the dual within- and between-covariate models are possible and there have been initial explorations of this.[58]

The availability of IPD for several different treatments would allow a much more thorough investigation of whether patient-level interactions are the same across treatments.

5. BIAS AND BIAS-ADJUSTMENT

In this section we examine approaches to bias adjustment for both internal and external biases. The difference between "bias adjustment" and the meta-regression models described above is slight but important. In meta-regression we concede that even within the formal scope of the decision problem there are distinct differences in relative treatment efficacy. In

bias adjustment, we have in mind a target population for decision making, but the evidence available, or at least some of the evidence, provides "biased", or potentially biased, estimates of the target parameter, perhaps because the trials have internal biases, perhaps because they concern different populations or settings, or both. Box 2 summarises four approaches to bias adjustment, which are discussed in more detail below.

Although regression in usually seen as a form of adjustment for differences in covariates, we still refer to it as a method for "bias adjustment" since covariates affect the 'external validity' of trials, which has been seen as a bias adjustment issue.[59]

Box 2

- Meta-regression (Section 4): A decision is required for a specific target population and specific treatments, but much of the evidence involves other populations, or other (similar) treatments. This approach is suitable for pair-wise meta-analysis, IC and Network meta-analysis of RCTs and works better with larger datasets.
- Adjustment for potential bias associated with trial-level markers:[60] The evidence base contains some studies with markers of potential bias, and a prior distribution for this bias can be estimated from external meta-epidemiological data. This approach is suitable for pair-wise meta-analysis, IC, Network meta-analysis and RCTs of mixed "quality", but could be extended to meta-analyses consisting of a mixture of trials and observational data. This approach is good for small datasets, including single trials, but depends on the relevance of the meta-epidemiological data used.
- Estimation and adjustment of bias associated with trial-level markers:[61-63] The extent of the bias can be estimated internally from the existing trial evidence. This approach is suitable for IC or Network meta-analysis of RCTs of mixed "quality", but could be extended to mixtures of trial and observational data. Works better with larger datasets.
- Elicitation of internal and external bias distributions from experts:[59] Can be applied to any of the situations above and is suitable for pair-wise meta-analysis, IC, Network meta-analysis of RCTs and/or observational studies.[64] This approach is good for small datasets, including single studies, but can be very time consuming.

Box 2 Different approaches to bias adjustment

5.1. COVARIATE ADJUSTMENT BY META-REGRESSION

This method uses the meta-regression models in Sections 4 or 5. The approach is an option in cases where evidence on the treatment effects in the target population or treatment is limited, but further information exists on other related populations or treatments. If it was felt that the treatment effects were systematically different in the two groups, then a meta-regression analysis would be a way in which to "borrow strength" from an additional set of related trials. For example, in the case of biologic treatments for RA, suppose a decision was required on treatments for patients who had failed on non-biologic DMARDs but who were unable to take MTX. Ideally, trials involving biologics and placebo would be needed. It might be felt that insufficient data was available in this patient group, but that the larger body of data on Biologics + MTX *vs* Placebo + MTX could be used. An interaction model could be used to borrow strength from this additional body of data, while adjusting for a common additional effect of biologics against placebo in the presence of MTX. Note that in this case the adjustment would only be relevant to the comparisons of biologics against placebo, *not* to the comparisons between biologics (see Box 1).

Investigators would also have the option of assuming that there was no interaction, i.e. that the effect of biologics against placebo was the same when taken with or without MTX. In this case the entire body of data could be used to estimate the treatment effects of biologics relative to placebo and relative to each other, without the introduction of any interaction terms.

5.2. ADJUSTMENT FOR BIAS BASED ON META-EPIDEMIOLOGICAL DATA

Schulz et al.[4] compared results from "high quality" RCTs to results from trials with certain indicators of potentially lower quality: lack of allocation concealment, or lack of double blinding. Their dataset included over 30 meta-analyses, in each of which both "high" and "low" quality trials were present. Their results suggested the relative treatment effect in favour of the newer treatment was, on average, higher in the lower quality studies. The effect was large, with odds ratios in favour of the newer treatment on average about 1.6 times higher.

Confronted by trial evidence of mixed quality, investigators have had two options: they can restrict attention to studies of high quality, or they can include all trials, of both high and low quality, in a single analysis. Both options have disadvantages: the first ignores what may be a

substantial proportion of the evidence; the second risks delivering a biased estimate of the treatment effect.

Welton et al.[60] suggest an approach that uses all the data, but simultaneously adjusts and down-weights the evidence from lower quality studies. For a pairwise meta-analysis, the model for the "high quality" data is the standard model introduced in TSD2:[8]

$$\theta_{ik} = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}} \tag{8}$$

For the lower quality data, the assumption is that each trial provides information, not on $\delta_{i,1k}$, but on a "biased" parameter $\delta_{i,1k} + \beta_i$, where the trial-specific bias terms β_i are drawn from a RE distribution, with a mean b_0 representing the expected bias, and a between-trials variance κ^2. Thus, for the lower quality trials:

$$\begin{aligned} \theta_{ik} &= \mu_i + (\delta_{i,1k} + \beta_i) I_{\{k \neq 1\}} \\ \beta_i &\sim N(b_0, \kappa^2) \end{aligned} \tag{9}$$

Values for b_0 and between trials variance are obtained from a Bayesian analysis of an external dataset, for example from collections of meta-analyses,[4,65] and these values can be plugged into the prior distribution in equation (9). This analysis must produce at least three estimates: a value for the expected bias b_0, a value for the standard error in the estimate of b_0, and a value for the between-study variability on bias. These values can then be used to inform priors for bias parameters to adjust and down-weight treatment effects for lower-quality trials in a new meta-analysis. Welton et al.[60] commented on the assumptions required by this method of bias adjustment. The analysis hinges critically on whether the study-specific biases in the dataset of interest can be considered exchangeable with those in the meta-epidemiological data used to provide the prior distributions used for adjustment, and indeed on whether they *would* be considered exchangeable by the relevant stakeholders in the decision.

At the time of writing, analyses of meta-epidemiological data are not yet available to inform priors while plausibly satisfying the exchangeability requirements. Nonetheless, one might take the view that any reasonable bias adjusted analysis is likely to give a better reflection of the true parameters than an unadjusted analysis. Welton et al.[60] suggest that, even when there are doubts about a particular set of values for the bias distribution, investigators may wish to run a series of sensitivity analyses to show that the presence of studies of lower quality, with potentially over-optimistic results, is not having an impact on the decision. Extensive meta-epidemiological analyses are currently an area of active research interest. It is already clear

that the degree of bias is dependent on the nature of the outcome measure, being greater with subjective (patient- or physician-reported) outcomes, and virtually undetectable with mortality.[5] Increasingly detailed information on quality-related bias is being published, and it is likely that sets of priors tailored for particular outcome types and disease conditions will become available in the future.

In principle the same form of bias adjustment could be extended to other type of bias, such as novelty bias, sponsorship bias, or small study bias, or to mixtures of RCTs and observational studies. Each of these extensions, however, depends on detailed and far-ranging analyses of very large meta-epidemiological datasets which have not yet been performed.

We turn next to a method that removes the difficulties associated with the strong "exchangeability" assumptions by estimating the parameters of the bias distribution, b_0 and variance κ^2, *internally*, within the dataset of interest.

5.3. ESTIMATION AND ADJUSTMENT FOR BIAS IN NETWORKS OF TRIALS

The bias model above (equations (8) and (9)) can also be estimated *internally*, without recourse to external data. Imagine a set of trials, some of which are "high" and some "low" quality. One can always use such analyses to learn about the size of bias and – with enough data – the variability in bias across studies, but one cannot always use them to borrow strength from biased data. For example, if there are only two treatments, the analysis would tell us about the bias distribution, but it would add nothing to our knowledge of the true treatment effect: for this we might just as well look at the high quality data alone.

However, with indirect comparisons or, in a network synthesis, if we assume that the mean and variance of the study-specific biases is the same for each treatment, then it is possible to simultaneously estimate the treatment effects and the bias effects in a single analysis, and thus to produce treatment effects that are based on the entire body of data, including both high and low quality studies, and also adjusted for bias.[61] The model is exactly the same as in the previous section

$$\theta_{ik} = \mu_i + (\delta_{i,1k} + \beta_{ik} x_i) I_{\{k \neq 1\}}$$

with x_i=1 if study i is considered to be at risk of bias and zero otherwise, and β_{ik} is the trial-specific bias of the treatment in arm k relative to the treatment in arm 1 of trial i. If A is placebo or standard treatment, and B,C,D are all active treatments, it would be reasonable to expect the same bias distribution to apply to the AB, AC, and AD trials. But it is less clear

how to code the bias model for BC, BD, and CD trials. We might make a distinction between active vs placebo/usual care and active vs active trials. If we assume that the average bias is always in favour of the newer treatment, then this becomes a model for novelty bias.[62] Another approach might be to propose a separate mean bias term for active vs active comparisons.[61] For example, the first type of trials would have a bias term which is assumed to follow a normal distribution with mean bias b_1: $\beta_{ik} \sim N(b_1, \kappa^2)$. Active vs active trials could be assumed to have a different mean bias b_2, $\beta_{ik} \sim N(b_2, \kappa^2)$, which could be assumed to favour the newest treatment or set to zero (see Dias et al.[61] for further details).

