

**NICE DSU TECHNICAL SUPPORT DOCUMENT 2:  
A GENERALISED LINEAR MODELLING FRAMEWORK  
FOR PAIRWISE AND NETWORK META-ANALYSIS OF  
RANDOMISED CONTROLLED TRIALS**

REPORT BY THE DECISION SUPPORT UNIT

August 2011

(last updated April 2014)

**Sofia Dias<sup>1</sup>, Nicky J Welton<sup>1</sup>, Alex J Sutton<sup>2</sup>, AE Ades<sup>1</sup>**

<sup>1</sup> School of Social and Community Medicine, University of Bristol, Canynge Hall, 39  
Whatley Road, Bristol BS8 2PS, UK

<sup>2</sup> 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](mailto: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](http://www.nicedsu.org.uk)

## **ABOUT THE TECHNICAL SUPPORT DOCUMENT SERIES**

The NICE Guide to the Methods of Technology Appraisal<sup>i</sup> 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.

---

<sup>i</sup> National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal, 2008 (updated June 2008), London.

## **Acknowledgements**

The DSU thanks Mike Campbell, Rachael Fleurence, Julian Higgins, Jeroen Jansen, Steve Palmer and the team at NICE, led by Zoe Garrett, for reviewing this document. The authors thank Gabriel Rogers for useful comments on the Appendix. 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 author 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., Welton, N.J., Sutton, A.J. & Ades, A.E. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. 2011; last updated April 2014; available from <http://www.nicedsu.org.uk>

## **EXECUTIVE SUMMARY**

This paper sets out a generalised linear model (GLM) framework for the synthesis of data from randomised controlled trials (RCTs). We describe a common model taking the form of a linear regression for both fixed and random effects synthesis, that can be implemented with Normal, Binomial, Poisson, and Multinomial data. The familiar logistic model for meta-analysis with Binomial data is a GLM with a logit link function, which is appropriate for probability outcomes. The same linear regression framework can be applied to continuous outcomes, rate models, competing risks, or ordered category outcomes, by using other link functions, such as identity, log, complementary log-log, and probit link functions. The common core model for the linear predictor can be applied to pair-wise meta-analysis, indirect comparisons, synthesis of multi-arm trials, and mixed treatment comparisons, also known as network meta-analysis, without distinction.

We take a Bayesian approach to estimation and provide WinBUGS program code for a Bayesian analysis using Markov chain Monte Carlo (MCMC) simulation. An advantage of this approach is that it is straightforward to extend to shared parameter models where different RCTs report outcomes in different formats but from a common underlying model. Use of the GLM framework allows us to present a unified account of how models can be compared using the Deviance Information Criterion (DIC), and how goodness of fit can be assessed using the residual deviance. WinBUGS code for model critique is provided. Our approach is illustrated through a range of worked examples for the commonly encountered evidence formats, including shared parameter models.

We give suggestions on computational issues that sometimes arise in MCMC evidence synthesis, and comment briefly on alternative software.

# CONTENTS

<b>1</b>	<b>INTRODUCTION TO PAIRWISE &amp; NETWORK META-ANALYSIS .....</b>	<b>9</b>
<b>2</b>	<b>DEVELOPMENT OF THE CORE MODELS: BINOMIAL DATA WITH LOGIT LINK .....</b>	<b>11</b>
2.1	WORKED EXAMPLE: A LOGIT MODEL FOR A META-ANALYSIS OF BINOMIAL DATA.....	11
2.1.1	<i>Model specification.....</i>	<i>12</i>
2.1.2	<i>Model fit and model comparison .....</i>	<i>13</i>
2.1.3	<i>WinBUGS implementation and illustrative results.....</i>	<i>16</i>
<b>3</b>	<b>GENERALISED LINEAR MODELS .....</b>	<b>19</b>
3.1	RATE DATA: POISSON LIKELIHOOD AND LOG LINK .....	20
3.2	RATE DATA: BINOMIAL LIKELIHOOD AND CLOGLOG LINK.....	21
3.3	COMPETING RISKS: MULTINOMIAL LIKELIHOOD AND LOG LINK .....	23
3.4	CONTINUOUS DATA: NORMAL LIKELIHOOD AND IDENTITY LINK.....	25
3.4.1	<i>Before/after studies: change from baseline measures .....</i>	<i>25</i>
3.5	TREATMENT DIFFERENCES .....	26
3.5.1	<i>Standardised mean differences .....</i>	<i>27</i>
3.6	ORDERED CATEGORICAL DATA: MULTINOMIAL LIKELIHOOD AND PROBIT LINK.....	29
3.7	ADDITIVE AND MULTIPLICATIVE EFFECTS WITH BINOMIAL DATA, AND OTHER NON-CANONICAL LINKS .....	31
<b>4</b>	<b>SHARED PARAMETER MODELS.....</b>	<b>31</b>
<b>5</b>	<b>EXTENSION TO INDIRECT COMPARISONS AND NETWORK META-ANALYSIS .....</b>	<b>33</b>
5.1	INCORPORATING MULTI-ARM TRIALS.....	35
5.1.1	<i>Multi-arm trials with treatment differences (trial-based summaries) .....</i>	<i>36</i>
<b>6</b>	<b>TECHNICAL ISSUES IN BAYESIAN MCMC .....</b>	<b>38</b>
6.1	CHOICE OF REFERENCE TREATMENT .....	38
6.2	CHOICE OF PRIORS .....	39
6.3	ZERO CELLS .....	40
<b>7</b>	<b>NON-BAYESIAN APPROACHES AND COMPUTATIONAL ISSUES .....</b>	<b>41</b>
7.1	BAYESIAN VERSUS FREQUENTIST APPROACHES IN THE CONTEXT OF DECISION MAKING ..	41
7.2.	COMPARISON OF META-ANALYTIC METHODS .....	42
7.3.	COMPARISON OF EVIDENCE SYNTHESIS SOFTWARE .....	43
<b>8.</b>	<b>FURTHER READING .....</b>	<b>44</b>
<b>9.</b>	<b>DISCUSSION .....</b>	<b>46</b>
<b>10.</b>	<b>REFERENCES.....</b>	<b>48</b>
<b>APPENDIX: ILLUSTRATIVE EXAMPLES AND WINBUGS CODE .....</b>		<b>56</b>
EXAMPLE 1.	BLOCKER.....	57
EXAMPLE 2.	DIETARY FAT.....	62
RESULTS	65	
EXAMPLE 3.	DIABETES .....	66
RESULTS	70	
EXAMPLE 4.	SCHIZOPHRENIA .....	72
RESULTS	77	
EXAMPLE 5.	PARKINSON’S .....	78
RESULTS	81	
EXAMPLE 6.	PSORIASIS.....	83
RESULTS	90	
EXAMPLE 7.	PARKINSON’S DIFFERENCE (TREATMENT DIFFERENCES AS DATA).....	92

RESULTS 95

EXAMPLE 8. PARKINSON'S SHARED PARAMETERS (MIXED TREATMENT DIFFERENCE AND ARM-LEVEL DATA) ..... 95

RESULTS 98

## TABLES AND FIGURES

Table 1 Blocker example: number of events and total number of patients in the control and beta-blocker groups for the 22 trials. <sup>31</sup> .....	12
Table 2 Blocker example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effect $d_{12}$ , absolute effects of the placebo ( $T_1$ ) and beta-blocker ( $T_2$ ) for a mean mortality of -2.2 and precision 3.3 on the logit scale; heterogeneity parameter $\sigma$ and model fit statistics.....	16
Table 3 Commonly used link functions and their inverse with reference to which likelihoods they can be applied to. ....	19
Table 4 Formulae for the residual deviance and model predictors for common likelihoods .....	20
Table A1 Index of WinBUGS code with details of examples and sections where they are described. ....	56
Table A2 Dietary fat example: Study names and treatment codes for the 10 included studies and person-years and total mortality observed in each study.....	62
Table A3 Dietary fat example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effect $d_{12}$ , absolute effects of the control diet ( $T_1$ ) and the reduced fat diet ( $T_2$ ) for a log-rate of mortality on the control diet with mean -3 and precision 1.77, heterogeneity parameter $\tau$ and model fit statistics. ....	66
Table A4 Diabetes example: study names, follow-up time in years, treatments compared, total number of new cases of diabetes and number of patients in each trial arm, where Diuretic = treatment 1, Placebo = treatment 2, $\beta$ blocker = treatment 3, CCB = treatment 4, ACE inhibitor = treatment 5 and ARB = treatment 6. <sup>100</sup> .....	67
Table A5 Diabetes example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effects of Placebo ( $d_{12}$ ), $\beta$ blocker ( $d_{13}$ ), CCB ( $d_{14}$ ), ACE inhibitor ( $d_{15}$ ) and ARB ( $d_{16}$ ) relative to Diuretic; absolute effects of diuretic ( $T_1$ ) Placebo ( $T_2$ ), $\beta$ blocker ( $T_3$ ), CCB ( $T_4$ ), ACE inhibitor ( $T_5$ ) and ARB ( $T_6$ ); heterogeneity parameter $\tau$ and model fit statistics .....	71
Table A6 Schizophrenia example: study names, follow-up time in weeks, treatments compared, total number of events for each of the four states and total number of patients in each trial arm, where Placebo = treatment 1, Olanzapine = 2, Amisulpride = 3, Zotepine = 4, Aripiprazole = 5, Ziprasidone = 6, Paliperidone = 7, Haloperidol = 8, Risperidone = 9. <sup>44</sup> .....	72
Table A7 Schizophrenia example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effects of Olanzapine ( $d_{12}$ ), Amisulpride ( $d_{13}$ ), Zotepine ( $d_{14}$ ), Aripiprazole ( $d_{15}$ ), Ziprasidone ( $d_{16}$ ), Paliperidone ( $d_{17}$ ), Haloperidol ( $d_{18}$ ) and Risperidone ( $d_{19}$ ) relative to Placebo, absolute probabilities of reaching each of the outcomes for Placebo ( $Pr_1$ ), Olanzapine ( $Pr_2$ ), Amisulpride ( $Pr_3$ ), Zotepine ( $Pr_4$ ), Aripiprazole ( $Pr_5$ ), Ziprasidone ( $Pr_6$ ), Paliperidone ( $Pr_7$ ), Haloperidol ( $Pr_8$ ) and Risperidone ( $Pr_9$ ); heterogeneity parameter $\tau$ for each of the three outcomes, and model fit statistics for the fixed and random effects models. ....	77
Table A8 Parkinson's example: study names, treatments compared, mean off-time reduction with its standard deviation, total number of patients in each trial arm; treatment differences and standard error of the differences; where treatment 1 is a placebo and treatments 2-5 are active drugs.....	79
Table A9 Parkinson example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effects of Treatments 2 to 5 ( $d_{12}$ to $d_{15}$ ) relative to Placebo, absolute effects of Placebo ( $T_1$ ) and treatments 2 to 5 ( $T_2$ to $T_5$ ), heterogeneity parameter $\tau$ and model fit statistics for different data types. ....	82
Table A10 Psoriasis example: study names, treatments compared, total number of patients with different percentage improvement and total number of patients in each trial arm, where Supportive Care = treatment 1, Etanercept 25mg = 2, Etanercept 50 mg = 3, Efalizumab = 4, Ciclosporin = 5, Fumaderm = 6, Infliximab = 7, Methotrexate = 8. <sup>51</sup> .....	86
Table A11 Psoriasis example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effects, on the probit scale, of Etanercept 25 mg ( $d_{12}$ ), Etanercept 50 mg ( $d_{13}$ ), Efalizumab ( $d_{14}$ ), Ciclosporin ( $d_{15}$ ), Fumaderm ( $d_{16}$ ), Infliximab	

( $d_{17}$ ), and Methotrexate ( $d_{18}$ ) relative to Supportive Care; absolute probabilities of achieving at least 50, 70 or 90% relief in symptoms for each treatment; heterogeneity parameter  $\tau$  and model fit statistics. .... 91

Figure 1 Blocker example: Plot of leverage versus Bayesian deviance residual  $w_{ik}$  for each data point, with curves of the form  $x^2+y=c$ , with  $c=1$  (solid),  $c=2$  (dashed),  $c=3$  (dotted) and  $c=4$  (dot-dashed), for the fixed effect model. .... 17

Figure 2 Blocker example: Plot of leverage versus Bayesian deviance residual  $w_{ik}$  for each data point, with curves of the form  $x^2+y=c$ , with  $c=1$  (solid),  $c=2$  (dashed),  $c=3$  (dotted) and  $c=4$  (dot-dashed), for the random effects model. .... 18

Figure A3 Diabetes network: each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison and the numbers by the treatment names are the treatment codes used in the modelling. .... 68

Figure A4 Schizophrenia network: each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison and the numbers by the treatment names are the treatment codes used in the modelling. .... 73

Figure A5 Parkinson network: each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison and the numbers by the treatment names are the treatment codes used in the modelling. .... 79

Figure A6 Psoriasis network: each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison and the numbers by the treatment names are the treatment codes used in the modelling. One trial compared two arms of Ciclosporin with Placebo and another compared two arms of Infliximab with placebo – these comparisons are not represented in the network. .... 83

## **Abbreviations and Definitions**

ACR	American College of Rheumatology
ANCOVA	Analysis of covariance
CEA	cost-effectiveness analysis
cloglog	complementary log-log
DIC	Deviance information criterion
$\Phi$	Normal cumulative distribution function
GLM	Generalised linear models
LRR	log-relative risk
MAR	missing at random
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
MTC	Mixed treatment comparisons
N	normal distribution
NNT	numbers needed to treat
PASI	Psoriasis area severity score
RCT	Randomised controlled trial
RD	risk difference
RR	relative risk
SMD	Standardised mean difference



# 1 INTRODUCTION TO PAIRWISE & NETWORK META-ANALYSIS

Meta-analysis, the pooling of evidence from independent sources, especially randomised controlled trials (RCTs) is now common in the medical research literature. There is a substantial literature on statistical methods for meta-analysis, going back to methods for combination of results from two-by-two tables,<sup>1</sup> with the introduction of random effects meta-analysis<sup>2</sup> a second important benchmark in the development of the field. Over the years methodological and software advances have contributed to the widespread use of meta-analytic techniques. A series of instructional texts and reviews have appeared,<sup>3-7</sup> and Sutton and Higgins<sup>8</sup> provide a review of recent developments.

With some exceptions,<sup>9,10</sup> there have been few attempts to systematise the field. A wide range of alternative methods are employed, mostly relevant to binary and continuous outcomes. Our purpose here is to present a single unified account of evidence synthesis of aggregate data from RCTs, specifically, but not exclusively, for use in probabilistic decision making.<sup>11</sup> In order to cover the variety of outcomes reported in trials and the range of data transformations required to achieve linearity, we adopt the framework of generalised linear modelling.<sup>12</sup> This provides for Normal, Binomial, Poisson and Multinomial likelihoods, with identity, logit, log, complementary log-log, and probit link functions, and common core models for the linear predictor in both fixed effects and random effects settings.

Indirect and mixed treatment comparisons (MTC), also known as network meta-analysis, represent a recent development in evidence synthesis, particularly in decision making contexts.<sup>13-23</sup> Rather than pooling information on trials comparing treatments A and B, network meta-analysis combines data from randomised comparisons, A vs B, A vs C, A vs D, B vs D, and so on, to deliver an internally consistent set of estimates while respecting the randomisation in the evidence.<sup>24</sup> Our common core models are designed for network meta-analysis, and can synthesise data from pair-wise meta-analysis, multi-arm trials, indirect comparisons and network meta-analysis without distinction. Indeed, pair-wise meta-analysis and indirect comparisons are special cases of network meta-analysis.

The common Generalised Linear Model (GLM) framework can, of course, be applied in either frequentist or Bayesian contexts. However, Bayesian Markov Chain Monte Carlo (MCMC) has for many years been the mainstay of “comprehensive decision analysis”,<sup>25</sup> because simulation from a Bayesian posterior distribution supplies both statistical estimation and inference, and a platform for probabilistic decision making under uncertainty. The freely available WinBUGS 1.4.3 MCMC package<sup>26</sup> takes full advantage of the modularity afforded

by a GLM approach to synthesis, allowing us to present a unified treatment of the fixed and random effects models for meta-analysis and model critique.

In Section 2 we present the standard Bayesian MCMC approach to pair-wise meta-analysis for binomial data, based on Smith et al.<sup>6</sup>. We then develop our approach to assessment of goodness of fit, model diagnostics and comparison based on the residual deviance and the Deviance Information Criterion (DIC).<sup>27</sup> In Section 3 the GLM framework for continuous, Poisson, and Multinomial likelihoods is developed with identity, log, complementary log-log and probit link functions, with an introduction to competing risks and ordered probit models. Section 3.4, on continuous outcomes, describes methods for “before-after” differences. All these models have a separate likelihood contribution for each trial arm: in Section 3.5 we develop a modified core model for forms of meta-analysis in which the likelihood is based on a summary treatment difference and its variance. Section 4 shows how different trial reporting formats can be accommodated within the same synthesis in shared parameter models. In Section 5 the core linear predictor models for pair-wise meta-analysis are shown to be immediately applicable to indirect comparisons, multi-arm trials, and network meta-analysis, without further extension.

An extensive appendix provides code to run a series of worked examples, and fully annotated WinBUGS code is also available at [www.nicedsu.org.uk](http://www.nicedsu.org.uk). Section 6 provides advice on formulation of priors and a number of technical issues in MCMC computation.

While Bayesian MCMC is surely the most convenient approach, particularly in decision making, it is certainly not the only one, and there have been a series of recent developments in frequentist software for evidence synthesis. These are briefly reviewed in Section 7, where we also outline the key issues in using frequentist methods in the context of probabilistic decision making. Section 8 provides some pointers to further reading, and more advanced extensions, and we conclude with a brief discussion.

This technical guide is the second in a series of technical support documents on methods for evidence synthesis in decision making. It focuses exclusively on synthesis of relative treatment effect data from randomised controlled trial (RCTs). Issues such as evidence consistency, and the construction of models for absolute treatment effects, are taken up in other guides in this series (see TSDs 4<sup>28</sup> and 5<sup>29</sup>).

## 2 DEVELOPMENT OF THE CORE MODELS: BINOMIAL DATA WITH LOGIT LINK

Consider a set of  $M$  trials comparing two treatments 1 and 2 in a pre-specified target patient population, which are to be synthesised in a meta-analysis. A fixed effect analysis would assume that each study  $i$  generates an estimate of the same parameter  $d_{12}$ , subject to sampling error. In a random effects model, each study  $i$  provides an estimate of the study-specific treatment effects  $\delta_{i,12}$  which are assumed not to be equal but rather *exchangeable*. This means that all  $\delta_{i,12}$  are ‘similar’ in a way which assumes that the trial labels,  $i$ , attached to the treatment effects  $\delta_{i,12}$  are irrelevant. In other words, the information that the trials provide is independent of the order in which they were carried out, over the population of interest.<sup>30</sup> The exchangeability assumption is equivalent to saying that the trial-specific treatment effects come from a common distribution with mean  $d_{12}$  and variance  $\sigma_{12}^2$ .

The common distribution is usually chosen to be a normal distribution, so that

$$\delta_{i,12} \sim N(d_{12}, \sigma_{12}^2) \tag{1}$$

It follows that the fixed effect model is a special case of this, obtained by setting the variance to zero.

Note that in the case of a meta-analysis of only two treatments the subscripts in  $d$ ,  $\delta$  and  $\sigma$  are redundant since only one treatment comparison is being made. We shall drop the subscripts for  $\sigma$ , but will keep the subscripts for  $\delta$  and  $d$ , to allow for extensions to multiple treatments in Section 5.

### 2.1 WORKED EXAMPLE: A LOGIT MODEL FOR A META-ANALYSIS OF BINOMIAL DATA

Carlin<sup>31</sup> and the WinBUGS user manual<sup>26</sup> consider a meta-analysis of 22 trials of beta-blockers to prevent mortality after myocardial infarction. The data available are the number of deaths in the treated and control arms, out of the total number of patients in each arm, for all 22 trials (Table 1).

**Table 1 Blocker example: number of events and total number of patients in the control and beta-blocker groups for the 22 trials.**<sup>31</sup>

study <i>i</i>	Control		Treatment	
	no. of events ( <i>r<sub>i1</sub></i> )	no. of patients ( <i>n<sub>i1</sub></i> )	no. of events ( <i>r<sub>i2</sub></i> )	no. of patients ( <i>n<sub>i2</sub></i> )
1	3	39	3	38
2	14	116	7	114
3	11	93	5	69
4	127	1520	102	1533
5	27	365	28	355
6	6	52	4	59
7	152	939	98	945
8	48	471	60	632
9	37	282	25	278
10	188	1921	138	1916
11	52	583	64	873
12	47	266	45	263
13	16	293	9	291
14	45	883	57	858
15	31	147	25	154
16	38	213	33	207
17	12	122	28	251
18	6	154	8	151
19	3	134	6	174
20	40	218	32	209
21	43	364	27	391
22	39	674	22	680

### 2.1.1 Model specification

Defining  $r_{ik}$  as the number of events (deaths), out of the total number of patients in each arm,  $n_{ik}$ , for arm  $k$  of trial  $i$ , we assume that the data generation process follows a Binomial likelihood i.e.

$$r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik}) \quad (2)$$

where  $p_{ik}$  represents the probability of an event in arm  $k$  of trial  $i$  ( $i=1, \dots, 22; k=1, 2$ ).

Since the parameters of interest,  $p_{ik}$ , are probabilities and therefore can only take values between 0 and 1, a transformation (link function) is used that maps these probabilities into a continuous measure between plus and minus infinity. For a Binomial likelihood the most commonly used link function is the logit link function (see Table 3). We model the probabilities of success  $p_{ik}$  on the logit scale as

$$\text{logit}(p_{ik}) = \mu_i + \delta_{i,1k} I_{\{k \neq 1\}} \quad (3)$$

where

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

In this setup,  $\mu_i$  are trial-specific baselines, representing the log-odds of the outcome in the ‘control’ treatment (i.e. the treatment indexed 1),  $\delta_{i,12}$  are the trial-specific log-odds ratios of success on the treatment group (2) compared to control (1). We can write equation (3) as

$$\begin{aligned} \text{logit}(p_{i1}) &= \mu_i \\ \text{logit}(p_{i2}) &= \mu_i + \delta_{i,12} \end{aligned}$$

where, for a random effects model the trial-specific log-odds ratios come from a common distribution:  $\delta_{i,12} \sim N(d_{12}, \sigma^2)$ . For a fixed effect model we replace equation (3) with

$$\text{logit}(p_{ik}) = \mu_i + d_{12} \times I_{\{k \neq 1\}}$$

which is equivalent to setting the between-trial heterogeneity  $\sigma^2$  to zero thus assuming homogeneity of the underlying true treatment effects.

An important feature of *all* the meta-analytic models presented here is that no model is assumed for the trial-specific baselines  $\mu_i$ . They are regarded as nuisance parameters which are estimated in the model. An alternative is to place a second hierarchical model on the trial baselines, or to put a bivariate normal model on both.<sup>32,33</sup> However, unless this model is correct, the estimated relative treatment effects will be biased. Our approach is therefore more conservative, and in keeping with the widely used frequentist methods in which relative effect estimates are treated as data (see Section 3.5) and baselines eliminated entirely. Baseline models are discussed in TSD5.<sup>29</sup>

### 2.1.2 Model fit and model comparison

To check formally whether a model’s fit is satisfactory, we will consider an absolute measure of fit: the overall residual deviance:  $\bar{D}_{res}$ . This is the posterior mean of the deviance under the current model, minus the deviance for the saturated model,<sup>12</sup> so that each data point should

contribute about 1 to the posterior mean deviance.<sup>27,34</sup> We can then compare the value of  $\bar{D}_{res}$  to the number of independent data points to check if the model fit can be improved. For a Binomial likelihood each trial arm contributes 1 independent data point and the residual deviance is calculated as

$$\begin{aligned} D_{res} &= \sum_i \sum_k 2 \left( r_{ik} \log \left( \frac{r_{ik}}{\hat{r}_{ik}} \right) + (n_{ik} - r_{ik}) \log \left( \frac{n_{ik} - r_{ik}}{n_{ik} - \hat{r}_{ik}} \right) \right) \\ &= \sum_i \sum_k dev_{ik} \end{aligned} \quad (4)$$

where  $r_{ik}$  and  $n_{ik}$  are the observed number of events and patients in each trial arm,  $\hat{r}_{ik} = n_{ik} p_{ik}$  is the expected number of events in each trial arm calculated at each iteration, based on the current model, and  $dev_{ik}$  is the deviance residual for each data point calculated at each iteration. This is then summarised by the posterior mean:  $\bar{D}_{res}$ .

Leverage statistics are familiar from frequentist regression analysis where they are used to assess the influence that each data point has on the model parameters. The leverage for each data point,  $leverage_{ik}$ , is calculated as the posterior mean of the residual deviance minus the deviance at the posterior mean of the fitted values. For a Binomial likelihood, letting  $\tilde{r}_{ik}$  be the posterior mean of  $\hat{r}_{ik}$ , and  $\overline{dev}_{ik}$  the posterior mean of  $dev_{ik}$ ,

$$p_D = \sum_i \sum_k leverage_{ik} = \sum_i \sum_k \left[ \overline{dev}_{ik} - \widetilde{dev}_{ik} \right]$$

where  $\widetilde{dev}_{ik}$  is the posterior mean of the deviance calculated by replacing  $\hat{r}_{ik}$  with  $\tilde{r}_{ik}$  in equation (4).

The Deviance Information Criterion (DIC)<sup>27</sup> is the sum of the posterior mean of the residual deviance,  $\bar{D}_{res}$ , and the leverage,  $p_D$ , (also termed the effective number of parameters). The DIC provides a measure of model fit that penalises model complexity – lower values of the DIC suggest a more parsimonious model. The DIC is particularly useful for comparing different parameter models for the same likelihood and data, for example fixed and random effects models or fixed effect models with and without covariates.

If the deviance residuals provide indications that the model does not fit the data well, leverage plots can give further information on whether poorly fitting data points are having a material effect on the model parameters. Leverage plots show each data point's contribution

to  $p_D$  ( $leverage_{ik}$ ) plotted against their contribution to  $\bar{D}_{res}$  ( $\overline{dev}_{ik}$ ) and can be used to check how each point is affecting the overall model fit and DIC. It is useful to display these summaries in a plot of  $leverage_{ik}$  vs  $w_{ik}$  for each data point, where  $w_{ik} = \pm\sqrt{\overline{dev}_{ik}}$ , with sign given by the sign of  $(r_{ik} - \hat{r}_{ik})$  to indicate whether the data is over- or under-estimated by the model. Curves of the form  $x^2 + y = c$ ,  $c=1,2,3,\dots$ , where  $x$  represents  $w_{ik}$  and  $y$  represents the leverage, are marked on the plots and points lying on such parabolas each contribute an amount  $c$  to the DIC.<sup>27</sup> Points which lie outside the lines with  $c=3$  can generally be identified as contributing to the model's poor fit. Points with a high leverage are influential, which means that they have a strong influence on the model parameters that generate their fitted values.

Leverage plots for the fixed and random effects models are presented in Figure 1 and Figure 2, respectively. From these the random effects model appears to be more appropriate as points lie closer to the centre of the plot. To further examine the model fit at individual data points, inspection of  $\overline{dev}_{ik}$  for all  $i$  and  $k$  will highlight points with a high residual deviance, over 2 say, as accounting for the lack of fit. This can help identify data points that fit poorly.

WinBUGS will calculate  $p_D$  and the posterior mean of the deviance for the current model  $\bar{D}$ , but will not output the contributions of the individual data points to the calculations. Furthermore, without subtracting the deviance for the saturated model,  $\bar{D}$  is hard to interpret and can only be useful for model comparison purposes and not to assess the fit of a single model. Therefore users wishing to produce leverage plots such as those in Figure 1 and Figure 2 need to calculate the contributions of individual studies to  $\bar{D}_{res}$  and to the leverage themselves. The latter needs to be calculated outside WinBUGS, for example in R or Microsoft Excel. The  $p_D$ , and therefore the DIC, calculated in the way we suggest is not precisely the same as that calculated in WinBUGS, except in the case of a normal likelihood. This is because WinBUGS calculates the fit at the mean value of the parameter values, while we propose the fit at the mean value of the fitted values. The latter is more stable in highly non-linear models with high levels of parameter uncertainty.

In this document we suggest that global DIC statistics and  $\bar{D}_{res}$  are consulted both to compare fixed and random effect models, and to ensure that overall fit is adequate. Leverage plots may be used to identify influential and/or poorly fitting observations. Guidance on choice of fixed or random effects model, an issue that is closely bound up with the impact of sparse data and

choice of prior distributions, is given in Section 6. In network meta-analysis there are additional issues regarding consistency between evidence sources on different contrasts. This is discussed fully in TSD4.<sup>28</sup>

### 2.1.3 WinBUGS implementation and illustrative results

Annotated WinBUGS 1.4.3 code is shown in the Appendix, for both a random effects model and a fixed effect model (Blocker Examples 1(c) and 1(d)). Included in the description of the code are some additional comments on alternative priors, and additional code that can be used when there are more than two treatments being compared, to rank the treatments, or compute the probability that each is the best treatment. We ran both fixed and random effects models, and some of the results, including the  $\bar{D}_{res}$  and DIC statistics, are shown in Table 2. All results are based on 20,000 iterations on 3 chains, after a burn-in of 10,000.

