

RESEARCH ARTICLE

Exploiting relationships between outcomes in Bayesian multivariate network meta-analysis with an application to relapsing-remitting multiple sclerosis

Ed Waddingham¹ | Paul M. Matthews¹ | Deborah Ashby²¹Division of Brain Sciences, Imperial College, London, UK²School of Public Health, Imperial College London, London, UK**Correspondence**

Ed Waddingham, Division of Brain Sciences, Imperial College, Burlington Danes, The Hammersmith Hospital, Du Cane Road, London W12 0NN, UK.
Email: e.waddingham@imperial.ac.uk

Funding information

Imperial College London, Grant/Award Number: Imperial College PhD Scholarship

In multivariate network meta-analysis (NMA), the piecemeal nature of the evidence base means that there may be treatment-outcome combinations for which no data is available. Most existing multivariate evidence synthesis models are either unable to estimate the missing treatment-outcome combinations, or can only do so under particularly strong assumptions, such as perfect between-study correlations between outcomes or constant effect size across outcomes. Many existing implementations are also limited to two treatments or two outcomes, or rely on model specification that is heavily tailored to the dimensions of the dataset. We present a Bayesian multivariate NMA model that estimates the missing treatment-outcome combinations via mappings between the population mean effects, while allowing the study-specific effects to be imperfectly correlated. The method is designed for aggregate-level data (rather than individual patient data) and is likely to be useful when modeling multiple sparsely reported outcomes, or when varying definitions of the same underlying outcome are adopted by different studies. We implement the model via a novel decomposition of the treatment effect variance, which can be specified efficiently for an arbitrary dataset given some basic assumptions regarding the correlation structure. The method is illustrated using data concerning the efficacy and liver-related safety of eight active treatments for relapsing-remitting multiple sclerosis. The results indicate that fingolimod and interferon beta-1b are the most efficacious treatments but also have some of the worst effects on liver safety. Dimethyl fumarate and glatiramer acetate perform reasonably on all of the efficacy and safety outcomes in the model.

KEYWORDS

correlated outcomes, evidence synthesis, multivariate network meta-analysis, relapsing-remitting multiple sclerosis, sparse data

1 | INTRODUCTION

Network meta-analysis (NMA) has established itself as a powerful evidence synthesis technique for combining summary data from multiple studies.¹⁻³ The technique is a generalization of standard (“pairwise”) meta-analysis that cannot only

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Statistics in Medicine* published by John Wiley & Sons, Ltd.

incorporate head-to-head studies comparing treatments directly, but also indirect evidence, that is, separate studies of each treatment against some other common comparator. The evidence base forms a connected network of treatments that are linked to one another either directly (via head-to-head studies) or indirectly (via a common comparator, or a chain of comparisons). NMA provides comparable treatment effect estimate for every treatment in the network, while pairwise meta-analysis usually only compares two treatments.

Meta-analyses are usually carried out on a single outcome measure of interest, but there are instances when multiple treatment outcomes must be estimated, including (but not limited to) regulatory benefit-risk assessments and health economic evaluations. Current practice in such situations would typically be to carry out separate NMAs for each outcome, but to do so is to ignore the possibility of correlations between outcomes, or in other words to make the (strong) assumption that outcomes occur independently. Where correlations do exist, this approach amounts to misspecification of the likelihood of the data, and will tend to distort the estimated standard errors of the study-specific conditional treatment effects. This in turn will impact on the ultimate population-average conditional estimates (which in essence are SE-weighted averages of the study-specific effects). Multivariate NMA models that allow for the correlations have therefore begun to appear in the literature. Correlations between outcomes can occur at the between-study level and the within-study level.^{4,5} Within-study correlations cannot usually be estimated from summary data, but they can be allowed for via assumed correlation coefficients (possibly combined with some form of sensitivity analysis). Allowing for correlations may be of particular relevance in situations where a decision variable is derived as a linear combination of outcomes, such as economic assessments or quantitative benefit-risk assessments,^{6,7} since the covariances between outcomes contribute additional terms to the variance of the composite variable.