The method can in principle be extended to include syntheses that are mixtures of trials and observational studies, but this does not appear to have been attempted yet. It can also be extended to any form of "internal" bias. Salanti et al.[62] adopted this model in their study of novelty bias in cancer trials. A particularly interesting application is to "small-study bias", which is one interpretation of "publication bias". The idea here is that the smaller the study the greater the bias. The "true" treatment effect can therefore be conceived as the effect that would be obtained in a study of infinite size. This, in turn, is taken to be the intercept in a regression of the treatment effect against the study variance. Moreno et al.[63,66] show that the bias-adjusted estimate from this approach approximates closely to the results found in a simple meta-analysis based on a register of prospectively reported data. Once again, in larger networks, some care would need to be exercised in how to code the direction of bias in "active-active" studies.

Like the methods described in Section 5.2, these methods may be considered by some as semi-experimental. There is certainly a great need for further experience with applications, and there is a particular need for further meta-epidemiological data on the relationships between the many forms of internal bias that have been proposed.[67] However, they appear to represent reasonable and valid methods for bias adjustment, and are likely to be superior to no bias adjustment in situations where data are of mixed quality. At the same time, the method is essentially a meta-regression based on "between-studies" comparisons. There is no direct evidence for a "causal" link between the markers of study quality and the size of the effect. It is therefore important to avoid using the method for small datasets, and to establish that the results are statistically robust, and not dependant on a small number of studies.

Because the underlying bias models in this section and the previous one are the same, it would be perfectly feasible to combine them, although this again has not been done before.

5.4. ELICITATION OF BIAS DISTRIBUTIONS FROM EXPERTS, OR BASED ON DATA

This method[59] is conceptually the simplest of all bias adjustment methods, applicable to trials and observational studies alike. It is also the most difficult and time-consuming to carry out. One advantage may be that it can be used when the number of trials is insufficient for meta-regression approaches (Sections 5.1, 5.3). Readers are referred to the original publication for details, but the essential ideas are as follows. Each study is considered by several independent experts using a pre-determined protocol. The protocol itemizes a series of potential internal and external biases, and each expert is asked to provide information that is used to develop a bias distribution. Among the internal biases that might be considered are selection biases (in observational studies), non-response bias, attrition bias, and so on. A study can suffer from both internal *and* external bias. When this process is complete the bias information on each study from each assessor is combined into a single bias distribution. The assessor distributions are then pooled mathematically. In the original publication the mean and variance of the bias distributions is statistically combined with the original study estimate and its variance, to create what is effectively a new, adjusted, estimate of the treatment effect in that study. The final stage is a conventional synthesis, in which the adjusted treatment effects from each study, and their variances, are treated as the data input for a standard pair-wise meta-analysis, indirect comparison or network synthesis. The methods in TSD2[8] (Section 3.5) can then be applied to the adjusted study-specific estimates.

This methodology[59] in its full form requires considerable time and care to execute. The key idea, replacing a potentially biased study estimate with an adjusted estimate based on expert opinion regarding bias, is one that can be carried out in many ways and with a degree of thoroughness that is commensurate with the sensitivity of the overall analysis to the parameters in question.

6. REFERENCES

1. National Institute for health and Clinical Excellence. Guide to the methods of technology appraisal (updated June 2008). 2008.

2. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2011; last updated April 2012; available from http://www.nicedsu.org.uk

3. Higgins, J.P.T., Green, S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, Wiley, Chichester; 2008.

4. Schulz, K.F., Chalmers, I., Hayes, R.J., Altman, D.G. Empirical Evidence of Bias. Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials. JAMA, J 1995; 273(5):408-412.

5. Wood, L., Egger, M., Gluud, L.L., Schulz, K., Juni, P., Altman, D. et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. British Medical Journal 2008; 336:601-605.

6. Higgins, J.P.T., Thompson, S.G., Spiegelhalter, D.J. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society (A) 2009; 172:137-159.

7. Dias, S., Welton, N.J., Sutton, A.J., Caldwell, D.M., Lu, G., Ades, A.E. NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials. 2011; last updated April 2012; available from http://www.nicedsu.org.uk

8. Dias, S., Welton, N.J., Sutton, A.J., Ades, A.E. NICE DSU Technical Support Document 2: A generalised linear modelling framework for pair-wise and network meta-analysis. 2011; last updated April 2012; available from http://www.nicedsu.org.uk

9. Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R. Introduction to Meta-Analysis. Wiley, Chichester; 2009.

10. Higgins, J.P.T., Thompson, S.G. Controlling the risk of spurious findings from meta-regression. Statistics In Medicine 2004; 23:1663-1682.

11. Rothman, K.J., Greenland, S., Lash, T.L. Modern Epidemiology. 3 ed. Lippincott, Williams & Wilkins, Philadelphia; 2008.

12. Govan, L., Ades, A.E., Weir, C.J., Welton, N.J., Langhorne, P. Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. Statistics In Medicine 2010; 29:1340-1356.

13. Dominici, F. Combining contingency tables with missing dimensions. Biometrics 2000; 56:546-553.

14. Welton, N.J., Johnstone, E.C., Munafo, M.R. A Cost-Effectiveness Analysis of Genetic Testing to Aid Treatment Choice for Smoking Cessation. Nicotine and Tobacco Research 2008; 10(1):231-240.

15. Glenny, A.M., Altman, D.G., Song, F., Sakarovitch, C., Deeks, J.J., D'Amico, R. et al. Indirect comparisons of competing interventions. Health Technology Assesment 2005; 9(26).

16. Lu, G., Welton, N.J., Higgins, J.P.T., White, I.R., Ades, A.E. Linear inference for Mixed Treatment Comparison Meta-analysis: A Two-stage Approach. Res Synth Method 2011; 2:43-60.

17. Sidik, K., Jonkman, J.N. A comparison of heterogeneity variance estimators in combining results of studies. Statistics In Medicine 2007; 26:1964-1981.

18. Rucker, G., Schwarzer, G., Carpenter, J.R., Schumacher, M. Undue reliance on I2 in assessing heterogeneity may mislead. BMC Medical Research Methodology 2008; 8:79.

19. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society (B) 2002; 64(4):583-616.

20. Higgins, J.P.T., Thompson, S.G. Quantifying heterogeneity in a meta-analysis. Statistics In Medicine 2002; 21:1539-1558.

21. Dakin, H., Fidler, C., Harper, C. Mixed Treatment Comparison Meta-Analysis Evaluating the Relative Efficacy of Nucleos(t)ides for Treatment of Nucleos(t)ide-Naive Patients with Chronic Hepatitis B. Value in Health 2010; 13:934-945.

22. Spiegelhalter, D.J., Abrams, K.R., Myles, J. Bayesian approaches to clinical trials and Health-Care Evaluation. Wiley, New York; 2004.

23. Ades, A.E., Lu, G., Higgins, J.P.T. The interpretation of random effects meta-analysis in decision models. Med Decis Making, MED 2005; 25(6):646-654.

24. Welton, N.J., White, I., Lu, G., Higgins, J.P.T., Ades, A.E., Hilden, J. Correction: Interpretation of random effects meta-analysis in decision models. Med Decis Making, MED 2007; 27:212-214.

25. Thompson, S.G. Why sources of heterogeneity in meta-analyses should be investigated. British Medical Journal 1994; 309:1351-1355.

26. Thompson, S.G. Why and how sources of heterogeneity should be invesitgated. In: Egger M., Davey Smith G., Altman D., eds. Systematic reviews in health care: meta-analysis in context. 2nd ed. BMJ Books; London: 2001; 157-175.

27. Sterne, J.A.C., Bradburn, M.J., Egger, M. Meta-analysis in Stata. In: Egger M, Davey Smith G, Altman DG, eds. Systematic reviews in health care: Meta-analysis in context. BMJ Books; London: 2001; 347-369.

28. Marshall, E.C., Spiegelhalter, D.J. Approximate cross-validatory predictive checks in disease mapping models. Stat Med 2003; 22(10):1649-1660.