**Table 2 Blocker example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effect  $d_{12}$ , absolute effects of the placebo ( $T_1$ ) and beta-blocker ( $T_2$ ) for a mean mortality of -2.2 and precision 3.3 on the logit scale; heterogeneity parameter  $\sigma$  and model fit statistics.**

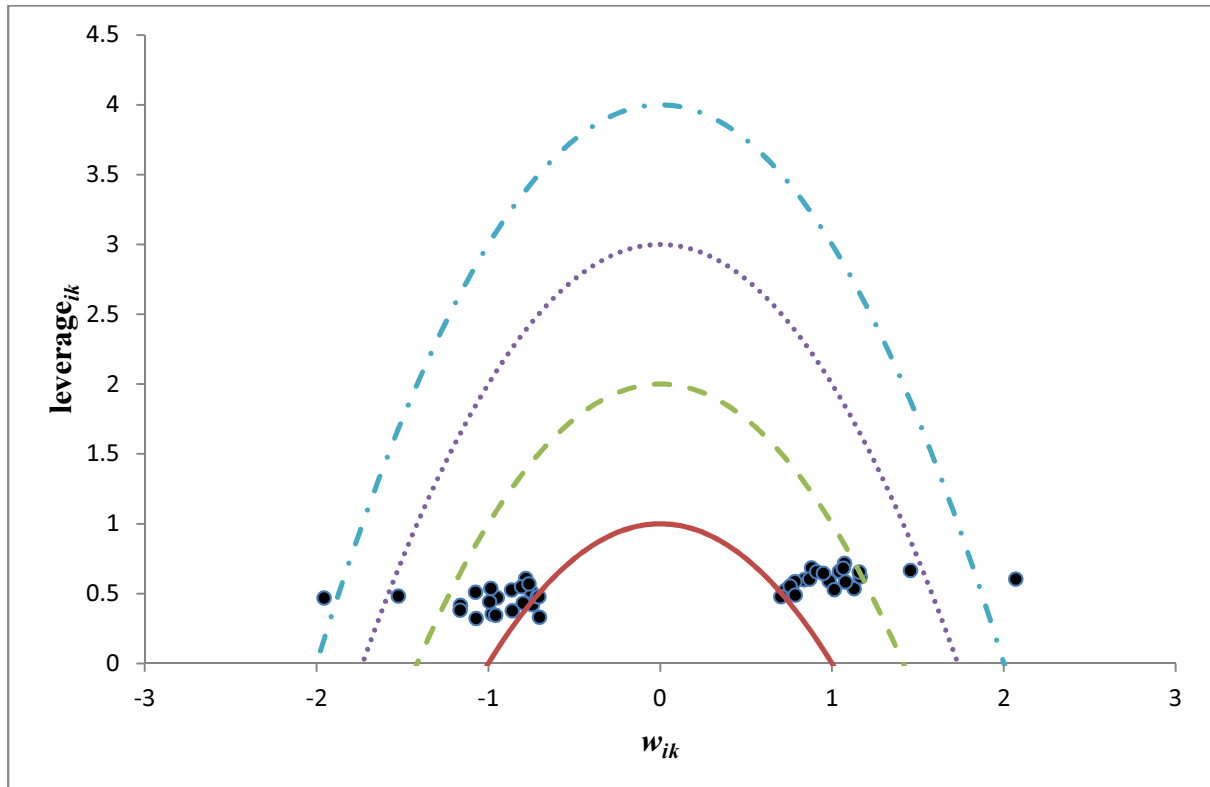
	Fixed Effect model				Random Effects model			
	mean	sd	median	CrI	mean	sd	median	CrI
$d_{12}$	-0.26	0.050	-0.26	(-0.36,-0.16)	-0.25	0.066	-0.25	(-0.38,-0.12)
$T_1$	0.11	0.055	0.10	(0.04,0.25)	0.11	0.055	0.10	(0.04,0.25)
$T_2$	0.09	0.045	0.08	(0.03,0.20)	0.09	0.046	0.08	(0.03,0.20)
$\sigma$	-	-	-	-	0.14	0.082	0.13	(0.01,0.32)
$\bar{D}_{res}$ *	46.8				41.9			
pD	23.0				28.1			
DIC	69.8				70.0			

\* compare to 44 data points

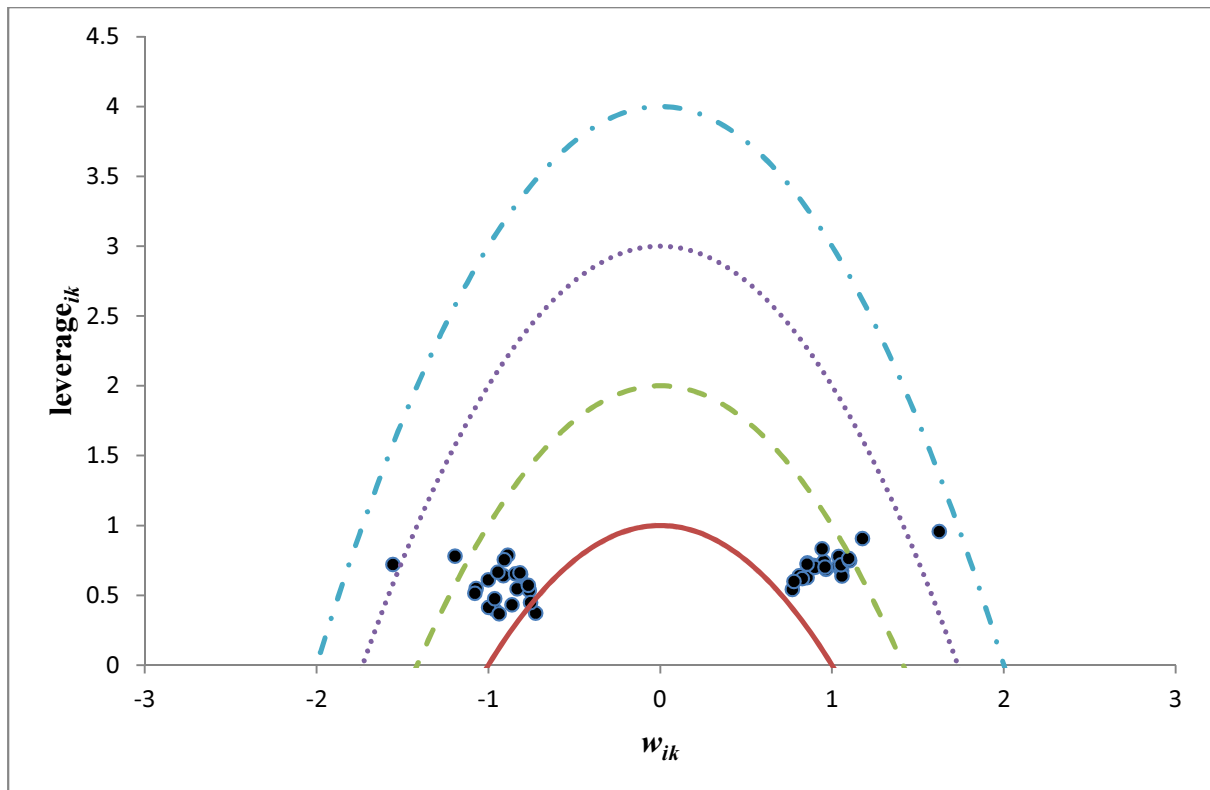
Comparing the fit of both these models using the posterior mean of the residual deviance indicates that although the random effects models is a better fit to the data, with a posterior mean of the residual deviance of 41.9 against 46.8 for the fixed effect model, this is achieved at the expense of more parameters. This better fit can also be seen in the leverage plots for the fixed and random effects model (Figure 1 and Figure 2), where two extreme points can be seen in Figure 1, at either side of zero. These points refer to the two arms of study 14 (Table 1) but are no longer so extreme in Figure 2. We would suggest careful re-examination of the evidence and consideration of issues such as the existence of important covariates. These and other issues are covered in TSD3.<sup>35</sup> The DIC suggests that there is little to choose between



the two models and the fixed effect model may be preferred since it is easier to interpret (Table 2). The posterior median of the pooled log odds ratio of beta-blockers compared to control in the fixed effect model is -0.26 with 95% Credible Interval (-0.36, -0.16) indicating a reduced mortality in the treatment group. The posterior medians of the probability of mortality on the control and treatment groups are 0.10 and 0.08, respectively (with credible intervals in Table 2). Results for the random effects model are similar.



**Figure 1 Blocker example:** Plot of leverage versus Bayesian deviance residual  $w_{ik}$  for each data point, with curves of the form  $x^2 + y = c$ , with  $c = 1$  (solid),  $c = 2$  (dashed),  $c = 3$  (dotted) and  $c = 4$  (dot-dashed), for the fixed effect model.



**Figure 2 Blocker example:** Plot of leverage versus Bayesian deviance residual  $w_{ik}$  for each data point, with curves of the form  $x^2+y=c$ , with  $c=1$  (solid),  $c=2$  (dashed),  $c=3$  (dotted) and  $c=4$  (dot-dashed), for the random effects model.

The logit model assumes linearity of effects on the logit scale. A number of authors, notably Deeks,<sup>36</sup> have rightly emphasised the importance of using a scale in which effects are additive, as is required by the linear model. Choice of scale can be guided by goodness of fit, or by lower between-study heterogeneity, but there is seldom enough data to make this choice reliably, and logical considerations (see below) may play a larger role. Quite distinct from choice of scale for modelling, is the issue of how to report treatment effects. Thus, while one might assume linearity of effects on the logit scale, the investigator, given information on the absolute effect of one treatment, is free to derive treatment effects on other scales, such as Risk Difference (RD), Relative Risk (RR), or Numbers Needed to Treat (NNT). The computer code provided in the Appendix shows how this can be done. An advantage of Bayesian MCMC is that appropriate distributions, and therefore credible intervals, are automatically generated for all these quantities.

### 3 GENERALISED LINEAR MODELS

We now extend our treatment to models other than the well-known logit link for data with a binomial likelihood. The essential idea is that the basic apparatus of the meta-analysis remains the same, but the *likelihood* and the *link function* can change to reflect the nature of the data (continuous, rate, categorical), and the sampling process that generated it (Normal, Poisson, Multinomial, etc). In GLM theory,<sup>12</sup> a likelihood is defined in terms of some unknown parameters  $\gamma$  (for example a Binomial likelihood as in Section 2), while a link function,  $g(\cdot)$ , maps the parameters of interest onto the plus/minus infinity range. Our meta-analysis model for the logit link in equation (3), now becomes a GLM taking the form

$$g(\gamma) = \theta_{ik} = \mu_i + \delta_{i,bk} I_{\{k \neq 1\}} \quad (5)$$

where  $g$  is an appropriate link function (for example the logit link), and  $\theta_{ik}$  is the linear predictor, usually a continuous measure of the treatment effect in arm  $k$  of trial  $i$  (for example the log-odds). As before,  $\mu_i$  are the trial-specific baseline effects in a trial  $i$ , treated as unrelated nuisance parameters. The  $\delta_{i,bk}$  are the trial-specific treatment effect of the treatment in arm  $k$  relative to the control treatment in arm  $b$  ( $b=1$ ) in that trial, and

$$\delta_{i,12} \sim N(d_{12}, \sigma^2) \quad (6)$$

as in equation (1).

**Table 3 Commonly used link functions and their inverse with reference to which likelihoods they can be applied to.**

<b>Link</b>	<b>Link function</b> $\theta = g(\gamma)$	<b>Inverse link function</b> $\gamma = g^{-1}(\theta)$	<b>Likelihood</b>
Identity	$\gamma$	$\theta$	Normal
Logit	$\ln\left(\frac{\gamma}{1-\gamma}\right)$	$\frac{\exp(\theta)}{1 + \exp(\theta)}$	Binomial Multinomial
Log	$\ln(\gamma)$	$\exp(\theta)$	Poisson
Complementary log-log (cloglog)	$\ln\{-\ln(1-\gamma)\}$	$1 - \exp\{-\exp(\theta)\}$	Binomial Multinomial
Reciprocal link	$1/\gamma$	$1/\theta$	Gamma
Probit	$\Phi^{-1}(\gamma)$	$\Phi(\theta)$	Binomial Multinomial

We now turn to consider the different types of outcome data generated in trials, and the GLMs required to analyse them. In each case, the basic model for meta-analysis remains the same (equations (5) and (6)). What changes are the likelihood and the link function. In a Bayesian framework, we also need to pay careful attention to the specification of the priors for the variance parameter. Table 3 has details of the most commonly used likelihoods, link and inverse link functions. The formulae for the residual deviance and the predicted values needed to calculate  $p_D$  for all the different likelihoods described are available in Table 4.

**Table 4** Formulae for the residual deviance and model predictors for common likelihoods

Likelihood	Model prediction	Residual Deviance
$r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik})$	$\hat{r}_{ik} = n_{ik} p_{ik}$	$\sum_i \sum_k 2 \left( r_{ik} \log \left( \frac{r_{ik}}{\hat{r}_{ik}} \right) + (n_{ik} - r_{ik}) \log \left( \frac{n_{ik} - r_{ik}}{n_{ik} - \hat{r}_{ik}} \right) \right)$
$r_{ik} \sim \text{Poisson}(\lambda_{ik} E_{ik})$	$\hat{r}_{ik} = \lambda_{ik} E_{ik}$	$\sum_i \sum_k 2 \left( (\hat{r}_{ik} - r_{ik}) + r_{ik} \log \left( \frac{r_{ik}}{\hat{r}_{ik}} \right) \right)$
$y_{ik} \sim N(\bar{y}_{ik}, se_{ik}^2)$ $se_{ik}$ assumed known	$\bar{y}_{ik}$	$\sum_i \sum_k \left( \frac{(y_{ik} - \bar{y}_{ik})^2}{se_{ik}^2} \right)$
$r_{i,k,1:J} \sim \text{Multinomial}(p_{i,k,1:J}, n_{ik})$	$\hat{r}_{ikj} = n_{ik} p_{ikj}$	$\sum_i \sum_k 2 \left( \sum_j r_{ikj} \log \left( \frac{r_{ikj}}{\hat{r}_{ikj}} \right) \right)$
Multivariate Normal $y_{i,1:k} \sim N_k(\bar{y}_{i,1:k}, \Sigma_{(k \times k)})$	$\bar{y}_{i,1:k}$	$\sum_i (y_{i,1:k} - \bar{y}_{i,1:k})^T \Sigma^{-1} (y_{i,1:k} - \bar{y}_{i,1:k})$

### 3.1 RATE DATA: POISSON LIKELIHOOD AND LOG LINK

When the data available for the RCTs included in the meta-analysis is in the form of counts over a certain time period (which may be different for each trial), a Poisson likelihood and a log link is used. Examples would be the number of deaths, or the number of patients in whom a device failed. But, rather than having a denominator *number* at risk, what is supplied is a total number of *person-years* at risk. For patients who do not reach the end event, the time at risk is the same as their follow-up time. For those that do, it is the time from the start of the trial to the event: in this way the method allows for censored observations.

Defining  $r_{ik}$  as the number of events occurring in arm  $k$  of trial  $i$  during the trial follow-up period,  $E_{ik}$  as the exposure time in person-years and  $\lambda_{ik}$  as the rate at which events occur in arm  $k$  of trial  $i$ , we can write the likelihood as

$$r_{ik} \sim \text{Poisson}(\lambda_{ik} E_{ik})$$

The parameter of interest is the hazard, the rate at which the events occur in each trial arm, and this is modelled on the log scale. The linear predictor in equation (5) is therefore on the log-rate scale:

$$\theta_{ik} = \log(\lambda_{ik}) = \mu_i + \delta_{i,bk} I_{\{k \neq 1\}} \quad (7)$$

A key assumption of this model is that in each arm of each trial the hazard is constant over the follow-up period. This can only be the case in homogeneous populations where all patients have the same hazard rate. In populations with constant but heterogeneous rates, the average hazard must necessarily decrease over time, as those with higher hazard rates tend to reach their end-points earlier and exit from the risk set.

These models are also useful for certain *repeated* event data. Examples would be the number of accidents, where each individual may have more than one accident. Here one would model the total number of accidents in each arm, that is, the average number of accidents multiplied by the number of patients. The Poisson model can also be used for observations repeated in space rather than time: for example the number of teeth requiring fillings. Using the Poisson model for repeated event data makes the additional assumption that the events are independent, so that, for example, an accident is no more likely in an individual who has already had an accident than in one who has not. Readers may consult previous work<sup>37-39</sup> for examples. Dietary Fat Examples 2(a) and 2(b) in the Appendix illustrate random and fixed effects meta-analyses of this sort.

### 3.2 RATE DATA: BINOMIAL LIKELIHOOD AND CLOGLOG LINK

In some meta-analyses, each included trial reports the proportion of patients reaching an end-point at a specified follow-up time, but the trials do not all have the same follow-up time. Defining  $r_{ik}$  as the number of events in arm  $k$  of trial  $i$ , with follow-up time  $f_i$  (measured in days, weeks etc), then the likelihood for the data generating process is Binomial, as in equation (2). Using a logit model implies one of the following assumptions: that all patients who reach the end-point do so by some specific follow-up time, and further follow-up would make no difference; or that the proportional odds assumption holds. This assumption implies a complex form for the hazard rates.<sup>40</sup> If longer follow-up results in more events, the standard

logit model is hard to interpret. The simplest way to account for the different length of follow-up in each trial, is to assume an underlying Poisson process for each trial arm, with a constant event rate  $\lambda_{ik}$ , so that  $T_{ik}$ , the time until an event occurs in arm  $k$  of trial  $i$ , has an exponential distribution

$$T_{ik} \sim \text{Exp}(\lambda_{ik})$$

The probability that there are no events by time  $f_i$  in arm  $k$  of trial  $i$ , the survival function, can be written as

$$\Pr(T_{ik} > f_i) = \exp(-\lambda_{ik} f_i)$$

Then, for each trial  $i$ ,  $p_{ik}$ , the probability of an event in arm  $k$  of trial  $i$  after follow-up time  $f_i$  can be written as

$$p_{ik} = 1 - \Pr(T_{ik} > f_i) = 1 - \exp(-\lambda_{ik} f_i) \quad (8)$$

which is time dependent.

We now model the event rate  $\lambda_{ik}$ , taking into account the different follow-up times  $f_i$ . Since equation (8) is a non-linear function of  $\log(\lambda_{ik})$  the complementary log-log (cloglog) link function<sup>41</sup> (Table 3) is used to obtain a generalised linear model for  $\log(\lambda_{ik})$  giving  $\theta_{ik} = \text{cloglog}(p_{ik}) = \log(f_i) + \log(\lambda_{i,bk})$ , and  $\log(\lambda_{ik})$  is modelled as in equation (7):

$$\theta_{ik} = \text{cloglog}(p_{ik}) = \log(f_i) + \mu_i + \delta_{i,bk} I_{\{k \neq 1\}}$$

with the treatment effects  $\delta_{i,bk}$  representing log-hazard ratios. The Diabetes Example, programs 3(a) and 3(b) in the Appendix, illustrates a cloglog meta-analysis.

The assumptions made in this model are the same as those for the Poisson rate models, namely that the hazards are constant over the entire duration of follow-up. This implies homogeneity of the hazard across patients in each trial, a strong assumption, as noted above. Nonetheless, this assumption may be preferable to assuming that the follow-up time makes no difference to the number of events. The clinical plausibility of these assumptions should be discussed and supported by citing relevant literature, or by examination of evidence of changes in outcome rates over the follow-up period in the included trials.

When the constant hazards assumption is not reasonable, but further follow-up time is believed to result in more events, extensions are available that allow for time-varying rates. One approach is to adopt piece-wise constant hazards. These models can be fitted if there is data reported at multiple follow-up times within the same study.<sup>42,43</sup> An alternative is to fit a Weibull model, which involves an additional “shape” parameter  $\alpha$ :

$$\Pr(T_{ik} > f_i) = \exp[-\lambda_{ik} f_i^\alpha]$$

which leads to:

$$\theta_{ik} = \text{cloglog}(p_{ik}) = \alpha(\log(f_i) + \mu_i + \delta_{i,bk} I_{\{k \neq 1\}})$$

Although no longer a GLM, since a non-linear predictor is used, these extensions lead to major liberalisation of modelling, but require more data. The additional Weibull parameter, for example, can only be adequately identified if there is data on a wide range of follow-up times, and if investigators are content to assume the *same* shape parameter for all treatments.

### 3.3 COMPETING RISKS: MULTINOMIAL LIKELIHOOD AND LOG LINK

A competing risk analysis is appropriate where multiple, mutually exclusive end-points have been defined, and patients leave the risk set if any one of them is reached. For example, in trials of treatments for schizophrenia<sup>44</sup> observations continued until patients either relapsed, discontinued treatment due to intolerable side effects, or discontinued for other reasons. Patients who remain stable to the end of the study are censored. The statistical dependencies between the competing outcomes need to be taken into account in the model. These dependencies are essentially within-trial, negative correlations between outcomes, applying in each arm of each trial. They arise because the occurrence of outcome events is a stochastic process, and if more patients should by chance reach one outcome, then fewer must reach the others.

Trials report  $r_{ikj}$ , the number of patients in arm  $k$  of trial  $i$  reaching each of the mutually exclusive end-points  $j=1,2,\dots,J$ , at the end of follow-up in trial  $i$ ,  $f_i$ . In this case the responses  $r_{ikj}$  will follow a multinomial distribution:

$$r_{i,k,j=1,\dots,J} \sim \text{Multinomial}(p_{i,k,j=1,\dots,J}, n_{ik}) \quad \text{with} \quad \sum_{j=1}^J p_{i,k,j} = 1 \quad (9)$$

and the parameters of interest are the rates (hazards) at which patients move from their initial state to any of the end-points  $j$ ,  $\lambda_{ikj}$ . Note that the  $J^{\text{th}}$  endpoint represents the censored observations, i.e. patients who do not reach any of the other end-points before the end of follow-up.

If we assume constant hazards  $\lambda_{ikj}$  acting over the period of observation  $f_i$  in years, weeks etc, the probability that outcome  $j$  has occurred by the end of the observation period for arm  $k$  in trial  $i$  is:

$$p_{ikj}(f_i) = \frac{\lambda_{ikj}}{\sum_{u=1}^{J-1} \lambda_{iku}} [1 - \exp(-f_i \sum_{u=1}^{J-1} \lambda_{iku})], \quad j = 1, 2, 3, \dots, J-1$$

The probability of remaining in the initial state, that is the probability of being censored, is simply 1 minus the sum of the probabilities of arriving at any of the  $J-1$  absorbing states, ie:

$$p_{ikJ}(f_i) = 1 - \sum_{u=1}^{J-1} p_{iku}(f_i)$$

The parameters of interest are the hazards,  $\lambda_{ikj}$ , and these are modelled on the log scale

$$\theta_{ikj} = \log(\lambda_{ikj}) = \mu_{ij} + \delta_{i,bk,j} I_{\{k \neq 1\}}$$

The trial-specific treatment effects  $\delta_{i,bk,j}$  of the treatment in arm  $k$  relative to the control treatment in arm  $b$  of that trial for outcome  $j$ , are assumed to follow a normal distribution

$$\delta_{i,12,j} \sim N(d_{12j}, \sigma_j^2)$$

The between-trials variance of the random effects distribution,  $\sigma_j^2$ , is specific to each outcome  $j$ . Three models for the variance can be considered: a fixed effect model, where  $\sigma_j^2 = 0$ ; a Random Effects Single Variance model where the between-trial variance  $\sigma_j^2 = \sigma^2$ , reflecting the assumption that the between-trials variation is the same for each outcome; and a Random Effect Different Variances model where  $\sigma_j^2$  denotes a different between-trials variation for each outcome  $j$ . See the Schizophrenia Example 4 in the Appendix for an illustration.

These competing risks models share the same assumptions as the cloglog models presented in Section 3.2 to which they are closely related: constant hazards over time, implying



proportional hazards, for each outcome. A further assumption is that the ratios of the risks attaching to each outcome must also remain constant over time (proportional competing risks). Further extensions where the assumptions are relaxed are available.<sup>45</sup>

### 3.4 CONTINUOUS DATA: NORMAL LIKELIHOOD AND IDENTITY LINK

With continuous outcome data the meta-analysis is based on the sample means,  $y_{ik}$ , with standard errors  $se_{ik}$ . As long as the sample sizes are not too small, the Central Limit Theorem allows us to assume that, even in cases where the underlying data are skewed, the sample means are approximately normally distributed, so that the likelihood can be written as

$$y_{ik} \sim N(\theta_{ik}, se_{ik}^2)$$

The parameter of interest is the mean,  $\theta_{ik}$ , of this continuous measure which is unconstrained on the real line. The identity link is used (Table 3) and the linear model can be written on the natural scale as

$$\theta_{ik} = \mu_i + \delta_{i,bk} I_{\{k \neq 1\}} \quad (10)$$

See the Parkinson's Example, programs 5(a) and 5(b) in the Appendix, for WinBUGS code.

#### 3.4.1 Before/after studies: change from baseline measures

In cases where the original trial outcome is continuous and measured at baseline and at a pre-specified follow-up point the most common method is to base the meta-analysis on the mean change from baseline for each patient and an appropriate measure of uncertainty (e.g. the variance or standard error) which takes into account any within-patient correlation. It should be noted that the most efficient and least biased statistic to use is the mean of the final reading, having adjusted for baseline via regression/ANCOVA. Although this is seldom reported, when available these should be the preferred outcome measures.<sup>5</sup>

The likelihood for the mean change from baseline in arm  $k$  of trial  $i$ ,  $y_{ik}^\Delta$ , with change variance  $V_{ik}^\Delta$  can be assumed normal such that

$$y_{ik}^\Delta \sim N(\theta_{ik}, V_{ik}^\Delta)$$

The parameter of interest is the mean,  $\theta_{ik}$ , of this continuous measure which is unconstrained on the real line. The identity link is used (Table 3) and the linear model can be written on the natural scale as in equation (10).

However, in practice many studies fail to report an adequate measure of the uncertainty for the before-after difference in outcome and instead report the mean and variance,  $y_{ik}^{(b)}$  and  $V_{ik}^{(b)}$ , (or other measure of uncertainty) at baseline (before), and at follow-up times (after),  $y_{ik}^{(a)}$  and  $V_{ik}^{(a)}$ , separately. While the mean change from baseline can be easily calculated as

$$y_{ik}^{\Delta} = y_{ik}^{(b)} - y_{ik}^{(a)}$$

to calculate  $V_{ik}^{\Delta}$  for such trials, information on the within-patient correlation  $\rho$  is required since

$$V_{ik}^{\Delta} = V_{ik}^{(b)} + V_{ik}^{(a)} - 2\rho\sqrt{V_{ik}^{(b)}V_{ik}^{(a)}}$$

Information on the correlation  $\rho$  is seldom available. It may be possible to obtain information from a review of similar trials using the same outcome measures, or else a reasonable value for  $\rho$ , often 0.5 (which is considered conservative) or 0.7,<sup>46</sup> can be used alongside sensitivity analyses.<sup>5,47</sup> A more sophisticated approach, which takes into account the uncertainty in the correlation, is to use whatever information is available within the dataset, from trials that report both the before/after variances and the change variance (see Section 4), and possibly external trials as well, to obtain an evidence-based prior distribution for the correlation, or even to estimate the correlation and the treatment effect simultaneously within the same analysis.<sup>48</sup>

### 3.5 TREATMENT DIFFERENCES

Trial results are sometimes only available as overall, trial-based summary measures, for example as mean differences between treatments, log-odds ratios, log-risk ratios, log-hazard ratios, risk differences, or some other trial summary statistic and its sample variance. In this case we can assume a normal distribution for the continuous measure of treatment effect of arm  $k$  relative to arm 1 in trial  $i$ ,  $y_{ik}$ , with variance  $V_{ik}$ , such that

$$y_{ik} \sim N(\theta_{ik}, V_{ik})$$

The parameters of interest are the trial-specific mean treatment effects  $\theta_{ik}$ . An identity link is used and since no trial-specific effects of the baseline or control treatment can be estimated the linear predictor is reduced to  $\theta_{ik} = \delta_{i,bk}$ . The trial baselines are eliminated and the  $\delta_{i,bk}$  are, exactly as in all previous models, assumed to come from a random effects distribution  $\delta_{i,12} \sim N(d_{12}, \sigma^2)$  or to be fixed  $\delta_{i,12} = d_{12}$ . Examples 7 (Parkinson's Differences) in the Appendix can be consulted.

Readers will recognise that this is overwhelmingly the most common form of meta-analysis, especially amongst the Frequentist methods. The case where the  $y_{ik}$  are log-odds ratios, and an inverse-variance weighting is applied, with variance based on the normal theory approximation, remains a main-stay in applied meta-analytic studies. We refer to some of the key literature comparing different meta-analytic estimators and methods in the discussion.

An important caveat about synthesis based on treatment differences relates to multi-arm trials. In Section 5 we show how the framework developed so far applies to syntheses that include multi-arm trials. However, trial-level data based on treatment differences present some special problems because, unlike data aggregated at the arm-level, there are correlations between the treatment differences *that require adjustment to the likelihood*. Details are given in Section 5.1. The WinBUGS coding we provide (Example 7) incorporates these adjustments. This point is also taken up in our discussion of alternative software (Section 7).