Another problem with multivariate evidence synthesis can arise due to sparsity of the evidence base. Not all of the source studies may have data on every outcome of interest. In some cases there may be treatment-outcome combinations for which neither direct nor indirect evidence is available, rendering the corresponding parameters inestimable. Allowing for correlations with other outcomes is not sufficient to fill in these gaps, as it still does not allow us to estimate the posterior means. However, we can estimate the missing parameters by specifying between-outcome relationships that are assumed to remain fixed (or at least similar) for all treatments and studies. Often outcomes will be related via underlying clinical pathways; hence they are not only statistically correlated in terms of between- or within- study variability but are also related at the level of the population-average treatment effect parameters. For example, if outcome A and outcome B are manifestations of the same underlying disease activity, then one would expect that a drug which performs well against outcome A will also perform well against outcome B. The relationship may be even closer: if the outcomes are in fact, different measures of the same underlying clinical concept (one being, say, an annualized event rate, and one a proportion over the entire study duration), or two slightly different definitions of the same clinical concept (eg, proportions defined using different thresholds), then clearly one would expect a link between the population-level treatment effects. It is to be expected that many systematic reviews (especially those considering multiple outcomes) will identify studies with a variety of outcome definitions, too different to be reliably pooled for a meta-analysis. Instead of choosing between definitions and discarding the remaining data (or perhaps trying multiple variations thereon), the ability to simultaneously model the relationship between outcomes could be useful in terms of maximizing the information that can be gleaned from such data. Even if only one outcome measure is of ultimate interest, using structural relationships (rather than just correlations) to borrow strength from other closely related outcomes in a sparse evidence network could be a valuable way to enhance a meta-analysis, effectively “filling in” the gaps in the data.

These issues all became evident in our earlier work applying multicriteria decision techniques to health decisions, specifically in the context of regulatory benefit-risk assessment of medicines.⁷⁻¹⁰ More established fields such as health economics are also sometimes concerned with modeling multiple clinical outcomes. As interest in multioutcome decisions increases, multivariate evidence synthesis methods are bound to be called upon; and since sparsity of outcomes (or heterogeneity of definitions) is a known problem in the clinical literature,^{11,12} we expect that there will be a demand for methods to deal with the issues highlighted above.

1.1 | Literature on multivariate NMA

Univariate meta-analyses have long been familiar in health research, but multivariate meta-analysis was rarely attempted until this century. In 2007 Riley et al published a bivariate random-effects pairwise meta-analysis of a dataset comparing two periodontal treatments.⁶ The analysis was frequentist and used iterative least squares estimation. The random effects model used a typical hierarchical Normal structure whereby the means of outcome distributions at the

within-study level are themselves drawn from a distribution at the between-study level. Such a model typically requires data on the within-study variances and covariances in order to estimate the overall population average (and perhaps the between-study variances/covariances). The authors noted the difficulty in obtaining the necessary correlation/covariance estimates from published summary data and, in a follow-up article,¹³ presented an alternative model with a nonhierarchical structure (based on the marginal distribution of the outcomes in terms of the overall population average) in order to avoid this issue; they acknowledged, however, that this model had a less realistic structure and provided less accurate estimates of between-study heterogeneity. Bayesian modeling may provide an alternative method for dealing with unknown within-study correlations; the uncertainty could simply be reflected in the model via an appropriate vague prior.

A 2013 article by Bujkiewicz et al presented a comprehensive Bayesian model for multivariate meta-analysis, illustrated with an example from rheumatoid arthritis,¹⁴ but the model is only suitable for pairwise meta-analyses comparing two treatments and not for more complex evidence networks.

One of the first attempts at multivariate NMA was made in 2013 by Hong et al, who presented two alternative models applied to an osteoarthritis dataset.¹⁵ Their “arm-based” model has proved controversial,^{16,17} as it models the absolute treatment outcome in each arm—an approach which may be susceptible to confounding between the treatment effects and differences in characteristics of trial populations.¹⁸ Their “contrast-based” model follows the more conventional meta-analytical principle of modeling the relative effects between treatment arms; we found, however, that the parameterization used was not able to estimate the treatment effects in our dataset and does not allow for within-study correlations. In another article, Hong et al used a slightly different contrast-based multivariate NMA model with a more conventional parameterization but still did not account for within-study correlations.¹⁹ It has been argued that ignoring these correlations can have a substantial impact on the results.²⁰

Two related articles by Efthimiou et al proposed a model for bivariate NMA, applied to an acute mania dataset.^{21,22} The authors showed in principle how the model could be extended to a greater number of dimensions. No generalized code was provided, however.