29. DuMouchel, W. Predictive cross-validation of Bayesian meta-analyses. In: Bernardo J.M., Berger J.O., Dawid A.P., Smith A.F.M., eds. Bayesian Statistics 5. Oxford University Press; Oxford: 1996; 107-127.

30. Egger, M., Davey-Smith, G. Misleading meta-analysis. British Medical Journal 1995; 310:752-754.

31. Higgins, J.P.T., Spiegelhalter, D.J. Being sceptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction. International Journal Of Epidemiology 2002; 31(1):96-104.

32. Li, J., Zhang, Q., Zhang, M., Egger, M. Intravenous magnesium for acute myocardial infarction. Cochrane Database of Systematic Reviews 2007; 2009, Issue 4:Art. No.: CD002755.

33. Raiffa, H., Schlaiffer, R. Applied statistical decision theory. Wiley Classics Library ed. Wiley Interscience, New York; 1967.

34. Dias, S., Welton, N.J., Caldwell, D.M., Ades, A.E. Checking consistency in mixed treatment comparison meta analysis. Statistics In Medicine 2010; 29:932-944.

35. Sutton AJ. Meta-analysis methods for combining information from different sources evaluating health interventions. Thesis/Dissertation: University of Leicester, UK; 2002.

36. Cooper, N.J., Sutton, A.J., Morris, D., Ades, A.E., Welton, N.J. Adressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. Statistics In Medicine 2009; 28:1861-1881.

37. Nixon, R., Bansback, N., Brennan, A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. Statistics In Medicine 2007; 26(6):1237-1254.

38. Berkey, C.S., Hoaglin, D.C., Mosteller, F., Colditz, G.A. A random effects regression model for meta-analysis. Statistics In Medicine 1995; 14:395-411.

39. National Institute for health and Clinical Excellence. Certolizumab pegol for the treatment of rheumatoid arthritis. 2010; TA186. National Institute for Health and Clinical Excellence. NICE technology appraisal guidance.

40. McIntosh, M.W. The population risk as an explanatory variable in research synthesis of clinical trials. Statistics In Medicine 1996; 15:1713-1728.

41. Thompson, S.G., Smith, T.C., Sharp, S.J. Investigating underlying risk as a source of heterogeneity in meta-analysis. Statistics In Medicine 1997; 16:2741-2758.

42. Stewart, L.A., Clarke, M.J. Practical methodology of meta-analyses (overviews) using updated individual patient data. Statistics In Medicine 1995; 14:2057-2079.

43. Riley, R.D., Lambert, P.C., Abo-Zaid, G. Meta-analysis of individual participant data: rationale, conduct, and reporting. British Medical Journal 2010; 340:c221.

44. Lambert, P.C., Sutton, A.J., Abrams, K.R., Jones, D.R. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. Journal of Clinical Epidemiology 2002; 55:86-94.

45. Berlin, J.A., Santanna, J., Schmid, C.H., Szczech, L.A., Feldman, H.I. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Statistics In Medicine 2002; 21:371-387.

46. Simmonds, M.C., Higgins, J.P.T., Stewart, L.A., Tierney, J.F., Clarke, M.J., Thompson, S.G. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. Clinical Trials 2005; 2:209-217.

47. Simmonds, M.C., Higgins, J.P.T. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. Statistics In Medicine 2007; 26:2982-2999.

48. Higgins, J.P.T., Whitehead, A., Turner, R.M., Omar, R.Z., Thompson, S.G. Meta-analysis of continuous outcome data from individual patients. Statistics In Medicine 2001; 20:2219-2241.

49. Goldstein, H., Yang, M., Omar, R.Z., Turner, R.M., Thompson, S.G. Meta-analysis using multilevel models with an application to the study of class size effects. Applied Statistics 2000; 49:399-412.

50. Turner, R.M., Omar, R.Z., Yang, M., Goldstein, H., Thompson, S.G. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. Statistics In Medicine 2000; 19:3417-3432.

51. Tudor Smith, C., Williamson, P.R., Marson, A.G. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. Statistics In Medicine 2005; 24:1307-1319.

52. Whitehead, A., Omar, R.Z., Higgins, J.P.T., Savaluny, E., Turner, R.M., Thompson, S.G. Meta-analysis of ordinal outcomes using individual patient data. Statistics In Medicine 2001; 20:2243-2260.

53. Tudor Smith, C., Marson, A.G., Chadwick, D.W., Williamson, P.R. Multiple treatment comparisons in epilepsy monotherapy trials. Trials 2007; 8:34.

54. Riley, R.D., Steyerberg, E.W. Meta-analysis of a binary outcome using individual participant data and aggregate data. Res Synth Method 2010; 1:17.

55. Riley, R.D., Simmonds, M.C., Look, M.P. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. Journal of Clinical Epidemiology 2007; 60:431-439.

56. Sutton, A.J., Kendrick, D., Coupland, C.A.C. Meta-analysis of individual- and aggregate-level data. Statistics In Medicine 2008; 27:651-669.

57. Riley, R.D., Lambert, P.C., Staessen, J.A., Wang, J., Gueyffier, F., Boutitie, F. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. Statistics In Medicine 2008; 27:1870-1893.

58. Saramago, P., Sutton, A.J., Cooper, N., Manca, A. Synthesizing effectiveness evidence from aggregate-and individual-patient level data for use in cost-effectiveness modelling. Presented at HESG Winter Conference, Centre for Health Economics, University of York. 2011.

59. Turner, R.M., Spiegelhalter, D.J., Smith, G.C.S., Thompson, S.G. Bias modelling in evidence synthesis. Journal of the Royal Statistical Society (A) 2009; 172:21-47.

60. Welton, N.J., Ades, A.E., Carlin, J.B., Altman, D.G., Sterne, J.A.C. Models for potentially biased evidence in meta-analysis using empirically based priors. Journal of the Royal Statistical Society (A) 2009; 172(1):119-136.

61. Dias, S., Welton, N.J., Marinho, V.C.C., Salanti, G., Higgins, J.P.T., Ades, A.E. Estimation and adjustment of bias in randomised evidence by using Mixed Treatment Comparison Meta-analysis. Journal of the Royal Statistical Society (A) 2010; 173(3):613-629.

62. Salanti, G., Dias, S., Welton, N.J., Ades, A.E., Golfinopoulos, V., Kyrgiou, M. et al. Evaluating novel agent effects in multiple treatments meta-regression. Statistics In Medicine 2010; 29:2369-2383.

63. Moreno, S.G., Sutton, A.J., Turner, E.H., Abrams, K.R., Cooper, N.J., Palmer, T.M. et al. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. BMJ 2009; 339:b2981.

64. Thompson, S., Ekelund, U., Jebb, S., Karin Lindroos, A., Mander, A., Sharp, S. et al. A proposed method of bias adjustment for meta-analyses of published observational studies. Internation Journal of Epidemiology 2010; doi: 10.1093/ije/dyq248.

65. Savovic, J., Harris, R.J., Wood, L., Beynon, R., Altman, D., Als-Nielsen, B. et al. Development of a combined database for meta-epidemiological research. Res Synth Method 2010; 1:212-225.

66. Moreno, S.G., Sutton, A.J., Ades, A.E., Stanley, T.D., Abrams, K.R., Peters, J.L. et al. Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Medical Research Methodology 2009; 9(2):doi:10.1186/1471-2288-9-2.

67. Dias, S., Welton, N.J., Ades, A.E. Study designs to detect sponsorship and other biases in systematic reviews. Journal of Clinical Epidemiology 2010; 63:587-588.

68. Madan, J., Stevenson, M.D., Ades, A.E., Cooper, K.L., Whyte, S., Akehurst, R. Consistency between direct and indirect trial evidence: Is direct evidence always more reliable? Value in Health 2011; (in press):DOI: 10.1016/j.jval.2011.05.042.

APPENDIX: ILLUSTRATIVE EXAMPLES AND WINBUGS CODE

This appendix gives illustrative WinBUGS code for all the examples presented in the main document. All programming code is fully annotated.

The program codes are printed here, but are also available as WinBUGS system files from http://www.nicedsu.org.uk. Users are advised to download the WinBUGS files from the website instead of copying and pasting from this document. We have provided the codes as complete programs. However, the majority of each meta-regression program is identical to the programs in TSD2.[8] We have therefore highlighted the main differences in blue and bold, to emphasise the modular nature of the code.

Table A1 gives an index of the programs and their relation to the descriptions in the text. Note that for each example there are random and fixed effects versions of the code except for the predictive cross-validation models which, by definition, only apply to RE models. All FE code can be run using the same data structure described for the random effects.

The code presented in programs 2 to 6 is completely general and will be suitable for fitting pairwise or network meta-analyses with any number of treatments and multi-arm trials. We also provide an indication of the relevant parameters to monitor for inference and model checking for the various programs. The nodes to monitor for the fixed effects models are the same as those for the random effects models, except that there is no heterogeneity parameter.

