

Evidence Synthesis for Decision Making 2: A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomized Controlled Trials

Sofia Dias, PhD, Alex J. Sutton, PhD, A. E. Ades, PhD, Nicky J. Welton, PhD

We set out a generalized linear model framework for the synthesis of data from randomized controlled trials. A common model is described, taking the form of a linear regression for both fixed and random effects synthesis, which can be implemented with normal, binomial, Poisson, and multinomial data. The familiar logistic model for meta-analysis with binomial data is a generalized linear model 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 pairwise meta-analysis, indirect comparisons, synthesis of multiarm 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 simulation. An advantage of this approach is that it is straightforward to extend to shared parameter models where different randomized controlled trials report outcomes in different formats but from a common underlying model. Use of the generalized linear model framework allows us to present a unified account of how models can be compared using the deviance information criterion and how goodness of fit can be assessed using the residual deviance. The approach is illustrated through a range of worked examples for commonly encountered evidence formats. **Key words:** generalized linear model; network meta-analysis; indirect evidence; meta-analysis. (*Med Decis Making* 2013;33:607–617)

Meta-analysis of randomized controlled trials is widely used in the medical research literature,^{1–3} and methodology for pairwise meta-analysis is well developed.^{4–9} Recently, meta-analysis methods have been extended to indirect and mixed treatment comparisons, also known as network meta-analysis (NMA), which combine data from

randomized comparisons, A v. B, A v. C, B v. D, etc., to deliver an internally consistent set of estimates while respecting the randomization in the evidence.¹⁰ NMA is particularly useful in decision-making contexts.^{11–21} We present a single unified approach to evidence synthesis of aggregate data from randomized controlled trials, specifically, but not exclusively, for use in probabilistic decision making.²² To cover the variety of outcomes reported in trials and the range of data transformations required to achieve linearity, we adopt the framework of generalized linear modeling (GLM).²³ 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 (FE) and random effects (RE) settings. Our common core models can synthesize data from pairwise meta-analysis, multiarm trials, indirect comparisons, and NMA without distinction. Indeed, pairwise meta-analysis and indirect comparisons are special cases of NMA.

Received 13 June 2012 from School of Social and Community Medicine, University of Bristol, Bristol, UK (SD, AEA, NJW); and Department of Health Sciences, University of Leicester, Leicester, UK (AJS). This series of tutorial papers was based on Technical Support Documents in Evidence Synthesis (available from <http://www.nicedsu.org.uk>), which were prepared with funding from the NICE Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. Revision accepted for publication 22 June 2012.

Supplementary material for this article is available on the *Medical Decision Making* Web site at <http://mdm.sagepub.com/supplemental>.

Address correspondence to Sofia Dias, School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK e-mail: s.dias@bristol.ac.uk.

DOI: 10.1177/0272989X12458724

We take a Bayesian Markov Chain Monte Carlo (MCMC) approach using the freely available software WinBUGS 1.4.3.²⁴ We include an extensive web appendix with fully annotated WinBUGS code for all models to run a series of worked examples. This code is also available at www.nicedsu.org.uk.²⁵

DEVELOPMENT OF THE CORE MODELS: BINOMIAL DATA WITH LOGIT LINK

We begin by presenting the standard Bayesian MCMC approach to pairwise meta-analysis for binomial data, based on Smith et al.,⁹ and develop our approach to assessment of goodness of fit, model diagnostics, and comparison, based on Spiegelhalter et al.²⁶ This approach can then be easily applied to other outcome types and to meta-analysis of multiple treatments.

Consider a set of M trials comparing two treatments, 1 and 2, in a prespecified target patient population, which are to be synthesized in a meta-analysis. An FE analysis assumes that each study generates an estimate of the same parameter d_{12} , subject to sampling error. In an RE model, each study i provides an estimate of the study-specific treatment effects $\delta_{i,12}$, which are assumed not to be equal but instead “similar” in a way that assumes that the information that the trials provide is independent of the order in which they were carried out (*exchangeable*), over the population of interest.^{27,28} The exchangeability assumption is equivalent to saying that the trial-specific treatment effects come from a common distribution with mean d_{12} and variance σ_{12}^2 .