### 3.5.1 Standardised mean differences

There are a series of standardised mean difference (SMD) measures commonly used with psychological or neurological outcome measures. These can be synthesised in exactly the same way as any other treatment effect summary. We include some specific comments here relating to the special issues they raise.

The main role of the SMD is to facilitate combining results from trials which have reported outcomes measured on different continuous scales. For example, some trials might use the Hamilton Depression scale, others the Montgomery-Asberg Depression Rating Scale. The idea is that the two scales are measuring essentially the same quantity, and that results can be placed on a common scale if the mean difference between the two arms in each trial is divided by its standard deviation. The best known SMD measures are Cohen's  $d$ <sup>49</sup>, and

Hedges' adjusted  $g$ ,<sup>50</sup> which differ only in how the pooled standard deviation is defined and the fact that Hedges'  $g$  is adjusted for small sample bias:

$$\text{Cohen's } d = \frac{\text{difference in means}}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2}}}$$

$$\text{Hedges' (adjusted) } g = \frac{\text{difference in means}}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}} \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right) \quad (11)$$

where  $n_1$  and  $n_2$  represent the sample sizes and  $s_1$  and  $s_2$  the standard errors of the means in arms 1 and 2 of a given trial.

However, dividing estimates through by the sample standard deviation introduces additional heterogeneity in two ways. First, standard deviations are themselves subject to sampling error, and secondly, the use of SMD opens the results to various kinds of distortion because trials vary in how narrowly defined the patient population is. For example we would expect trials with narrow inclusion criteria such as “severe depression”, to have smaller sample standard deviations, and thus larger SMDs, than trials on patients with “severe to moderate depression”. A procedure that would produce more interpretable results would be to divide all estimates from a given test instrument by the standard deviation obtained in a representative population sample, external to the trial.

The Cochrane Collaboration recommends the use of Hedges'  $g$  (equation (11)), while noting that interpretation of the overall intervention effect is difficult.<sup>5</sup> It recommends re-expressing the pooled SMD in terms of effect sizes as small, medium or large (according to some rules of thumb), transforming the pooled SMD into an Odds Ratio, or re-expressing the SMD in the units of one or more of the original measurement instruments,<sup>5</sup> although it is conceded none of these manoeuvres mitigates the drawbacks mentioned above.

SMDs are sometimes used for non-continuous outcomes. For example in a review of topical fluoride therapies to reduce caries in children and adolescents, the outcomes were the number of new caries observed but the mean number of caries in each trial arm were modelled as SMD.<sup>51</sup> Where possible, it is preferable to use the appropriate GLM, in this case a Poisson likelihood and log link, as this is likely to reduce heterogeneity.<sup>38</sup>

### 3.6 ORDERED CATEGORICAL DATA: MULTINOMIAL LIKELIHOOD AND PROBIT LINK

In some applications, the data generated by the trial may be continuous but the outcome measure categorised, using *one or more* pre-defined cut-offs. Examples include the PASI (Psoriasis Area Severity Index) and the ACR (American College of Rheumatology) scales, where it is common to report the percentage of patients who have improved by more than certain benchmark relative amounts. Thus ACR-20 would represent the proportion of patients who have improved by at least 20% on the ACR scale, PASI-75 the proportion who have improved by at least 75% on the PASI scale. Trials may report ACR-20, ACR-50 and ACR-70, or only one or two of these end-points. We can provide a coherent model and make efficient use of such data by assuming that the treatment effect is the same regardless of the cut-off. This assumption can be checked informally by examining the relative treatment effects at different cut-offs in each trial and seeing if they are approximately the same. In particular, there should not be a systematic relationship between the relative effects at different cut-off points. The residual deviance check of model fit is also a useful guide.

The likelihood is the same as in the competing risk analysis: trials report  $r_{ikj}$ , the number of patients in arm  $k$  of trial  $i$  belonging to different, mutually exclusive categories  $j=1,2,\dots,J$ , where these categories represent the different thresholds (e.g. 20%, 50% or 70% improvement), on a common underlying continuous scale. The responses for each arm  $k$  of trial  $i$  in category  $j$  will follow a multinomial distribution as defined in equation (9) and the parameters of interest are the probabilities,  $p_{ikj}$ , that a patient in arm  $k$  of trial  $i$  belongs to category  $j$ . We may use the probit link function to map  $p_{ikj}$  onto the real line. This is the inverse of the normal cumulative distribution function  $\Phi$  (see Table 3). The model can be written as

$$\theta_{ikj} = \Phi^{-1}(p_{ikj}) = \mu_{ij} + \delta_{i,bk} I_{\{k \neq 1\}}$$

or equivalently

$$p_{ikj} = \Phi(\mu_{ij} + \delta_{i,bk} I_{\{k \neq 1\}})$$

In this setup, the pooled effect of taking the experimental treatment instead of the control is to change the probit score (or  $Z$  score) of the control arm, by  $\delta_{i,bk}$  standard deviations. This can be translated back into probabilities of events by noting that when the pooled treatment effect

$d_{12} > 0$ , then for a patient population with an underlying probability  $\pi_j$  of an event in category  $j$ , the experimental treatment will increase this probability to  $\Phi(\Phi^{-1}(\pi_j) + d_{12})$ .

The model is set-up with the assumption that there is an underlying continuous variable which has been categorised by specifying different cut-offs,  $z_{ij}$ , which correspond to the point at which an individual moves from one category to the next in trial  $i$ . Several options are available regarding the relationship between outcomes within each arm. Re-writing the model as

$$p_{ikj} = \Phi(\mu_i + z_{ij} + \delta_{i,bk} I_{\{k \neq 1\}})$$

we can consider the terms  $z_{ij}$  as the differences on the standard normal scale between the response to category  $j$  and the response to category  $j-1$  in all the arms of trial  $i$ . One option is to assume a ‘fixed effect’  $z_{ij} = z_j$  for each of the  $J-1$  categories over all trials  $i$ , or a ‘random effect’ in which the trial-specific terms are drawn from a distribution, but are the same for each arm within a trial, taking care to ensure that the  $z_j$  are increasing with category (i.e. are ordered). Choice of model can be made on the basis of DIC.

Example 6 (Psoriasis) in the Appendix, illustrates fixed and random effects meta-analyses with fixed effects  $z_j$ . Examples of very similar analyses can be found in the health technology assessment literature on psoriasis,<sup>52</sup> psoriatic arthritis<sup>53</sup> and rheumatoid arthritis,<sup>54</sup> although in some cases random effect models were placed on baselines, which is not the practice we recommend. The model, and the WinBUGS coding, are appropriate in cases where different trials use different thresholds, or when different trials report different numbers of thresholds, as is the case in the Psoriasis Example 6. There is, in fact, no particular requirement for trials to even use the same underlying scale, in this case the PASI: this could however require an expansion of the number of categories.

Unless the response probabilities are very extreme the probit model will be undistinguishable from the logit model in terms of model fit or DIC. Choosing which link function to use should therefore be based on the data generating process and on the interpretability of the results.

### 3.7 ADDITIVE AND MULTIPLICATIVE EFFECTS WITH BINOMIAL DATA, AND OTHER NON-CANONICAL LINKS

It was mentioned earlier (Section 2.1) that the appropriate scale of measurement, and thus the appropriate link function, was the one in which effects were linear. It is common to see Log Relative Risks (LRR) and Risk Differences (RD) modelled using the treatment difference approach (Section 3.4), but there are advantages to adopting an arm-based analysis with Binomial likelihoods (see discussion). To perform an arm-based analysis using the RD or LRR requires special programming, because, unlike the “canonical”<sup>12</sup> logit models, there is otherwise nothing to prevent the fitted probabilities in a risk difference or log risk model from being outside the natural zero-to-one range for probabilities. Suitable adjustments to coding have been published for Frequentist software,<sup>55</sup> or more recently for WinBUGS.<sup>56</sup> A Risk Difference model would be:

$$\begin{aligned}\mu_i &\sim \text{Uniform}(0,1) \\ p_{ik} &= \mu_i + \min(\max(\delta_{i,bk}, -\mu_i), (1 - \mu_i))I_{\{k \neq 1\}}\end{aligned}$$

The effect of this construction is to guarantee that *both* the baseline probability  $\mu_i$  and  $\mu_i + \delta_{i,bk}$  remain in the interval (0,1) with  $\delta_{i,bk}$  interpreted as a Risk Difference. For a Relative Risk model:

$$\begin{aligned}\exp(\mu_i) &\sim \text{Uniform}(0,1) \\ \log(p_{ik}) &= \mu_i + \min(\delta_{i,bk}, -\mu_i)I_{\{k \neq 1\}}\end{aligned}$$

Here,  $\delta_{i,bk}$  is a Log Relative Risk. Warn et al.<sup>56</sup> should be consulted for further details of the WinBUGS coding and considerations on prior distributions.

Our experience with these models is that they can sometimes be less stable, and issues of convergence and starting values need especially close attention. One can readily avoid their use, of course, by using estimates of Relative Risk or Risk Difference as data. But this approach runs into difficulties when multi-arm trials are included (see Sections 5.1 and 7).

## 4 SHARED PARAMETER MODELS

Shared parameter models allow the user to generate a single coherent synthesis when trials report results in different formats. For example some trials may report binomial data for each

arm, while others report only the estimated log odds ratios and their variances; or some may report numbers of events and time at risk, while others give binomial data at given follow-up times. In either case the trial-specific relative effects  $\delta_{i,bk}$  represent the shared parameters, which are generated from a common distribution regardless of which format trial  $i$  is reported in.

So if in a meta-analysis of  $M$  trials,  $M_1$  trials report the mean of a continuous outcome for each arm of the trial, and the remaining trials report only the difference in the means of each experimental arm relative to control, a shared parameter model to obtain a single pooled estimate, can be written as a combination of the models presented in Section 3.4 such that

$$y_{ik} \sim N(\theta_{ik}, se_{ik}^2)$$

where

$$\theta_{ik} = \begin{cases} \mu_i + \delta_{i,bk} I_{\{k \neq 1\}} & \text{for } i = 1, \dots, M_1; k = 1, 2, \dots, a_i \\ \delta_{i,bk} & \text{for } i = M_1 + 1, \dots, M; k = 2, \dots, a_i \end{cases}$$

and  $a_i$  represents the number of arms in trial  $i$  ( $a_i=2,3,\dots$ ). The trial-specific treatment effects  $\delta_{i,bk}$  come from a common random effects distribution  $\delta_{i,12} \sim N(d_{12}, \sigma^2)$  as before.

Separate likelihood statements could also be defined, so for example in a meta-analysis with a binomial outcome, the  $M_1$  trials reporting the binomial counts in each trial arm could be combined with the trials reporting only the log-odds ratio of each experimental treatment relative to control and its variance. In this case the binomial data would be modelled as in Section 2.1 and the continuous log-odds ratio data could be modelled as in Section 3.5, with the shared parameter being the trial-specific treatment effects  $\delta_{i,bk}$  as before. For a fixed effect model,  $\delta_{i,12}$  can be replaced by  $d_{12}$  in the model specification.

These models can be easily coded in WinBUGS by having different loops for each of the data types, taking care to index the trial-specific treatment effects appropriately.

Examples of shared parameter models will primarily include cases where some trials report results for each arm, whether proportions, rates, or continuous outcomes, and other trials report only the between-arm differences. A common model for log rates could be shared between trials with Poisson outcomes and time-at-risk and trials with Binomial data with a cloglog link; log rate ratios with identity link and normal approximation sample variance could form a third type of data for a shared log rate model. These models can be used to



combine studies reporting outcomes as mean differences or as binomial data<sup>57</sup> and to combine data on survival endpoints which have been summarised either by using a hazard ratio or as number of events out of the total number of patients.<sup>58</sup> Another possibility would be to combine trials reporting test results at one or more cut-points using a probit link with binomial or multinomial likelihoods, with data on continuous outcomes transformed to a standard normal deviate scale.

To combine trials which report continuous outcome measures on different scales with trials reporting binary outcomes created by dichotomising the underlying continuous scale, authors have suggested converting the odds ratios calculated from the dichotomous response into a SMD,<sup>5,59</sup> or converting both the binary and continuous measures into log-odds ratios for pooling.<sup>60</sup> These methods could be used within a shared parameter model.

Examples 7 and 8 (Parkinson's differences and shared parameter) in the Appendix are shared parameter models.

## 5 EXTENSION TO INDIRECT COMPARISONS AND NETWORK META-ANALYSIS

In Section 2 we defined a set of  $M$  trials over which the study-specific treatment effects of treatment 2 compared to treatment 1,  $\delta_{i,12}$ , were exchangeable with mean  $d_{12}$  and variance  $\sigma_{12}^2$ . We now suppose that, within the same set of trials (i.e. trials which are relevant to the same research question), comparisons of treatments 1 and 3 are also made. To carry out a pairwise random effects meta-analysis of treatment 1 v 3, we would now assume that the study-specific treatment effects of treatment 3 compared to treatment 1,  $\delta_{i,13}$ , are also exchangeable such that  $\delta_{i,13} \sim N(d_{13}, \sigma_{13}^2)$ . If so, it can then be shown that the study-specific treatment effects of treatment 3 compared to 2,  $\delta_{i,23}$ , are also exchangeable:

$$\delta_{i,23} \sim N(d_{23}, \sigma_{23}^2)$$

This follows from the transitivity relation  $\delta_{i,23} = \delta_{i,13} - \delta_{i,12}$ . It can further be shown<sup>61</sup> that this implies

$$d_{23} = d_{13} - d_{12} \tag{12}$$

and

$$\sigma_{23}^2 = \sigma_{12}^2 + \sigma_{13}^2 - 2\rho_{23}^{(1)}\sigma_{12}\sigma_{13}$$

where  $\rho_{23}^{(1)}$  represents the correlation between the relative effects of treatment 3 compared to treatment 1, and the relative effect of treatment 2 compared to treatment 1 within a trial (see Lu & Ades<sup>61</sup>). For simplicity we will assume equal variances in all subsequent methods, i.e.  $\sigma_{12}^2 = \sigma_{13}^2 = \sigma_{23}^2 = \sigma^2$ , and this implies that the correlation between any two treatment contrasts in a multi-arm trial is 0.5.<sup>19</sup> For heterogeneous variance models see Lu & Ades.<sup>61</sup>

The exchangeability assumptions regarding the treatment effects  $\delta_{i,12}$  and  $\delta_{i,13}$  therefore make it possible to derive indirect comparisons of treatment 3 vs treatment 2, from trials of treatment 1 vs 2 and 1 vs 3, and also allow us to include trials of treatments 2 vs 3 in a coherent synthesis with the 1 vs 2 and 1 vs 3 trials.

Note the relationship between the standard assumptions of pair-wise meta-analysis, and those required for indirect and mixed treatment comparisons. For a random effects pair-wise meta-analysis, we need to assume exchangeability of the effects  $\delta_{i,12}$  over the 1 vs 2 trials, and also exchangeability of the effects  $\delta_{i,13}$  over the 1 vs 3 trials. For network meta-analysis, we must assume the exchangeability of both treatment effects over both 1 vs 2 *and* 1 vs 3 trials. The theory extends readily to additional treatments  $k = 4, 5, \dots, S$ . In each case we must assume the exchangeability of the  $\delta$ 's across the entire set of trials. Then the within-trial transitivity relation is enough to imply the exchangeability of all the treatment effects  $\delta_{i,xy}$ . The *consistency equations*<sup>21</sup>

$$\begin{aligned} d_{23} &= d_{13} - d_{12} \\ d_{24} &= d_{14} - d_{12} \\ &\vdots \\ d_{(s-1),s} &= d_{1s} - d_{1,(s-1)} \end{aligned}$$

are also therefore implied; they are assumptions required by indirect comparisons and MTC, but, given that we are assuming that all trials are relevant to the same research question, they are not *additional* assumptions. However, whilst in theory, consistency of the treatment effects must hold, there may be inconsistency in the evidence. Methods to assess evidence consistency are addressed in TSD4.<sup>28</sup> The consistency equations can also be seen as an example of the distinction between the  $(s-1)$  *basic* parameters<sup>62</sup>  $d_{12}, d_{13}, d_{14}, \dots, d_{1s}$  on which

prior distributions are placed, and the *functional* parameters which are functions of the basic parameters, and represent the remaining contrasts. It is precisely the reduction in the number of dimensions, from the number of functions on which there are data to the number of basic parameters, that allows all data, whether directly informing basic or functional parameters, to be combined within a coherent (internally consistent) model.

Now that several treatments are being compared, we clarify our notation a little, and have the trial-specific treatment effects of the treatment in arm  $k$ , relative to the control treatment (in arm 1), drawn from a common random effects distribution:

$$\delta_{i,1k} \sim N(d_{t_{i1},t_{ik}}, \sigma^2)$$

where  $d_{t_{i1},t_{ik}}$  represents the mean effect of the treatment in arm  $k$  in trial  $i$ ,  $t_{ik}$ , compared to the treatment in arm 1 of trial  $i$ ,  $t_{i1}$ , and  $\sigma^2$  represents the between-trial variability in treatment effects (heterogeneity). For trials that compare treatments 1 and 2  $d_{t_{i1},t_{ik}} = d_{12}$ , for trials that compare treatments 2 and 3  $d_{t_{i1},t_{ik}} = d_{23}$  and so on. The pooled treatment effect of treatment 3 compared to treatment 2,  $d_{23}$ , is then obtained from equation (12).

The WinBUGS code provided in the Appendix will extend to MTC. Examples 3 to 8 in the Appendix illustrate analyses with multiple treatments.

## 5.1 INCORPORATING MULTI-ARM TRIALS

Suppose we have a number of multi-arm trials involving the treatments of interest, 1,2,3,4,... Among commonly suggested stratagems are combining all active arms into one, or splitting the control group between all relevant experimental groups, or ignoring all but two of the trial arms.<sup>5</sup> None of these are satisfactory. The question of how to conduct a meta-analysis of the multi-arms trials has been considered in a Bayesian framework by Lu & Ades,<sup>20</sup> and in a frequentist framework by Lumley<sup>22</sup> and Chootrakool & Shi.<sup>63</sup>

Based on the same exchangeability assumptions above, a single multi-arm trial will estimate a vector of random effects  $\delta_i$ . For example a three-arm trial will produce two random effects and a four-arm trial three. Assuming, as before, that the relative effects all have the same between-trial variance we have

$$\boldsymbol{\delta}_i = \begin{pmatrix} \delta_{i,12} \\ \vdots \\ \delta_{i,1a_i} \end{pmatrix} \sim N_{a_i-1} \left( \begin{pmatrix} d_{1,t_{i2}} \\ \vdots \\ d_{1,t_{ia_i}} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \dots & \sigma^2/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \dots & \sigma^2 \end{pmatrix} \right) \quad (13)$$

where  $\boldsymbol{\delta}_i$  is the vector of random effects, which follows a multivariate normal distribution,  $a_i$  represents the number of arms in trial  $i$  ( $a_i=2,3,\dots$ ) and  $d_{1,t_{ik}} = d_{1,t_k} - d_{1,t_1}$ . Then the conditional univariate distributions for the random effect of arm  $k>2$ , given all arms from 2 to  $k-1$ , is (see eg. Raiffa & Schlaiffer<sup>64</sup>)

$$\delta_{i,1k} \mid \begin{pmatrix} \delta_{i,12} \\ \vdots \\ \delta_{i,1(k-1)} \end{pmatrix} \sim N \left( \left( d_{1,t_k} - d_{1,t_1} \right) + \frac{1}{k-1} \sum_{j=1}^{k-1} \left[ \delta_{i,1j} - \left( d_{1,t_j} - d_{1,t_1} \right) \right], \frac{k}{2(k-1)} \sigma^2 \right) \quad (14)$$

Either the multivariate distribution in equation (13) or the conditional distributions in equation (14) must be used to estimate the random effects for each multi-arm study so that the between-arm correlations between parameters are taken into account. The code presented in the Appendix uses the formulation in equation (14) as it allows for a more generic code which works for trials with any number of arms.

This formulation provides another interpretation of the exchangeability assumptions made in the previous section, and indeed another way of deducing the consistency relations. This is that we may consider a connected network of  $M$  trials involving  $S$  treatments to originate from  $M$   $S$ -arm trials, but that some of the arms are missing at random (MAR). (Note that MAR does not mean that the choice of arms is random, but that the missingness of arms is unrelated to the efficacy of the treatment). It should be noted that the general formulation is no different from the model presented by Higgins & Whitehead.<sup>19</sup>

The WinBUGS code provided in the Appendix is based on a conditional distribution formulation of the multivariate normal distribution. It therefore exactly instantiates the theory behind network meta-analysis that relates it to pair-wise meta-analysis. The code in the Appendix will analyse pair-wise meta-analysis, indirect comparisons, network meta-analysis (MTC) and multi-arm trials without distinction.

### 5.1.1 Multi-arm trials with treatment differences (trial-based summaries)

As mentioned in Section 3.5, when results from multi-arm trials are presented as (continuous) treatment differences relative to the control arm (arm 1), a correlation between the treatment

differences is induced, since all differences are taken relative to the same control arm. Unlike the correlations between the relative effect parameters, this correlation is inherent in the *data*, and so requires an adjustment to the likelihood. A trial with  $a_i$  arms produces  $a_i-1$  treatment differences which are correlated. The covariance between differences taken with respect to the same control arm is equal to the observed variance for the common control arm. So for example in a three arm trial comparing treatments A, B and C, letting  $y_{AB}$  and  $y_{AC}$  represent the treatment differences of treatments B and C relative to treatment A, we know that

$$Var(y_{AB} - y_{AC}) = Var(y_{AB}) + Var(y_{AC}) - 2Cov(y_{AB}, y_{AC}) \quad (15)$$

and,

$$\begin{aligned} Var(y_{AB}) &= Var(y_A) + Var(y_B) \\ Var(y_{AC}) &= Var(y_A) + Var(y_C) \end{aligned} \quad (16)$$

with,  $y_A$ ,  $y_B$  and  $y_C$  representing the original measurements on each arm of the trial, because, in a randomised controlled trial the measurements in each trial arm are independent. By successive replacement of the expressions in equation (16) into equation (15), we have:

$$Cov(y_{AB}, y_{AC}) = Var(y_A)$$

So, the likelihood for a trial  $i$  with  $a_i$  arms would be defined as multivariate normal

$$\begin{pmatrix} y_{i2} \\ y_{i3} \\ \vdots \\ y_{i,a_i} \end{pmatrix} \sim N_{a_i-1} \left( \begin{pmatrix} \theta_{i2} \\ \theta_{i3} \\ \vdots \\ \theta_{i,a_i} \end{pmatrix}, \begin{bmatrix} V_{i2} & se_{i1}^2 & \cdots & se_{i1}^2 \\ se_{i1}^2 & V_{i3} & \cdots & se_{i1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ se_{i1}^2 & se_{i1}^2 & \cdots & V_{i,a_i} \end{bmatrix} \right)$$

where the diagonal elements in the variance-covariance matrix represent the variances of the treatment differences and the off-diagonal elements represent the observed variance in the control arm in trial  $i$ , denoted by  $se_{i1}^2$  (see Section 3.4). For example, when the treatment differences are given as log-odds ratios,  $se_{i1}^2$  is the variance of the log-odds for arm 1 of trial  $i$ . Example 7 (Parkinson's Difference) in the Appendix includes a three-arm trial in a meta-analysis of treatment differences where the variance of the common arm,  $se_{i1}^2$ , is known. When  $se_{i1}^2$  is not reported, an approximation should be made, perhaps based on the variances

of the differences.<sup>58</sup> If the value of the control variance is available only for some of the included trials, that information can be used to estimate the parameters of a distribution for the control variance (assumed to be common). This estimated distribution can then be used to predict the variance of the control arm, in the trials where it is missing. This method has been used, in a slightly different context, to predict missing variances.<sup>65</sup> Riley<sup>66</sup> provides a review of methods to impute unknown within-study correlations within the context of multivariate meta-analysis. These methods can also be applied to network meta-analysis with multi-arm trials.

## 6 TECHNICAL ISSUES IN BAYESIAN MCMC

The use of the WinBUGS Bayesian MCMC software has advantages (Section 7), but it also requires some care. Users are strongly advised to acquire a good understanding of Bayesian theory,<sup>67</sup> and to follow advice given in the WinBUGS manual. Particular care must be taken in checking convergence, and we suggest that at least three chains are run, starting from widely different (yet sensible) initial values. The diagnostics recommended in the literature should be used to check convergence.<sup>68,69</sup> Users should *also* ensure that, after convergence, each chain is sampling from the same posterior. Posteriors should be examined visually for spikes and unwanted peculiarities, and both the initial “burn-in” and the posterior samples should be conservatively large. The number of iterations for both must be reported in the analysis. An often cited guideline suggests that the Monte Carlo error, which reflects both the number of simulations and the degree of autocorrelation, should be no more than 5% of the posterior standard deviation of the parameters of interest.

Beyond these warnings, which apply to all Bayesian MCMC analyses, evidence synthesis models have particular properties which may require careful examination: choice of reference treatment, choice of prior distributions and zero counts in binomial and Poisson data.

### 6.1 CHOICE OF REFERENCE TREATMENT

While the likelihood is not altered by a change in which treatment is taken to be “Treatment 1”, the choice of the reference treatment can affect the posterior estimates because priors cannot be totally non-informative. However, for the vague priors we suggest throughout for  $\mu_i$  and  $d_{1k}$  (see below) we expect the effect to be negligible. Choice should therefore be based on ease of interpretation, with placebo or standard treatment usually taken as Treatment 1. In

larger networks, it is preferable to choose as Treatment 1 a treatment that is in the “centre” of the network. In other words, choose the treatment that has been trialled against the highest number of other treatments. The purpose of this is to reduce strong correlations that may otherwise be induced between mean treatment effects for each pair of treatments  $k$  and  $h$ ,  $d_{1k}$  and  $d_{1h}$ : these can slow convergence and make for inefficient sampling from the posterior.

## 6.2 CHOICE OF PRIORS

We recommend vague or flat priors, such as  $N(0, 100^2)$ , throughout for  $\mu_i$  and  $d_{1k}$ . Informative priors for relative effect measures would require special justification.

It has become standard practice to also set vague priors for the between-trial variances. For binomial with logit links models, the usual practice is to place a Uniform prior on the standard deviation, for example  $\sigma \sim \text{Uniform}(0,2)$ . The upper limit of 2 represents a huge range of trial-specific treatment effects. For example if the median treatment effect was an odds ratio of 1.5, then we would expect 95% of trials to have *true* odds ratios between 0.2 and 11. For rate models, whether with log or cloglog linking functions, uniform priors on  $\sigma$  may also be used, but investigators need to be aware of the scale: a prior that is vague for a rate per year may not be so vague for a rate per month. Similarly, for continuous outcomes close attention to the scale of measurement is essential. For trials with blood pressure as the outcome  $\sigma \sim \text{Uniform}(0,100)$  may be considered vague. The posterior distribution of  $\sigma$  should always be inspected to ensure that it is sufficiently different from the prior as this would otherwise indicate that the prior is dominating the data and no posterior updating has taken place.

An alternative approach, which was once popular but has since fallen out of favour, is to set a vague Gamma prior on the precision, for example  $1/\sigma^2 \sim \text{Gamma}(.001,.001)$ . This approach gives a low prior weight to unfeasibly large  $\sigma$  on the logit scale. The disadvantage is that this puts more weight on values of  $\sigma$  near zero. On the other hand, there are occasions where it may be an advantage that this prior rules out values of  $\sigma$  at zero, because it is not uncommon, particularly when data is sparse, that MCMC sampling can “get stuck” at  $\sigma = 0$ , leading to spikes in the posterior distribution of both  $\sigma$  and the treatment effect parameters  $d_{1k}$ . In these cases a Gamma prior may improve numerical stability and speed convergence.

However they are formulated, there are major disadvantages in routinely using vague priors, although this has become a widely accepted practice. In the absence of large numbers of large

trials, the posterior distribution of  $\sigma$  will be poorly identified and likely to include values that, on reflection, are implausibly high or possibly implausibly low. Two further alternatives may be found useful when there is insufficient data to adequately estimate the between-trials variation. The first is the use of external data.<sup>19</sup> If there is insufficient data in the meta-analysis, it may be reasonable to use an estimate for  $\sigma$  from a larger meta-analysis on the same trial outcome involving a similar treatment for the same condition. The posterior distribution, or a posterior predictive distribution, from such an analysis could be used to approximate an informative prior. The derivation of suitable priors from large numbers of meta-analyses is currently being actively researched.

If there is no data on similar treatments and outcomes that can be used, an informative prior can be elicited from a clinician who knows the field. This can be done by posing the question in this way. “Suppose we accept that different trials, even if infinitely large, can produce different effect sizes. If the average effect was an odds ratio of 1.8 (choose a plausible average), what do you think an extremely high and an extremely low effect would be, in a very large trial?” Based on the answer to this it should be possible, by trial and error, to construct an informative Gamma prior for  $1/\sigma^2$ , or a Normal prior for  $\sigma$ , subject to  $\sigma > 0$ . For further discussion of priors for variance parameters see Lambert et al.<sup>70</sup> and Spiegelhalter et al.<sup>67</sup>

### 6.3 ZERO CELLS

Because Binomial and Poisson likelihoods with zero cells are allowed, special precautions do not usually need to be taken in the case of the occasional trial with a zero cell count. This is a major strength of the Bayesian MCMC approach, because some popular frequentist approaches for log odds ratios or log relative risks have to add an arbitrary constant, usually 0.5, to cells in order to obtain non-infinite estimates of treatment effects and non-infinite variance, but in so doing they generate biased estimates of effect size.<sup>71,72</sup>

However, in extreme cases where several trials have zero cells and many of the trials are small, the models we have recommended can be numerically unstable, either failing to converge, or converging to a posterior with very high standard deviation on some of the treatment effects. This is unlikely to happen with fixed effect models, and it can often be remedied in random effects models by using a (more) informative prior on the variance



parameter. A last resort, recognising the assumptions being made, is to put a random effect model on the treatment baselines  $\mu_i$  as well as the relative treatment effects  $d_{lk}$ .