Jackson et al set out a multivariate NMA model and a method-of-moments technique for estimating the key parameters; the authors acknowledge, however, that this estimation technique may lack efficiency.²³

Although the models above all make allowance for correlations between outcomes, they lack the capability to estimate any treatment effects that are completely missing from the dataset. As such these models cannot fill in the gaps in a sparse evidence network; rather, their “borrowing of strength” is limited to a reduction in posterior variance for those treatment-outcome combinations where data is available. In complex evidence networks involving multiple outcomes, some treatment-outcome combinations may not be represented in the data and to estimate these missing parameters, a model needs to incorporate some kind of structural link between the treatment effects that allows an indirect estimate to be made.

One approach to this problem is to consider how the outcomes can be logically linked within the context of the particular dataset under study. In a survival analysis context, for example, a 2-year survival probability p_2 and a 1-year survival probability p_1 can be linked by the equation $p_2 = p_1^2$ under the assumption of a constant hazard. Context-specific structural relationships such as this have previously been used in multivariate NMA to estimate missing treatment effects^{24,25} and this may often be a sensible approach when the logical relations between outcomes can be easily derived without making too many assumptions. There are likely to be occasions, however, when two or more outcomes are clearly (or possibly) related but the precise form of the relationship is not clear. For example, the effect of a treatment on the incidence of (i) myocardial infarction and (ii) stroke are likely to be related due to the shared cardiovascular pathways involved, but deriving an explicit link between the two would be very difficult. In order to exploit any such relationships, one may wish to link outcomes in a more generic way.

Achana et al proposed a model along these lines which assumes that the (additive) difference in the effect size for any two treatments is constant across all outcomes.²⁶ While a constant additive difference may be suitable in some situations, we believe that a constant proportional relationship is often likely to be more appropriate since effect sizes may vary considerably from one outcome to another. We propose therefore to use proportional mappings (ie, linear mappings with no intercept term) between the population-average conditional treatment effect parameters. This echoes a model used by Lu and Kounali and Ades et al,^{27,28} who carried out pairwise multivariate meta-analyses using similar proportional mappings. In both examples the mappings were applied to the study-level random treatment effects, since the underlying outcomes were believed to be perfectly correlated at the between-study level. Our approach allows for a somewhat looser proportional relationship that can be (to some extent) obscured by between-study heterogeneity. The model will also be applicable to more complex networks of interventions. A similar model has recently been published by Bujkiewicz et al,²⁹

but is restricted to two outcomes (with a focus on deriving surrogacy relations) and two-arm trials, while our model is intended to be applicable in the more general multivariate NMA setting.

1.2 | Aim of the research

The aim of this work is to develop a multivariate extension of NMA based on summary data from published clinical studies, that can (i) account for correlations between outcomes (at both the within- and between- study levels) in a flexible manner and (ii) effectively borrow strength between outcomes via mappings in order to allow the estimation of treatment effects that would otherwise remain inestimable. The models and code presented here are intended to be applicable to a wide range of datasets with minimal recoding required, and to allow for all structural correlations that may exist.

2 | DATASET: FIRST-LINE THERAPIES FOR RRMS

Relapsing-remitting multiple sclerosis (RRMS) is an autoimmune disorder with a wide range of disabling neurological symptoms, which come in the form of highly symptomatic episodes (relapses) with periods of remission in between. Remission is sometimes only partial, however, and patients generally undergo a gradual permanent worsening of overall disability in addition to the pattern of relapse and remission.³⁰ There is no cure for RRMS but a number of disease-modifying therapies exist that have varying effectiveness in controlling the symptoms and varying risk profiles. For a long time the standard first-line treatments have been injectable drugs, but recently a number of oral therapies have appeared on the market that show promise as first-line therapies, as they can be easily self-administered and are reasonably well tolerated: dimethyl fumarate, fingolimod, teriflunomide, and laquinimod. It is anticipated that many patients on the older injectable therapies will switch to one of these new drugs.³¹ However, there is a lack of direct trial evidence comparing these drugs.

A recent Cochrane review of RRMS therapies³² presented an NMA of these drugs alongside a number of other RRMS treatments (generally either older treatments, or those which are reserved for more aggressive disease cases due to their less acceptable risk profile). The review, while thoroughly assessing the available trials for inclusion, was limited to two basic efficacy outcomes (the proportion of patients avoiding relapse and disability progression) and one safety/acceptability outcome (the proportion of patients adhering to treatment), and did not make any attempt to allow for any correlations between these outcomes. Furthermore, there were different definitions of the efficacy outcomes among the source studies; where two definitions conflicted, either one version was dropped or both were presented as separate sparsely populated NMAs. We have based our dataset on the studies in the Cochrane review, but in order to illustrate our method's ability to model the relationship between similar outcomes, we have used an expanded set of efficacy outcomes and a set of safety outcomes concerning liver toxicity. We also restrict the treatment options to only the established first-line treatments and second-generation oral drugs specified above, and only the usual prescribed dosages. In total 16 source studies were used,^{33–48} providing data on eight active treatments. The precise outcome definitions, treatment list, and data table are given in the Supporting Information. A summary is shown in Table 1 and a network diagram for each outcome is shown in Figure 1.