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

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

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

Note that for pairwise meta-analysis, the subscripts in d , δ , and σ are redundant since only one treatment comparison is being made. We shall drop the subscripts for σ but keep the subscripts for δ and d , to allow for extensions to multiple treatments.

Worked Example: Binomial Likelihood, Logit Link (Appendix: Example 1)

We consider a meta-analysis of 22 trials of beta-blockers to prevent mortality after myocardial infarction.^{28,29} 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 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 Trials²⁹

Study i	Control		Treatment	
	No. of Events (r_{i1})	No. of Patients (n_{i1})	No. of Events (r_{i2})	No. of Patients (n_{i2})
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

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:

$$r_{ik} \sim \text{Binomial}(p_{ik}, n_{ik}), \quad (1)$$

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 take only values between 0 and 1, a transformation (link function) is used that maps these into a continuous measure between plus and minus infinity. For a binomial likelihood, the most commonly used link function is the logit (Table 2). We model the probabilities of success p_{ik} on the logit scale:

$$\text{logit}(p_{ik}) = \mu_i + \delta_{ik} I_{\{k \neq 1\}}, \quad (2)$$

where

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

Table 2 Commonly Used Link Functions and Their Inverse With Reference to Which Likelihoods They Can Be Applied

Link	Link Function, $\theta=g(\gamma)$	Inverse Link Function, $\gamma=g^{-1}(\theta)$	Likelihood
Identity	γ	θ	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

μ_i are trial-specific baselines, representing the log-odds of the outcome in the “control” treatment (i.e., treatment 1), and $\delta_{i,12}$ are trial-specific log-odds ratios of success for the treatment group (indexed 2) compared to control (indexed 1). We can write equation 2 as

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

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

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

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

An important feature of *all* the models presented here is that no assumptions are made about the trial-specific baselines μ_i . They are regarded as nuisance parameters that 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.^{30,31} 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 and baselines eliminated entirely. Baseline models are discussed by Dias et al.³²

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,²³ so that each data point should contribute about 1 to the posterior mean deviance.^{26,33} 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 binomial likelihoods, each trial arm contributes 1 independent data point, and the residual deviance is calculated (for each iteration of the MCMC simulation) as

$$\begin{aligned} D_{res} &= \sum_i \sum_k 2 \left(r_{ik} \log\left(\frac{r_{ik}}{\hat{r}_{ik}}\right) + (n_{ik} - r_{ik}) \log\left(\frac{n_{ik} - r_{ik}}{n_{ik} - \hat{r}_{ik}}\right) \right) \\ &= \sum_i \sum_k dev_{ik} \end{aligned} \tag{3}$$

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

Leverage statistics are used in regression analysis to assess the influence that each data point has on the model parameters. In a Bayesian framework, 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 \widetilde{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 calculated by replacing \hat{r}_{ik} with \tilde{r}_{ik} in equation 3.

The deviance information criterion (DIC)²⁶ 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 penalizes model complexity—lower

Table 3 Blocker Example

	Fixed Effects Model				Random Effects Model			
	Mean	s	Median	CrI	Mean	s	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
σ^a	—	—	—	—	0.14	0.082	0.13	0.01, 0.32
\bar{D}_{res}^b	46.8				41.9			
p_D	23.0				28.1			
DIC	69.8				70.0			

Note: Posterior mean, standard deviation (s), 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 σ and model fit statistics. Results are based on 20,000 iterations on 3 chains, after a burn-in of 10,000.

^aBased on a Uniform(0,5) prior distribution.

^bCompare to 44 data points.

values of the DIC suggest a more parsimonious model. The DIC can be used to compare different models for the same likelihood and data—for example, FE v. RE models or FE models with and without covariates. \bar{D}_{res} should also be consulted to ensure that overall fit is adequate.

WinBUGS will automatically calculate the posterior mean of the deviance for the current model \bar{D} but not \bar{D}_{res} . The former can be useful for model comparison purposes only and not to assess the fit of a single model. Further, the p_D , and therefore the DIC, calculated in the way that we suggest, is not precisely the same as that calculated in WinBUGS, except in the case of a normal likelihood. The reason is that 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.³⁴ The latter is more stable in highly nonlinear models with high levels of parameter uncertainty.