A specific problem arises in sparse networks, in which for example there is only one trial making the comparison X vs Y, and treatment Y only appears in this one trial. If the trial contains a zero cell, it may not be possible to estimate a treatment effect. One solution is to revert to the practice of adding 1 to the denominator and 0.5 to the numerator, or the 0.5 can be replaced with a fraction that is closer to the expected treatment effect to reduce bias.<sup>72</sup> The problem can also be solved by placing a distribution on the baseline model. Trials with zero cells in *both* arms do not contribute evidence in the treatment effect and can be excluded, unless a model has been assumed for the baselines (see TSD5<sup>29</sup>).

Readers should be aware that, when evaluating model fit using the residual deviance, this will always appear large (i.e. >1) for individual data points where there are zero cells. This is because none of the models presented can actually predict a zero cell since probabilities at zero or one are ruled out. Also no leverage can be calculated for these points.

## **7 NON-BAYESIAN APPROACHES AND COMPUTATIONAL ISSUES**

In this section we briefly review the advantages of the Bayesian MCMC approach to synthesis for probabilistic decision modelling. Bayesian methods are clearly convenient in this context, but frequentist analyses can be used to approximate Bayesian posterior sampling, as we describe below. We then provide some pointers to the literature that examines the statistical properties of alternative Bayesian and frequentist meta-analytic estimators. Finally, we describe frequentist software that can be used for evidence synthesis. The statistical reliability of the different synthesis methods is, of course, an entirely separate issue from the accuracy of different computational approaches to implementing the synthesis method.

### **7.1 BAYESIAN VERSUS FREQUENTIST APPROACHES IN THE CONTEXT OF DECISION MAKING**

One of the advantages of the Bayesian MCMC approach is that sampling from the posterior distribution fulfils at the same time the need for posterior inference and the MC approach to probabilistic modelling. Although inference in the form of significance tests and interval

estimation tends to be set aside when a decision maker follows the rules of cost effectiveness analysis (CEA),<sup>73</sup> conclusions from CEA may be highly sensitive to the model that is assumed from the data. Statistical model critique and model choice are, therefore critical to CEA, whether conducted in a Bayesian or frequentist framework.

Posterior sampling in addition retains the correlation between parameters that is induced by their joint estimation from the same data. For this reason, when there are closed loops in the evidence structure, it is essential to either use the posterior samples from WinBUGS in the decision model or to take steps to propagate the correlations through the model (see below). Distributions based on the posterior marginal summaries are not adequate.

Perfectly valid evidence synthesis is also, of course, produced by frequentist software (see below), and the question then arises of how the results from such analyses can be used in a probabilistic decision modelling context. For pair-wise meta-analysis and indirect comparisons, as long as they are restricted to two-arm trials, and do not involve covariates on baselines or treatment effects, it is simple to use the maximum likelihood (ML) estimates and their standard errors to create parameter distributions for forward MC sampling. Where multi-arm trials or MTC evidence structures (in other words where there are loops in the evidence structure), or baseline models, or covariates are involved, two approaches are possible. One is the bootstrap,<sup>74</sup> which requires re-sampling from the data, and the other is to use the ML parameter estimates of all the parameters and their covariance matrix to form a multivariate normal distribution from which to carry out MC sampling. The latter is technically easier, and the bootstrap runs into difficulties when there are zero cells. This multivariate distribution obtained should be *approximately* the same as a Bayesian posterior distribution.

## **7.2. COMPARISON OF META-ANALYTIC METHODS**

While there is no *technical* reason why frequentist methods cannot be used, there are a wide variety of estimators to choose between for count data, which in specific circumstances can produce different results. There is a useful literature comparing meta-analytic estimators for binomial data, based on simulation studies. Bradburn et al<sup>71</sup> discuss the biases arising from the common practice of adding 0.5, or other amounts, to zero cells, and it is an undoubted advantage of methods using exact binomial and Poisson likelihoods, like Bayesian MCMC, that these problems are very largely avoided (though see Section 6.3). Simulation studies on

fixed effect estimators<sup>72</sup> have shown that Bayesian MCMC has performed well, and ranks with Mantel Haenszel<sup>75</sup> and Exact method<sup>76</sup> estimators, and is superior to the Peto method<sup>77</sup> and inverse-variance weighting in a wide range of situations. There is a useful simulation study of tests of heterogeneity for binomial data.<sup>78</sup>

There has, however, been little work comparing Bayesian and non-Bayesian approaches in the context of random effect models, nor on our proposal that a choice can be made between random and fixed effect models on the basis of the DIC statistics.

It should be emphasised that except for the Exact method (which only applies to fixed effects models), *none* of the frequentist methods use the Binomial, Multinomial or Poisson likelihoods, but instead rely on normal approximations. This should not create problems as long as low cells counts are rare.

Furthermore, in random effects models, the MCMC implementation automatically takes into account the uncertainty in the between-study heterogeneity parameter  $\sigma^2$ . Whilst this is also possible using frequentist approaches,<sup>79,80</sup> it is rarely done in practice, possibly because such models are not currently implemented in user-friendly frequentist software.

### **7.3. COMPARISON OF EVIDENCE SYNTHESIS SOFTWARE**

In this section we provide a brief review of frequentist software for evidence synthesis, and where possible a comparison with the Bayesian MCMC WinBUGS software. We will refer to existing literature where possible. Alternative software for pair-wise meta-analysis has been developed over the past two decades, both as stand-alone applications and as ‘macro’ routines for pre-existing packages. Probably the most comprehensive and up-to-date routines for existing software are those developed and freely distributed for STATA<sup>81</sup> and R.<sup>82</sup> Stand alone packages include Comprehensive Meta-Analysis,<sup>83</sup> Meta-Analyst<sup>84</sup> and MIX<sup>85</sup> as well as the Cochrane Collaboration software RevMan.<sup>86</sup> If pair-wise meta-analysis is all that is required, and appropriate estimators are used, this software is perfectly adequate to produce estimates and corresponding standard errors to inform a distribution in a probabilistic decision model.

These approaches can also be used to carry out two pairwise meta-analyses from which an “indirect” comparison can be formed. This is effectively an implementation of the Bucher method for indirect comparisons<sup>87</sup> (and a simple example of an MTC network). Note that in a random effects context, this approach allows for separate and unrelated meta-analyses for AB

and AC whereas all models considered in this guide (and software examples in the Appendix) have considered a common random effect. (Note that this could also be achieved in any simulation package, including WinBUGS, by setting up distributions for the AB and AC effects, and then subtracting one from the other to form a distribution for the BC effect).

It would appear no stand-alone software to fit MTC models to arbitrarily complex networks exists. Perhaps the most general code for frequentist software is that developed by Lumley.<sup>22</sup> However, this fits a different model from the one described in this paper which does not assume consistency of direct and indirect evidence and will be discussed in TSD4.<sup>28</sup>

It is also important to note that most frequentist software for MTC is based on data on trial-level differences, and not data aggregated at arm-level. As we saw in Section 5.1, when multi-arm trials are involved, it is essential to take account of both the correlation in parameters in random effect models, but also the correlations in the likelihood, which affect both fixed and random effect models. To the best of our knowledge, there is only one frequentist software module available for STATA, `mvmeta`, that takes account of the correlations at the data level.<sup>81,88</sup> Frequentist analysis can also be correctly carried out in SAS.<sup>89</sup> In our experience, failure to make the adjustment can give materially different results whenever multi-arm trials form a substantial proportion of the evidence for any treatment contrast.

## 8. FURTHER READING

This brief introduction raises many further questions. Some of these are taken up in detail in the other technical guides in this series. TSD3<sup>35</sup> for example covers issues of heterogeneity, looking at meta-regression and treatment effects in subgroups. Individual patient data has a particularly important role in meta-regression. Regression and other methods for bias adjustment, including publication bias and so-called “small-study bias” will also be covered. Similarly, the introduction of network meta-analysis (Section 5) raises questions about inconsistency between “direct” and “indirect” evidence, which will be addressed in TSD4.<sup>28</sup> Here we will restrict attention to further extensions involving multiple outcomes.

There are a huge variety of multiple outcome structures, only a minority of which have been examined in the context of evidence synthesis. One particularly important area for multiple outcome synthesis is where a set of eligible trials have been identified, but some report one outcome, some another, and perhaps others report both. The function of multiple outcome

synthesis in this case is to use all the available data, taking into account the correlation between outcomes at the within-trial (between-patient) level, and the between-trial level. Methods for doing this have been described by Riley and others.<sup>90,91</sup> Readers may also refer to earlier work also based on multivariate normal methods.<sup>92,93</sup>

A quite different form of multiple outcome synthesis occurs when different trials report a single outcome, but at different follow-up times. Further, some may report at several follow-up times. Lu et al<sup>42</sup> describe a series of piece-wise-constant hazard models that extend the cloglog models of Section 3.2, in that treatment effects are allowed to vary between time intervals.

Multiple outcomes are often structurally related. For example, time-to-progression cannot post-date overall survival in cancer studies, and time to end of ‘flu cannot post-date time to the end of symptoms in studies of influenza treatments. Where possible, these structural constraints should be built into the synthesis. Burch et al<sup>94</sup> and Welton et al<sup>95,96</sup> provide examples. A further form of structural relation can be seen in trials on treatments for myocardial infarction, where trial evidence on an early intermediary outcome, coronary patency, can be combined with trial evidence on mortality.<sup>13,97</sup>

Multiple reporting formats present further challenges for synthesis. Many of these are covered in the earlier section on shared parameter models (Section 4). A more complex example<sup>94,95</sup> involved a synthesis that combined trials reporting median time to an event, mean time to an event, and proportion experiencing the event before a certain time. Results may also be reported separately for different subgroups, or collapsed over subgroups. Here, too, methods exist that make it possible to combine all the available information.<sup>98-101</sup> Another example of multiple reporting formats would be trials reporting on binary, categorical or continuous scales.<sup>57</sup>

A final special topic is synthesis for Markov models. It seems clear that, for purposes of synthesis, it may be preferable to express Markov models in terms of transition rates rather than transition probabilities.<sup>102</sup> This facilitates combining information from studies run over different follow-up periods, as explained in Section 3.2. It also gives the flexibility to model different treatment effects on different transitions (e.g. competing risks, Section 3.3), which is difficult to achieve with logit models. A further option is to combine data from studies where certain transitions are unobserved.<sup>102</sup> Finally, it is worth mentioning that it is often open to question whereabouts in a Markov model a treatment effect is operating. Readers are referred

to recent work<sup>103</sup> showing how treatment effects can be parameterised, synthesised and estimated in Markov rate models.

## 9. DISCUSSION

Our objective in this document has been to present a single unified account of evidence synthesis of aggregate data from RCTs, specifically but not exclusively for use in probabilistic decision making. In order to cover the variety of outcomes that are reported, and the range of data transformations required to obtain approximate linearity, we have set this within the familiar framework of Generalised Linear Models. This leads to a modular approach: different likelihoods and link functions may be employed, but the “synthesis” operation, which occurs at the level of the linear predictor, takes the exact same form in every case. The linear predictor, furthermore, is a regression model of a very specific type, with  $K-1$  treatment effect parameters for any  $K$  treatment network, offering a single model for pair-wise meta-analysis, indirect comparisons, network meta-analysis (mixed treatment comparisons), and synthesis of multi-arm trials in any combination. This has all been presented in a Bayesian MCMC context and supported by code for WinBUGS 1.4.3.<sup>26</sup> The use of WinBUGS dovetails with the GLM approach, as it allows us to take full advantage of the modularity implied by GLMs.

This document brings together a great deal of previous work on Bayesian meta-analysis and MTC methods, and on meta-analysis methods for particular outcomes, into a single accessible and unified treatment and with a modular approach to computation. One aspect that may be considered novel is the application to synthesis of multi-arm trials. Although it would be relatively simple to extend all the commonly used estimators for use in meta-analysis of multiple trials on the same  $K$  comparators, this does not appear to have been done. In our framework, however, the transitivity and exchangeability assumptions required for MTC automatically deliver multiple-treatment meta-analysis, and vice versa.

In the course of the document we have raised the question: what is the relation between the arm-based analyses assumed in our models and software, with the far more common trial-based summaries where the difference between treatments is taken as the data. With count data, there has always been a strong case for adopting an arm-based approach because it avoids normal approximations for the likelihood, and difficulties with zero cells. We have, however, brought to light a further problem with the arm-based approach, which is that with

multi-arm trials a correction must be made to the likelihood to account for correlations in the data. This applies to continuous outcome data as well as count data, and to both fixed and random effect models. With the appropriate correction to the likelihood, the results from arm-based and treatment difference-based analyses are identical. Since all frequentist software for synthesis currently adopts the trial-level treatment difference approach, we can give a quite specific recommendation that this software should be avoided for multi-treatment syntheses with multi-arm trials, unless the correlation between relative treatment effects from the same trial can be appropriately accounted for. For syntheses where all trials are two-arm, there is no reason why frequentist methods should not be used, as long as statistically sound estimators are used and appropriate steps are taken to propagate parameter uncertainty, including correlations, through the decision model.

Bayesian analysis is by no means a panacea: one area that clearly deserves more work is how to specify a “vague” prior distribution for the variance parameter. There can be little doubt that the vague priors that are generally recommended produce posteriors that are biased upwards. The extent of the bias is likely to be greater when the true variance is low, and when there is little data: either few trials or small trials, but this is also a problem when using frequentist estimators. The question could perhaps be resolved through a comprehensive simulation exercise. Although we can be reassured that the bias tends to be conservative, ultimately it may be preferable to use informative priors, perhaps tailored to particular outcomes and disease areas, based on studies of many hundreds of meta-analyses. This is currently an active research area.

## 10. REFERENCES

1. Zelen, M. The analysis of several 2 x 2 contingency tables. *Biometrika* 1971; 58(1):129-137.
2. DerSimonian, R., Laird, N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; 7(3):177-188.
3. Cooper, H., Hedges, L. The handbook of research synthesis. Russell Sage Foundation, New York, 1994.
4. Egger, M., Davey Smith, G., Altman, D.G. Systematic reviews in health care: meta-analysis in context. Second ed. BMJ publishing group, London; 2001.
5. Higgins, J.P.T., Green, S.E. Cochrane handbook for systematic reviews of interventions version 5.0.0 [updated February 2008]. *The Cochrane Collaboration* 2008.
6. Smith, T.C., Spiegelhalter, D.J., Thomas, A. Bayesian approaches to random effects meta analysis: A comparative study. *Statistics in Medicine* 1995; 14(24):2685-2699.
7. Sutton, A.J., Abrams, K.R. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; 10(4):277-303.
8. Sutton, A.J., Higgins, J. Recent developments in meta analysis. *Statistics in Medicine* 2008; 27(5):625-650.
9. Whitehead, A. Meta-analysis of controlled clinical trials. Wiley, Chichester; 2002.
10. Whitehead, A., Whitehead, J. A general parametric approach to the meta analysis of randomized clinical trials. *Statistics in Medicine* 1991; 10(11):1665-1677.
11. National Institute for health and Clinical Excellence. Guide to the methods of technology appraisal (updated June 2008). 2008. NICE, London.
12. McCullagh, P., Nelder, J.A. Generalized linear models. Chapman & Hall/CRC, 1989.
13. Ades, A.E. A chain of evidence with mixed comparisons: models for multi parameter synthesis and consistency of evidence. *Statistics in Medicine* 2003; 22(19):2995-3016.
14. Ades, A.E., Sculpher, M., Sutton, A.J., Abrams, K.R., Cooper, N., Welton, N.J. et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics* 2006; 24(1):1-19.
15. Caldwell, D.M., Ades, A.E., Higgins, J.P.T. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *Bmj* 2005; 331(7521):897-900.
16. Gleser, L.J., Olkin, I. Stochastically dependent effect sizes. In: Cooper H., Hedges L., eds. *The handbook of research synthesis*. Russell Sage Foundation, New York; 1994; 339-355.



17. Gleser, L.J., Olkin, I. Meta-analysis for 2 x 2 Tables with Multiple Treatment Groups. In: Stangl D.K., Berry D.A., eds. *Meta-analysis in medicine and health policy*. Marcel Dekker, New York; 2000.
18. Hasselblad, V. Meta-analysis of multi-treatment studies. *Medical Decision Making* 1998; 18:37-43.
19. Higgins, J.P.T., Whitehead, A. Borrowing strength from external trials in a meta analysis. *Statistics in Medicine* 1996; 15(24):2733-2749.
20. Lu, G., Ades, A.E. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* 2004; 23(20):3105-3124.
21. Lu, G., Ades, A.E. Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association* 2006; 101(474):447-459.
22. Lumley, T. Network meta analysis for indirect treatment comparisons. *Statistics in Medicine* 2002; 21(16):2313-2324.
23. Salanti, G., Higgins, J.P.T., Ades, A.E., Ioannidis, J.P.A. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 2008; 17(3):279-301.
24. 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 Assessment* 2005; 9(26):1-149.
25. Parmigiani, G. Modeling in medical decision making: a Bayesian approach. Wiley, Chichester; 2002.
26. Spiegelhalter, D., Thomas, A., Best, N., Lunn, D. WinBUGS user manual version 1.4. 2003; available from <http://www.mrc-bsu.cam.ac.uk/bugs/> (accessed 2007).
27. 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 Series B (Statistical Methodology)* 2002; 64(4):583-639.
28. 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.nicesu.org.uk>
29. 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.nicesu.org.uk>
30. Bernardo, J.M., Smith, A.F.M. Bayesian theory. Fourth ed. John Wiley and Sons, New York; 1994.
31. Carlin, J.B. Meta-analysis for 2 x 2 tables: a Bayesian approach. *Statistics in Medicine* 1992; 11(2):141-158.

32. Van Houwelingen, H.C., Zwinderman, K.H., Stijnen, T. A bivariate approach to meta analysis. *Statistics in Medicine* 1993; 12(24):2273-2284.
33. Van Houwelingen, H.C., Arends, L.R., Stijnen, T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; 21:589-624.
34. Dempster, A.P. The direct use of likelihood for significance testing. *Statistics and Computing* 1997; 7(4):247-252.
35. 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>
36. Deeks, J.J. Issues in the selection of a summary statistic for meta analysis of clinical trials with binary outcomes. *Statistics in Medicine* 2002; 21(11):1575-1600.
37. Cooper, N.J., Sutton, A.J., Lu, G., Khunti, K. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. *Archives of Internal Medicine* 2006; 166(12):1269-1275.
38. Dias, S., Welton, N.J., Marinho, V.C.C., Salanti, G., Higgins, J.P.T., Ades, A.E. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. *Journal of the Royal Statistical Society: Series A(Statistics in Society)* 2010; 173(3):613-629.
39. Guevara, J.P., Berlin, J.A., Wolf, F.M. Meta-analytic methods for pooling rates when follow-up duration varies: a case study. *BMC Medical Research Methodology* 2004; 4(1):17.
40. Collett, D. Modelling survival data in medical research. Chapman & Hall, London; 1994.
41. Prentice, R.L., Gloeckler, L.A. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 1978;57-67.
42. Lu, G., Ades, A.E., Sutton, A.J., Cooper, N.J., Briggs, A.H., Caldwell, D.M. Meta-analysis of mixed treatment comparisons at multiple follow up times. *Statistics in Medicine* 2007; 26(20):3681-3699.
43. Stettler, C., Wandel, S., Allemann, S., Kastrati, A., Morice, M.C., Schomig, A. et al. Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. *The Lancet* 2007; 370(9591):937-948.
44. NICE Collaborating Centre for Mental Health. Schizophrenia: core interventions in the treatment and management of schizophrenia in adults in primary and secondary care (update). 2010. NICE, London.
45. Ades, A.E., Mavranouzouli, I., Dias, S., Welton, N.J., Whittington, C., Kendall, T. Network Meta Analysis with Competing Risk Outcomes. *Value in Health* 2010; 13:976-983.

46. Frison, L., Pocock, S.J. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Statistics in Medicine* 1992; 11(13):1685-1704.
47. Follmann, D., Elliott, P., Suh, I., Cutler, J. Variance imputation for overviews of clinical trials with continuous response. *Journal of Clinical Epidemiology* 1992; 45(7):769-773.
48. Abrams, K.R., Gillies, C.L., Lambert, P.C. Meta analysis of heterogeneously reported trials assessing change from baseline. *Statistics in Medicine* 2005; 24(24):3823-3844.
49. Cohen, J. Statistical power analysis for the behavioral sciences. Academic Press, New York, 1969.
50. Hedges, L.V., Olkin, I. Statistical methods for meta-analysis. Academic Press New York: 1985.
51. Salanti, G., Marinho, V., Higgins, J.P.T. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *Journal of Clinical Epidemiology* 2009; 62(8):857-864.
52. Woolacott, N., Hawkins, N., Mason, A., Kainth, A., Khadjesari, Z., Vergel, Y.B. et al. Etanercept and efalizumab for the treatment of psoriasis: a systematic review. *Health Technology Assessment* 2006; 10(46):1-252.
53. Woolacott, N., Bravo Vergel, Y., Hawkins, N., Kainth, A., Khadjesari, Z., Misso, K. et al. Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health Technology Assessment* 2006; 10(31):1-258.
54. Jobanputra, P., Barton, P., Bryan, S., Burls, A. The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation. *Health Technology Assessment* 2002; 6(21):1-110.
55. WACHOLDER, S. Binomial regression in GLIM: estimating risk ratios and risk differences. *American Journal of Epidemiology* 1986; 123(1):174-184.
56. Warn, D.E., THOMPSON, S.G., Spiegelhalter, D.J. Bayesian random effects meta-analysis of trials with binary outcomes: methods for the absolute risk difference and relative risk scales. *Statistics in Medicine* 2002; 21(11):1601-1623.
57. Dominici, F., Parmigiani, G., Wolpert, R.L., Hasselblad, V. Meta-Analysis of Migraine Headache Treatments: Combining Information from Heterogeneous Designs. *Journal of the American Statistical Association* 1999; 94(445):16-17.
58. Woods, B.S., Hawkins, N., Scott, D.A. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: A tutorial. *BMC Medical Research Methodology* 2010; 10(1):54.
59. Chinn, S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine* 2000; 19(22):3127-3131.

60. Whitehead, A., Bailey, A.J., Elbourne, D. Combining summaries of binary outcomes with those of continuous outcomes in a meta-analysis. *Journal of Biopharmaceutical Statistics* 1999; 9(1):1-16.
61. Lu, G., Ades, A. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics* 2009; 10(4):792-805.
62. Eddy, D.M., Hasselblad, V., Shachter, R.D. Meta-analysis by the confidence profile method. Academic Press, London, 1992.
63. Chootrakool, H., Shi, J.Q. Meta-analysis of multi-arm trials using empirical logistic transform. *The Open Medical Informatics Journal* 2008; 2:112-116.
64. Raiffa, H., Schlaiffer, R. Applied statistical decision theory, Wiley Classics Library edn. Wiley Interscience, New York, 1967.
65. Dakin, H.A., Welton, N.J., Ades, A.E., Collins, S., Orme, M., Kelly, S. Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. *Statistics in Medicine* 2011; 30:2511-2535.
66. Riley, R. Multivariate meta-analysis: the effect of ignoring with-study correlation. *Journal of the Royal Statistical Society Series A (Statistics in Society)* 2012; 172:789-811.
67. Spiegelhalter, D.J., Abrams, K.R., Myles, J.P. Bayesian approaches to clinical trials and health-care evaluation. Wiley, New York; 2004.
68. Brooks, S.P., Gelman, A. Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; 7(4):434-455.
69. Gelman, A. Inference and monitoring convergence. In: Gilks W.R., Richardson S., Spiegelhalter D.J., eds. *Markov chain Monte Carlo in practice*. Chapman & Hall, London; 1996; 131-143.
70. Lambert, P.C., Sutton, A.J., Burton, P.R., Abrams, K.R., Jones, D.R. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* 2005; 24(15):2401-2428.
71. Bradburn, M.J., Deeks, J.J., Berlin, J.A., Russell Localio, A. Much ado about nothing: a comparison of the performance of meta analytical methods with rare events. *Statistics in Medicine* 2007; 26(1):53-77.
72. Sweeting, M., Sutton, A.J., Lambert, P.C. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine* 2004; 23(9):1351-1375.
73. Claxton, K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics* 1999; 18(3):341-364.

74. Efron, B., Tibshirani, R. An introduction to the bootstrap. Chapman & Hall/CRC, New York; 1993.
75. Mantel, N., Haenszel, W. Statistical aspects of the analysis of data from retrospective studies. *Journal of the National Cancer Institute* 1959; 22(4):719-748.
76. Fleiss, J.L. Measures of effect size for categorical data. In: Cooper H., Hedges L., eds. *The handbook of research synthesis*. Russell Sage Foundation, New York; 1994; 245-260.
77. Yusuf, S., Petro, R., Lewis, J., Collins, R., Sleight, P. Beta blockade before and after myocardial infarction: an overview of randomised trials. *Progress in Cardiovascular Diseases* 1985; 27:335-371.
78. Gavaghan, D.J., Moore, R.A., McQuay, H.J. An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. *Pain - Journal of the International Association for the Study of Pain* 2000; 85(3):415-424.
79. Biggerstaff, B.J., Tweedie, R.L. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Statistics in Medicine* 1997; 16(7):753-768.
80. Hardy, R., Thompson, S. A likelihood approach to meta analysis with random effects. *Statistics in Medicine* 1996; 15(6):619-629.
81. Sterne, J. Meta-analysis in Stata: An updated collection from the Stata Journal. Stata Press Copyright, 2009.
82. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* 2010.
83. Borenstein, M., Hedges, L., Higgins, J., Rothstein, H. Comprehensive meta-analysis version 2. Englewood, NJ: Biostat 2005.
84. Wallace, B.C., Schmid, C.H., Lau, J., Trikalinos, T.A. Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. *BMC Medical Research Methodology* 2009; 9:80.
85. Bax, L., Yu, L.M., Ikeda, N., Tsuruta, H., Moons, K.G.M. Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. *BMC Medical Research Methodology* 2006; 6(1).
86. The Nordic Cochrane Centre. Review Manager (RevMan). Version 5.0. *The Cochrane Collaboration* 2008.
87. Bucher, H.C., Guyatt, G.H., Griffith, L.E., Walter, S.D. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *Journal of Clinical Epidemiology* 1997; 50(6):683-691.
88. White, I.R. Multivariate random-effects meta-analysis: Updates to mvmeta. *Stata Journal* 2011; 11:255-270.

89. Jones, B., Roger, J., Lane, P.W., Lawton, A., Fletcher, C., Cappellen, J.C. et al. Statistical approaches for conducting network meta-analysis in drug development. *Pharmaceutical Statistics* 2011; 10:523-531.
90. Riley, R.D., Abrams, K.R., Lambert, P.C., Sutton, A.J., Thompson, J.R. An evaluation of bivariate random effects meta analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine* 2007; 26(1):78-97.
91. Riley, R.D., Thompson, J.R., Abrams, K.R. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 2008; 9(1):172-186.
92. Arends, L.R., Hoes, A.W., Lubsen, J., Grobbee, D.E., Stijnen, T. Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Statistics in Medicine* 2000; 19(24):3497-3518.
93. Nam, I.S., Mengersen, K., Garthwaite, P. Multivariate meta analysis. *Statistics in Medicine* 2003; 22(14):2309-2333.
94. Burch, J., Paulden, M., Conti, S., Stock, C., Corbett, M., Welton, N.J. et al. Influenza - zanamivir, amantadine and oseltamivir (review): assessment report. 2008. NICE, London.
95. Welton, N.J., Cooper, N.J., Ades, A.E., Lu, G., Sutton, A.J. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. *Statistics in Medicine* 2008; 27(27):5620-5639.
96. Welton, N.J., Willis, S.R., Ades, A.E. Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. *Research Synthesis Methods* 2010; 1:239-257.
97. Eddy, D.M., Hasselblad, V., Shachter, R.D. Analysis of alcohol and breast cancer. *Meta-analysis by the confidence profile method*. Academic Press, London; 1992; 271-282.
98. Dominici, F. Combining contingency tables with missing dimensions. *Biometrics* 2000; 56(2):546-553.
99. 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(12):1340-1356.
100. Salanti, G., Higgins, J., White, I.R. Bayesian synthesis of epidemiological evidence with different combinations of exposure groups: application to a gene-gene-environment interaction. *Statistics in Medicine* 2006; 25(24):4147-4163.
101. Welton, N.J., Johnstone, E.C., David, S.P., Munafo, M.R. A cost-effectiveness analysis of genetic testing of the DRD2 Taq1A polymorphism to aid treatment choice for smoking cessation. *Nicotine & Tobacco Research* 2008; 10(1):231-240.