Table A1 Index of WinBUGS code with details of examples and sections where they are described.

Program	Fixed or Random Effects	Example name	Model specification
1	RE	Magnesium (Section 3.1)	Predictive cross-validation for pairwise meta-analysis
2	RE	Adverse events in Chemotherapy (Section 3.2)	Predictive cross-validation for network meta-analysis
3 (a) (b)	RE FE	Statins (Section 4.1.1)	Meta-regression with subgroups
4 (a) (b)	RE FE	BCG Vaccine (Section 4.3.1) and Certolizumab (Section 4.3.2)	Meta-regression with continuous covariate
6 (a) (b)	RE FE	Certolizumab (Section 4.4.1)	Meta-regression with adjustment for baseline risk

EXAMPLE 1. MAGNESIUM: PREDICTIVE CROSS-VALIDATION

This example and results are described in Section 3.1. The WinBUGS code for predictive cross-validation in a pairwise meta-analysis is given in program 1. The code is identical to the simple code for pairwise meta-analysis presented in TSD2[8] (program 1(a)), apart from the lines highlighted below.

Program 1: Binomial likelihood, logit link, predictive cross-validation, two-treatments (Magnesium example). Two-arm trials only.

```
# Binomial likelihood, logit link, pairwise meta-analysis (2 treatments)
# Random effects model with Predictive Cross-validation
model{                                  # *** PROGRAM STARTS
for(i in 1:ns){                         # LOOP THROUGH STUDIES
   delta[i,1] <- 0                      # treatment effect is zero for control arm
   mu[i] ~ dnorm(0,.0001)               # vague priors for all trial baselines
   for (k in 1:2) {                     # 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]))   #Deviance contribution
        +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))       }
   resdev[i] <- sum(dev[i,])            #  summed residual deviance contribution for this trial
   delta[i,2] ~ dnorm(d[2],tau)         # trial-specific LOR distributions
 }
totresdev <- sum(resdev[])              #Total Residual Deviance
d[1]<- 0                                # treatment effect is zero for reference treatment
d[2] ~ dnorm(0,.0001)                   # vague prior for treatment effect
sd ~ dunif(0,5)                         # vague prior for between-trial SD
tau <- pow(sd,-2)                       # between-trial precision = (1/between-trial variance)
delta.new ~ dnorm(d[2],tau)             # predictive distribution for future trial
p.base ~ dbeta(a,b)                     # draw baseline (control group) effect
a <- r[ns+1,1]                          # no events in control group
b <- n[ns+1,1]-r[ns+1,1]                # no of non-events in control group
logit(p.new) <- logit(p.base) + delta.new    # predictive prob of event in treatment group
r.new ~ dbin(p.new, n[ns+1,2])          # draw predicted number of events in treatment group
# Bayesian p-value: probability of obtaining a value as extreme as the value
# observed (r[ns+1,2]), given the model and the remaining data
p.cross <- step(r.new - r[ns+1,2]) - 0.5*equals(r.new,r[ns+1,2])  # extreme value "larger"
}                                       # *** PROGRAM ENDS
```

The cross-validation p-value is obtained by monitoring p.cross. To obtain posterior summaries for other parameters of interest, the nodes d, delta.new and sd need to be monitored. To obtain the posterior means of the parameters required to assess model fit and model comparison, dev, totresdev and the DIC (from the WinBUGS DIC tool), need to be monitored. In addition, to produce plots of the "shrunken" estimates such as those in Figure 2 and Figure 3, the node delta needs to be monitored.

The data structure is identical to that presented in TSD2,[8] but the last row of data represents the trial for which we want to calculate the cross-validation p-value (ISIS-4 in this example). Briefly, ns is the number of studies in which the model is to be based, and in the main body of

data r[,1] and n[,1] are the numerators and denominators for the first treatment; r[,2] and n[,2], the numerators and denominators for the second listed treatment, and the trial to be excluded is given at the end. Text is included after the hash symbol (#) for ease of reference to the original data source.

```
# Data (Magnesium Example)
list(ns=15)

r[,1]	n[,1]	r[,2]	n[,2]	#	ID
2	36	1	40	#	1
23	135	9	135	#	2
7	200	2	200	#	3
1	46	1	48	#	4
8	148	10	150	#	5
9	56	1	59	#	6
3	23	1	25	#	7
1	21	0	22	#	8
11	75	6	76	#	9
7	27	1	27	#	10
12	80	2	89	#	11
13	33	5	23	#	12
8	122	4	130	#	13
118	1157	90	1159	#	14
17	108	4	107	#	15
2103	29039	2216	29011	#	16
END


# Initial values
# Initial values for delta and other variables can be generated by WinBUGS.
#chain 1
list(d=c( NA, 0), sd=1, mu=c(0,0,0,0,0,    0,0,0,0,0,    0,0,0,0,0), p.base=0.5)
#chain 2
list(d=c( NA, -1), sd=4, mu=c(-3,-3,-3,-3,-3,    -3,-3,-3,-3,-3,    -3,-3,-3,-3,-3), p.base=.2)
#chain 3
list(d=c( NA, 2), sd=2, mu=c(-3,5,-1,-3,7,    -3,-4,-3,-3,0,    -3,-3,0,3,5), p.base=.8)
```

EXAMPLE 2. PREDICTIVE CROSS-VALIDATION IN NETWORK META-ANALYSIS

A synthesis of evidence on three treatments to reduce the incidence of febrile neutropenia (FN), an adverse event during chemotherapy, was carried out for a cost-effectiveness analysis.[68] We will take 'No Treatment', coded 1, as the reference for the analysis. The three treatments of interest, filgrastim, pegfilgrastim and lenograstim are coded 2 to 4. Table A2 shows the number of patients with FN, r_{ik}, out of all included patients, n_{ik}, and the treatments compared, t_{ik}, in each arm of the included trials (i=1,…,25; k=1,2). The network diagram is presented in Figure A1.

Table A2 Number of adverse events r_{ik}, out of the total number of patients receiving chemotherapy n_{ik}, in arms 1 and 2 of 25 trials for the 4 treatments t_{ik}.

	Treatments		Number of events		Number of patients	
Study ID	arm 1 t_{i1}	arm 2 t_{i2}	arm 1 r_{i1}	arm 2 r_{i2}	arm 1 n_{i1}	arm 2 n_{i2}
1	2	3	15	10	75	77
2	2	3	27	14	147	149
3	2	3	2	5	25	46
4	2	3	6	6	31	29
5	2	3	1	0	13	14
6	1	2	26	34	72	276
7	1	2	17	9	39	41
8	1	2	15	4	72	77
9	1	2	86	72	192	197
10	1	2	52	34	104	101
11	1	2	62	40	125	125
12	1	2	27	16	85	90
13	1	2	80	38	104	95
14	1	2	34	17	64	65
15	1	2	38	25	130	129
16	1	4	18	5	28	23
17	1	4	42	36	59	61
18	1	4	15	5	26	22
19	1	4	62	52	80	82
20	1	4	14	5	43	43
21	1	3	27	11	73	73
22	1	3	34	14	343	343
23	1	3	5	4	29	30
24	1	3	10	3	118	123
25	1	3	78	6	465	463

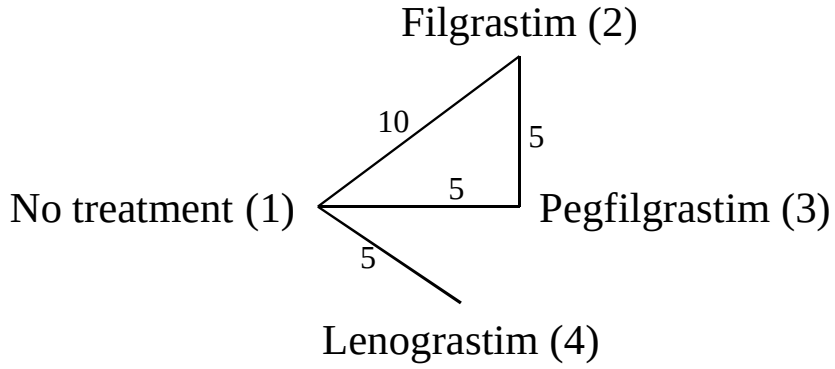

Figure A1 Adverse events in Chemotherapy: Treatment network. Lines connecting two treatments indicate that a comparison between these treatments has been made. The numbers on the lines indicate how many RCTs compare the two connected treatments.