Examining the contribution of each data point to p_D in leverage plots can help identify influential and/or poorly fitting observations.^{25,26} However, deciding between FE or RE models can be highly dependent on the impact of sparse data and choice of prior distributions. In NMA there are additional issues regarding consistency among evidence sources on different contrasts that need to be taken into account.³⁵

WinBUGS implementation and illustrative results. In comparison of the fit of the FE and RE models (Table 3), the posterior mean of the residual deviance indicates that although the RE models is a better fit to the data, with $\bar{D}_{res} = 41.9$ against 46.8 for the FE model; this is achieved at the expense of more parameters (p_D is higher in the RE model).

The DIC suggests that there is little to choose between the two models—differences of less than 3 or 5 are not considered important—and the FE model may be preferred since it is easier to interpret (Table 3). The posterior median of the pooled log odds ratio of beta-blockers compared to control in the FE model is -0.26 with 95% credible interval (CrI; $-0.36, -0.16$), indicating a reduced mortality in the treatment group. The posterior medians of the absolute probability of mortality on the control and treatment groups are 0.10 and 0.08, respectively (CrIs in Table 3). Results for the RE model are similar.

The logit model assumes additivity of effects on the logit scale.³⁶ Choice of scale can be guided by goodness of fit or by lower between-study heterogeneity, but there are seldom enough data to make this choice reliably, and logical considerations may play a larger role.³⁷ Quite distinct from choice of scale for modeling is the issue of how to report treatment effects. Thus, while one might assume linearity of effects on the logit scale, given information on the absolute effect of one treatment, it is possible to derive treatment effects on other scales, such as risk difference, relative risk, or numbers needed to treat. This is illustrated in the appendix. An advantage of Bayesian MCMC is that appropriate distributions and, therefore, CrIs are automatically generated for all these quantities.

GENERALIZED LINEAR MODELS

We now extend our approach to models for other data types. The core models remain the same, but the likelihood and the link function change to reflect

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}	$\sum_i \sum_k \left(\frac{(y_{ik} - \bar{y}_{ik})^2}{se_{ik}^2} \right)$
$r_{i,k,1:J} \sim \text{Multinomial}(p_{i,k,1:J}, n_{ik})$	$\hat{r}_{ikj} = n_{ik} p_{ikj}$	$\sum_i \sum_k 2 \left(\sum_j r_{ikj} \log \left(\frac{r_{ikj}}{\hat{r}_{ikj}} \right) \right)$

the nature of the data (continuous, rate, categorical) and the sampling process that generated them (normal, Poisson, multinomial, etc.). In GLM theory,²³ a likelihood is defined in terms of some unknown parameters γ , 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 2 now becomes a GLM where

$$g(\gamma) = \theta_{ik} = \mu_i + \delta_{i,bk} I_{\{k \neq 1\}} \quad (4)$$

is an appropriate link function (e.g., the logit link) and θ_{ik} is the linear predictor, usually a continuous measure of the treatment effect in arm k of trial i (e.g., in log-odds form); μ_i is defined as before, and $\delta_{i,bk}$ is the trial-specific treatment effect of the treatment in arm k relative to the control treatment in arm b (we assume $b = 1$ throughout) so that for an RE model

$$\delta_{i,1k} \sim N(d_{1k}, \sigma^2). \quad (5)$$

Table 2 has details of the most commonly used likelihoods and link and inverse link functions, and Table 4 provides the formulae for the residual deviance and the predicted values needed to calculate \bar{D}_{res} and p_D for different likelihoods.

Whatever the type of outcome data and GLMs used to analyze them, the basic model for meta-analysis remains the same (equations 4 and 5); however, in a Bayesian framework, specification of the range for the prior for the heterogeneity parameter requires care.

Rate Data: Poisson Likelihood and Log Link (Appendix: Example 2)

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 λ_{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 modeled on the log scale. The linear predictor in equation 4 is therefore on the log-rate scale:

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

The key assumption of this model is that in each arm of each trial, the hazard is constant over the follow-up period, implying a homogeneous population where all patients have the same hazard rate.