102. Welton, N.J., Ades, A.E. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: Uncertainty propagation, evidence synthesis, and model calibration. *Medical Decision Making* 2005; 25(6):633-645.
103. Price, M.J., Welton, N.J., Ades, A.E. Parameterization of treatment effects for meta analysis in multi state Markov models. *Statistics in Medicine* 2011; 30(2):140-151.
104. Hooper, L., Summerbell, C.D., Higgins, J.P., Thompson, R.L., Clements, G., Capps, N. et al. Reduced or modified dietary fat for preventing cardiovascular disease. *Cochrane Database of Systematic Reviews* 2001;(2):-CD002137.
105. Elliott, W.J., Meyer, P.M. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *The Lancet* 2007; 369(9557):201-207.

## APPENDIX: ILLUSTRATIVE EXAMPLES AND WINBUGS CODE

This appendix gives illustrative WinBUGS code for all the link functions and likelihoods, as well as example code for shared parameter models. All programming code is fully annotated. The program codes are printed here, but are also available as WinBUGS system files at [www.nicedsu.org.uk](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 RE program is identical to other RE programs, and similarly for the FE programs. We have therefore highlighted the linear predictor in blue, and the likelihood and deviance calculations in red to emphasise the modular nature of the code. Tables A1 gives an index of the programmes and their relation to the descriptions in the text. Note that for each example there are random and fixed effects versions of the code. All fixed effects code can be run using the same data structure described for the random effects.

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

<b>Program number</b>	<b>Fixed or Random Effects</b>	<b>Likelihood</b>	<b>Link Function</b>	<b>Example name</b>	<b>Model specification</b>
1	(a) Random (2-arm) (b) Fixed (2-arm) (c) Random (d) Fixed	Binomial	logit	Blocker	Section 2.1
2	(a) Random (b) Fixed	Poisson	log	Dietary fat	Section 3.1
3	(a) Random (b) Fixed	Binomial	cloglog	Diabetes	Section 3.2
4	(a) Random (b) Fixed	Multinomial	log	Schizophrenia	Section 3.3
5	(a) Random (b) Fixed	Normal	identity	Parkinson's	Section 3.4
6	(a) Random (b) Fixed	Multinomial	probit	Psoriasis	Section 3.6
7	(a) Random (b) Fixed	Normal (difference data)	identity	Parkinson's	Section 3.5
8	(a) Random (b) Fixed	Normal (shared parameter)	identity	Parkinson's	Section 4



## EXAMPLE 1. BLOCKER

The first two programmes (Blocker 1(a) and 1(b)) are somewhat apart from the other programmes: these programmes are only capable of processing syntheses of two treatments and 2-arm trials. We include them for the benefit of readers who may wish to start with the simplest possible case and see how the more general code that allows incorporation of multi-arm trials is related to the simpler code. The Blocker example is described in Section 2.1.

### Program 1(a): Binomial likelihood, logit link, Random Effects, two treatments (Blocker example). Two-arm trials Only

```
# Binomial likelihood, logit link, pairwise meta-analysis (2 treatments)
# Random effects model
model{
  for(i in 1:ns){
    delta[i,1] <- 0
    mu[i] ~ dnorm(0, .0001)
    for (k in 1:2) {
      r[i,k] ~ dbin(p[i,k], n[i,k])
      logit(p[i,k]) <- mu[i] + delta[i,k]
      rhat[i,k] <- p[i,k] * n[i,k]
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k]) - log(rhat[i,k]))
        + (n[i,k] - r[i,k]) * (log(n[i,k] - r[i,k]) - log(n[i,k] - rhat[i,k])))
    }
    resdev[i] <- sum(dev[i,])
    delta[i,2] ~ dnorm(d[2], tau)
  }
  totresdev <- sum(resdev[])
  d[1] <- 0
  d[2] ~ dnorm(0, .0001)
  sd ~ dunif(0, 5)
  tau <- pow(sd, -2)
}
# *** PROGRAM STARTS
# LOOP THROUGH STUDIES
# treatment effect is zero for control arm
# vague priors for all trial baselines
# LOOP THROUGH ARMS
# binomial likelihood
# model for linear predictor
# expected value of the numerators
# Deviance contribution
# summed residual deviance contribution for this trial
# trial-specific LOR distributions
# Total Residual Deviance
# treatment effect is zero for reference treatment
# vague prior for treatment effect
# vague prior for between-trial SD
# between-trial precision = (1/between-trial variance)
# *** PROGRAM ENDS
```

The data structure has two components: a list specifying the number of studies  $ns$  and the main body of data which is in a column format:  $r_{[,1]}$  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. Text can be included after the hash symbol (#) for ease of reference to the original data source.

```
# Data (Blocker example)
list(ns=22)
```

$r_{[,1]}$	$n_{[,1]}$	$r_{[,2]}$	$n_{[,2]}$	#	Study ID
3	39	3	38	#	1
14	116	7	114	#	2
11	93	5	69	#	3
127	1520	102	1533	#	4
27	365	28	355	#	5
6	52	4	59	#	6
152	939	98	945	#	7
48	471	60	632	#	8
37	282	25	278	#	9
188	1921	138	1916	#	10
52	583	64	873	#	11

```

47      266      45      263      #      12
16      293      9      291      #      13
45      883      57      858      #      14
31      147      25      154      #      15
38      213      33      207      #      16
12      122      28      251      #      17
6       154      8       151      #      18
3       134      6       174      #      19
40      218      32      209      #      20
43      364      27      391      #      21
39      674      22      680      #      22
END

```

```

# Initial values
#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, 0, 0, 0, 0, 0))
#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, -3, -3, -3, -3, -3))
#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, -3, -3, -1, -3, -7, -3, -3))

```

### Program 1(b): Binomial likelihood, logit link, Fixed Effects, two treatments (Blocker example), Two-arm trials only.

```

# Binomial likelihood, logit link, pairwise meta-analysis (2 treatments)
# Fixed effect model
model{
  for(i in 1:ns){
    mu[i] ~ dnorm(0,.0001)
    for(k in 1:2){
      r[i,k] ~ dbin(p[i,k],n[i,k])
      logit(p[i,k]) <- mu[i] + d[k]
      rhat[i,k] <- p[i,k] * n[i,k]
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k])))
      + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
    }
    resdev[i] <- sum(dev[i,])
  }
  toresdev <- sum(resdev[])
  d[1] <- 0
  d[2] ~ dnorm(0,.0001)
}

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

```

All code presented below 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.

This example and results are described in Section 2.1 of the paper (Table 1 and Table 2). The WinBUGS code for random effects is given in program 1(c) and the fixed effects code is given in program 1(d).

### Program 1(c): Binomial likelihood, logit link, Random Effects (Blocker example)

```

# Binomial likelihood, logit link
# Random effects model for multi-arm trials
model{
  for(i in 1:ns){
    w[i,1] <- 0
    delta[i,1] <- 0
    mu[i] ~ dnorm(0,.0001)
    for (k in 1:na[i]) {
      r[i,k] ~ dbin(p[i,k],n[i,k])
      logit(p[i,k]) <- mu[i] + delta[i,k]
      rhat[i,k] <- p[i,k] * n[i,k]
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k])))
        + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
    }
    resdev[i] <- sum(dev[i,1:na[i]])
    for (k in 2:na[i]) {
      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]
      taud[i,k] <- tau *2*(k-1)/k
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
    }
  }
  totesdev <- sum(resdev[])
  d[1] <- 0
  for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }
  sd ~ dunif(0,5)
  tau <- pow(sd,-2)
}

```

```

# *** PROGRAM STARTS
# LOOP THROUGH STUDIES
# adjustment for multi-arm trials is zero for control arm
# treatment effect is zero for control arm
# vague priors for all trial baselines
# LOOP THROUGH ARMS
# binomial likelihood
# model for linear predictor
# expected value of the numerators
#Deviance contribution
# summed residual deviance contribution for this trial
# LOOP THROUGH ARMS
# mean of LOR distributions (with multi-arm trial correction)
# precision of LOR distributions (with multi-arm trial correction)
# adjustment for multi-arm RCTs
# cumulative adjustment for multi-arm trials
#Total Residual Deviance
# treatment effect is zero for reference treatment
# vague priors for treatment effects
# vague prior for between-trial SD. ALTERNATIVES BELOW
# between-trial precision = (1/between-trial variance)
# *** PROGRAM ENDS

```

Alternative prior distributions can be used for the Random Effects Variance. For example, the last two lines above can be replaced by a vague Gamma prior on the precision parameter, which is sometimes also referred to as a vague inverse Gamma prior on the variance:

```

tau ~ dgamma(.001,.001) # vague gamma prior on the precision
sd <- pow(tau,-0.5)

```

See Section 6.2 in the main document for further discussion of prior distributions.

Additional code can be added before the closing brace to estimate all the pair-wise Log Odds Ratios and Odds Ratios, to generate ranking statistics and the probability that each treatment is the best treatment, and to produce estimates of absolute effects, given additional information on the absolute treatment effect on one of the treatments. In addition, given an

assumption about the absolute effect of one treatment, it is possible to express the treatment effect on other scales (risk difference, relative risk), or number needed to treat, *and* to obtain confidence intervals for all these quantities. This is illustrated below.

```
#####
# Extra code for all odds ratios and log odds ratios, ranking, and absolute effects, and relative effects
# on alternative scales: Numbers Needed to Treat, Risk Difference, Relative Risks
#####
# 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) {
    or[c,k] <- exp(d[k] - d[c])
    lor[c,k] <- (d[k]-d[c])
  }
}

# ranking on relative scale
for (k in 1:nt) {
  rk[k] <- nt+1-rank(d[],k) # assumes events are "good"
# rk[k] <- rank(d[],k) # assumes events are "bad"
  best[k] <- equals(rk[k],1) #calculate probability that treat k is best
  for (h in 1:nt){ prob[h,k] <- equals(rk[k],h) } # calculates probability that treat k is h-th best
}

# 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
A ~ dnorm(meanA,precA)
for (k in 1:nt) { logit(T[k]) <- A + d[k] }

# Provide estimates of number needed to treat NNT[k], Risk Difference RD[k],
# and Relative Risk RR[k], for each treatment, relative to treatment 1
for (k in 2:nt) {
  NNT[k] <- 1/(T[k] - T[1]) # assumes events are "good"
# NNT[k] <- 1/(T[1]- T[k]) # assumes events are "bad"
  RD[k] <- T[k] - T[1]
  RR[k] <- T[k]/T[1]
}

```

The data structure has two components: a list specifying the number of treatments  $nt$  and number of studies  $ns$ . Both data components need to be loaded into WinBUGS for the program to run. The main body of data is in a vector format, in the order  $r_{[,1]}$  then  $n_{[,1]}$ , the numerators and denominators for the first treatment, then  $r_{[,2]}$  then  $n_{[,2]}$ , the numerators and denominators for the second listed treatment, then  $t_{[,1]}$  and  $t_{[,2]}$ , the treatment number identifiers for the first and second listed treatments, and finally the number of arms in each trial,  $na[]$ . The purpose for this structure becomes clearer in datasets with multi-arm trials. An important feature of the code presented is the assumption that the treatments are always presented in ascending (numerical) order and that treatment 1 is taken as the reference treatment. This rule is crucial when conducting network meta-analysis to ensure the correct relative effects are estimated.

We strongly recommend the use of the column data format shown here in preference to the list format that WinBUGS also allows, and the use of comments to add trial names or references. This facilitates data checking.

```
# Data (Blocker example)
list(nt=2, ns=22)
```

r[,1]	n[,1]	r[,2]	n[,2]	t[,1]	t[,2]	na[]
3	39	3	38	1	2	2
14	116	7	114	1	2	2
11	93	5	69	1	2	2
127	1520	102	1533	1	2	2
27	365	28	355	1	2	2
6	52	4	59	1	2	2
152	939	98	945	1	2	2
48	471	60	632	1	2	2
37	282	25	278	1	2	2
188	1921	138	1916	1	2	2
52	583	64	873	1	2	2
47	266	45	263	1	2	2
16	293	9	291	1	2	2
45	883	57	858	1	2	2
31	147	25	154	1	2	2
38	213	33	207	1	2	2
12	122	28	251	1	2	2
6	154	8	151	1	2	2
3	134	6	174	1	2	2
40	218	32	209	1	2	2
43	364	27	391	1	2	2
39	674	22	680	1	2	2

END

```
# Initial values
# Initial values for delta 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, 0, 0, 0, 0, 0, 0, 0))
#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, -3, -3, -3, -3, -3, -3, -3))
#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, -3, -3, -1, -3, -7, -3, -3))
```

To obtain the posterior summaries of the parameters of interest for inference, the nodes  $d$  and  $sd$  (in the random effects model only) 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 calculate the leverage for each data point and to draw leverage plots,  $rhat$  needs to be monitored.

### Program 1(d): Binomial likelihood, logit link, Fixed Effects (Blocker example)

```
# Binomial likelihood, logit link
# Fixed effects model
model{
  for(i in 1:ns){
    mu[i] ~ dnorm(0, .0001)
    for(k in 1:na[i]) {
      # *** PROGRAM STARTS
      # LOOP THROUGH STUDIES
      # vague priors for all trial baselines
      # LOOP THROUGH ARMS
```

```

r[i,k] ~ dbin(p[i,k],n[i,k]) # binomial likelihood
logit(p[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]] # 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
}
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
} # *** 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, 0, 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, -3, -3))
#chain 3
list(d=c( NA, 2), mu=c(-3, 5, -1, -3, 7, -3, -4, -3, -3, 0, -3, -3, 0, 3, 5, -3, -3, -1, -3, -7, -3, -3))

```

## EXAMPLE 2. DIETARY FAT

In a Cochrane Review of randomised controlled trials to assess the effect of change in dietary fats on total and cardiovascular mortality,<sup>104</sup> data extracted was in the form of rates and given as the number of events per person-years observed (Table A2).

**Table A2 Dietary fat example: Study names and treatment codes for the 10 included studies and person-years and total mortality observed in each study.**

Study name and ID	Treatment			Person-yrs obs			Total mortality			Number randomised		
	control	diet	diet 2	control	diet	diet 2	control	diet	diet 2	control	diet	diet 2
	t[,1]	t[,2]	t[,3]	E[,1]	E[,2]	E[,3]	r[,1]	r[,2]	r[,3]	n[,1]	n[,2]	n[,3]
1.DART	1	2		1917	1925		113	111		1015	1018	
2.London Corn /Olive	1	2	2	43.6	41.3	38	1	5	3	26	28	26
3.London Low Fat	1	2		393.5	373.9		24	20		129	123	
4.Minnesota Coronary	1	2		4715	4823		248	269		4516	4516	
5.MRC Soya	1	2		715	751		31	28		194	199	
6.Oslo Diet-Heart	1	2		885	895		65	48		206	206	
7.STARS	1	2		87.8	91		3	1		30	30	
8.Sydney Diet-Heart	1	2		1011	939		28	39		237	221	
9.Veterans Administration	1	2		1544	1588		177	174		422	424	
10.Veterans Diet & Skin CA	1	2		125	123		2	1		67	66	

Most of the trials compared only one reduced fat dietary intervention with a control diet (non-reduced fat). However, the ‘London Corn/Olive’ trial compared two types of reduced fat diets

against control (for more details see Hooper et al.<sup>104</sup>). For the purpose of this example we considered the two different types of diet as the same intervention (treatment 2), but kept the treatment arms separately, so that in a random effects model, this trial will provide two correlated estimates of the trial-specific treatment effect  $\delta_{i,12}$  and in the fixed effects model, both arms will contribute to the estimate of the common treatment effect  $d_{12}$ .

The model for this type of data is outlined in Section 3.1. The WinBUGS code for random effects is given in program 2(a) and the fixed effects code is given in program 2(b).

### Program 2(a): Poisson likelihood, log link, Random Effects (Dietary fat example)

```
# Poisson likelihood, log link
# Random effects model for multi-arm trials
model{
  for(i in 1:ns){
    w[i,1] <- 0
    delta[i,1] <- 0
    mu[i] ~ dnorm(0,.0001)
    for (k in 1:na[i]) {
      r[i,k] ~ dpois(theta[i,k])
      theta[i,k] <- lambda[i,k]*E[i,k]
      log(lambda[i,k]) <- mu[i] + delta[i,k]
      dev[i,k] <- 2*((theta[i,k]-r[i,k]) + r[i,k]*log(r[i,k]/theta[i,k])) #Deviance contribution
    }
    resdev[i] <- sum(dev[i,1:na[i]])
    for (k in 2:na[i]) {
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
      md[i,k] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]
      taud[i,k] <- tau *2*(k-1)/k
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
    }
  }
  totesdev <- sum(resdev[])
  d[1]<-0
  for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }
  sd ~ dunif(0,5)
  tau <- pow(sd,-2)
}

# *** PROGRAM STARTS
# LOOP THROUGH STUDIES
# adjustment for multi-arm trials is zero for control arm
# treatment effect is zero for control arm
# vague priors for all trial baselines
# LOOP THROUGH ARMS
# Poisson likelihood
# failure rate * exposure
# model for linear predictor
# summed residual deviance contribution for this trial
# LOOP THROUGH ARMS
# trial-specific LOR distributions
# mean of LOR distributions (with multi-arm trial correction)
# precision of LOR distributions (with multi-arm trial correction)
# adjustment for multi-arm RCTs
# cumulative adjustment for multi-arm trials
#Total Residual Deviance
# treatment effect is zero for reference treatment
# vague priors for treatment effects
# vague prior for between-trial SD
# between-trial precision = (1/between-trial variance)
# *** PROGRAM ENDS
```

As before, additional code can be added to monitor all the  $K(K-1)/2$  log hazard ratios and hazard ratios when there are more than 2 treatments:

```
# pairwise HRs and LHRs for all possible pair-wise comparisons, if nt>2
for (c in 1:(nt-1)) {
  for (k in (c+1):nt) {
    lhr[c,k] <- (d[k]-d[c])
    log(lhr[c,k]) <- lhr[c,k]
  }
}
```

or to rank treatments and monitor the probabilities that each is best, and to generate absolute rates for each treatment. For example if, on the basis of some external data we believe that

the log-rate for Treatment 1 has mean -3, and precision 1.77, then we can generate absolute rates for other treatments as follows:

```
A ~ dnorm(-3,1.77)
for (k in 1:nt) { log(T[k]) <- A + d[k] }
```

A further variable that may be required for cost-effectiveness modelling might be the proportion of patients that would be expected to have an event, after a follow-up of, say, 3 months, under each treatment. In this example the rates are per year, so:

```
for (k in 1:nt) { p[k] <- 1-exp(-T[k]*0.25) }
```

The data structure again has two components: a list specifying the number of treatments  $nt$  and number of studies  $ns$ . Both data components need to be loaded into WinBUGS for the program to run. The main body of data is in a vector format, and we need to allow for a 3-arm trial. Three places are therefore required to specify the treatments  $t_{[,]}$ , the exposure times  $E_{[,]}$  and the number of events  $r_{[,]}$  in each arm; “NA” indicates that the data is missing for a particular cell. As before  $na[]$  is the number of arms in each study. Text can be included after a hash symbol for ease of reference to the original data source.

```
# Data (Dietary fat example)
list(ns=10, nt=2)

t[,1]  t[,2]  t[,3]  E[,1]  E[,2]  E[,3]  r[,1]  r[,2]  r[,3]  na[]  #ID
1      2      NA    1917   1925   NA     113    111    NA     2     #2 DART
1      2      2     43.6   41.3   38     1      5      3     3     #10 London Corn /Olive
1      2      NA    393.5  373.9  NA     24     20     NA     2     #11 London Low Fat
1      2      NA    4715   4823   NA     248    269    NA     2     #14 Minnesota Coronary
1      2      NA    715    751    NA     31     28     NA     2     #15 MRC Soya
1      2      NA    885    895    NA     65     48     NA     2     #18 Oslo Diet-Heart
1      2      NA    87.8   91     NA     3      1      NA     2     #22 STARS
1      2      NA    1011   939    NA     28     39     NA     2     #23 Sydney Diet-Heart
1      2      NA    1544   1588   NA     177    174    NA     2     #26 Veterans Administration
1      2      NA    125    123    NA     2      1      NA     2     #27 Veterans Diet & Skin CA
END
```

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



To get the posterior summaries of the parameters of interest for inference, the nodes  $d$  and  $sd$  (in the random effects model only) 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 the DIC tool), need to be monitored. In addition, to calculate the leverage for each data point and to draw leverage plots,  $theta$  needs to be monitored.

### Program 2(b): Poisson likelihood, log link, Fixed Effects (Dietary fat)

```

# Poisson likelihood, log link
# Fixed effects model for multi-arm trials
model{
  for(i in 1:ns){
    mu[i] ~ dnorm(0,.0001)
    for (k in 1:na[i]) {
      r[i,k] ~ dpois(theta[i,k])
      theta[i,k] <- lambda[i,k]*E[i,k]
      log(lambda[i,k]) <- mu[i] + d[t[i,k]] - d[t[i,1]]
      dev[i,k] <- 2*((theta[i,k]-r[i,k]) + r[i,k]*log(r[i,k]/theta[i,k]))
    }
    resdev[i] <- sum(dev[i,1:na[i]])
  }
  totresdev <- sum(resdev[])
  d[1]<-0
  for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }
}

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

```

## RESULTS

The results from the two models (3 chains: 20,000 iterations after a burn-in of 20,000 for the FE model and 100,000 iterations after a burn-in of 100,000 for the RE model) are compared in Table A3. The random and fixed effects models are indistinguishable in terms of model fit, and both appear to fit the data well in that  $\bar{D}_{res}$  is close to 21, the number of data points. The posterior median of the pooled log-rate of a reduced fat diet, compared to the control diet is -0.01 in the FE model with 95% Credible Interval (-0.11, 0.10) suggesting no difference in the number of cardiovascular mortalities in each group. The posterior medians of the absolute rates of mortality (and their 95% Credible intervals), having assumed that the log-rate of mortality on the control diet has mean -3 and precision 1.77, on the control and reduced fat diets are the same (Table A3).

**Table A3 Dietary fat example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effect  $d_{12}$ , absolute effects of the control diet ( $T_1$ ) and the reduced fat diet ( $T_2$ ) for a log-rate of mortality on the control diet with mean - 3 and precision 1.77, heterogeneity parameter  $\tau$  and model fit statistics.**

	FE model				RE model			
	mean	sd	median	CrI	mean	sd	median	CrI
$d_{12}$	-0.01	0.054	-0.01	(-0.11,0.10)	-0.02	0.09	-0.01	(-0.19,0.16)
$T_1$	0.06	0.04	0.05	(0.01,0.18)	0.06	0.04	0.05	(0.01,0.18)
$T_2$	0.06	0.04	0.05	(0.01,0.18)	0.06	0.04	0.05	(0.01,0.18)
$\tau$	-	-	-	-	0.13	0.12	0.10	(0.00,0.43)
$\bar{D}_{res}^*$	22.32				21.5			
p <sub>D</sub>	10.9				13.3			
DIC	33.2				34.8			

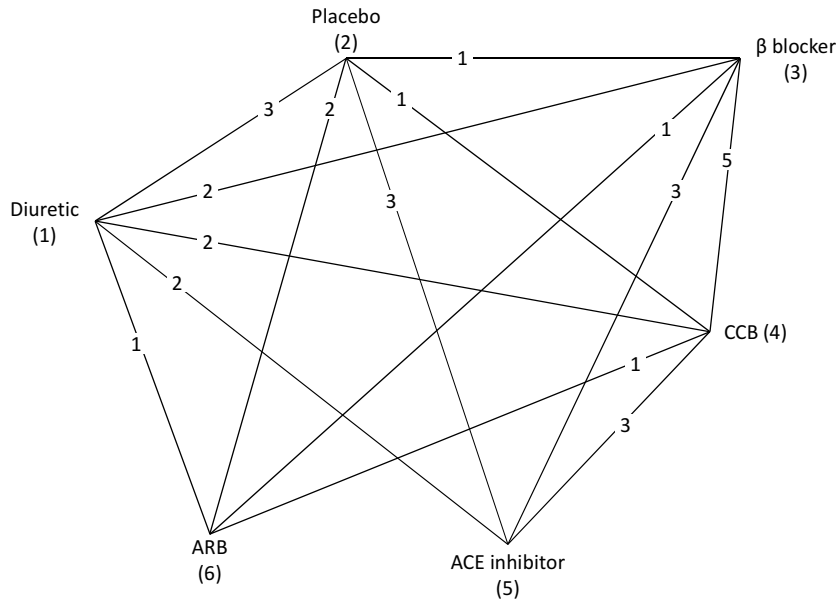
\* Compare to 21 data points

### EXAMPLE 3. DIABETES

Here we show code for a linear model on the log rate scale based on binomial data gathered at different follow-up times. We use as an illustration a network meta-analysis to assess the incidence of diabetes in randomised controlled trials of antihypertensive drugs.<sup>105</sup> The outcome was new cases of diabetes observed over the trial duration period (measured in years) for 6 different drugs: Diuretic (treatment 1), Placebo (treatment 2),  $\beta$  blocker (treatment 3), CCB (treatment 4), ACE inhibitor (treatment 5) and ARB (treatment 6). In this example of a network meta-analysis, the reference treatment chosen was diuretic, as recommended in this field – for more details see Elliott & Meyer.<sup>105</sup> The data are presented in Table A4 and the network diagram in Figure A3.

**Table A4 Diabetes example: study names, follow-up time in years, treatments compared, total number of new cases of diabetes and number of patients in each trial arm, where Diuretic = treatment 1, Placebo = treatment 2,  $\beta$  blocker = treatment 3, CCB = treatment 4, ACE inhibitor = treatment 5 and ARB = treatment 6.**<sup>105</sup>

Study ID	Study Follow-up (in years)	Treatment			New cases of diabetes			Total number of patients		
		arm 1	arm 2	arm 3	arm 1	arm 2	arm 3	arm 1	arm 2	arm 3
	time[]	t[,1]	t[,2]	t[,3]	r[,1]	r[,2]	r[,3]	n[,1]	n[,2]	n[,3]
1.MRC-E	5.8	1	2	3	43	34	37	1081	2213	1102
2.EWPH	4.7	1	2		29	20		416	424	
3.SHEP	3	1	2		140	118		1631	1578	
4.HAPPHY	3.8	1	3		75	86		3272	3297	
5.ALLHAT	4	1	4	5	302	154	119	6766	3954	4096
6.INSIGHT	3	1	4		176	136		2511	2508	
7.ANBP-2	4.1	1	5		200	138		2826	2800	
8.ALPINE	1	1	6		8	1		196	196	
9.FEVER	3.3	2	4		154	177		4870	4841	
10.DREAM	3	2	5		489	449		2646	2623	
11.HOPE	4.5	2	5		155	102		2883	2837	
12.PEACE	4.8	2	5		399	335		3472	3432	
13.CHARM	3.1	2	6		202	163		2721	2715	
14.SCOPE	3.7	2	6		115	93		2175	2167	
15.AASK	3.8	3	4	5	70	32	45	405	202	410
16.STOP-2	4	3	4	5	97	95	93	1960	1965	1970
17.ASCOT	5.5	3	4		799	567		7040	7072	
18.NORDIL	4.5	3	4		251	216		5059	5095	
19.INVEST	4	3	4		665	569		8078	8098	
20.CAPPP	6.1	3	5		380	337		5230	5183	
21.LIFE	4.8	3	6		320	242		3979	4020	
22.VALUE	4.2	4	6		845	690		5074	5087	



**Figure A3 Diabetes network:** each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison and the numbers by the treatment names are the treatment codes used in the modelling.

The model for this type of data is outlined in Section 3.2. The WinBUGS code for random effects is given in program 3(a) and the fixed effects code is given in program 3(b).