3 | METHODS

Our model was developed in line with the general framework for contrast-based NMA models described by Dias et al⁴⁹ but uses a slightly different parameterization, which has been previously used in univariate NMA⁵⁰ but is here extended into the multivariate domain. We assume that, within each study arm, outcomes follow a multivariate normal distribution. Binary outcomes are thus incorporated on the logit scale, and count (Poisson) outcomes on the log rate scale.

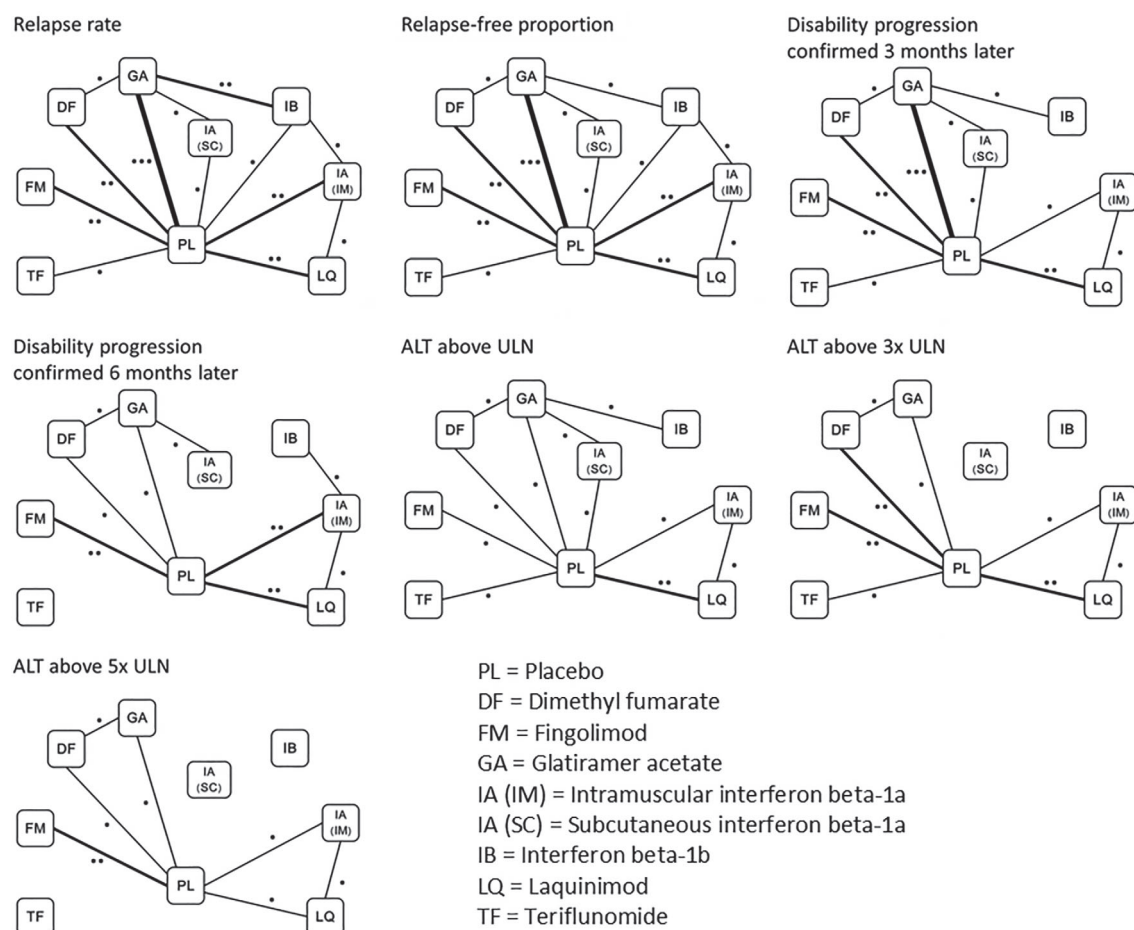
3.1 | Model structure

In what follows $t = 1, \dots, NT$ indexes treatments, $\omega = 1, \dots, \Omega$ indexes outcomes, $i = 1, \dots, NS$ indexes studies, $k = 1, \dots, NA_i$ indexes treatment arms within study i , and $j = 1, \dots, NO_i$ indexes outcomes in study i . $t = 1$ is the reference treatment