These models are also useful for repeated event data. Examples include (1) a model for the total number of accidents in each arm where each individual may have more than one accident and (2) observations repeated in space rather than time, such as the number of teeth requiring fillings. Using the Poisson model for repeated event data makes the additional assumption that 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.^{38–40}

Rate Data: Binomial Likelihood and Cloglog Link (Appendix: Example 3)

In some meta-analyses, each trial reports the proportion of patients reaching an endpoint at a specified follow-up time, but the trials do not all have the same follow-up time. By defining r_{ik} as the number of events in arm k of trial i , with follow-up time f_i , the likelihood for the data-generating process is binomial, as in equation 1.

One way to take the length of follow-up in each trial into account in the analysis is to assume an

underlying Poisson process for each trial arm, with a constant event rate λ_{ik} . The linear model becomes

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

with the treatment effects $\delta_{i,1k}$ representing log-hazard ratios. The assumptions made in this model are as for the Poisson rate models—namely, that the hazards are constant and homogeneous in each trial. Relaxation of this assumption is possible when suitable data are available.²⁵

Logit models for probability outcomes in studies with different follow-up times are also possible. One option is to assume that all outcome events that are going to occur will have occurred before the observation period in the trial has ended, regardless of variation between studies in follow-up time. Another is to assume a proportional odds model, which implies a complex form for the hazard rates.⁴¹ The clinical plausibility of these assumptions should be discussed and supported either by citation of relevant literature or by examination of evidence on changes in outcome rate over the period of follow-up.

Competing Risks: Multinomial Likelihood and Log Link (Appendix: Example 4)

A competing risk analysis is appropriate where multiple, mutually exclusive endpoints have been defined and patients leave the risk set if any one of them is reached. For example, in trials of treatments for schizophrenia,⁴² observations continued until patients 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.

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

$$r_{i,k,j=1,\dots,J} \sim \text{Multinomial}(p_{i,k,j=1,\dots,J}, n_{ik}) \text{ with } \sum_{j=1}^J p_{i,k,j} = 1, \quad (6)$$

and the parameters of interest are the rates (hazards) at which patients move from their initial state to any of the endpoints j , λ_{ikj} . Note that the J^{th} endpoint represents the censored observations—that is, patients not reaching any of the other endpoints before end of follow-up.

If we assume constant hazards λ_{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 j is

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

The probability of remaining in the initial state (i.e., being censored) is

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

The hazards, λ_{ikj} , are modeled on the log scale:

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

The trial-specific treatment effects $\delta_{i,1k,j}$ of the treatment in arm k relative to the control treatment in arm 1 of that trial for outcome j is assumed to follow a normal distribution:

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

The between-trials variance of the RE distribution, σ_j^2 , is specific to each outcome j . Different models for σ_j^2 can be considered.²⁵

These competing risks models share the same assumptions as the cloglog models presented above 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). Extensions where the assumptions are relaxed are possible.⁴³

Continuous Data: Normal Likelihood and Identity Link (Appendix: Example 5)

With continuous outcome data, meta-analysis is often 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

$$y_{ik} \sim N(\theta_{ik}, se_{ik}^2). \quad (7)$$

The parameter of interest is the mean, θ_{ik} , of this continuous measure, which is unconstrained on the real line. The identity link is used (Table 2), and the linear model can be written as

$$\theta_{ik} = \mu_i + \delta_{i,1k} \mathbf{I}_{\{k \neq 1\}}. \tag{8}$$

Before/after studies: Change from baseline measures. In cases where the original continuous trial outcome is measured at baseline and at a prespecified follow-up point, meta-analysis can be based on the mean change from baseline and an appropriate measure of uncertainty (e.g., the variance or standard error), which takes into account any within-patient correlation. It is preferable to use the mean of the final reading, having adjusted for baseline via regression/ANCOVA, if available.⁸

The likelihood for the mean change from baseline in arm k of trial i , y_{ik} , with change variance se_{ik}^2 is given in equation 7, and θ_{ik} is modeled on the natural scale as in equation 8. Various workarounds are commonly used when information on the change variance is lacking.^{8,25,44–46}

Treatment Differences (Appendix: Example 7)

Trial results are sometimes available only 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 relative measure of treatment effect of arm k relative to arm 1 in trial i , y_{ik} , with variance V_{ik} , $k \geq 2$, such that

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

This is overwhelmingly the most common form of meta-analysis, especially among 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 mainstay in applied meta-analytic studies.