### Program 3(a): Binomial likelihood, cloglog link, Random Effects (Diabetes example)

```

# Binomial likelihood, cloglog link
# Random effects model for multi-arm trials
model{
  for(i in 1:ns){
    w[i,1] <- 0
    delta[i,1] <- 0
    mu[i] ~ dnorm(0,.0001)
    for(k in 1:na[i]) {
      r[i,k] ~ dbin(p[i,k],n[i,k])
      cloglog(p[i,k]) <- log(time[i]) + mu[i] + delta[i,k]
      rhat[i,k] <- p[i,k] * n[i,k]
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
        + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k]))) #Deviance contribution
    }
    resdev[i] <- sum(dev[i,1:na[i]])
    for(k in 2:na[i]) {
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
      md[i,k] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]
      taud[i,k] <- tau * 2*(k-1)/k
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
    }
  }
  totesdev <- sum(resdev[])
  d[1]<-0
  for(k in 2:nt){ d[k] ~ dnorm(0,.0001) }
}

```

```

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

```

Additional code to generate all the treatment contrasts, absolute effects, ranking can be added, as with the Poisson – log link models. To generate absolute probabilities for each treatment, if, on the basis of some external data, we believe that the cloglog of the probability of an event for Treatment 1, after a time period of 3 years, has mean -4.2, and precision 1.11, then we can generate absolute probabilities for other treatments as follows:

```

A ~ dnorm(-4.2,1.11)
for (k in 1:nt) { cloglog(T[k]) <- log(3) + A + d[k] }

```

The main body of data is in the same format as the binomial likelihood with logit link in Example 1, with an additional vector for follow-up time, `time[]`. Note that treatments are ordered numerically so that the treatment in arm 2 has a higher code than the treatment in arm 1, and the treatment in arm 3 has a higher code than the treatment in arm 2. This is essential to ensure the correct relative effects are obtained.

```

# Data (Diabetes example)
list(ns=22, nt=6)

```

time[]	t[,1]	r[,1]	n[,1]	t[,2]	r[,2]	n[,2]	t[,3]	r[,3]	n[,3]	na[]	#Study
5.8	1	43	1081	2	34	2213	3	37	1102	3	#MRC-E 38
4.7	1	29	416	2	20	424	NA	NA	NA	2	#EWPH 32
3	1	140	1631	2	118	1578	NA	NA	NA	2	#SHEP 42
3.8	1	75	3272	3	86	3297	NA	NA	NA	2	#HAPPHY33
4	1	302	6766	4	154	3954	5	119	4096	3	#ALLHAT26
3	1	176	2511	4	136	2508	NA	NA	NA	2	#INSIGHT35
4.1	1	200	2826	5	138	2800	NA	NA	NA	2	#ANBP-2 18
1	1	8	196	6	1	196	NA	NA	NA	2	#ALPINE 27
3.3	2	154	4870	4	177	4841	NA	NA	NA	2	#FEVER 20
3	2	489	2646	5	449	2623	NA	NA	NA	2	#DREAM 31
4.5	2	155	2883	5	102	2837	NA	NA	NA	2	#HOPE 34
4.8	2	399	3472	5	335	3432	NA	NA	NA	2	#PEACE 40
3.1	2	202	2721	6	163	2715	NA	NA	NA	2	#CHARM 30
3.7	2	115	2175	6	93	2167	NA	NA	NA	2	#SCOPE 41
3.8	3	70	405	4	32	202	5	45	410	3	#AASK 25
4	3	97	1960	4	95	1965	5	93	1970	3	#STOP-2 43
5.5	3	799	7040	4	567	7072	NA	NA	NA	2	#ASCOT 28
4.5	3	251	5059	4	216	5095	NA	NA	NA	2	#NORDIL 39
4	3	665	8078	4	569	8098	NA	NA	NA	2	#INVEST 36
6.1	3	380	5230	5	337	5183	NA	NA	NA	2	#CAPPP 29
4.8	3	320	3979	6	242	4020	NA	NA	NA	2	#LIFE 37
4.2	4	845	5074	6	690	5087	NA	NA	NA	2	#VALUE 44

```

# Initial Values
# In this case it is advisable to initialise delta to avoid numerical errors
#chain 1
list(d=c(NA,0,0,0,0,0), 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),

```

```

delta= structure(.Data= c(NA, 0, 0, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, 0, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA,
0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0,0, NA, 0, 0, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA,
NA, 0, NA), .Dim=c(22, 3)))
#chain 2
list(d=c(NA,-1,4,-1,2,3), sd=3, mu=c(1,1,0,1,0, 0,1,0,0,0, 1,1,0,0,0, 0,1,0,0,0, 1,1),
delta= structure(.Data= c(NA, 0, 0, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, 0, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA,
0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0,0, NA, 0, 0, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA,
NA, 0, NA), .Dim=c(22, 3)))
#chain 3
list(d=c(NA,1,4,-3,-2,3), sd=4.5, mu=c(1,1,0,1,0, 0,1,0,0,0, 1,1,0,-2,0, 0,1,0,-2,0, 1,1),
delta= structure(.Data= c(NA, 0, 0, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, 0, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA,
0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0,0, NA, 0, 0, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA, NA, 0, NA,
NA, 0, NA), .Dim=c(22, 3)))

```

### Program 3(b): Binomial likelihood, cloglog link, Fixed Effects (Diabetes example)

```

# Binomial likelihood, cloglog link
# Fixed effects model
model{
  for(i in 1:ns){
    mu[i] ~ dnorm(0,.0001)
    for (k in 1:na[i]) {
      r[i,k] ~ dbin(p[i,k],n[i,k])
      cloglog(p[i,k]) <- log(time[i]) + mu[i] + d[t[i,k]] - d[t[i,1]]
      rhat[i,k] <- p[i,k] * n[i,k]
      dev[i,k] <- 2 * (r[i,k] * (log(r[i,k])-log(rhat[i,k]))
        + (n[i,k]-r[i,k]) * (log(n[i,k]-r[i,k]) - log(n[i,k]-rhat[i,k])))
    }
    resdev[i] <- sum(dev[i,1:na[i]])
  }
  totresdev <- sum(resdev[])
  d[1]<-0
  for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }
}

# Initial Values
#chain 1
list(d=c(NA,0,0,0,0,0), mu=c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0) )
#chain 2
list(d=c(NA,1,1,1,1,1), mu=c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0) )
#chain 3
list(d=c(NA,1,1,1,1,2), mu=c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0,0) )

```

## RESULTS

Both fixed and random effects models were fitted (3 chains: 100,000 iterations after a burn-in of 50,000). From the results presented in Table A5, we see that the fixed effects model has a very poor fit and the random effects model should be preferred for inference.

**Table A5 Diabetes example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effects of Placebo ( $d_{12}$ ),  $\beta$  blocker ( $d_{13}$ ), CCB ( $d_{14}$ ), ACE inhibitor ( $d_{15}$ ) and ARB ( $d_{16}$ ) relative to Diuretic; absolute effects of diuretic ( $T_1$ ) Placebo ( $T_2$ ),  $\beta$  blocker ( $T_3$ ), CCB ( $T_4$ ), ACE inhibitor ( $T_5$ ) and ARB ( $T_6$ ); heterogeneity parameter  $\tau$  and model fit statistics.**

	FE model				RE model			
	mean	sd	median	CrI	mean	sd	median	CrI
$d_{12}$	-0.25	0.06	-0.25	(-0.36,-0.14)	-0.29	0.09	-0.29	(-0.47,-0.12)
$d_{13}$	-0.06	0.06	-0.06	(-0.17,0.05)	-0.07	0.09	-0.07	(-0.25,0.10)
$d_{14}$	-0.25	0.05	-0.25	(-0.36,-0.15)	-0.24	0.08	-0.24	(-0.41,-0.08)
$d_{15}$	-0.36	0.05	-0.36	(-0.46,-0.25)	-0.40	0.09	-0.40	(-0.58,-0.24)
$d_{16}$	-0.45	0.06	-0.45	(-0.58,-0.33)	-0.47	0.11	-0.47	(-0.70,-0.27)
$T_1$	0.065	0.067	0.044	(0.01,0.25)	0.065	0.067	0.044	(0.01,0.25)
$T_2$	0.052	0.055	0.034	(0.01,0.20)	0.050	0.053	0.033	(0.01,0.20)
$T_3$	0.062	0.064	0.042	(0.01,0.24)	0.061	0.064	0.041	(0.01,0.24)
$T_4$	0.051	0.055	0.034	(0.01,0.20)	0.052	0.056	0.035	(0.01,0.20)
$T_5$	0.047	0.050	0.031	(0.00,0.18)	0.045	0.048	0.030	(0.00,0.18)
$T_6$	0.043	0.046	0.028	(0.00,0.17)	0.042	0.046	0.028	(0.00,0.17)
$\tau$					0.13	0.04	0.12	(0.05,0.23)
$\bar{D}_{res}^*$	78.25				53.7			
pD	27.0				38.0			
DIC	105.2				91.7			

\* Compare to 48 data points

The posterior median of the pooled treatment effects, on the complementary log-log scale, of treatments 2 to 6 relative to the reference treatment show a beneficial effect of all the treatment with the exception of treatment 3 (Table A5).

The posterior medians of the absolute probabilities of developing diabetes after a period of three years, assuming that the cloglog of the probability of developing diabetes on Placebo has mean -4.2 and precision 1.11, on each of the treatments are between 3 and 4% (Table A5).

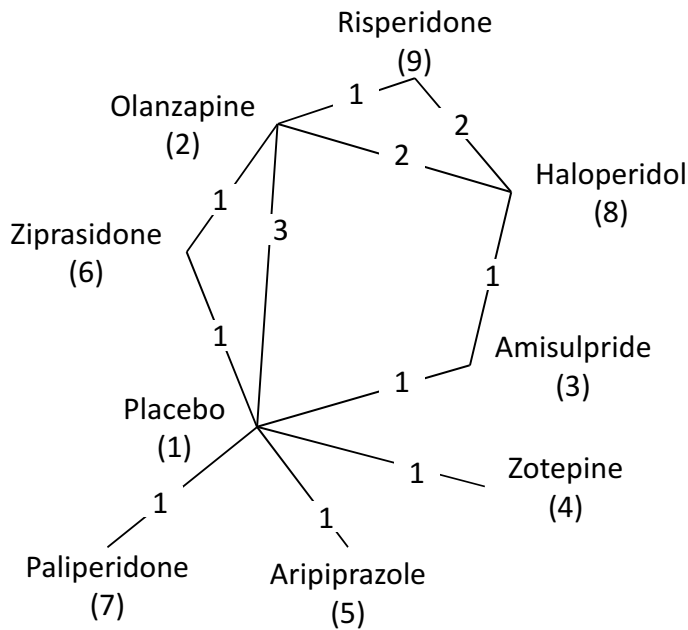
#### EXAMPLE 4. SCHIZOPHRENIA

In a network meta-analysis of trials of antipsychotic medication for the prevention of relapse in people with schizophrenia, 17 trials comparing 9 treatments including placebo were included.<sup>45</sup> The data available from each trial are the number of patients in each of three outcome states at the end of follow-up. The outcome states are: relapse ( $j=1$ ), discontinuation of treatment due to intolerable side effects ( $j=2$ ), and discontinuation for other reasons ( $j=3$ ), which might include inefficacy of treatment that did not fulfil all criteria for relapse, or loss to follow-up. Patients not reaching any of these end-points at the end of follow-up were considered as censored observations, and still in remission ( $j=4$ ) (for more details see Ades et al.<sup>45</sup>). The data are presented in Table A6 and the network diagram in Figure A4.

**Table A6 Schizophrenia example: study names, follow-up time in weeks, treatments compared, total number of events for each of the four states and total number of patients in each trial arm, where Placebo = treatment 1, Olanzapine = 2, Amisulpride = 3, Zotepine = 4, Aripiprazole = 5, Ziprasidone = 6, Paliperidone = 7, Haloperidol = 8, Risperidone = 9.<sup>45</sup>**

Study	follow-up (weeks)	treatment		number of events								total no of patients	
				relapse		discontinuation due to intolerable side effects		discontinuation for other reasons		Patient still in remission			
				arm 1	arm 2	arm 1	arm 2	arm 1	arm 2	arm 1	arm 2		
f[]	t[,1]	t[,2]	r[,1,1]	r[,2,1]	r[,1,2]	r[,2,2]	r[,1,3]	r[,2,3]	r[,1,4]	r[,2,4]	n[,1]	n[,2]	
1	42	1	2	28	9	12	2	15	19	47	194	102	224
2	46	1	2	7	10	0	2	4	16	2	17	13	45
3	46	1	2	5	6	2	10	5	15	2	17	14	48
4	26	1	3	5	4	5	1	39	26	23	38	72	69
5	26	1	4	21	4	4	16	24	21	9	20	58	61
6	26	1	5	85	50	13	16	12	18	45	71	155	155
7	52	1	6	43	71	11	19	7	28	10	88	71	206
8	47	1	7	52	23	1	3	7	17	41	61	101	104
9	28	2	6	11	8	6	5	44	33	10	9	71	55
10	52	2	8	87	34	54	20	170	50	316	76	627	180
11	52	2	8	28	29	9	14	26	25	78	66	141	134
12	28	2	9	20	53	17	17	36	18	99	79	172	167
13	52	3	8	5	9	3	5	2	2	19	15	29	31
14	52	8	9	65	41	29	22	80	60	14	54	188	177
15	104	8	9	8	4	0	3	4	4	18	22	30	33





**Figure A4 Schizophrenia network:** each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison and the numbers by the treatment names are the treatment codes used in the modelling.

A random effects model with different between-trial variation for each outcome and a fixed effects model were fitted, as outlined in Section 3.3. The WinBUGS code for the random effects model is given in program 4(a), and the fixed effect code is given in program 4(b).

**Program 4(a): Multinomial likelihood (with competing risks), log link, Random Effects (Schizophrenia example)**

```
# Multinomial likelihood, log link (competing risks)
# Random effects model for multi-arm trials
model{
  for(i in 1:ns){
    for(m in 1:3){
      w[i,1,m] <- 0
      delta[i,1,m] <- 0
      mu[i,m] ~ dnorm(0, .0001)
    }
    for(k in 1:na[i]){
      r[i,k,1:4] ~ dmulti(p[i,k,1:4],n[i,k])
      p[i,k,4] <- 1 - sum(p[i,k,1:3])
      slam[i,k] <- sum(lamda[i,k,])
      for(m in 1:4){
        rhat[i,k,m] <- p[i,k,m]*n[i,k]
        dv[i,k,m] <- 2*r[i,k,m]*log(r[i,k,m]/rhat[i,k,m])
      }
      dev[i,k] <- sum(dv[i,k,])
      for(m in 1:3){
        # cumulative pr(failed) at each end point (per year), data in weeks
        p[i,k,m] <- lamda[i,k,m] * (1-exp(-slam[i,k]*r[i]/52))/slam[i,k]
      }
    }
  }
}

# *** PROGRAM STARTS
# LOOP THROUGH STUDIES
# LOOP OVER 3 ENDPOINTS
# adjustment for multi-arm trials is zero for control arm
# treatment effect is zero for control arm
# vague priors for all trial baselines

# LOOP THROUGH ARMS
# multinomial likelihood

# sum of the 3 hazard rates
# LOOP OVER ALL ENDPOINTS
# predicted number events
#Deviance contribution

# deviance contribution for arm
# LOOP THROUGH 3 ENDPOINTS
```

```

log(lamda[i,k,m]) <- mu[i,m] + delta[i,k,m] # model for linear predictor for each outcome
for (k in 2:na[i]) { # LOOP THROUGH ARMS
  delta[i,k,m] ~ dnorm(md[i,k,m],taud[i,k,m])
  md[i,k,m] <- d[t[i,k],m] - d[t[i,1],m] + sw[i,k,m] # mean of LHR distributions (with multi-arm correction)
  tau[i,k,m] <- tau[m] *2*(k-1)/k # precision of LHR distributions with multi-arm correction
  w[i,k,m] <- (delta[i,k,m] - d[t[i,k],m] + d[t[i,1],m]) # adjustment for multi-arm RCTs
  sw[i,k,m] <- sum(w[i,1:k-1,m])/(k-1) # cumulative adjustment for multi-arm trials
}
}
resdev[i] <- sum(dev[i,1:na[i]]) # summed residual deviance contribution for this trial
}
#Total Residual Deviance
totresdev <- sum(resdev[]) # LOOP THROUGH 3 END-POINTS
for (m in 1:3) { # treatment effect is zero for reference treatment
  d[1,m]<-0 # vague priors for treatment effects
  for (k in 2:nt){ d[k,m] ~ dnorm(0,.0001) } # vague prior for between-trial SD
  sd[m] ~ dunif(0,5) # between-trial precision = (1/between-trial variance)
  tau[m] <- pow(sd[m],-2)
}
} # *** PROGRAM ENDS

```

Additional code to monitor all treatment contrasts and rank treatments can be added as before. Given values for the mean for each outcome,  $\text{meanA} = c(-0.078, -1.723, -0.7185)$ , and precision,  $\text{precA} = c(1.6, 1.05, 0.61)$ , of the hazards for each endpoint on Treatment 1, from external sources, absolute effects, and absolute probabilities of the competing outcomes occurring within a given time period,  $\text{timeA}$ , say, a 1-month ( $1/12=0.083$  years) interval, could be monitored as follows:

```

for (k in 1:nt) {pslam[k] <- sum(T[k,]) } #LOOP THROUGH TREATMENTS, summing the 3 rates
for (m in 1:3) { # LOOP THROUGH 3 END-POINTS
  A[m] ~ dnorm(meanA[m],precA[m])
  for (k in 1:nt) { # LOOP THROUGH TREATMENTS
    log(T[k,m]) <- A[m] + d[k,m]
    cumpr[k,m] <- T[k,m] * (1-exp(-pslam[k]*timeA))/pslam[k] # cumulative pr(failed) at each end point
  }
}
}

```

The data structure again consists of a list specifying the number of treatments  $n_t$  and number of studies  $n_s$ , with the main body of data in a vector format;  $t_{ij}$  represents the follow-up time in that trial. Only two columns are required for each arm variable since there are no multi-arm trials:  $t_{i,1}$  then  $t_{i,2}$  the treatment codes; then  $r_{i,k,j}$  the number of events for the  $k$ -th treatment, outcome  $j$ ; then  $n_{i,1}$  and  $n_{i,2}$  the total number of individuals in each trial arm; and finally the number of arms in the study,  $na_i$ , and the study identifiers commented out. Both data components need to be loaded into WinBUGS for the program to run.

```

# Data (Schizophrenia example)
list(ns=15, nt=9)

t[i,1]   t[i,2]   r[i,1,1]  r[i,2,1]  r[i,1,2]  r[i,2,2]  r[i,1,3]  r[i,2,3]  r[i,1,4]  r[i,2,4]  n[i,1]  n[i,2]
na[i]    #study ID

```



Readers experimenting with this example need to be aware of difficulties with starting values. We have found one set of starting values which converges, but to a different posterior. Examples like this remind us of the importance of careful attention to the technical aspects of fitting models by Bayesian MCMC, and the need to look at the results obtained with different starting values. This is also an example where inverse gamma priors on the between trial variance leads to faster convergence, and avoids spikes in the posterior distributions.<sup>45</sup>

### Program 4(b): Multinomial likelihood (with competing risks), log link, Fixed Effects (Schizophrenia example)

```
# Multinomial likelihood, log link (competing risks)
# Fixed effects model for multi-arm trials
model{
  for(j in 1:ns){
    for(m in 1:3) {
      w[i,1,m] <- 0
      delta[i,1,m] <- 0
      mu[i,m] ~ dnorm(0, .0001)
    }
    for(k in 1:na[i]) {
      r[i,k,1:4] ~ dmulti(p[i,k,1:4], n[i,k])
      p[i,k,4] <- 1 - sum(p[i,k,1:3])
      slam[i,k] <- sum(lamda[i,k,])
      for(m in 1:4) {
        rhat[i,k,m] <- p[i,k,m]*n[i,k]
        dv[i,k,m] <- 2*r[i,k,m]*log(r[i,k,m]/rhat[i,k,m])
      }
      dev[i,k] <- sum(dev[i,k,])
      for(m in 1:3) {
        p[i,k,m] <- lamda[i,k,m] * (1-exp(-slam[i,k]*f[i]/52))/slam[i,k] # cumulative pr(failed) at each end point
        log(lamda[i,k,m]) <- mu[i,m] + d[t[i,k],m] - d[t[i,1],m] # model for linear predictor
      }
    }
    resdev[i] <- sum(dev[i,1:na[i]])
  }
  totesdev <- sum(resdev[])
  for(m in 1:3) {
    d[1,m] <- 0
    for(k in 2:nt){ d[k,m] ~ dnorm(0, .0001) }
  }
}

# Initial Values
#chain 1
list(d= structure(.Data= c(NA, NA, NA, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0), .Dim=c(9, 3)),
mu= structure(.Data= 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, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0), .Dim=c(15, 3)) )

#chain 2
list(d= structure(.Data= c(NA, NA, NA, 1,0,1, 0,0,1, 0,0,0, 0,0,1, 0,0,0, 0,0,0, 0,1,0, 0,0,1), .Dim=c(9, 3)),
mu= structure(.Data= 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, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0, 0,0,0), .Dim=c(15, 3)) )

#chain 3
list(d= structure(.Data= c(NA, NA, NA, -1,0,-1, 0,0,-1, 0,0,0, 0,0,-1, 0,0,0, 0,0,0, 0,-1,0, 0,0,-1), .Dim=c(9, 3)),
mu= structure(.Data= c(0,5,0, 0,5,0, 0,0,0, 0,0,5, 0,0,0, 0,0,5, 0,0,0, 0,0,5, 0,0,0, 0,5,0, 0,5,0, 0,0,0, 0,0,0, 3,0,3, 0,0,0), .Dim=c(15, 3)))
```

## RESULTS

Results (based on 3 chains: 100,000 iterations after a burn-in of 50,000 for the FE model and 100,000 iterations after a burn-in of 10,000 for the RE model) are presented in Table A7. Note that the follow-up time data are entered in weeks, while the analysis delivers annual rates. The model fit statistics suggest that the random effects model is a better fit to the data and the DIC shows that this model should be preferred. The log-hazard rates for each of the competing events and the absolute probabilities of the competing outcomes occurring within 1 month (assuming a baseline log-hazard as detailed above) are given in Table A7. For a graphical representation of which treatment is best for each of the competing outcomes and further comments see Ades et al.<sup>45</sup>

**Table A7 Schizophrenia example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effects of Olanzapine ( $d_{12}$ ), Amisulpride ( $d_{13}$ ), Zotepine ( $d_{14}$ ), Aripripazole ( $d_{15}$ ), Ziprasidone ( $d_{16}$ ), Paliperidone ( $d_{17}$ ), Haloperidol ( $d_{18}$ ) and Risperidone ( $d_{19}$ ) relative to Placebo, absolute probabilities of reaching each of the outcomes for Placebo ( $Pr_1$ ), Olanzapine ( $Pr_2$ ), Amisulpride ( $Pr_3$ ), Zotepine ( $Pr_4$ ), Aripripazole ( $Pr_5$ ), Ziprasidone ( $Pr_6$ ), Paliperidone ( $Pr_7$ ), Haloperidol ( $Pr_8$ ) and Risperidone ( $Pr_9$ ); heterogeneity parameter  $\tau$  for each of the three outcomes, and model fit statistics for the fixed and random effects models.**

	FE model				RE model			
	mean	sd	median	CrI	mean	sd	median	CrI
<b>Relapse</b>								
$d_{12}$	-1.57	0.23	-1.57	(-2.01,-1.13)	-1.46	0.48	-1.47	(-2.39,-0.50)
$d_{13}$	-1.15	0.48	-1.14	(-2.10,-0.22)	-0.97	0.80	-0.97	(-2.53,0.63)
$d_{14}$	-2.10	0.23	-0.52	(-0.96,-0.06)	-2.10	0.99	-2.08	(-4.11,-0.19)
$d_{15}$	-0.73	0.19	-0.72	(-1.10,-0.36)	-0.73	0.82	-0.73	(-2.37,0.92)
$d_{16}$	-1.12	0.20	-1.12	(-1.50,-0.73)	-1.23	0.65	-1.23	(-2.56,0.05)
$d_{17}$	-1.02	0.26	-1.01	(-1.53,-0.53)	-1.02	0.84	-1.02	(-2.70,0.66)
$d_{18}$	-0.90	0.26	-0.90	(-1.41,-0.38)	-0.69	0.64	-0.71	(-1.93,0.63)
$d_{19}$	-1.27	0.28	-1.27	(-1.82,-0.71)	-1.12	0.74	-1.13	(-2.59,0.37)
$Pr_1$	0.09	0.07	0.07	(0.02,0.29)	0.09	0.07	0.07	(0.02,0.29)
$Pr_2$	0.02	0.02	0.02	(0.00,0.08)	0.03	0.03	0.02	(0.00,0.10)
$Pr_3$	0.03	0.04	0.02	(0.00,0.14)	0.05	0.07	0.03	(0.00,0.23)
$Pr_4$	0.01	0.02	0.01	(0.00,0.06)	0.02	0.04	0.01	(0.00,0.10)
$Pr_5$	0.05	0.04	0.03	(0.01,0.16)	0.06	0.08	0.03	(0.00,0.27)
$Pr_6$	0.03	0.03	0.02	(0.00,0.11)	0.03	0.05	0.02	(0.00,0.15)
$Pr_7$	0.03	0.03	0.02	(0.00,0.12)	0.04	0.06	0.02	(0.00,0.20)
$Pr_8$	0.04	0.04	0.03	(0.01,0.14)	0.06	0.07	0.04	(0.00,0.24)
$Pr_9$	0.03	0.03	0.02	(0.00,0.10)	0.04	0.06	0.02	(0.00,0.19)
$\tau$					0.73	0.32	0.66	(0.30,1.53)

<b>Discontinuation due to side effects</b>								
$d_{12}$	-1.33	0.34	-1.33	(-2.01,-0.67)	-1.07	0.89	-1.14	(-2.67,0.97)
$d_{13}$	-1.67	0.68	-1.64	(-3.09,-0.41)	-1.75	1.39	-1.72	(-4.63,0.99)
$d_{14}$	1.15	0.59	1.12	(0.06,2.41)	1.14	1.56	1.12	(-2.05,4.34)
$d_{15}$	0.02	0.38	0.02	(-0.72,0.78)	0.02	1.49	0.02	(-3.10,3.12)
$d_{16}$	-1.05	0.34	-1.06	(-1.71,-0.35)	-0.98	1.16	-1.02	(-3.28,1.52)
$d_{17}$	1.31	1.43	1.16	(-1.10,4.58)	1.35	2.05	1.24	(-2.48,5.76)
$d_{18}$	-0.90	0.39	-0.90	(-1.67,-0.15)	-0.84	1.14	-0.86	(-3.16,1.56)
$d_{19}$	-1.34	0.42	-1.33	(-2.17,-0.52)	-0.71	1.40	-0.89	(-3.08,2.63)
$Pr_1$	0.02	0.02	0.01	(0.00,0.09)	0.02	0.03	0.01	(0.00,0.09)
$Pr_2$	0.01	0.01	0.00	(0.00,0.03)	0.01	0.04	0.00	(0.00,0.07)
$Pr_3$	0.01	0.01	0.00	(0.00,0.03)	0.01	0.05	0.00	(0.00,0.06)
$Pr_4$	0.08	0.10	0.04	(0.00,0.35)	0.12	0.19	0.04	(0.00,0.82)
$Pr_5$	0.02	0.03	0.01	(0.00,0.10)	0.05	0.12	0.01	(0.00,0.38)
$Pr_6$	0.01	0.01	0.00	(0.00,0.04)	0.02	0.06	0.01	(0.00,0.11)
$Pr_7$	0.13	0.20	0.04	(0.00,0.84)	0.16	0.26	0.05	(0.00,0.97)
$Pr_8$	0.01	0.01	0.01	(0.00,0.04)	0.02	0.06	0.01	(0.00,0.11)
$Pr_9$	0.01	0.01	0.00	(0.00,0.03)	0.03	0.10	0.01	(0.00,0.26)
$\tau$					1.20	0.80	1.03	(0.14,3.31)
<b>Discontinuation due to other reasons</b>								
$d_{12}$	-0.52	0.23	-0.52	(-0.96,-0.06)	-0.51	0.26	-0.52	(-0.99,0.03)
$d_{13}$	-0.60	0.25	-0.60	(-1.11,-0.11)	-0.59	0.33	-0.60	(-1.23,0.07)
$d_{14}$	-0.48	0.32	-0.48	(-1.10,0.15)	-0.48	0.40	-0.49	(-1.25,0.30)
$d_{15}$	0.24	0.38	0.23	(-0.50,1.00)	0.23	0.46	0.24	(-0.64,1.11)
$d_{16}$	-0.44	0.27	-0.44	(-0.96,0.10)	-0.44	0.33	-0.45	(-1.06,0.24)
$d_{17}$	0.74	0.47	0.73	(-0.13,1.70)	0.75	0.52	0.76	(-0.26,1.79)
$d_{18}$	-0.43	0.26	-0.44	(-0.94,0.08)	-0.43	0.32	-0.43	(-1.04,0.22)
$d_{19}$	-1.07	0.29	-1.07	(-1.64,-0.51)	-1.08	0.36	-1.10	(-1.76,-0.33)
$Pr_1$	0.07	0.10	0.04	(0.00,0.37)	0.07	0.10	0.04	(0.00,0.37)
$Pr_2$	0.05	0.08	0.02	(0.00,0.26)	0.05	0.08	0.02	(0.00,0.27)
$Pr_3$	0.05	0.07	0.02	(0.00,0.24)	0.05	0.07	0.02	(0.00,0.25)
$Pr_4$	0.05	0.08	0.02	(0.00,0.27)	0.05	0.08	0.02	(0.00,0.27)
$Pr_5$	0.10	0.13	0.05	(0.00,0.49)	0.09	0.13	0.05	(0.00,0.49)
$Pr_6$	0.05	0.08	0.03	(0.00,0.28)	0.05	0.08	0.02	(0.00,0.29)
$Pr_7$	0.14	0.17	0.07	(0.00,0.66)	0.13	0.17	0.07	(0.00,0.66)
$Pr_8$	0.05	0.08	0.03	(0.00,0.28)	0.05	0.08	0.02	(0.00,0.29)
$Pr_9$	0.03	0.05	0.01	(0.00,0.16)	0.03	0.05	0.01	(0.00,0.16)
$\tau$					0.17	0.16	0.13	(0.00,0.59)
$\bar{D}_{res}^*$	120.0				89.1			
pD	68.5				80.2			
DIC	188.5				169.3			

\*compare to 90 data points

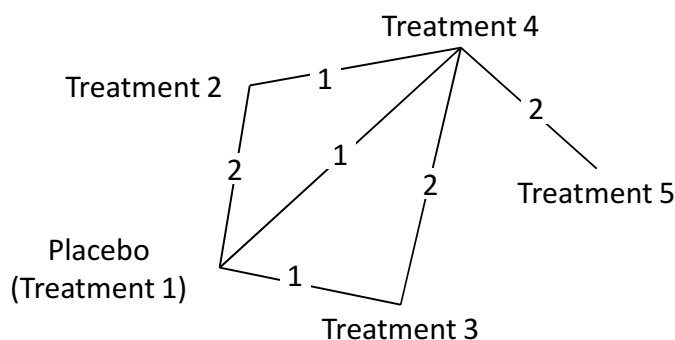
## EXAMPLE 5. PARKINSON'S

The data presented in Table A8 are the mean off-time reduction in patients given dopamine Agonists as adjunct therapy in Parkinson's disease. The data available are the mean, standard

deviation and number of patients in each trial arm, for 7 studies of five different drugs: placebo, coded 1, and five active drugs coded 2 to 5. The network diagram is presented in Figure A5. The authors are grateful to Jeroen Jansen of Mapi Values for giving us access to this data.