TABLE 1 Outcomes and data completeness for the relapsing-remitting multiple sclerosis case study

Outcome	Definition	No. studies reporting outcome	Treatments with data								
			DF	FM	GA	IA (IM)	IA (SC)	IBB	LQ	TF	
Relapse	Log annualized relapse rate	16	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Log odds of avoiding relapse	15	✓	✓	✓	✓	✓	✓	✓	✓	✓
Disability progression	Log odds of progressing; confirmed 3 months later	12	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Log odds of progressing; confirmed 6 months later	8	✓	✓	✓	✓	✓	✓	✓	✓	✓
Liver toxicity	Log odds of ALT above upper limit of normal range	9	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Log odds of ALT above 3× upper limit of normal range	7	✓	✓	✓	✓			✓	✓	✓
	Log odds of ALT above 5× upper limit of normal range	4	✓	✓	✓	✓			✓		

Abbreviations: ALT, alanine aminotransferase; DF, dimethyl fumarate; FM, fingolimod; GA, glatiramer acetate; IA (IM), intramuscular interferon beta-1a; IA (SC), subcutaneous interferon beta-1a; IB, interferon beta-1b; LQ, laquinimod; TF, teriflunomide.

**FIGURE 1** Network diagrams for the relapsing-remitting multiple sclerosis case study. A line between two treatments indicates the existence of direct trial evidence comparing them; the number of head-to-head trials is indicated by the number of dots (and the thickness of the line)

relative to which all the other treatments' effects are expressed. This will usually be a placebo/no treatment option. We use the notation t_{ik} to refer to the treatment in arm k of study i , and ω_{ij} to refer to the j th outcome in study i . The parameters of primary interest will usually be the treatment effects $d_{\omega t}$ representing the difference between the mean value of outcome ω on treatment t compared with its mean value on treatment 1. Within an individual study i , δ_{ijk} represents the effect (on ω_{ij} , the j th outcome in the study) of t_{ik} , the treatment in the k th arm of the study, again expressed relative to treatment 1, and (in the random effects model) has mean $d_{\omega_{ij}t_{ik}}$.

The formal specification of the model is set out below. The model is designed to work with trial data aggregated at the treatment arm level. The $NO_i \times NA_i$ -length vector $\mathbf{y}_i$ contains the observed means of outcomes 1,..., NO_i in arms 1,..., NA_i of study i . The convention we will follow is to index the elements of $\mathbf{y}_i$ (and other vectors of the same length) by pairs $(j,k) \in \{1, \dots, NO_i\} \times \{1, \dots, NA_i\}$ so that y_{ijk} refers to the observed mean of outcome j in arm k of study i . A multivariate Normal likelihood is assigned to $\mathbf{y}_i$ as follows:

$$\mathbf{y}_i \sim MVN(\boldsymbol{\mu}_i + \boldsymbol{\delta}_i^*, \mathbf{CV}_i), \quad (1)$$

where $\boldsymbol{\delta}_i^*$ is a vector of length $NO_i \times NA_i$ whose (j,k) th element is given by $\delta_{ijk}^* = \delta_{ijk} - \frac{1}{NA_i} \sum_{m=1}^{NA_i} \delta_{ijm}$ and $\boldsymbol{\delta}_i$ is the vector of length $NO_i \times NA_i$ containing the effects on outcomes $j = 1, \dots, NO_i$ in treatment arms $k = 1, \dots, NA_i$ of study i relative to treatment 1, $\boldsymbol{\mu}_i$ is the study-specific NO_i -length vector of outcome values as averages across all treatment arms in study i , and $\mathbf{CV}_i$ is the within-study covariance matrix. Outcomes occur in one treatment arm independently of the other arms, so (if the elements of $\mathbf{y}_i$ are ordered lexicographically, advancing at the first level through the outcomes j , then at the second level through the arms k) $\mathbf{CV}_i$ takes block diagonal form, with zero covariances between treatment arms but nonzero covariances for different outcomes within each arm. Within each arm or block, the diagonal terms (within-study outcome variances) are provided by the data. The off-diagonal terms, however, cannot be estimated from typical arm-level summary data; we construct them within the model by using assumed values for the within-study correlations. Initially we assume these correlations are equal to a fixed value ρ_w for all pairs of outcomes; later we look at relaxing this assumption. Other strategies for dealing with the within-study correlations have been discussed elsewhere.²¹

This specification makes use of a parameterization that has previously been proposed (albeit in a univariate context)⁵⁰ where the nuisance parameter μ_{ij} represents the average value of outcome j across all arms of study i . That is, $\mu_{ij} = \frac{1}{NA_i} \sum_k y_{ijk}$.