The parameters of interest are the trial-specific mean treatment effects θ_{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 $\theta_{ik} = \delta_{i,1k}$.

Standardized mean differences. There are a series of standardized mean difference measures commonly used with psychological or neurological outcome measures. These can be synthesized in exactly the same way as any other treatment effect summary. The idea is that the two scales are measuring essentially the same quantity and that results from

differences 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 standardized mean difference measures are Cohen’s d ⁴⁷ and Hedges’s adjusted g .⁴⁸

However, dividing estimates through by the sample standard deviation introduces additional heterogeneity and distortion⁴⁹ and produces results that some find less interpretable.⁸ 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.

Standardized mean differences are sometimes used for noncontinuous outcomes. However, this is not recommended; use of the appropriate GLM is likely to reduce heterogeneity.³⁹

Ordered Categorical Data: Multinomial Likelihood and Probit Link (Appendix: Example 6)

In some applications, the data generated by the trial may be continuous but the outcome measure categorized, using one or more predefined cutoffs. Examples include the PASI (Psoriasis Area Severity Index) and the ACR scale (American College of Rheumatology), where it is common to report the percentage of patients who have improved by more than certain benchmark relative amounts. Thus, ACR-20 represents the proportion of patients who have improved by at least 20% on the ACR scale. Trials may report ACR-20, ACR-50, and ACR-70 or only one or two of these endpoints. A coherent model that makes efficient use of such data is obtained by assuming that the treatment effect is the same regardless of the cutoff.

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 6, and the parameters of interest are the probabilities, p_{ikj} , that a patient in arm k of trial i belongs to category j . We use the probit link function, the inverse of the normal cumulative distribution function Φ , to map p_{ikj} onto the real line (Table 2). The model is written as

$$\theta_{ikj} = \Phi^{-1}(p_{ikj}) = \mu_{ij} + \delta_{i,1k} \mathbf{I}_{\{k \neq 1\}}.$$

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 $\delta_{i,1k}$ standard deviations.

The model assumes that there is an underlying continuous variable that has been categorized by specifying different cutoffs, 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. Rewriting the model as

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

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. Unless the response probabilities are very extreme, the probit model will be indistinguishable from the logit model in terms of model fit or DIC. Choice of link function can be based on the data generating process and on the interpretability of the results.

Other Link Functions and Shared Parameter Models (Appendix: Example 8)

Risk differences and relative risks are usually modeled by using the difference-based methods described previously. However, an arm-based analysis can be performed with a binomial likelihood.⁵⁰

The WinBUGS platform makes it particularly easy to implement different GLMs that include a “shared parameter.” For example, some trials might report time at risk and number of events, while others report only the hazard ratios.

Extension to NMA (Appendix: Examples 3–8)

We now show how the core GLMs for pairwise meta-analysis are immediately applicable to indirect comparisons, multiarm trials, and NMA, without further extension.

We have 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 σ_{12}^2 . We now suppose that, within the same set of trials (i.e., trials relevant to the same research question), comparisons of treatments 1 and 3 are also made. To carry out a pairwise RE 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 $\delta_{i,13} \sim N(d_{13}, \sigma_{13}^2)$. It follows from the transitivity relation, $\delta_{i,23} = \delta_{i,13} - \delta_{i,12}$, that the study-specific treatment effects of treatment 3 compared to 2, $\delta_{i,23}$, are also exchangeable⁵¹:

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

It can further be shown that this implies

$$d_{23} = d_{13} - d_{12} \tag{9}$$

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.⁵¹ For simplicity, we assume equal variances in all subsequent methods so that $\sigma_{12}^2 = \sigma_{13}^2 = \sigma_{23}^2 = \sigma^2$, which implies that the correlation between any two treatment contrasts in a multiarm trial is 0.5.¹⁷ For heterogeneous variance models, see Lu and Ades.⁵¹