**Table A8 Parkinson’s example: study names, treatments compared, mean off-time reduction with its standard deviation, total number of patients in each trial arm; treatment differences and standard error of the differences; where treatment 1 is a placebo and treatments 2-5 are active drugs.**

Study	Treatment	y	sd	n	diff	se(diff)
1	1	-1.22	3.7	54		
	3	-1.53	4.28	95	-0.31	0.668
2	1	-0.7	3.7	172		
	2	-2.4	3.4	173	-1.7	0.383
3	1	-0.3	4.4	76		
	2	-2.6	4.3	71	-2.3	0.718
	4	-1.2	4.3	81	-0.9	0.695
4	3	-0.24	3	128		
	4	-0.59	3	72	-0.35	0.442
5	3	-0.73	3	80		
	4	-0.18	3	46	0.55	0.555
6	4	-2.2	2.31	137		
	5	-2.5	2.18	131	-0.3	0.274
7	4	-1.8	2.48	154		
	5	-2.1	2.99	143	-0.3	0.320



**Figure A5 Parkinson network: each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison and the numbers by the treatment names are the treatment codes used in the modelling.**

The model for this type of data is outlined in Section 3.4. The WinBUGS code for random effects is given in program 5(a) and the fixed effects code is given in program 5(b).

## Program 5(a): Normal likelihood, identity link, Random Effects (Parkinson's example)

```

# Normal likelihood, identity link
# Random effects model for multi-arm trials
model{
  for(i in 1:ns){
    w[i,1] <- 0
    delta[i,1] <- 0
    mu[i] ~ dnorm(0,.0001)
    for (k in 1:na[i]) {
      var[i,k] <- pow(se[i,k],2)
      prec[i,k] <- 1/var[i,k]
      y[i,k] ~ dnorm(theta[i,k],prec[i,k])
      theta[i,k] <- mu[i] + delta[i,k]
      dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k] #Deviance contribution
    }
    resdev[i] <- sum(dev[i,1:na[i]])
    for (k in 2:na[i]) {
      delta[i,k] ~ dnorm(md[i,k],taud[i,k])
      md[i,k] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]
      taud[i,k] <- tau *2*(k-1)/k
      w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
    }
  }
  totresdev <- sum(resdev[])
  d[1]<-0
  for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }
  sd ~ dunif(0,5)
  tau <- pow(sd,-2)
}
# *** PROGRAM ENDS
# *** PROGRAM STARTS
# LOOP THROUGH STUDIES
# adjustment for multi-arm trials is zero for control arm
# treatment effect is zero for control arm
# vague priors for all trial baselines
# LOOP THROUGH ARMS
# calculate variances
# set precisions
# normal likelihood
# model for linear predictor
# summed residual deviance contribution for this trial
# LOOP THROUGH ARMS
# trial-specific LOR distributions
# mean of treat effects distributions (with multi-arm trial correction)
# precision of treat effects distributions (with multi-arm trial correction)
# adjustment for multi-arm RCTs
# cumulative adjustment for multi-arm trials
#Total Residual Deviance
# treatment effect is zero for reference treatment
# vague priors for treatment effects
# vague prior for between-trial SD.
# between-trial precision = (1/between-trial variance)

```

Additional code to monitor all treatment contrasts and rank treatments can be added as before. Given values for the mean, -0.73, and precision, 21, of the outcome on Treatment 1, from external sources, absolute effects, and the absolute treatment effect, could be monitored by adding the following code:

```

A ~ dnorm(-.73,21)
for (k in 1:nt) { T[k] <- A + d[k] }

```

The maximum number of arms is 3, so 3 vectors are needed for the treatment indicators,  $t_{[,1]}$ ,  $t_{[,2]}$ ,  $t_{[,3]}$ ; the continuous outcomes  $y_{[,]}$ ; and their standard errors  $se_{[,]}$ ; and finally the number of arms,  $na_{[,]}$ .

```

# Data (Parkinson's example)
list(ns=7, nt=5)

```

$t_{[,1]}$	$t_{[,2]}$	$t_{[,3]}$	$y_{[,1]}$	$y_{[,2]}$	$y_{[,3]}$	$se_{[,1]}$	$se_{[,2]}$	$se_{[,3]}$	$na_{[,]}$
1	3	NA	-1.22	-1.53	NA	0.504	0.439	NA	2
1	2	NA	-0.7	-2.4	NA	0.282	0.258	NA	2
1	2	4	-0.3	-2.6	-1.2	0.505	0.510	0.478	3
3	4	NA	-0.24	-0.59	NA	0.265	0.354	NA	2
3	4	NA	-0.73	-0.18	NA	0.335	0.442	NA	2
4	5	NA	-2.2	-2.5	NA	0.197	0.190	NA	2
4	5	NA	-1.8	-2.1	NA	0.200	0.250	NA	2

END



```

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

```

### Program 5(b): Normal likelihood, identity link, Fixed Effects (Parkinson's example)

```

# Normal likelihood, identity link
# Fixed effects model for multi-arm trials
model{
  for(i in 1:ns){
    mu[i] ~ dnorm(0,.0001)
    for (k in 1:na[i]) {
      var[i,k] <- pow(se[i,k],2)
      prec[i,k] <- 1/var[i,k]
      y[i,k] ~ dnorm(theta[i,k],prec[i,k])
      theta[i,k] <- mu[i] + d[t[i,k]] - d[t[i,1]]
      dev[i,k] <- (y[i,k]-theta[i,k])*(y[i,k]-theta[i,k])*prec[i,k]
    }
    resdev[i] <- sum(dev[i,1:na[i]])
  }
  toresdev <- sum(resdev[])
  d[1]<-0
  for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }
}

# *** PROGRAM STARTS
# LOOP THROUGH STUDIES
# vague priors for all trial baselines
# LOOP THROUGH ARMS
# calculate variances
# set precisions
# normal likelihood
# model for linear predictor
#Deviance contribution
# summed residual deviance contribution for this trial

#Total Residual Deviance
# treatment effect is zero for reference treatment
# vague priors for treatment effects
# *** PROGRAM ENDS

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

```

## RESULTS

Results (based on 3 chains: 100,000 iterations after a burn-in of 50,000) are presented in Table A9. The random and fixed effects model both fit the data well, and since the random effects model has a higher DIC (due to having a higher effective number of parameters) the FE model should be preferred. The difference in mean of symptoms for each of the treatments compared to placebo and the absolute mean reduction in symptoms (assuming a baseline treatment effect as detailed above) are given in Table A9.

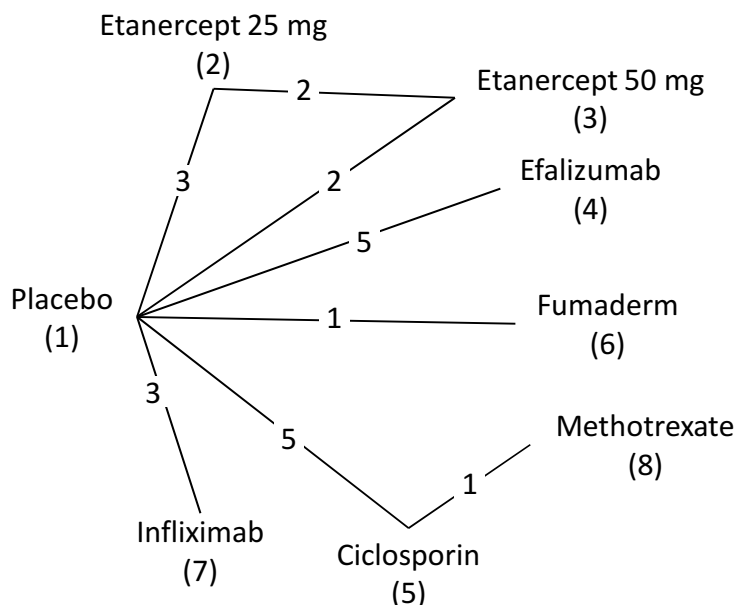
**Table A9 Parkinson example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for both the fixed and random effects models for the treatment effects of Treatments 2 to 5 ( $d_{12}$  to  $d_{15}$ ) relative to Placebo, absolute effects of Placebo ( $T_1$ ) and treatments 2 to 5 ( $T_2$  to  $T_5$ ), heterogeneity parameter  $\tau$  and model fit statistics for different data types.**

	FE model				RE model			
	mean	sd	median	CrI	mean	sd	median	CrI
<b>Arm-level data: Example 5</b>								
$d_{12}$	-1.81	0.33	-1.81	(-2.46,-1.16)	-1.85	0.54	-1.84	(-2.91,-0.85)
$d_{13}$	-0.47	0.49	-0.47	(-1.43,0.49)	-0.50	0.66	-0.50	(-1.78,0.75)
$d_{14}$	-0.52	0.48	-0.52	(-1.46,0.43)	-0.53	0.65	-0.53	(-1.77,0.71)
$d_{15}$	-0.82	0.52	-0.82	(-1.84,0.22)	-0.83	0.80	-0.83	(-2.35,0.69)
$T_1$	-0.73	0.22	-0.73	(-1.16,-0.30)	-0.73	0.22	-0.73	(-1.16,-0.30)
$T_2$	-2.54	0.40	-2.54	(-3.32,-1.76)	-2.58	0.58	-2.57	(-3.72,-1.50)
$T_3$	-1.21	0.53	-1.20	(-2.25,-0.15)	-1.23	0.70	-1.23	(-2.57,0.10)
$T_4$	-1.25	0.53	-1.25	(-2.28,-0.21)	-1.26	0.69	-1.26	(-2.57,0.05)
$T_5$	-1.55	0.57	-1.55	(-2.66,-0.43)	-1.57	0.83	-1.56	(-3.14,0.02)
$\tau$	-	-	-	-	0.40	0.43	0.28	(0.01,1.55)
$\bar{D}_{res}^*$	13.3				13.6			
pD	11.0				12.4			
DIC	24.3				26.0			
<b>Trial-level data (differences): Example 7</b>								
$d_{12}$	-1.81	0.33	-1.81	(-2.47,-1.16)	-1.85	0.54	-1.84	(-2.92,-0.84)
$d_{13}$	-0.48	0.49	-0.48	(-1.43,0.47)	-0.50	0.66	-0.49	(-1.79,0.75)
$d_{14}$	-0.52	0.48	-0.52	(-1.46,0.42)	-0.53	0.65	-0.53	(-1.79,0.72)
$d_{15}$	-0.82	0.52	-0.82	(-1.84,0.20)	-0.83	0.81	-0.83	(-2.38,0.69)
$T_1$	-0.73	0.22	-0.73	(-1.16,-0.30)	-0.73	0.22	-0.73	(-1.16,-0.30)
$T_2$	-2.54	0.40	-2.54	(-3.33,-1.76)	-2.58	0.58	-2.57	(-3.72,-1.49)
$T_3$	-1.21	0.53	-1.21	(-2.25,-0.17)	-1.23	0.70	-1.22	(-2.59,0.10)
$T_4$	-1.25	0.53	-1.25	(-2.28,-0.22)	-1.26	0.69	-1.26	(-2.59,0.06)
$T_5$	-1.55	0.56	-1.55	(-2.66,-0.44)	-1.56	0.83	-1.56	(-3.17,0.01)
$\tau$	-	-	-	-	0.41	0.44	0.28	(0.01,1.56)
$\bar{D}_{res}^\dagger$	6.3				6.6			
pD	4.0				5.5			
DIC	10.3				12.1			
<b>Arm and Trial-level data (shared parameter model): Example 8</b>								
$d_{12}$	-1.81	0.33	-1.81	(-2.46,-1.16)	-1.85	0.54	-1.83	(-2.91,-0.86)
$d_{13}$	-0.48	0.49	-0.48	(-1.43,0.48)	-0.51	0.66	-0.50	(-1.79,0.75)
$d_{14}$	-0.52	0.48	-0.52	(-1.47,0.41)	-0.54	0.65	-0.54	(-1.78,0.70)
$d_{15}$	-0.82	0.52	-0.82	(-1.85,0.20)	-0.84	0.80	-0.84	(-2.35,0.69)
$T_1$	-0.73	0.22	-0.73	(-1.16,-0.30)	-0.73	0.22	-0.73	(-1.16,-0.30)
$T_2$	-2.54	0.40	-2.54	(-3.32,-1.77)	-2.57	0.58	-2.57	(-3.71,-1.49)
$T_3$	-1.21	0.53	-1.21	(-2.25,-0.17)	-1.23	0.70	-1.23	(-2.58,0.09)
$T_4$	-1.26	0.53	-1.25	(-2.29,-0.23)	-1.27	0.68	-1.27	(-2.58,0.05)
$T_5$	-1.56	0.57	-1.56	(-2.67,-0.45)	-1.57	0.83	-1.57	(-3.14,0.02)
$\tau$	-	-	-	-	0.40	0.43	0.28	(0.01,1.53)
$\bar{D}_{res}^\ddagger$	9.3				9.6			
pD	7.0				8.5			
DIC	16.3				18.1			

\* compare to 15 data points; † compare to 8 data points; ‡ compare to 11 data points

## EXAMPLE 6. PSORIASIS

In an HTA report to evaluate the effectiveness of treatments for moderate to severe chronic plaque psoriasis,<sup>52</sup> 16 trials, comparing 8 treatments were identified: Supportive care (coded 1); Etanercept 25 mg (2); Etanercept 50 mg (3); Efalizumab (4); Ciclosporin (5); Fumaderm (6); Infliximab (7) and Methotrexate (8). The network diagram is presented in Figure A6. Each trial reported the number of patients in mutually exclusive categories representing the percentage improvement in symptoms as measured by the PASI score. Different trials reported on different categories defining 3 cut-points, 50, 75 and 90% improvement, in addition to the scale's lower and upper bounds (0 and 100% improvement, respectively). In the code below, we define: C=1 representing 0% improvement (the scale's lower bound); C=2 representing 50% improvement; C=3 representing 75% improvement; and C=4 representing 90% improvement. The data is presented in Table A10.



**Figure A6 Psoriasis network: each edge represents a treatment, connecting lines indicate pairs of treatments which have been directly compared in randomised trials. The numbers on the lines indicate the numbers of trials making that comparison and the numbers by the treatment names are the treatment codes used in the modelling. One trial compared two arms of Ciclosporin with Placebo and another compared two arms of Infliximab with placebo – these comparisons are not represented in the network.**

The model for this type of data is outlined in Section 3.6. The likelihood contribution of each trial is multinomial and this can be used to model the data directly in WinBUGS. However, since the reported categories are different in different studies and overlap, it is helpful to re-write the multinomial likelihood as a series of conditional Binomials.

So, for a trial  $i$  reporting the number of patients  $r_{ikj}$  in category  $j=1, \dots, J_i-1$ , we can write

$$r_{ikj} \sim \text{Binomial}(q_{ikj}, N_{ikj}) \quad j=1, \dots, J_i-1$$

where

$$q_{ik1} = \Pr(\text{PASI score in category 1 of trial } i)$$

$$q_{ik2} = \Pr(\text{PASI score in category 2 of trial } i \mid \text{not in category 1})$$

...

$$q_{ikj} = \Pr(\text{PASI score in category } j \text{ of trial } i \mid \text{not in categories } 1, 2, \dots, j-1)$$

and  $N_{ikj} = n_{ik} - \sum_{u=1}^{j-1} r_{iku}$ .

Noting that the lower and upper bounds of each mutually exclusive category are defined by the cut-points above and the scale's lower and upper bounds, for arm  $k$  of trial  $i$  we can define  $q_{ikj}$  as the probability of belonging to category  $j$ , in arm  $k$  of trial  $i$ ,

$$q_{ikj} = \Pr(\text{PASI score in category } j) = \Pr(L_j < \text{PASI score} < U_j)$$

where  $L_j$  and  $U_j$  define the lower and upper bounds of the interval defining category  $j$ . So, for example, for arm 1 of study 1 in Table A10 category 1 is 0-50% improvement so

$$q_{111} = \Pr(\text{having less than 50\% improvement in PASI score}) = \Pr(0 < \text{PASI score} < 50).$$

Letting  $p_{ikc}$  denote the probability of achieving a PASI score of at least  $c$ , in arm  $k$  of trial  $i$ , for  $c=50, 75, 90$  we model

$$\pi_{ikc} = \Pr(\text{PASI score} > c) = \Phi(\theta_{ik} + z_c)$$

where  $\theta_{ik}$  is the linear predictor and  $\Phi$  is the standard normal cumulative distribution function. Cut- points  $z_{50}$ ,  $z_{75}$  and  $z_{90}$  have been coded  $z_1$ ,  $z_2$  and  $z_3$ , respectively, in the code below. We set  $z_1=0$  and give non-informative priors to  $z_2$  and  $z_3$ .

The terms  $z_c$  can be thought of as the distance on the standard normal scale between the category boundaries. The ‘‘fixed effect’’ model above assumes that these distances are the same in every trial and for every treatment. An alternative might be that they differ between trials, but that within a trial the distances between categories are the same. This leads us to a ‘‘random effects’’ model in which for each trial  $i$ ,  $z_{ic}$  varies around a mean

$$z_{ic} \sim N(\nu_c, \sigma_z^2).$$

The mean and variance are then given vague priors in the usual way. One interpretation of this model, which can be used with a Fixed or Random treatment effects, is that there may be

differences between trials in the way that the underlying symptoms are scored, in this case on the PASI scale. It can be shown that, for  $j=1, \dots, J_{i-1}$

$$q_{ikj} = 1 - \frac{\Pr(\text{PASI} > C_{j+1})}{\Pr(\text{PASI} > C_j)} = 1 - \frac{\pi_{ikC_{j+1}}}{\pi_{ikC_j}}$$

with  $C_j$  and  $C_{j+1}$  representing the lower and upper bounds of the interval defining category  $j$ , respectively. Using these relationships simplifies the code and makes it general for any number of categories and cut-off points.

Table A10 Psoriasis example: study names, treatments compared, total number of patients with different percentage improvement and total number of patients in each trial arm, where Supportive Care = treatment 1, Etanercept 25mg = 2, Etanercept 50 mg = 3, Efalizumab = 4, Ciclosporin = 5, Fumaderm = 6, Infliximab = 7, Methotrexate = 8.<sup>52</sup>

	arm 1 arm 2 arm 3			Outcomes presented in arm 1 for each category	total number of patients in arm 1	Outcomes presented in arm 2 for each category	total number of patients in arm 2	Outcomes presented in arm 3 for each category	total number of patients in arm 3									
<b>Trials presenting outcomes in 4 categories</b>																		
Trial	t[,1]	t[,2]	t[,3]	0-50	50-75	75-90	90-100	n[,1]	0-50	50-75	75-90	90-100	n[,2]	0-50	50-75	75-90	90-100	n[,3]
1.Elewski 2004	1	2	3	175	12	5	1	193	70	59	46	21	196	44	54	56	40	194
2.Gottlieb 2003	1	2		49	5	1	0	55	17	23	11	6	57					
3.Lebwohl 2003	1	4		103	13	5	1	122	112	68	42	10	232					
4.Leonardi 2003	1	2	3	142	18	5	1	166	68	39	36	19	162	43	40	45	36	164
<b>Trials presenting outcomes in 3 categories</b>																		
Trial	t[,1]	t[,2]	t[,3]	0-50	50-75	75-100	n[,1]	0-50	50-75	75-100	n[,2]	0-50	50-75	75-100	n[,3]			
5.Gordon 2003	1	4		161	18	8	187	153	118	98	369							
<b>Trials presenting outcomes in 2 categories (PASI 50)</b>																		
Trial	t[,1]	t[,2]	t[,3]	0-50	50-100	n[,1]	0-50	50-100	n[,2]	0-50	50-100	n[,3]						
6.ACD2058g	1	4		145	25	170	63	99	162									
7.ACD2600g	1	4		230	33	263	216	234	450									
8.Guenther 1991	1	5		10	1	11	0	12	12									
9.IMP24011	1	4		226	38	264	245	284	529									
<b>Trials presenting outcomes in 2 categories (PASI 75)</b>																		
Trial	t[,1]	t[,2]	t[,3]	0-75	75-100	n[,1]	0-75	75-100	n[,2]	0-75	75-100	n[,3]						
10.Altmeyer 1994	1	6		50	1	51	37	12	49									
11.Chaudari 2001	1	7		9	2	11	2	9	11									
12.Ellis 1991	1	5	5	25	0	25	16	9	25	7	13	20						
13.Gottlieb 2004	1	7	7	48	3	51	28	71	99	12	87	99						
14.Heydendael 2003	5	8		12	30	42	17	26	43									
15.Meffert 1997	1	5		41	2	43	37	4	41									
16.Van Joost 1988	1	5		10	0	10	3	7	10									

The WinBUGS code for random effects is given in program 6(a) and the fixed effects code is given in program 6(b).

### Program 6(a): Conditional Binomial likelihood, probit link, Random Effects (Psoriasis example)

```
# Binomial likelihood, probit link (different categories)
# Random effects model for multi-arm trials
model{
for(i in 1:ns){
  w[i,1] <- 0
  delta[i,1] <- 0
  mu[i] ~ dnorm(0,.0001)
  for(k in 1:na[i]) {
    p[i,k,1] <- 1
    for(j in 1:nc[i]-1) {
      r[i,k,j] ~ dbin(q[i,k,j],n[i,k,j])
      q[i,k,j] <- 1-(p[i,k,C[i,j+1]]/p[i,k,C[i,j]])
      theta[i,k,j] <- mu[i] + delta[i,k] + z[j]
      rhat[i,k,j] <- q[i,k,j] * n[i,k,j]
      dv[i,k,j] <- 2 * (r[i,k,j]*(log(r[i,k,j])-log(rhat[i,k,j]))
      +(n[i,k,j]-r[i,k,j])*(log(n[i,k,j]-r[i,k,j]) - log(n[i,k,j]-rhat[i,k,j])))
    }
    dev[i,k] <- sum(dv[i,k,1:nc[i]-1])
    for(j in 2:nc[i]) {
      p[i,k,C[i,j]] <- 1 - phi.adj[i,k,j]
    }
  }
  # adjust link function phi(x) for extreme values that can give numerical errors
  # when x < -5, phi(x)=0, when x > 5, phi(x)=1
  phi.adj[i,k,j] <- step(5+theta[i,k,j-1])
  * (step(theta[i,k,j-1]-5)
  + step(5-theta[i,k,j-1])*phi(theta[i,k,j-1]) )
}
}
for(k in 2:na[i]) {
  delta[i,k] ~ dnorm(md[i,k],taud[i,k])
  md[i,k] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]
  taud[i,k] <- tau *2*(k-1)/k
  w[i,k] <- (delta[i,k] - d[t[i,k]] + d[t[i,1]])
  sw[i,k] <- sum(w[i,1:k-1])/(k-1)
}
resdev[i] <- sum(dev[i,1:na[i]])
}
}
z[1] <- 0
for(j in 2:Cmax-1) {
  z.aux[j] ~ dunif(0,5)
  z[j] <- z[j-1] + z.aux[j]
}
totresdev <- sum(resdev[])
d[1] <- 0
for(k in 2:nt){ d[k] ~ dnorm(0,.0001) }
sd ~ dunif(0,5)
tau <- pow(sd,-2)
}
# *** PROGRAM STARTS
# LOOP THROUGH STUDIES
# adjustment for multi-arm trials is zero for control arm
# treatment effect is zero for control arm
# vague priors for all trial baselines
# LOOP THROUGH ARMS
# Pr(PASI >0)
# LOOP THROUGH CATEGORIES
# binomial likelihood
# conditional probabilities
# linear predictor
# predicted number events
#Deviance contribution of each category
# deviance contribution of each arm
# LOOP THROUGH CATEGORIES
# link function
# Total Residual Deviance
# treatment effect is zero for reference treatment
# vague priors for treatment effects
# vague prior for between-trial SD
# between-trial precision = (1/between-trial variance)
# *** PROGRAM ENDS
```

Additional code to monitor all treatment contrasts and rank treatments can be added as before. Given values for the mean, 1.097, and precision, 123, of the effects on Treatment 1 on the probit scale, from external sources, absolute effects, and absolute probabilities  $T[j,k]$  of

having over 50, 75 or 90% improvement ( $j=1,2,3$  respectively) on treatment  $k$ , could be monitored as follows:

```
A ~ dnorm(1.097,123)
# calculate prob of achieving PASI50,75,90 on treat k
for (k in 1:nt) {
  for (j in 1:Cmax-1) { T[j,k] <- 1 - phi(A + d[k] + z[j]) }
}
```

The data structure again consists of a list specifying the number of treatments  $nt$ , number of studies  $ns$  and total number of categories  $c_{max}$ , with the main body of data in a vector format. Both data components need to be loaded into WinBUGS for the program to run. Three columns are required for each arm variable since there are four three-arm trials:  $t_{[,1]}$ ,  $t_{[,2]}$  and  $t_{[,3]}$  are the treatment codes;  $na[]$  represents the number of arms and  $nc[]$  the number of cut-offs in that trial;  $c_{[,1]}$ ,  $c_{[,2]}$ ,  $c_{[,3]}$ ,  $c_{[,4]}$  represent the cut-offs used to define the categories reported in each trial – four columns are needed as the maximum number of cut-offs given in a trial is four – these cut-offs are coded 1 to 4 as described above; then  $r_{[,k,j]}$  the number of events for the  $k$ -th treatment, in category  $j$  given the number of events in categories 1 to  $j-1$ ; then  $n_{[,k,1]}$  represents the total number of individuals in trial arm  $k$ ,  $n_{[,k,2]}$  represents the total number of individuals in trial arm  $k$ , which were not in category 1 of that trial and  $n_{[,k,3]}$  represents the total number of individuals in trial arm  $k$ , which were not in categories 1 or 2 of that trial. So, for example, for the first trial in Table A10, Elewski 2004, 193 patients were included in arm 1 of the trial. These patients were split between the four categories as follows: 175 out of 193 patients had between 0 and 50% improvement, leaving 18 patients who could belong to any of the other categories, thus  $r_{111}=173$ ,  $n_{111}=193$ ; 12 out of 18 patients had between 50 and 75% improvement, leaving 6 patients who could belong to any of the other categories, thus  $r_{112}=12$ ,  $n_{112}=193-175=18$ ; 5 out of 6 patients had between 75 and 90% improvement, leaving 1 patient (who necessarily had over 90% improvement), thus  $r_{113}=5$ ,  $n_{112}=18-12=6$ . All other  $n_{ikj}$  were similarly calculated for the other trials.

Note that for trials reporting  $j$  categories,  $r$  and  $n$  only need to be defined for the first  $j-1$  categories with the remaining columns coded NA, and some of the data given below is redundant.

```
# Data (Psoriasis example)
list(ns=16, nt=8, Cmax=4)
```



t[,1]	t[,2]	t[,3]	na[]	nc[]	C[,1]	C[,2]	C[,3]	C[,4]	r[,1,1]	r[,1,2]	r[,1,3]	
	n[,1,1]	n[,1,2]	n[,1,3]	r[,2,1]	r[,2,2]	r[,2,3]	n[,2,1]	n[,2,2]	n[,2,3]	r[,3,1]	r[,3,2]	
	r[,3,3]	n[,3,1]	n[,3,2]	n[,3,3]	#	Trial	Year					
1	2	3	3	4	1	2	3	4	175	12	5	193
	18	6	70	59	46	196	126	67	44	54	56	194
	150	96	#	Elewski	2004							
1	2	NA	2	4	1	2	3	4	49	5	1	55
	6	1	17	23	11	57	40	17	NA	NA	NA	NA
	NA	NA	#	Gottlieb	2003							
1	4	NA	2	4	1	2	3	4	103	13	5	122
	19	6	112	68	42	232	120	52	NA	NA	NA	NA
	NA	NA	#	Lebwohl	2003							
1	2	3	3	4	1	2	3	4	142	18	5	166
	24	6	68	39	36	162	94	55	43	40	45	164
	121	81	#	Leonardi	2003							
1	4	NA	2	3	1	2	3	NA	161	18	8	187
	26	8	153	118	98	369	216	98	NA	NA	NA	NA
	NA	NA	#	Gordon	2003							
1	4	NA	2	2	1	2	NA	NA	145	25	NA	170
	25	NA	63	99	NA	162	99	NA	NA	NA	NA	NA
	NA	NA	#	ACD2058g	2004							
1	4	NA	2	2	1	2	NA	NA	230	33	NA	263
	33	NA	216	234	NA	450	234	NA	NA	NA	NA	NA
	NA	NA	#	ACD2600g	2004							
1	5	NA	2	2	1	2	NA	NA	10	1	NA	11
	1	NA	0	12	NA	12	12	NA	NA	NA	NA	NA
	NA	NA	#	Guenther	1991							
1	4	NA	2	2	1	2	NA	NA	226	38	NA	264
	38	NA	245	284	NA	529	284	NA	NA	NA	NA	NA
	NA	NA	#	IMP24011	2004							
1	6	NA	2	2	1	3	NA	NA	50	1	NA	51
	1	NA	37	12	NA	49	12	NA	NA	NA	NA	NA
	NA	NA	#	Altmeyer	1994							
1	7	NA	2	2	1	3	NA	NA	9	2	NA	11
	2	NA	2	9	NA	11	9	NA	NA	NA	NA	NA
	NA	NA	#	Chaudari	2001							
1	5	5	3	2	1	3	NA	NA	25	0	NA	25
	0	NA	16	9	NA	25	9	NA	7	13	NA	20
	13	0	#	Ellis	1991							
1	7	7	3	2	1	3	NA	NA	48	3	NA	51
	3	NA	28	71	NA	99	71	NA	12	87	NA	99
	87	0	#	Gottlieb	2004							
5	8	NA	2	2	1	3	NA	NA	12	30	NA	42
	30	NA	17	26	NA	43	26	NA	NA	NA	NA	NA
	NA	NA	#	Heydendael	2003							
1	5	NA	2	2	1	3	NA	NA	41	2	NA	43
	2	NA	37	4	NA	41	4	NA	NA	NA	NA	NA
	NA	NA	#	Meffert	1997							
1	5	NA	2	2	1	3	NA	NA	10	0	NA	10
	0	NA	3	7	NA	10	7	NA	NA	NA	NA	NA
	NA	NA	#	Van Joost	1988							

END

# Initial Values

# Initial values for delta can be generated by WinBUGS.

#chain 1

list(d = c(NA,0,0,0,0,0,0,0), mu = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0), sd = 1, z.aux=c(NA, 0.66,1.3))

#chain 2

list(d = c(NA,0,0,1,0,0,2,0), mu = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0), sd = 1.5, z.aux=c(NA, 0.5,1))

#chain 3

list(d = c(NA,1,1,1,1,1,1,1), mu = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0), sd = 3, z.aux=c(NA, 0.1,2))

The parameters to monitor are the same as in Example 1. We may in addition want to monitor node  $z$  to obtain the posterior summaries for the different cut-off points.