It may be helpful to further illustrate this parameterization by means of some simple examples. In a two-arm study i the mean of outcome j in each arm is as follows:

$$\text{Arm 1: } \mu_{ij} + \delta_{ij1} - \frac{1}{2}(\delta_{ij1} + \delta_{ij2}) = \mu_{ij} + \frac{1}{2}\delta_{ij1} - \frac{1}{2}\delta_{ij2}.$$

$$\text{Arm 2: } \mu_{ij} + \delta_{ij2} - \frac{1}{2}(\delta_{ij1} + \delta_{ij2}) = \mu_{ij} - \frac{1}{2}\delta_{ij1} + \frac{1}{2}\delta_{ij2}.$$

In a three-arm study i the mean of outcome j in each arm is as follows:

$$\text{Arm 1: } \mu_{ij} + \frac{2}{3}\delta_{ij1} - \frac{1}{3}\delta_{ij2} - \frac{1}{3}\delta_{ij3}$$

$$\text{Arm 2: } \mu_{ij} - \frac{1}{3}\delta_{ij1} + \frac{2}{3}\delta_{ij2} - \frac{1}{3}\delta_{ij3}$$

$$\text{Arm 3: } \mu_{ij} - \frac{1}{3}\delta_{ij1} - \frac{1}{3}\delta_{ij2} + \frac{2}{3}\delta_{ij3}$$

In either case (and in general) it can be seen that $\delta_{ijk_2} - \delta_{ijk_1}$ corresponds to the treatment effect in arm k_2 relative to arm k_1 and is thus estimated by $y_{ijk_2} - y_{ijk_1}$.

It is more common in NMA to use a different parameterization where the mean of the distribution in (1) is set to $\boldsymbol{\mu}_i + \boldsymbol{\delta}_i$, with δ_{ijk} set to zero for $k = 1$ rather than for $t_{ik} = 1$; this means that the $\boldsymbol{\mu}_i$ corresponds to the outcome values in the arbitrary selected first arm of each trial, and the δ_{ijk} ($k > 1$) are expressed relative to this first arm. For example, under this parameterization, in a two-arm study i the mean of outcome j is μ_{ij} in arm 1 and $\mu_{ij} + \delta_{ij2}$ in arm 2.

The key difference between the two parameterizations emerges when one considers the (prior) variance of the mean of y_{ijk} . Consider, for example, a two-arm study. Under the usual NMA parameterization, the variance of the mean of outcome j is $\text{var}(\mu_{ij})$ in arm 1 and $\text{var}(\mu_{ij}) + \text{var}(\delta_{ij2})$ in arm 2. In other words the variance is always higher in arm 2 than in arm 1, even though the numbering of the arms (at least in studies with no placebo arm) is arbitrary.

Under our parameterization, by contrast, the variance is equal in all study arms. In the two-arm study example, the variance of the mean of outcome j is $\text{var}(\mu_{ij}) + \text{var}\left(\frac{1}{2}\delta_{ij1}\right) + \text{var}\left(\frac{1}{2}\delta_{ij2}\right)$ in both arms 1 and 2.

In the random effects version of the model, the study-specific treatment effects $\boldsymbol{\delta}$ also follow a multivariate normal distribution in order to allow for between-study correlations. Here, however, in addition to the correlations between

outcomes, there are also correlations between treatment arms, as the δ estimates obtained within different arms of a trial are linked by their common baseline. The correlation between δ in different trial arms is usually taken to be 0.5, under the assumption of equal variance of outcome across trial arms.⁵⁰ If the (between-study) correlation of different outcomes in the same trial arm is assumed to be ρ_b , then the correlation of different outcomes in different trial arms is the product $0.5\rho_b$.^{21,27,28} Initially we assume that ρ_b is a constant for all pairs of outcomes; later we will show how this can be extended to a broader class of correlation structures.