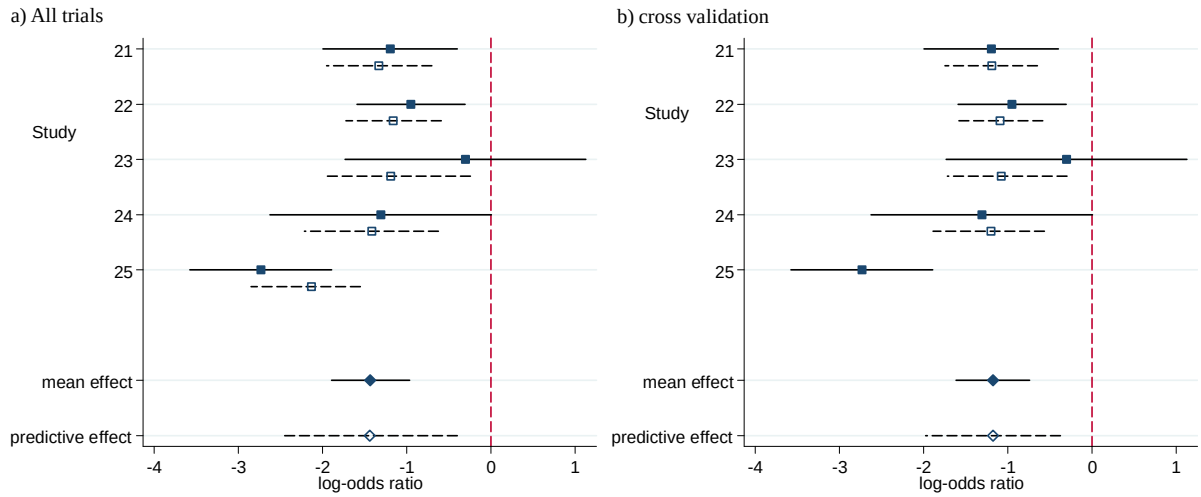

Figure A2 Adverse events in Chemotherapy: Crude log-odds ratios with 95% CI (filled squares, solid lines); posterior mean with 95% CrI of the trial-specific log-odds ratios, "shrunken" estimates, (open squares, dashed lines); posterior mean with 95% CrI of the posterior (filled diamond, solid line) and predictive distribution (open diamond, dashed line) of the pooled treatment effect for a RE model a) including all the trials and b) excluding trial 25 (cross-validation model).

Figure A2(a) shows a forest plot with the crude log-odds ratios calculated from the data and the "shrunken" estimates (i.e. the trial-specific treatment effects, assumed to be exchangeable) for the trials comparing treatments 1 and 3, along with the posterior and predictive effects of treatment 1 compared to 3, from a RE model including all the trials in Table A2. Although the RE network meta-analysis fits the data well (posterior mean of the residual deviance is 49.6, compared to 50 data points), trial 25 has an estimated trial-specific log-odds ratio which is somewhat different from the other trials and may be contributing to the high estimated heterogeneity in this network (posterior median of σ=0.42 with 95% CrI (0.20, 0.73)). To investigate whether this trial is an "outlier", cross-validation, based on a "leave one out" approach, was used as described in Section 3. The result is a p-value of 0.004, indicating that a trial with a results as extreme as trial 25 would be very unlikely, given our model for the remaining data (convergence was achieved after 60,000 burn-in iterations and results are based on 100,000 samples from three independent chains).

The WinBUGS code to fit the standard RE model is given in TSD2[8] (Program 1(c)). The WinBUGS code for predictive cross-validation in a network meta-analysis is given in Program 2. Note that this code is completely general and can be used for predictive cross-validation in networks with or without multi-arm trials and in pairwise meta-analysis.

We have picked the most extreme of 25 trials, so there is an implication that n=25 tests could be performed. Taking into account the effective number of tests that could be undertaken, we need to compare the observed p-value to its expected value, which is $1/(n+1) = 0.038$, the value of the n-th Uniform order statistic. The observed p-value is substantially less than this, indicating that trial 25 may be an "outlier". This can also be seen in Figure A2(b) which now presents the "shrunken" estimates mean and predictive treatment effects for the trials comparing treatments 1 and 3, along with the posterior and predictive effects of treatment 1 compared to 3, from a RE model excluding trial 25 (but including the observed log-odds ratio and CI for this trial). The 95% CI for the observed log-odds ratio from trial 25 is (-3.57, -1.89) which is well outside the 95% CrI for the posterior mean (-1.61, -0.74) and only marginally within the bounds of the 95% CrI for the predictive mean treatment effect (-1.98, -0.38), which is the basis for predictive cross-validation. The posterior median for the between-trials heterogeneity for the RE network meta-analysis excluding trial 25 is 0.29 with 95% CrI (0.05, 0.58), smaller than for the model with the full data.

Program 2: Binomial likelihood, logit link, predictive cross-validation, network meta-analysis with multi-arm trials (Adverse Events in Chemotherapy example).

```
# Binomial likelihood, logit link, network meta-analysis (multi-arm trials)
# Random effects model with Predictive Cross-validation
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,.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]))  #Deviance contribution
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
       }
   resdev[i] <- sum(dev[i,1:na[i]])     #  summed residual deviance contribution for this trial
   for (k in 2:na[i]) {                 # LOOP THROUGH ARMS
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])  # trial-specific LOR distributions
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k] # mean of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k       # precision of LOR distributions (with multi-arm trial correction)
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])   # adjustment for multi-arm RCTs
      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
for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }  # vague priors for treatment effects
sd ~ dunif(0,5)                         # vague prior for between-trial SD
tau <- pow(sd,-2)                       # between-trial precision = (1/between-trial variance)
# predictive distribution for future trial is multivariate normal
delta.new[1] <- 0                       # treatment effect is zero for reference treatment
w.new[1] <- 0                           # adjustment for conditional mean is zero for ref. treat.
for (k in 2:nt) {                       # LOOP THROUGH TREATMENTS
   delta.new[k] ~ dnorm(m.new[k],tau.new[k])  # conditional distribution of each delta.new
```

```
   m.new[k] <-  d[k] + sw.new[k]          # conditional mean of delta.new
   tau.new[k] <- tau *2*(k-1)/k            # conditional precision of delta.new
   w.new[k] <- delta.new[k] - d[k]        # adjustment for conditional mean
   sw.new[k] <- sum(w.new[1:k-1])/(k-1) # cumulative adjustment for cond. mean
  }
p.base ~ dbeta(a,b)                       # draw baseline (control group) effect
a <- r[ns+1,1]                            # no. of events in control group
b <- n[ns+1,1]-r[ns+1,1]                  # no of non-events in control group
for (k in 2:na[ns+1]) {                   # LOOP THROUGH ARMS
# predictive prob of event for each treatment arm of the new trial
   logit(p.new[k]) <- logit(p.base) + (delta.new[t[ns+1,k]]- delta.new[t[ns+1,1]])
   r.new[k] ~ dbin(p.new[k], n[ns+1,k])  # draw predicted number of events for each arm of the new trial
# Bayesian p-value: probability of obtaining a value as extreme as the
# value observed (r[ns+1,2]), given the model and the remaining data
   p.cross[k] <- step(r[ns+1,2] - r.new[k]) - 0.5*equals(r.new[k],r[ns+1,2])  # extreme value "smaller"
  }
}              # *** PROGRAM ENDS
```

The relevant nodes to monitor are the same as in Program 1.

The code below can be added before the closing brace to predict all pairwise log-odds ratios and odds ratios in a new trial.

```
# pairwise ORs and LORs for all possible pair-wise comparisons, if nt>2
for (c in 1:(nt-1)) {
   for (k in (c+1):nt)  {
     lor.new[c,k] <- delta.new[k]- delta.new[c]
     or.new[c,k] <- exp(lor.new[c,k])
     }
}
```