### Program 6(b): Conditional Binomial likelihood, probit link, Fixed Effects (Psoriasis example)

```
# Binomial likelihood, probit link (different categories)
# Fixed effects model for multi-arm trials
model{
  for(i in 1:ns){
    mu[i] ~ dnorm(0, .0001)
    for(k in 1:na[i]) {
      p[i,k,1] <- 1
      for(j in 1:nc[i]-1) {
        r[i,k,j] ~ dbin(q[i,k,j], n[i,k,j])
        q[i,k,j] <- 1 - (p[i,k,C[i,j]+1])/p[i,k,C[i,j]]
        theta[i,k,j] <- mu[i] + d[t[i,k,j]] - d[t[i,1]] + z[j]
        rhat[i,k,j] <- q[i,k,j] * n[i,k,j]
        dv[i,k,j] <- 2 * (r[i,k,j]*(log(r[i,k,j])-log(rhat[i,k,j]))
          + (n[i,k,j]-r[i,k,j])*(log(n[i,k,j]-r[i,k,j]) - log(n[i,k,j]-rhat[i,k,j])))
        }
        dev[i,k] <- sum(dv[i,k,1:nc[i]-1])
        for(j in 2:nc[i]) {
          p[i,k,C[i,j]] <- 1 - phi.adj[i,k,j]
        }
      }
      resdev[i] <- sum(dev[i,1:na[i]])
    }
  }
  z[1] <- 0
  for(j in 2:Cmax-1) {
    z.aux[j] ~ dunif(0,5)
    z[j] <- z[j-1] + z.aux[j]
  }
  totesdev <- sum(resdev[])
  d[1] <- 0
  for(k in 2:nt){ d[k] ~ dnorm(0, .0001) }
}

# Initial Values
#chain 1
list(d = c(NA,0,0,0,0,0,0,0), mu = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0), z.aux=c(NA, 0.66,1.3))

#chain 2
list(d = c(NA,0,0,1,0,0,2,0), mu = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0), z.aux=c(NA, 0.5,1) )

#chain 3
list(d = c(NA,1,1,1,1,1,1,1), mu = c(0,0,0,0,0, 0,0,0,0,0, 0,0,0,0,0, 0), z.aux=c(NA, 0.1,2))
```

## RESULTS

Results for the fixed and random effects models are presented in Table A11 (results based on 3 chains: 100,000 iterations after a burn-in of 40,000 and 50,000 for the FE and RE models, respectively). From the residual deviance and DIC we conclude that the random effects

model should be preferred as it is a better fit to the data and has a smaller DIC. The treatment effects relative to Supportive care (treatment 1) are all below zero which suggests that all treatments are better than Supportive care at increasing the probability of a reduction in symptoms on the probit scale. The absolute probabilities of achieving a reduction on at least 50, 75 or 90% in symptoms show that, for example, there is on average 0% probability of achieving at least a 90% reduction in symptoms with Supportive care, but this probability is on average 37% with Infliximab. A model which assumes the cut-points differ between trials and come from a common distribution was also fitted and gave very similar results.

**Table A11 Psoriasis example: posterior mean, standard deviation (sd), median and 95% Credible interval (CrI) for the fixed and random effects models for the treatment effects, on the probit scale, of Etanercept 25 mg ( $d_{12}$ ), Etanercept 50 mg ( $d_{13}$ ), Efalizumab ( $d_{14}$ ), Ciclosporin ( $d_{15}$ ), Fumaderm ( $d_{16}$ ), Infliximab ( $d_{17}$ ), and Methotrexate ( $d_{18}$ ) relative to Supportive Care; absolute probabilities of achieving at least 50, 70 or 90% relief in symptoms for each treatment; heterogeneity parameter  $\tau$  and model fit statistics.**

	FE model				RE model			
	mean	sd	median	CrI	mean	sd	median	CrI
$d_{12}$	-1.51	0.10	-1.51	(-1.70,-1.33)	-1.53	0.25	-1.52	(-2.06,-1.02)
$d_{13}$	-1.92	0.10	-1.92	(-2.12,-1.72)	-1.93	0.29	-1.93	(-2.53,-1.33)
$d_{14}$	-1.19	0.06	-1.19	(-1.30,-1.08)	-1.19	0.19	-1.19	(-1.58,-0.80)
$d_{15}$	-1.92	0.34	-1.90	(-2.64,-1.28)	-2.05	0.45	-2.01	(-3.07,-1.29)
$d_{16}$	-1.48	0.48	-1.45	(-2.53,-0.63)	-1.50	0.64	-1.46	(-2.84,-0.32)
$d_{17}$	-2.33	0.27	-2.33	(-2.89,-1.83)	-2.31	0.39	-2.31	(-3.08,-1.53)
$d_{18}$	-1.61	0.45	-1.60	(-2.52,-0.75)	-1.75	0.66	-1.70	(-3.22,-0.58)
$\tau$	-	-	-	-	0.32	0.23	0.27	(0.03,0.90)
<b>Probability of achieving at least 50% relief in symptoms (PASI50)</b>								
Supportive Care	0.14	0.02	0.14	(0.10,0.18)	0.14	0.02	0.14	(0.10,0.18)
Etanercept 25 mg	0.66	0.05	0.66	(0.56,0.75)	0.66	0.09	0.67	(0.46,0.84)
Etanercept 50 mg	0.79	0.04	0.79	(0.71,0.86)	0.79	0.08	0.80	(0.59,0.93)
Efalizumab	0.54	0.04	0.54	(0.45,0.62)	0.54	0.08	0.54	(0.37,0.70)
Ciclosporin	0.78	0.10	0.79	(0.56,0.94)	0.81	0.11	0.82	(0.57,0.98)
Fumaderm	0.63	0.16	0.64	(0.31,0.93)	0.63	0.20	0.64	(0.21,0.96)
Infliximab	0.88	0.05	0.89	(0.76,0.97)	0.87	0.08	0.89	(0.66,0.98)
Methotrexate	0.68	0.15	0.69	(0.36,0.92)	0.70	0.18	0.73	(0.30,0.98)
<b>Probability of achieving at least 75% relief in symptoms (PASI75)</b>								
Supportive Care	0.03	0.01	0.03	(0.02,0.05)	0.03	0.01	0.03	(0.02,0.05)
Etanercept 25 mg	0.37	0.05	0.37	(0.27,0.47)	0.38	0.10	0.37	(0.20,0.59)
Etanercept 50 mg	0.53	0.05	0.53	(0.42,0.63)	0.53	0.11	0.53	(0.30,0.76)
Efalizumab	0.25	0.03	0.25	(0.19,0.33)	0.26	0.07	0.25	(0.14,0.41)
Ciclosporin	0.52	0.13	0.52	(0.27,0.79)	0.57	0.16	0.56	(0.28,0.89)
Fumaderm	0.37	0.17	0.34	(0.11,0.76)	0.38	0.21	0.35	(0.06,0.84)
Infliximab	0.68	0.10	0.68	(0.48,0.86)	0.67	0.13	0.68	(0.37,0.89)
Methotrexate	0.41	0.16	0.40	(0.13,0.75)	0.46	0.21	0.44	(0.10,0.92)

Probability of achieving at least 90% relief in symptoms (PASI90)								
Supportive Care	0.00	0.00	0.00	(0.00,0.01)	0.00	0.00	0.00	(0.00,0.01)
Etanercept 25 mg	0.13	0.03	0.13	(0.08,0.19)	0.14	0.06	0.13	(0.05,0.28)
Etanercept 50 mg	0.23	0.04	0.23	(0.16,0.32)	0.24	0.09	0.23	(0.09,0.46)
Efalizumab	0.07	0.02	0.07	(0.04,0.11)	0.08	0.03	0.07	(0.03,0.15)
Ciclosporin	0.24	0.11	0.22	(0.08,0.50)	0.29	0.15	0.26	(0.08,0.66)
Fumaderm	0.15	0.11	0.11	(0.02,0.46)	0.16	0.15	0.12	(0.01,0.58)
Infliximab	0.38	0.11	0.37	(0.19,0.60)	0.37	0.14	0.36	(0.13,0.67)
Methotrexate	0.17	0.11	0.15	(0.03,0.45)	0.22	0.18	0.17	(0.02,0.72)
$\bar{D}_{res}$ *	74.9				62.5			
pD	25.0				33.5			
DIC	99.9				95.9			

\* compare with 58 data points

## EXAMPLE 7. PARKINSON'S DIFFERENCE (TREATMENT DIFFERENCES AS DATA)

We now assume that the data available for the Parkinson's example were not the mean off-time reduction for patients in each arm of the trial, but rather the differences in off-time reduction, and their standard errors, between the intervention and control arms for each trial, as presented in the last two columns of Table A8. The data available are therefore the differences, their standard errors and the treatments compared in each trial, coded as before.

Random and fixed effects models were fitted, as outlined in Section 3.4. The code is given below. As explained in Section 5.1 the coding for the likelihood has been modified to allow for the 3-arm trial. This requires users to set up the data file with all two-arm trials first, then 3-arm trials, then – if any were present – 4-arm trials, and so on.

### Program 7(a): Normal likelihood, identity link, treatment differences, Random Effects (Parkinson's Differences)

```
# Normal likelihood, identity link, trial-level data given as treatment differences
# Random effects model for multi-arm trials
model{
  for(i in 1:ns2) {
    y[i,2] ~ dnorm(delta[i,2],prec[i,2])
    resdev[i] <- (y[i,2]-delta[i,2])*(y[i,2]-delta[i,2])*prec[i,2] #Deviance contribution for trial i
  }
  for(i in (ns2+1):(ns2+ns3)) {
    for(k in 1:(na[i]-1)) {
      for(j in 1:(na[i]-1)) { Sigma[i,j,k] <- V[i]*(1>equals(j,k)) + var[i,k+1]*equals(j,k) }
    }
    Omega[i,1:(na[i]-1),1:(na[i]-1)] <- inverse(Sigma[i,,]) #Precision matrix
  }
  # multivariate normal likelihood for 3-arm trials
  y[i,2:na[i]] ~ dnmnorm(delta[i,2:na[i]],Omega[i,1:(na[i]-1),1:(na[i]-1)])
  #Deviance contribution for trial i
  for(k in 1:(na[i]-1)){
    ydiff[i,k] <- y[i,(k+1)] - delta[i,(k+1)]
    z[i,k] <- inprod2(Omega[i,k,1:(na[i]-1)], ydiff[i,1:(na[i]-1)])
  }
}
```

```

    }
    resdev[i] <- inprod2(ydiff[i, 1:(na[i]-1)], z[i, 1:(na[i]-1)])
  }
  for(i in 1:(ns2+ns3)){
    w[i, 1] <- 0 # LOOP THROUGH ALL STUDIES
    delta[i, 1] <- 0 # adjustment for multi-arm trials is zero for control arm
    for(k in 2:na[i]) { # treatment effect is zero for control arm
      # LOOP THROUGH ARMS
      var[i, k] <- pow(se[i, k], 2) # calculate variances
      prec[i, k] <- 1/var[i, k] # set precisions
    }
    resdev[i] <- sum(dev[i, 2: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 treat effects distributions
      md[i, k] <- d[t[i, 1]] - d[t[i, k]] + sw[i, k] # mean of treat effects distributions (with multi-arm trial correction)
      taud[i, k] <- tau * 2 * (k-1) / k # precision of treat effects 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)
} # *** PROGRAM ENDS

```

Additional code to monitor all treatment contrasts and rank treatments can be added as before. Given values for the mean, meanA=-0.73, and precision, precA=21, of the effects on Treatment 1, from external sources, absolute effects, could be monitored as follows:

```

A ~ dnorm(meanA, precA)
for(k in 1:nt) { T[k] <- A + d[k] }

```

If trials with four or more arms were included, a further multivariate normal likelihood statement would need to be added and the corresponding variance-covariance and precision matrices built (Sigma2 and Omega2, say). So, for example if ns4 4-arm trials were available, we would add the following lines of code to the above, taking care to change all the relevant loops to go through all trials:

```

for(i in (ns2+ns3+1):(ns2+ns3+ns4)) { # LOOP THROUGH 4-ARM STUDIES
  for(k in 1:(na[i]-1)) { # set variance-covariance matrix
    for(j in 1:(na[i]-1)) { Sigma2[i, j, k] <- V[i] * (1 - equals(j, k)) + var[i, k+1] * equals(j, k) }
  }
  Omega2[i, 1:(na[i]-1), 1:(na[i]-1)] <- inverse(Sigma2[i, ,]) # Precision matrix
} # multivariate normal likelihood for 4-arm trials
y[i, 2:na[i]] ~ dnmnorm(delta[i, 2:na[i]], Omega2[i, 1:(na[i]-1), 1:(na[i]-1)])
}

```

If no multi-arm trials are included, the code simplifies to:

```

# Normal likelihood, identity link, trial-level data given as treatment differences
# Random effects model for two-arm trials
model{
  # *** PROGRAM STARTS
  for(i in 1:ns2) { # LOOP THROUGH 2-ARM STUDIES
    y[i, 2] ~ dnorm(delta[i, 2], prec[i, 2]) # normal likelihood for 2-arm trials
  }
}

```

```

var[i,2] <- pow(se[i,2],2)           # calculate variances
prec[i,2] <- 1/var[i,2]             # set precisions
dev[i,2] <- (y[i,2]-delta[i,2])*(y[i,2]-delta[i,2])*prec[i,2] #Deviance contribution
delta[i,2] ~ dnorm(md[i,2],tau)     # trial-specific treat effects distributions
md[i,2] <- d[t[i,2]] - d[t[i,1]]    # mean of treat effects distributions
}
totresdev <- sum(dev[,2])           #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)
}
# *** PROGRAM ENDS

```

The data structure is similar to that of Example 5 but we now have to specify the number of two-arm trials  $ns_2$  and the number of three-arm trials  $ns_3$ . The maximum number of arms is 3, so 3 vectors are needed for the treatment indicators,  $t_{[,1]}$   $t_{[,2]}$ ,  $t_{[,3]}$ ; for a trial with 3 treatment arms, two treatment differences will be available, so 2 vectors of differences (the continuous outcomes)  $y_{[,1]}$  and their standard errors  $se_{[,1]}$  are needed; and finally the number of arms,  $na[]$  and  $v[]$  the variance of the baseline treatment in that trial (needed to adjust for the correlation in multi-arm trials – note that this variable only need to have values assigned when there are multi-arm trials), with NA denoting a missing observation. Note that any three-arm trials need to appear at the end of the column format data.

```
# Data (Parkinson's example – trial-level data: treatment differences)
```

```
list(ns2=6, ns3=1, nt=5)
```

$t_{[,1]}$	$t_{[,2]}$	$t_{[,3]}$	$y_{[,2]}$	$y_{[,3]}$	$se_{[,2]}$	$se_{[,3]}$	$na[]$	$V[]$
1	3	NA	-0.31	NA	0.668089651	NA	2	NA
1	2	NA	-1.7	NA	0.382640605	NA	2	NA
3	4	NA	-0.35	NA	0.441941738	NA	2	NA
3	4	NA	0.55	NA	0.555114559	NA	2	NA
4	5	NA	-0.3	NA	0.274276316	NA	2	NA
4	5	NA	-0.3	NA	0.320087245	NA	2	NA
1	2	4	-2.3	-0.9	0.71774604	0.694988091	3	0.254736842

```
END
```

```
# Initial Values
```

```
# Initial values for delta can be generated by WinBUGS.
```

```
#chain 1
```

```
list(d=c( NA, 0,0,0,0), sd=1)
```

```
#chain 2
```

```
list(d=c( NA, -1,-3,-1,1), sd=4)
```

```
#chain 3
```

```
list(d=c( NA, 2,2,2,2), sd=2)
```

The parameters to monitor are the same as in Example 5.

### **Program 7(b): Normal likelihood, identity link, treatment differences, Fixed Effects (Parkinson's Differences)**

```
# Normal likelihood, identity link, trial-level data given as treatment differences
```

```
# Fixed effects model
```

```

model{
# *** PROGRAM STARTS
for(i in 1:ns2) {
# LOOP THROUGH 2-ARM STUDIES
y[i,2] ~ dnorm(delta[i,2],prec[i,2]) # normal likelihood for 2-arm trials
resdev[i] <- (y[i,2]-delta[i,2])*(y[i,2]-delta[i,2])*prec[i,2] #Deviance contribution for trial i
}
for(i in (ns2+1):(ns2+ns3)) {
# LOOP THROUGH MULTI-ARM STUDIES
for (k in 1:(na[i]-1)) {
# set variance-covariance matrix
for (j in 1:(na[i]-1)) { Sigma[i,j,k] <- V[i]*(1>equals(j,k)) + var[i,k+1]*equals(j,k) }
}
Omega[i,1:(na[i]-1),1:(na[i]-1)] <- inverse(Sigma[i,,]) #Precision matrix
# multivariate normal likelihood for 3-arm trials
y[i,2:na[i]] ~ dnmnorm(delta[i,2:na[i]],Omega[i,1:(na[i]-1),1:(na[i]-1)])
#Deviance contribution for trial i
for (k in 1:(na[i]-1)){
# multiply vector & matrix
ydiff[i,k]<- y[i,(k+1)] - delta[i,(k+1)]
z[i,k]<- inprod2(Omega[i,k,1:(na[i]-1)], ydiff[i,1:(na[i]-1)])
}
resdev[i]<- inprod2(ydiff[i,1:(na[i]-1)], z[i,1:(na[i]-1)])
}
}
for(i in 1:(ns2+ns3)){
# LOOP THROUGH ALL STUDIES
for (k in 2:na[i]) {
# LOOP THROUGH ARMS
var[i,k] <- pow(se[i,k],2) # calculate variances
prec[i,k] <- 1/var[i,k] # set precisions
delta[i,k] <- d[t[i,k]] - d[t[i,1]]
}
resdev[i] <- sum(dev[i,2: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
for (k in 2:nt){ d[k] ~ dnorm(0, .0001) } # vague priors for treatment effects
}
# *** PROGRAM ENDS

# Initial Values
#chain 1
list(d=c( NA, 0,0,0,0))
#chain 2
list(d=c( NA, -1,-3,-1,1))
#chain 3
list(d=c( NA, 2,2,2,2))

```

## RESULTS

Results (based on 3 chains: 100,000 iterations after a burn-in of 50,000) are presented in Table A9 and are the same as the results obtained using the model in Example 5.

### EXAMPLE 8. PARKINSON'S SHARED PARAMETERS (MIXED TREATMENT DIFFERENCE AND ARM-LEVEL DATA)

To illustrate a meta-analysis with a shared parameter model (Section 4) we will assume that the data available for the Parkinson's example were the mean off-time reduction for patients in each arm of the trial for the first three trials, but only the differences between the intervention and control arms (and their standard errors) were available for the remaining trials (Table A8).

Random and fixed effects models were fitted. The code below consists of a combination of the code used in Example 5, for the arm-level data, and the code used in Example 7, for the trial level data.

### Program 8(a): Normal likelihood, identity link, shared parameter model, Random Effects (Parkinson's shared parameters)

```
# Normal likelihood, identity link, Arm and Trial-level data (treatment differences)
# Random effects model for multi-arm trials
model{
  for(i in 1:ns.a){
    w.a[i,1] <- 0
    delta[i,1] <- 0
    mu[i] ~ dnorm(0,.0001)
    for (k in 1:na.a[i]) {
      var.a[i,k] <- pow(se.a[i,k],2)
      prec.a[i,k] <- 1/var.a[i,k]
      y.a[i,k] ~ dnorm(theta[i,k],prec.a[i,k])
      theta[i,k] <- mu[i] + delta[i,k]
      dev[i,k] <- (y.a[i,k]-theta[i,k])*(y.a[i,k]-theta[i,k])*prec.a[i,k] #Deviance contribution
    }
    resdev[i] <- sum(dev[i,1:na.a[i]])
    for (k in 2:na.a[i]) {
      delta[i,k] ~ dnorm(md[i,k],taud.a[i,k])
# mean of LOR distributions, with multi-arm trial correction
      md[i,k] <- d[t.a[i,k]] - d[t.a[i,1]] + sw.a[i,k]
      taud.a[i,k] <- tau *2*(k-1)/k
      w.a[i,k] <- (delta[i,k] - d[t.a[i,k]] + d[t.a[i,1]])
      sw.a[i,k] <- sum(w.a[i,1:k-1])/(k-1)
    }
  }
  for(i in 1:ns.t){
    w[i,1] <- 0
    delta[i+ns.a,1] <- 0
    for (k in 2:na[i]) {
      var[i,k] <- pow(se[i,k],2)
      prec[i,k] <- 1/var[i,k]
      y[i,k] ~ dnorm(delta[i+ns.a,k],prec[i,k])
      dev[i+ns.a,k] <- (y[i,k]-delta[i+ns.a,k])*(y[i,k]-delta[i+ns.a,k])* prec[i,k]
    }
    resdev[i+ns.a] <- sum(dev[i+ns.a,2:na[i]])
    for (k in 2:na[i]) {
      delta[i+ns.a,k] ~ dnorm(md[i+ns.a,k],taud[i,k])
      md[i+ns.a,k] <- d[t[i,k]] - d[t[i,1]] + sw[i,k]
      taud[i,k] <- tau *2*(k-1)/k
      w[i,k] <- (delta[i+ns.a,k] - d[t[i,k]] + d[t[i,1]])
      sw[i,k] <- sum(w[i,1:k-1])/(k-1)
    }
  }
  totresdev <- sum(resdev[])
  d[1]<-0
  for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }
  sd ~ dunif(0,5)
  tau <- pow(sd,-2)
}
# *** PROGRAM STARTS
# LOOP THROUGH STUDIES WITH ARM DATA
# adjustment for multi-arm trials is zero for control arm
# treatment effect is zero for control arm
# vague priors for all trial baselines
# LOOP THROUGH ARMS
# calculate variances
# set precisions
# normal likelihood
# model for linear predictor
# summed residual deviance contribution for this trial
# LOOP THROUGH ARMS
# trial-specific LOR distributions
# precision of LOR distributions (with multi-arm trial correction)
# adjustment, multi-arm RCTs
# cumulative adjustment for multi-arm trials
# LOOP THROUGH STUDIES WITH TRIAL DATA
# adjustment for multi-arm trials is zero for control arm
# treatment effect is zero for control arm
# LOOP THROUGH ARMS
# calculate variances
# set precisions
# normal likelihood
#Deviance contribution
# summed residual deviance contribution for this trial
# LOOP THROUGH ARMS
# trial-specific LOR distributions
# mean of LOR distributions, with multi-arm trial correction
# precision of LOR distributions (with multi-arm trial correction)
# adjustment, multi-arm RCTs
# cumulative adjustment for multi-arm trials
#Total Residual Deviance
# treatment effect is zero for reference treatment
# vague priors for treatment effects
# vague prior for between-trial SD
# between-trial precision = (1/between-trial variance)
# *** PROGRAM ENDS
```

Additional code to monitor all treatment contrasts, rank treatments and obtain absolute treatment effects can be added as before.



The data structure for this code consists of three parts. First, a list giving the number of studies with arm-level information, `ns.a`, the number of studies with trial-level information, `ns.t`, and the number of treatments, `nt`. Two sections of column format data follow: one with the arm-level data, with the structure described for program 5(a), and another with the trial-level data and the same structure described for program 7(a). All three data components need to be loaded into WinBUGS for the program to run. Note that, because two separate sets of data are being read into WinBUGS, the variable names referring to the arm-level data have the added suffix `.a`, to distinguish them from the trial-level data.

```
# Data (Parkinson's example: Arm and Trial-level data)
list(ns.a=3, ns.t=4, nt=5)

# Arm-level data
t.a[,1]  t.a[,2]  t.a[,3]  y.a[,1]  y.a[,2]  y.a[,3]  se.a[,1]  se.a[,2]  se.a[,3]  na.a[]  #  study
1        3      NA    -1.22   -1.53   NA      0.504    0.439    NA       2      #    1
1        2      NA    -0.7    -2.4    NA      0.282    0.258    NA       2      #    2
1        2      4     -0.3    -2.6   -1.2    0.505    0.510    0.478   3      #    3
END

# Trial-level data
t[,1]  t[,2]  y[,2]  se[,2]  na[]  #  study
3      4    -0.35  0.441941738  2  #    4
3      4    0.55  0.555114559  2  #    5
4      5   -0.3  0.274276316  2  #    6
4      5   -0.3  0.320087245  2  #    7
END

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

The parameters to monitor are the same as in Example 5.

### Program 8(b): Normal likelihood, identity link, shared parameter model, Fixed Effects (Parkinson's shared parameters)

```
# Normal likelihood, identity link, Arm and Trial-level data (treatment differences)
# Fixed effects model
model{
  for(i in 1:ns.a){
    mu[i] ~ dnorm(0,.0001)
    for(k in 1:na.a[i]) {
      var.a[i,k] <- pow(se.a[i,k],2)
      prec.a[i,k] <- 1/var.a[i,k]
      y.a[i,k] ~ dnorm(theta[i,k],prec.a[i,k])
      theta[i,k] <- mu[i] + d[t.a[i,k]] - d[t.a[i,1]]
      dev[i,k] <- (y.a[i,k]-theta[i,k])*(y.a[i,k]-theta[i,k])*prec.a[i,k] #Deviance contribution
    }
    resdev[i] <- sum(dev[i,1:na.a[i]])
  }
  for(i in 1:ns.t){
    # *** PROGRAM STARTS
    # LOOP THROUGH STUDIES WITH ARM DATA
    # vague priors for all trial baselines
    # LOOP THROUGH ARMS
    # calculate variances
    # set precisions
    # normal likelihood
    # model for linear predictor
    # summed residual deviance contribution for this trial
    # LOOP THROUGH STUDIES WITH TRIAL DATA
```

```

for (k in 2:na[i]) {
  var[i,k] <- pow(se[i,k],2)
  prec[i,k] <- 1/var[i,k]
  y[i,k] ~ dnorm(theta[i+ns.a,k],prec[i,k])
  theta[i+ns.a,k] <- d[t[i,k]] - d[t[i,1]]
  dev[i+ns.a,k] <- (y[i,k]-theta[i+ns.a,k])*
    (y[i,k]-theta[i+ns.a,k])* prec[i,k]
}
resdev[i+ns.a] <- sum(dev[i+ns.a,2:na[i]])
}
totresdev <- sum(resdev[])
d[1]<-0
for (k in 2:nt){ d[k] ~ dnorm(0,.0001) }
}

# LOOP THROUGH ARMS
# calculate variances
# set precisions
# normal likelihood
# model for linear predictor
#Deviance contribution

# summed residual deviance contribution for this trial

#Total Residual Deviance
# treatment effect is zero for reference treatment
# vague priors for treatment effects
# *** PROGRAM ENDS

# Initial Values
#chain 1
list(d=c( NA, 0,0,0,0), mu=c(0, 0, 0))
#chain 2
list(d=c( NA, -1,-3,-1,1), mu=c(-3, -3, -3))
#chain 3
list(d=c( NA, 2,2,2,2), mu=c(-3, 5, -1))

```

## RESULTS

Results (based on 3 chains: 100,000 iterations after a burn-in of 20,000 and 50,000 for the FE and Re models, respectively) are presented in Table A9 and are the same as the results obtained using the model in Example 5.