Our parameterization requires that the treatment effect $\delta_{ijk} = 0$ whenever $t_{ik} = 1$, since by definition the treatment effects are expressed relative to the reference treatment. We therefore override the random effects (which will generally be nonzero) with zeroes in study arms where the reference treatment was used. This is achieved by defining the random effects distribution on an accessory variable θ_i with the same dimensions as δ_i , then setting $\delta_{ijk} = 0$ for $t_{ik} = 1$ and $\delta_{ijk} = \theta_{ijk}$ for all other values of t_{ik} .

Let the population mean effect of treatment t on outcome ω be denoted by $d_{\omega t}$ (expressed relative to treatment 1, so $d_{\omega 1} = 0$ for all outcomes ω). Then we have:

$$\theta_i \sim MVN(\mathbf{d}_i, \Sigma_i) \quad (2)$$

where $\mathbf{d}_i$ is a vector of length $NO_i \times NA_i$ (also indexed by pairs $(j, k) \in \{1, \dots, NO_i\} \times \{1, \dots, NA_i\}$) that picks out the appropriate mean treatment effects for the outcomes and treatments in study i (ie, the (j, k) th element of $\mathbf{d}_i$ is $d_{\omega_{jk} t_{ik}}$) and Σ_i is a $(NO_i \times NA_i) \times (NO_i \times NA_i)$ between-study treatment effects covariance matrix. The diagonal elements of Σ_i are equal to the random-effects variance σ^2 (assumed equal for all outcomes and treatments) and the off-diagonal elements are equal to either $0.5\sigma^2$ (same j different k), $\rho_b\sigma^2$ (different j same k), or $0.5\rho_b\sigma^2$ (different j different k). The Supporting Information shows the form of Σ_i based on a lexicographic ordering of the elements of θ_i .

For the fixed-effect model, we simply set $\delta_{ijk} = d_{\omega_{jk} t_{ik}}$.

3.2 | Coding via variance decomposition

Bayesian NMAs to date have mostly been carried out in either the WinBUGS or OpenBUGS software programs. In existing versions of these programs, however, implementing an indexed multivariate normal distribution with arbitrary dimensions poses difficulties, which is one reason why existing multivariate evidence syntheses have relied on some degree of hard-coding of data features within the model code. Of the generalizable Bayesian models that fully allow for correlations, only Lu and Kounali,²⁷ Ades et al,²⁸ Efthimiou et al,²² and Achana et al²⁶ provide BUGS code—but their code constructs covariance matrices according to the dimensions of the dataset, and hence would require rewriting for a new dataset. We avoid this problem by taking advantage of the particular correlation structure within Σ . Assuming for now that $0 \leq \rho_b \leq 1$, a consequence is that the random element of δ_{ijk} decomposes proportionally into the following components:

- a fixed proportion $0.5\rho_b$ that is shared with $\delta_{ij'k'}$ for all j', k' ;
- an additional proportion $(\rho_b - 0.5\rho_b) = 0.5\rho_b$ that is shared with $\delta_{ij'k}$ for all j' ;
- an additional proportion $(0.5 - 0.5\rho_b)$ that is shared with $\delta_{ijk'}$ for all k' ; and
- a remaining proportion of $1 - 0.5\rho_b - (0.5 - 0.5\rho_b) - 0.5\rho_b = 0.5 - 0.5\rho_b$ that is unique to δ_{ijk} .

We can use this to express the multivariate normal distribution as a combination of independent univariate normal distributions. (2) is equivalent to

$$\delta_{ijk} \sim N(d_{\omega_{jk} t_{ik}} + E_i + F_{ik} + G_{ij}, (0.5 - 0.5\rho_b)\sigma^2), \quad (3)$$

where $E_i \sim N(0, 0.5\rho_b\sigma^2)$ is component (i) of the variation in δ_{ijk} , $F_{ik} \sim N(0, (\rho_b - 0.5\rho_b)\sigma^2)$ is component (ii), $G_{ij} \sim N(0, (0.5 - 0.5\rho_b)\sigma^2)$ is component (iii), and the remaining variance $(0.5 - 0.5\rho_b)\sigma^2$ is component (iv).

Equivalently, we can set $E_i, F_{ik}, G_{ij} \sim N(0, 1)$ and rescale them within the definition of δ_{ijk} by multiplying by the required SD (ie, $\sqrt{(0.5\rho_b\sigma^2)}$, and so on). This way the variance can be allowed to vary by arm or by outcome, if desired. We choose to keep σ^2 constant in this instance, but apply a similar decomposition to the multivariate within-trial distribution of $\mathbf{y}_{ik}$ shown in (1), with a SD drawn from the data that varies by arm and by outcome.