The data structure is identical to that presented in TSD2[8] (Program 1(c)), but the last row of data represents the trial for which we want to calculate the cross-validation p-value for.

```
# Data (Adverse events in Chemotherapy)
list(ns=24, nt=4)
```

t[,1]	t[,2]	na[]	r[,1]	r[,2]	n[,1]	n[,2]	#	ID
2	3	2	15	10	75	77	#	1
2	3	2	27	14	147	149	#	2
2	3	2	2	5	25	46	#	3
2	3	2	6	6	31	29	#	4
2	3	2	1	0	13	14	#	5
1	2	2	26	34	72	276	#	6
1	2	2	17	9	39	41	#	7
1	2	2	15	4	72	77	#	8
1	2	2	86	72	192	197	#	9
1	2	2	52	34	104	101	#	10
1	2	2	62	40	125	125	#	11
1	2	2	27	16	85	90	#	12
1	2	2	80	38	104	95	#	13
1	2	2	34	17	64	65	#	14
1	2	2	38	25	130	129	#	15
1	4	2	18	5	28	23	#	16
1	4	2	42	36	59	61	#	17
1	4	2	15	5	26	22	#	18
1	4	2	62	52	80	82	#	19
1	4	2	14	5	43	43	#	20
1	3	2	27	11	73	73	#	21
1	3	2	34	14	343	343	#	22
1	3	2	5	4	29	30	#	23

```
1        3        2        10       3        118      123      #        24
1        3        2        78       6        465      463      #        25
END


# Initial values
# Initial values for delta and other variables can be generated by WinBUGS.
#chain 1
list(d=c( NA, 0,0,0), sd=1, mu=c(0,0,0,0,0,    0,0,0,0,0,    0,0,0,0,0,    0,0,0,0,0,    0,0,0,0), p.base=0.5)
#chain 2
list(d=c( NA, -1,-2,1), sd=4, mu=c(-3,-3,-3,-3,-3,-3,    -3,-3,-3,-3,-3,-3,    -3,-3,-3,-3,-3,-3,    -3,-3,-3,-3,-3,-3,    -3,-3,-3,-3), p.base=.2)
#chain 3
list(d=c( NA, 2,3,-3), sd=2, mu=c(-3,5,-1,-3,7,    -3,-4,-3,-3,0,    -3,5,-1,-3,7,    -3,-4,-3,-3,0,    -3,-3,0,3), p.base=.8)
```

EXAMPLE 3. STATINS: META-REGRESSION WITH SUBGROUPS

This example and results are described in Section 4.1.1. Although this example only included 2 treatments, the code presented below can also be used for subgroup analysis with multiple treatments and including multi-arm trials. The WinBUGS code for random effects subgroup meta-regression model is given in program 3(a) and the fixed effects code is given in program 3(b).

Program 3(a): Binomial likelihood, logit link, Random Effects, meta-regression with subgroups (Statins example)

```
# Binomial likelihood, logit link, subgroup
# Random effects model for multi-arm trials
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,.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
# model for linear predictor, covariate effect relative to treat in arm 1
      logit(p[i,k]) <- mu[i] + delta[i,k] + (beta[t[i,k]]-beta[t[i,1]]) * x[i]
      rhat[i,k] <- p[i,k] * n[i,k]      # expected value of the numerators
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))  #Deviance contribution
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
       }
   resdev[i] <- sum(dev[i,1:na[i]])     # summed residual deviance contribution for this trial
   for (k in 2:na[i]) {                 # LOOP THROUGH ARMS
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])  # trial-specific LOR distributions
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k] # mean of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k       # precision of LOR distributions (with multi-arm trial correction)
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])  # adjustment for multi-arm RCTs
      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
beta[1] <- 0                            # covariate effect is zero for reference treatment
for (k in 2:nt){                        # LOOP THROUGH TREATMENTS
   d[k] ~ dnorm(0,.0001)                # vague priors for treatment effects
   beta[k] <- B                         # common covariate effect
```

```
 }
B ~ dnorm(0,.0001)              # vague prior for covariate effect
sd ~ dunif(0,5)                 # vague prior for between-trial SD
tau <- pow(sd,-2)               # between-trial precision = (1/between-trial variance)
}                               # *** PROGRAM ENDS
```

To obtain posterior summaries for other parameters of interest, the nodes d, B and sd need to be monitored. To obtain the posterior means of the parameters required to assess model fit and model comparison, dev, totresdev and the DIC (from the WinBUGS DIC tool), need to be monitored.

Additional code can be added before the closing brace to estimate all the pair-wise log odds ratios and odds ratios and to produce estimates of absolute effects, given additional information on the absolute treatment effect on one of the treatments, for given covariate values. For further details on calculating other summaries from the results and on converting the summaries onto other scales, refer to the Appendix in TSD2.[8]

```
#############################################################################
# Extra code for calculating all odds ratios and log odds ratios, and absolute effects, for covariate
# values in vector z, with length nz (given as data)
#############################################################################
for (k in 1:nt){
   for (j in 1:nz) { dz[j,k] <- d[k] + (beta[k]-beta[1])*z[j] }  # treatment effect when covariate = z[j]
 }
# pairwise ORs and LORs for all possible pair-wise comparisons
for (c in 1:(nt-1)) {
   for (k in (c+1):nt)  {
# when covariate is zero
      or[c,k] <- exp(d[k] - d[c])
      lor[c,k] <- (d[k]-d[c])
# at covariate=z[j]
      for (j in 1:nz) {
         orz[j,c,k] <- exp(dz[j,k] - dz[j,c])
         lorz[j,c,k] <- (dz[j,k]-dz[j,c])
       }
   }
}
# Provide estimates of treatment effects T[k] on the natural (probability) scale
# Given a Mean Effect, meanA, for 'standard' treatment 1, with precision (1/variance) precA, and covariate value z[j]
A ~ dnorm(meanA,precA)
for (k in 1:nt) {
   for (j in 1:nz){
      logit(T[j,k]) <- A + d[k] + (beta[k]-beta[1]) * z[j]
    }
 }
```

For a meta-regression with two subgroups vector z would be added to the list data statement as list(z=c(1), nz=1).

The data structure is identical to that presented in TSD2,[8] but now has an added column x[] which represents the value of the covariate (taking values 0 or 1) for each trial. The

remaining variables represent the number of treatments, nt, the number of studies, ns, r[,1] and n[,1] are the numerators and denominators for the first treatment, r[,2] and n[,2], the numerators and denominators for the second listed treatment, t[,1] and t[,2] are the treatment number identifiers for the first and second listed treatments, and na[] is the number of arms in each trial.

```
# Data (Statins example)
list(ns=19, nt=2)
```

t[,1]	t[,2]	na[]	r[,1]	n[,1]	r[,2]	n[,2]	x[]	#	ID	name
1	2	2	256	2223	182	2221	1	#	1	4S
1	2	2	4	125	1	129	1	#	2	Bestehorn
1	2	2	0	52	1	94	1	#	3	Brown
1	2	2	2	166	2	165	1	#	4	CCAIT
1	2	2	77	3301	80	3304	0	#	5	Downs
1	2	2	3	1663	33	6582	0	#	6	EXCEL
1	2	2	8	459	1	460	1	#	7	Furberg
1	2	2	3	155	3	145	1	#	8	Haskell
1	2	2	0	42	1	83	1	#	9	Jones
1	2	2	4	223	3	224	0	#	10	KAPS
1	2	2	633	4520	498	4512	1	#	12	LIPID
1	2	2	1	124	2	123	1	#	13	MARS
1	2	2	11	188	4	193	1	#	14	MAAS
1	2	2	5	78	4	79	1	#	15	PLAC 1
1	2	2	6	202	4	206	1	#	16	PLAC 2
1	2	2	3	532	0	530	0	#	17	PMSGCRP
1	2	2	4	178	2	187	1	#	18	Riegger
1	2	2	1	201	3	203	1	#	19	Weintraub
1	2	2	135	3293	106	3305	0	#	20	Wscotland

```
END


# Initial values
# Initial values for delta can be generated by WinBUGS.
#chain 1
list(d=c( NA, 0), mu=c(0,0,0,0,0,   0,0,0,0,0,   0,0,0,0,0,   0,0,0,0), B=0, sd=1)
#chain 2
list(d=c( NA, -1), mu=c(-3,-3,3,-3,3,    -3,3,-3,3,-3,    -3,-3,3,3,-3,   3,-3,-3,3), B=-1, sd=3)
#chain 3
list(d=c( NA, 2), mu=c(-3,5,-1,-3,7,    -3,-4,-3,-3,0,    5,0,-2,-5,1,   -2,5,3,0), B=1.5, sd=0.5)
```

Program 3(b): Binomial likelihood, logit link, Fixed Effects, meta-regression with subgroups (Statins example)

```
# Binomial likelihood, logit link, subgroup
# Fixed effects model with one covariate
model{                                  # *** PROGRAM STARTS
for(i in 1:ns){                         # LOOP THROUGH STUDIES
   mu[i] ~ dnorm(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
# model for linear predictor, covariate effect relative to treat in arm 1
      logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]] + (beta[t[i,k]]-beta[t[i,1]]) * x[i]
      rhat[i,k] <- p[i,k] * n[i,k]      # expected value of the numerators
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))  #Deviance contribution
          +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
    }
   resdev[i] <- sum(dev[i,1:na[i]])     # summed residual deviance contribution for this trial
   }
totresdev <- sum(resdev[])              # Total Residual Deviance
```

```
d[1] <- 0                          # treatment effect is zero for reference treatment
beta[1] <- 0                       # covariate effect is zero for reference treatment
for (k in 2:nt){                   # LOOP THROUGH TREATMENTS
   d[k] ~ dnorm(0,.0001)           # vague priors for treatment effects
   beta[k] <- B                    # common covariate effect
 }
B ~ dnorm(0,.0001)                 # vague prior for covariate effect
}                                  # *** PROGRAM ENDS


# Initial values
#chain 1
list(d=c( NA, 0), mu=c(0,0,0,0,0,   0,0,0,0,0,   0,0,0,0,0,   0,0,0,0), B=0)
#chain 2
list(d=c( NA, -1), mu=c(-3,-3,3,-3,3,    -3,3,-3,3,-3,    -3,-3,3,3,-3,   3,-3,-3,3), B=-1)
#chain 3
list(d=c( NA, 2), mu=c(-3,5,-1,-3,7,    -3,-4,-3,-3,0,    5,0,-2,-5,1,   -2,5,3,0), B=1.5)
```

EXAMPLE 4. BCG VACCINE

This example and results are described in Section 4.3.1. The WinBUGS code for random effects meta-regression model with a continuous covariate is given in program 4(a) and the fixed effects code is given in program 4(b). This code can also be used for networks with multiple treatments and including multi-arm trials (see Example 5).