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

Note the relationship between the standard assumptions of pairwise meta-analysis and those required for NMA. For an RE pairwise meta-analysis, we need to assume exchangeability of the effects $\delta_{i,12}$ over the 1 v. 2 trials and also exchangeability of the effects $\delta_{i,13}$ over the 1 v. 3 trials. For NMA, we must assume the exchangeability of both treatment effects over both 1 v. 2 and 1 v. 3 trials. The theory extends readily to additional treatments where, in each case, we must assume the exchangeability of the δ 's across the entire set of trials. Then the within-trial transitivity relation is enough to imply the exchangeability of all the treatment effects $\delta_{i,XY}$ and the consistency equations¹⁹:

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

are also therefore implied, where $s > 2$ is the number of treatments being compared. These assumptions are

required by indirect comparisons and NMA, but given that we are already assuming that all trials are relevant to the same research question, they are not additional assumptions.

While consistency of the treatment effects must hold for a given patient population, inconsistency in the evidence can be created by trial-level effect modifiers. Evidence consistency needs to be checked in all networks.³⁵

Now that several treatments are being compared, the notation needs to be clarified. The trial-specific treatment effects of the treatment in arm k , relative to the treatment in arm 1, are drawn from a common RE 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 σ^2 represents the between-trial variability in treatment effects (heterogeneity). For trials that compare treatments 1 and 2, $d_{t_{i1},t_{i2}} = d_{12}$, for 2 v. 3 trials, $d_{t_{i1},t_{i3}} = d_{23}$, and so on. The pooled treatment effect of treatment 3 compared to treatment 2, d_{23} , is obtained from equation 9.

Incorporating Multiarm Trials

Suppose that we have a number of trials with more than two arms (multiarm trials) involving the treatments of interest. Among commonly suggested stratagems for synthesis are (1) combining all active arms into one, (2) splitting the control group between all relevant experimental groups, and (3) ignoring all but two of the trial arms,⁸ but none of these are satisfactory.

Based on the same exchangeability assumptions above, a single multiarm trial will estimate a vector of correlated RE δ_i , so a three-arm trial will produce two RE and a four-arm trial, three. Assuming, as before, homogeneous between-trial variance, σ^2 , we have

$$\delta_i = \begin{pmatrix} \delta_{i,12} \\ \vdots \\ \delta_{i,1a_i} \end{pmatrix} \sim N_{a_i-1} \left(\begin{pmatrix} d_{t_{i1},t_{i2}} \\ \vdots \\ d_{t_{i1},t_{ia_i}} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \dots & \sigma^2/2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \dots & \sigma^2 \end{pmatrix} \right), \tag{10}$$

where δ_i is the vector of RE, which follows a multivariate normal distribution; a_i represents the number of arms in trial i ($a_i = 2, 3, \dots$); and $d_{t_{i1},t_{ik}} = d_{1,t_{ik}} - d_{1,t_{i1}}$. Equivalently, the conditional

univariate distributions for the random effect of arm $k > 2$, given all arms from 2 to $k - 1$, are as follows⁵²:

$$\delta_{i,1k} \mid \begin{pmatrix} \delta_{i,12} \\ \vdots \\ \delta_{i,1(k-1)} \end{pmatrix} \sim N \left((d_{1,t_{ik}} - d_{1,t_{i1}}) + \frac{1}{k} \sum_{j=1}^{k-1} [\delta_{i,1j} - (d_{1,t_{ij}} - d_{1,t_{i1}})], \frac{k}{2(k-1)} \sigma^2 \right). \tag{11}$$

Either the multivariate distribution in equation 10 or the conditional distributions in equation 11 must be used to estimate the RE for each multiarm study so that the between-arm correlations are taken into account.

This formulation provides another interpretation of the exchangeability assumptions. 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. (Note: *Missing at random* does not mean that the choice of arms is random but that the missingness of arms is unrelated to the efficacy of the treatment.)

The WinBUGS code provided in the appendix is based on equation 11. It therefore exactly instantiates the theory behind NMA that relates it to pairwise meta-analysis. The code provided will analyze pairwise meta-analysis, indirect comparisons, and NMA with and without multiarm trials without distinction.