Program 4(a): Binomial likelihood, logit link, Random Effects, meta-regression with a continuous covariate (BCG vaccine example)

```
# Binomial likelihood, logit link, continuous covariate
# Random effects model for multi-arm trials
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,.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
# model for linear predictor, covariate effect relative to treat in arm 1 (centring)
      logit(p[i,k]) <- mu[i] + delta[i,k] + (beta[t[i,k]]-beta[t[i,1]]) * (x[i]-mx)
      rhat[i,k] <- p[i,k] * n[i,k]     # expected value of the numerators
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))  #Deviance contribution
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
       }
   resdev[i] <- sum(dev[i,1:na[i]])    #  summed residual deviance contribution for this trial
   for (k in 2:na[i]) {                # LOOP THROUGH ARMS
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])  # trial-specific LOR distributions
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k] # mean of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k      # precision of LOR distributions (with multi-arm trial correction)
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])  # adjustment for multi-arm RCTs
      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
beta[1] <- 0                           # covariate effect is zero for reference treatment
for (k in 2:nt){                       # LOOP THROUGH TREATMENTS
   d[k] ~ dnorm(0,.0001)               # vague priors for treatment effects
```

```
   beta[k] <- B                    # common covariate effect
  }
B ~ dnorm(0,.0001)                 # vague prior for covariate effect
sd ~ dunif(0,5)                    # vague prior for between-trial SD
tau <- pow(sd,-2)                  # between-trial precision = (1/between-trial variance)
}                                  # *** PROGRAM ENDS
```

The relevant nodes to monitor are the same as in Program 3.

The data structure is the same as in Example 3, but now we add the mean covariate value mx to the list data, for centring.

```
# Data (BCG vaccine example)
list(ns=13, nt=2, mx=33.46)

t[,1]    t[,2]    na[]    r[,1]    n[,1]    r[,2]    n[,2]    x[]    #    ID
1        2        2       11       139      4        123      44     #    1
1        2        2       29       303      6        306      55     #    2
1        2        2       11       220      3        231      42     #    3
1        2        2       248      12867    62       13598    52     #    4
1        2        2       47       5808     33       5069     13     #    5
1        2        2       372      1451     180      1541     44     #    6
1        2        2       10       629      8        2545     19     #    7
1        2        2       499      88391    505      88391    13     #    8
1        2        2       45       7277     29       7499     27     #    9
1        2        2       65       1665     17       1716     42     #    10
1        2        2       141      27338    186      50634    18     #    11
1        2        2       3        2341     5        2498     33     #    12
1        2        2       29       17854    27       16913    33     #    13
END
```

To estimate all the pair-wise Log Odds Ratios, Odds Ratios and absolute effects, for covariate values 0, 13 and 50, vector z could added to the list data as list(z=c(0,13,50), nz=3).

```
# Initial values
# Initial values for delta can be generated by WinBUGS.
#chain 1
list(d=c( NA, 0), mu=c(0,0,0,0,0,    0,0,0,0,0,    0,0,0), sd=1, B=0, sd=1)
#chain 2
list(d=c( NA, -1), mu=c(-3,-3,-3,3,-3,    -3,3,-3,-3,-3,    -3,-3,3), B=-2, sd=3)
#chain 3
list(d=c( NA, 2), mu=c(-3,5,-1,-3,7,    -3,-4,-3,-3,0,    5,0,-5), B=5, sd=0.5)
```

Program 4(b): Binomial likelihood, logit link, Fixed Effects, meta-regression with a continuous covariate (BCG vaccine example)

```
# Binomial likelihood, logit link
# Fixed effects model with continuous covariate
model{                                 # *** PROGRAM STARTS
for(i in 1:ns){                        # LOOP THROUGH STUDIES
   mu[i] ~ dnorm(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
# model for linear predictor, covariate effect relative to treat in arm 1
      logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]] + (beta[t[i,k]]-beta[t[i,1]]) * (x[i]-mx)
      rhat[i,k] <- p[i,k] * n[i,k]     # expected value of the numerators
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))  #Deviance contribution
         +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
```

```
    }
   resdev[i] <- sum(dev[i,1:na[i]])        # summed residual deviance contribution for this trial
    }
totresdev <- sum(resdev[])                 # Total Residual Deviance
d[1] <- 0                                  # treatment effect is zero for reference treatment
beta[1] <- 0                               # covariate effect is zero for reference treatment
for (k in 2:nt){                           # LOOP THROUGH TREATMENTS
   d[k] ~ dnorm(0,.0001)                   # vague priors for treatment effects
   beta[k] <- B                            # common covariate effect
 }
B ~ dnorm(0,.0001)                         # vague prior for covariate effect
}                                          # *** PROGRAM ENDS


# Initial values
#chain 1
list(d=c( NA, 0), mu=c(0,0,0,0,0,    0,0,0,0,0,    0,0,0), B=0)
#chain 2
list(d=c( NA, -1), mu=c(-3,-3,-3,3,-3,    -3,3,-3,-3,-3,    -3,-3,3), B=-2)
#chain 3
list(d=c( NA, 2), mu=c(-3,5,-1,-3,7,    -3,-4,-3,-3,0,    5,0,-5), B=5)
```

EXAMPLE 5. CERTOLIZUMAB: CONTINUOUS COVARIATE

This example and results are described in Section 4.3.2. The WinBUGS code for random effects meta-regression model with a continuous covariate and non-informative priors is given in program 4(a) and the fixed effects code is given in program 4(b). The relevant nodes to monitor are the same as in Program 3. The data structure is the same as in Example 4, but now we have more than 2 treatments being compared.

```
# Data (Certolizumab example – covariate is disease duration)
list(ns=12, nt=7, mx=8.21)
```

t[,1]	t[,2]	na[]	n[,1]	n[,2]	r[,1]	r[,2]	x[]	#	ID	Study name
1	3	2	63	65	9	28	6.85	#	1	Kim 2007 (37)
1	3	2	200	207	19	81	10.95	#	2	DE019 Trial (36)
1	3	2	62	67	5	37	11.65	#	3	ARMADA Trial (34)
1	2	2	199	393	15	146	6.15	#	4	RAPID 1 Trial (40)
1	2	2	127	246	4	80	5.85	#	5	RAPID 2 Trial (41)
1	5	2	363	360	33	110	8.10	#	6	START Study (57)
1	5	2	110	165	22	61	7.85	#	7	ATTEST Trial (51)
1	5	2	47	49	0	15	8.30	#	8	Abe 2006 (50)
1	4	2	30	59	1	23	13.00	#	9	Weinblatt 1999 (49)
1	6	2	40	40	5	5	11.25	#	11	Strand 2006 (62)
1	7	2	49	50	14	26	0.92	#	12	CHARISMA Study (64)
1	7	2	204	205	22	90	7.65	#	13	OPTION Trial (67)

```
END


# Initial values for RE model
#chain 1
list(d=c( NA, 0,0,0,0,0,0), mu=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0,0), sd=1, B=0)
#chain 2
list(d=c( NA, -1,1,-1,1,-1,1), mu=c(-3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3), sd=0.5, B=-1)
#chain 3
list(d=c( NA, 2,-2,2,-2,2,-2), mu=c(-3, 5, -1, -3, 7, -3, -4, -3, -3, 0, 5, 0), sd=3, B=5)


# Initial values for FE model
```

```
#chain 1
list(d=c( NA, 0,0,0,0,0,0), mu=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0), B=0)
#chain 2
list(d=c( NA, -1,1,-1,1,-1,1), mu=c(-3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3), B=-2)
#chain 3
list(d=c( NA, 2,-2,2,-2,2,-2), mu=c(-3, 5, -1, -3, 7, -3, -4, -3, -3, 0, 5, 0), B=5)
```

A RE model with Uniform(0,5) prior for σ, the heterogeneity parameter is not identifiable. This is because there is a trial with a zero cell and not many replicates of each comparison. Due to the paucity of information from which the between-trial variation can be estimated, in the absence of an informative prior on σ, the relative treatment effect for this trial will tend towards infinity. We have therefore used an informative half-normal prior, represented by the solid line in Figure A3, which ensures stable computation:

$$\sigma \sim \text{Half-Normal}(0, 0.32^2)$$

This prior distribution was chosen to ensure that, *a priori*, 95% of the trial-specific ORs lie within a factor of 2 from the median OR for each comparison. Under this prior the mean σ is 0.26. To fit the RE meta-regression model with this prior distribution, the line of code annotated as 'vague prior for between-trial SD' in Program 4(a) should be replaced with the two lines below:

```
sd ~ dnorm(0,prec)I(0,)     # prior for between-trial SD
prec <- pow(0.32,-2)
```

This prior should not be used unthinkingly. Informative prior distributions allowing wider or narrower ranges of values can be used by changing the value of `prec` in the code above.

In this example, the posterior distribution obtained for σ is given by the dotted line in Figure A3, and shows that the range plausible values for σ has not changed much, but the probability that σ will have values close to zero has decreased.

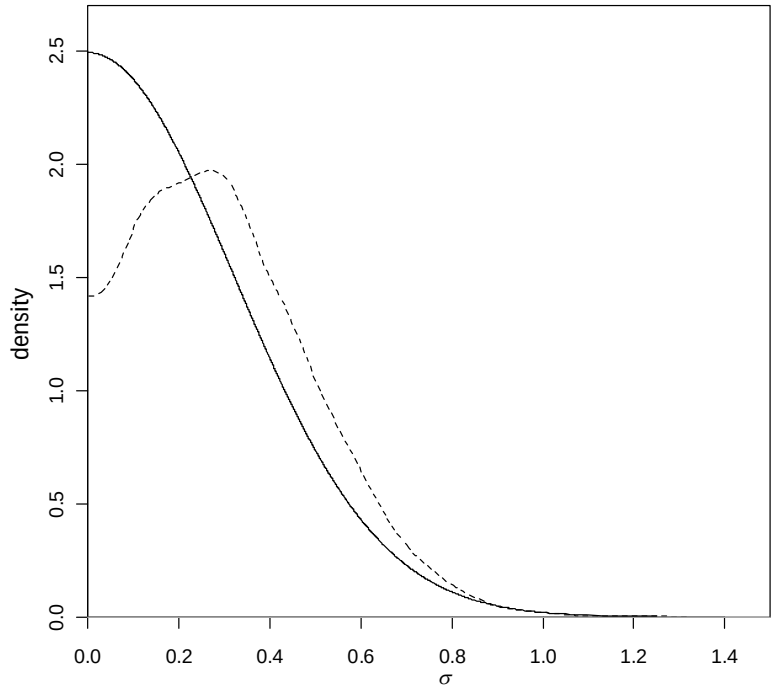

Figure A3 Certolizumab: meta-regression with informative Half-Normal(0,0.32^2) prior distribution. Probability density function of the prior distribution is given by the solid line and the posterior density by the dotted line.

EXAMPLE 6. CERTOLIZUMAB: BASELINE RISK

This example and results are described in Section 4.4.1. The WinBUGS code for the meta-regression model with adjustment for baseline risk for random and fixed treatment effects is similar to programs 4(a) and 4(b), respectively, but now x[i] is replaced with mu[i] in the definitions of the linear predictor. The variability of the normal prior distribution needs to be reduced to avoid numerical errors (this only minimally affects the posterior results).

Program 6(a): Binomial likelihood, logit link, Random Effects, meta-regression with adjustment for baseline risk (Certolizumab example)

```
# Binomial likelihood, logit link
# Random effects model for multi-arm trials
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,.001)                # 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
# model for linear predictor, covariate effect relative to treat in arm 1
      logit(p[i,k]) <- mu[i] + delta[i,k] + (beta[t[i,k]]-beta[t[i,1]]) * (mu[i]-mx)
      rhat[i,k] <- p[i,k] * n[i,k]      # expected value of the numerators
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))  #Deviance contribution
```

```
        + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
      }
   resdev[i] <- sum(dev[i,1:na[i]])          #  summed residual deviance contribution for this trial
   for (k in 2:na[i]) {                      # LOOP THROUGH ARMS
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])  # trial-specific LOR distributions
      md[i,k] <-  d[t[i,k]] - d[t[i,1]] + sw[i,k] # mean of LOR distributions (with multi-arm trial correction)
      taud[i,k] <- tau *2*(k-1)/k            # precision of LOR distributions (with multi-arm trial correction)
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])  # adjustment for multi-arm RCTs
      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
beta[1] <- 0                                 # covariate effect is zero for reference treatment
for (k in 2:nt){                             # LOOP THROUGH TREATMENTS
   d[k] ~ dnorm(0,.0001)                     # vague priors for treatment effects
   beta[k] <- B                              # common covariate effect
 }
B ~ dnorm(0,.0001)                           # vague prior for covariate effect
sd ~ dunif(0,5)                              # vague prior for between-trial SD
tau <- pow(sd,-2)                            # between-trial precision = (1/between-trial variance)
}                                            # *** PROGRAM ENDS
```

The relevant nodes to monitor are the same as in Program 3.

The data structure is the same as Example 4, but without variable x[].

```
# Data (Certolizumab, baseline risk)
list(ns=12, nt=7, mx=-2.421)

t[,1]   t[,2]   na[]   n[,1]   n[,2]   r[,1]   r[,2]   #   ID   Study name
1       3       2      63      65      9       28      #   1    Kim 2007 (37)
1       3       2      200     207     19      81      #   2    DE019 Trial (36)
1       3       2      62      67      5       37      #   3    ARMADA Trial (34)
1       2       2      199     393     15      146     #   4    RAPID 1 Trial (40)
1       2       2      127     246     4       80      #   5    RAPID 2 Trial (41)
1       5       2      363     360     33      110     #   6    START Study (57)
1       5       2      110     165     22      61      #   7    ATTEST Trial (51)
1       5       2      47      49      0       15      #   8    Abe 2006 (50)
1       4       2      30      59      1       23      #   9    Weinblatt 1999 (49)
1       6       2      40      40      5       5       #   11   Strand 2006 (62)
1       7       2      49      50      14      26      #   12   CHARISMA Study (64)
1       7       2      204     205     22      90      #   13   OPTION Trial (67)
END


# Initial values
# Initial values for delta and other variables can be generated by WinBUGS.
#chain 1
list(d=c( NA, 0,0,0,0,0,0), mu=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0,0), sd=1, B=0)
#chain 2
list(d=c( NA, -1,1,-1,1,-1,1), mu=c(-3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3), sd=0.5, B=-1)
#chain 3
list(d=c( NA, 2,-2,2,-2,2,-2), mu=c(-3, 5, -1, -3, 7, -3, -4, -3, -3, 0, 5, 0), sd=3, B=5)
```

Program 6(b): Binomial likelihood, logit link, Fixed Effects, meta-regression with adjustment for baseline risk (Certolizumab example)

```
# Binomial likelihood, logit link
# Fixed effects model with one covariate (independent covariate effects)
model{                                       # *** PROGRAM STARTS
for(i in 1:ns){                              # LOOP THROUGH STUDIES
   mu[i] ~ dnorm(0,.001)                     # 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
# model for linear predictor, covariate effect relative to treat in arm 1
      logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]] + (beta[t[i,k]]-beta[t[i,1]]) * (mu[i]-mx)
      rhat[i,k] <- p[i,k] * n[i,k]          # expected value of the numerators
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))  #Deviance contribution
          +  (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
    }
   resdev[i] <- sum(dev[i,1:na[i]])         # summed residual deviance contribution for this trial
    }
totresdev <- sum(resdev[])                  # Total Residual Deviance
d[1] <- 0                                   # treatment effect is zero for reference treatment
beta[1] <- 0                                # covariate effect is zero for reference treatment
for (k in 2:nt){
   d[k] ~ dnorm(0,.0001)                    # vague priors for treatment effects
   beta[k] <- B                             # common covariate effect
 }
B ~ dnorm(0,.0001)                          # vague prior for covariate effect
}                                           # *** PROGRAM ENDS


# Initial values
#chain 1
list(d=c( NA, 0,0,0,0,0,0), mu=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0,0), B=0)
#chain 2
list(d=c( NA, -1,1,-1,1,-1,1), mu=c(-3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3, -3), B=-2)
#chain 3
list(d=c( NA, 2,-2,2,-2,2,-2), mu=c(-3, 5, -1, -3, 7, -3, -4, -3, -3, 0, 5, 0), B=5)
```