When results from multiarm 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 additional adjustment to the likelihood.^{25,53}

DISCUSSION

We have presented a single unified account for evidence synthesis of aggregate data from randomized controlled trials. 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 GLM. 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. Furthermore, the linear predictor is a regression model with $K-1$ treatment effect parameters for any K treatment network, offering a single model for pairwise meta-analysis, indirect comparisons, NMA, and synthesis of multiarm trials in any combination. This has all been presented in a Bayesian MCMC context and supported by code for WinBUGS that allows us to take full advantage of the modularity implied by GLMs.

The conceptual and practical advantages of Bayesian MCMC in the context of probabilistic decision making are well known,^{54–57} although alternative software can be used.²⁵ However, there are a series of technical issues that need careful attention, including convergence, Monte Carlo error, and parameterization.²⁵ Two particular issues that always deserve care are zero cells and choice of prior distributions. Generally, no special precautions are needed for zero cells, but in sparse data sets, they may result in instability. One solution is to put informative priors on the between-trial variance in RE models, although this may not always suffice.²⁵ We have recommended vague uniform priors for the heterogeneity standard deviation, but with sparse data this may result in clinically unrealistic posterior distributions for between-study variation and treatment effects. One option is to formulate priors based on clinical opinion or on other meta-analyses with similar outcomes.^{17,25} However, 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,⁵⁸ and this is currently an active research area.

ACKNOWLEDGMENTS

We thank Jenny Dunn at NICE Decision Support Unit and Mike Campbell, Rachael Fleurence, Julian Higgins, Jeroen Jansen, Steve Palmer, and the team at NICE, led by Zoe Garrett, for reviewing earlier versions of this article.

REFERENCES

- Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Stat Med*. 2008;27:625–50.
- Whitehead A. *Meta-analysis of Controlled Clinical Trials*. Chichester, UK: Wiley; 2002.
- Whitehead A, Whitehead J. A general parameteric approach to the meta-analysis of randomised clinical trials. *Stat Med*. 1991;10:1665–77.
- Zelen M. The analysis of several 2×2 tables. *Biometrika*. 1971;58:129–37.
- DerSimonian R, Laird N. Meta-analysis of clinical trials. *Control Clin Trials*. 1986;7:177–88.
- Cooper H, Hedges L. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation; 1994.
- Egger M, Davey-Smith G, Altman D. *Systematic Reviews in Health Care: Meta-analysis in Context*. 2nd ed. London: BMJ; 2001.
- Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.0. Chichester, UK: Cochrane Collaboration; 2008.
- Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med*. 1995;14:2685–99.
- Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. *Health Technol Assess*. 2005;9(26):1–134.
- Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter evidence synthesis and consistency of evidence. *Stat Med*. 2003;22:2995–3016.
- Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*. 2005;331:897–900.
- Welton NJ, Ades AE. A model of toxoplasmosis incidence in the UK: evidence synthesis and consistency of evidence. *Appl Stat*. 2005;54:385–404.
- Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper H, Hedges LV, eds. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation; 1994. p 339–55.
- Gleser L, Olkin I. Meta-analysis for 2×2 tables with multiple treatment groups. In: Stangl DK, Berry DA, eds. *Meta-analysis in Medicine and Health Policy*. New York: Marcel Dekker; 2000.
- Hasselblad V. Meta-analysis of multi-treatment studies. *Med Decis Making*. 1998;18:37–43.
- Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15:2733–49.
- Lu G, Ades A. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23:3105–24.
- Lu G, Ades A. Assessing evidence consistency in mixed treatment comparisons. *J Am Stat Assoc*. 2006;101:447–59.
- Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21:2313–24.
- Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomised trials. *Stat Methods Med Res*. 2008;17:279–301.
- National Institute for Health and Clinical Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE; 2008.
- McCullagh P, Nelder JA. *Generalised Linear Models*. 2nd ed. London: Chapman & Hall; 1989.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput*. 2000;10:325–37.
- Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 2: a generalised linear modelling framework for pair-wise and network meta-analysis of randomised controlled trials. Available from: <http://www.nicedsu.org.uk>

26. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *J Roy Stat Soc B*. 2002; 64(4):583–616.
27. Bernardo JM, Smith AFM. *Bayesian Theory*. New York: Wiley; 1994.
28. Spiegelhalter D, Thomas A, Best N, Lunn D. WinBUGS user manual. Version 1.4, January 2003. Version 1.4.3, 2007. <http://www.mrc-bsu.cam.ac.uk/bugs>.
29. Carlin JB. Meta-analysis for 2×2 tables: a Bayesian approach. *Stat Med*. 1992;11:141–58.
30. van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. *Stat Med*. 1993;12:2273–84.
31. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multi-variate approach and meta-regression. *Stat Med*. 2002;21:589–624.
32. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: the baseline natural history model. *Med Decis Making*. 2013;33(5):657–670.
33. Dempster AP. The direct use of likelihood for significance testing. *Stat Comput*. 1997;7:247–52.
34. Spiegelhalter D. Two brief topics on modelling with WinBUGS. IceBugs: a workshop about the development and use of the BUGS programme. Available from: <http://www.math.helsinki.fi/openbugs/IceBUGS/Presentations/SpiegelhalterIceBUGS.pdf>
35. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making*. 2013;33(5):641–656.
36. Deeks JJ. Issues on the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med*. 2002;21:1575–600.
37. Caldwell DM, Welton NJ, Dias S, Ades AE. Selecting the best scale for measuring treatment effect in a network meta-analysis: a case study in childhood nocturnal enuresis. *Res Syn Meth*. 2012; 3:111–25.
38. Cooper NJ, Sutton AJ, Lu G, Khunti K. Mixed comparison of stroke prevention treatments in individuals with nonrheumatic atrial fibrillation. *Arch Intern Med*. 2006;166:1269–75.
39. Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomised evidence by using mixed treatment comparison meta-analysis. *J Roy Stat Soc A*. 2010;173:613–29.
40. Guevara JP, Berlin JA, Wolf FM. Meta-analytic methods for pooling rates when follow-up duration varies: a case study. *BMC Med Res Methodol*. 2004;4:17–23.
41. Collett D. *Modelling Survival Data in Medical Research*. London: Chapman & Hall; 1994.
42. National Collaborating Centre for Mental Health. *Schizophrenia: Core Interventions in the Treatment and Management of Schizophrenia in Adults in Primary and Secondary Care*. London: NICE; 2010.
43. Ades AE, Mavranouzouli I, Dias S, Welton NJ, Whittington C, Kendall T. Network meta-analysis with competing risk outcomes. *Value Health*. 2010;13(8):976–83.
44. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med*. 1992;11:1685–704.
45. Follmann D, Elliott P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. *J Clin Epidemiol*. 1992;45:769–73.
46. Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. *Stat Med*. 2005;24:3823–44.
47. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press; 1969.
48. Hedges LV, Olkin I. *Statistical Methods for Meta-analysis*. London: Academic Press; 1985.
49. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia: Lippincott, Williams & Wilkins; 2008.
50. Warn DE, Thompson SG, Spiegelhalter DJ. Bayesian random effects meta-analysis of trials with binary outcomes: method for absolute risk difference and relative risk scales. *Stat Med*. 2002; 21:1601–23.
51. Lu G, Ades AE. Modelling between-trial variance structure in mixed treatment comparisons. *Biostatistics*. 2009;10:792–805.
52. Raiffa H, Schlaiffer R. *Applied Statistical Decision Theory*. New York: Wiley Interscience; 1967.
53. Franchini A, Dias S, Ades AE, Jansen J, Welton N. Accounting for correlation in mixed treatment comparisons with multi-arm trials. *Res Syn Meth*. 2012;3:142–60.
54. Parmigiani G. *Modeling in Medical Decision Making: a Bayesian Approach*. San Francisco: Wiley; 2002.
55. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ*. 2005;14:339–47.
56. Ades AE, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health Econ*. 2005;14:373–81.
57. Briggs A. Handling uncertainty in economic evaluation. *Br Med J*. 1999;319:120.
58. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JPT. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *Int J Epidemiol*. 2012;41:818–27.
59. Hooper L, Summerbell CD, Higgins JPT, et al. Reduced or modified dietary fat for preventing cardiovascular disease. *Cochrane Database Syst Rev*. 2000;2:CD002137.
60. Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *Lancet*. 2007; 369:201–7.
61. Woolacott N, Hawkins NS, Mason A, et al. Etanercept and efalizumab for the treatment of psoriasis: a systematic review. *Health Technol Assess*. 2006;10(46):1–233.