This method allows us to avoid explicitly specifying the covariance matrices $\mathbf{CV}_i$ and Σ_i or using the standard BUGS multivariate normal distribution when computing the likelihood. Our code does, however, include calculation of $\mathbf{CV}_i$ in order to compute the mean residual deviance $\bar{D}$, effective number of parameters p_D and deviance information criterion DIC, which are used to assess model fit and complexity.⁵¹

The residual deviance $D(\boldsymbol{\varphi})$, where $\boldsymbol{\varphi}$ represents the set of parameters of interest to be estimated, is a standardized measure of model fit, calculated for the multivariate Normal likelihood as

$$D(\boldsymbol{\varphi}) = \sum_i (\mathbf{y}_i - E[\mathbf{Y}_i|\boldsymbol{\varphi}]) \mathbf{CV}_i^{-1} (\mathbf{y}_i - E[\mathbf{Y}_i|\boldsymbol{\varphi}]),$$

where $\mathbf{CV}_i^{-1}$ is the within-study coprecision matrix, the inverse of $\mathbf{CV}_i$.

The effective number of parameters p_D measures the complexity of the model and is calculated as the mean residual deviance less the residual deviance evaluated at the mean, that is, $p_D = \bar{D}(\boldsymbol{\varphi}) - D(\bar{\boldsymbol{\varphi}})$. The deviance information DIC = $\bar{D}(\boldsymbol{\varphi}) + p_D$ is used to compare models in terms of both their fit and complexity; models with lower DIC are favored.

See the Supporting Information for more details and proofs of the key results in this section.

3.3 | Mappings

The final element of the model provides structural links between the mean treatment effects $d_{\omega t}$ by means of linear mappings between different outcomes. We use straightforward multiplicative mappings with no additive/intercept terms, which seems appropriate as a null treatment would have zero effect on any outcome. Implicitly this model assumes that if a treatment has a nonzero effect on any one outcome then it must have a nonzero effect on all outcomes, and the model's applicability is therefore limited to situations where this assumption holds (see the Discussion for further comments on this point). The approach is similar to that employed by Lu and Kounali and Ades et al,^{27,28} except that their mappings were applied to the study-specific random effects δ . The mappings help to borrow strength in sparse data networks, given the fairly strong assumption that the treatment effects on different outcomes occur in consistent proportions. If preferred, they can be omitted from the model, and each $d_{\omega t}$ simply given an identical prior.

To assess the validity of the proportionality assumption in the RRMS dataset, two-way plots were drawn of the posterior mean treatment effects $d_{\omega t}$ estimated in a univariate NMA model for each outcome. These suggest that the majority of estimates did occur in broadly proportionate fashion, albeit imperfectly and with several clear outliers. See the Supporting Information for details.

For the “random-mapping” model, mappings are allowed to vary between treatments (but remain broadly similar), amounting to a relaxed version of the proportionality assumption. For $\omega > 1$, the mapping equation for treatment t is $d_{\omega t} = \beta_{\omega t} d_{1t}$ where $\beta_{\omega t}$ maps the treatment's effect on outcome 1 to its effect on outcome ω .

In a similar manner to the model of Lu and Kounali and Ades et al,^{27,28} we only allow the absolute value of the mappings to be random—in other words, the sign of each treatment effect $d_{\omega t}$ is known in advance for each outcome ω and is the same for all treatments. When the reference treatment is a placebo/no treatment option this should not usually be too restrictive, at least for treatments that have already been shown to be viable in pivotal trials, as active treatments will tend to have a nonnegative effect on efficacy and a negative effect on safety.

It is necessary to define a distribution by which mappings vary between treatments. We choose to define this distribution on the logarithmic scale—that is, by specifying a distribution for $\log(\beta_{\omega t})$ —for two reasons. The first relates to the variability: a constant SD of mappings on the log scale corresponds to linear mappings with SDs proportional to their magnitude, which has intuitive mathematical appeal and prevents the distribution of low-valued mappings from taking negative values.

The second reason for the log transformation relates to correlations between mappings. For outcomes $\omega_1, \omega_2 > 1$, the estimated mappings $\widehat{\beta}_{\omega_1 t}$ and $\widehat{\beta}_{\omega_2 t}$ will be correlated (as t varies) since for a given value of t , they are estimated by the absolute ratios $|\widehat{d}_{\omega_1 t}/\widehat{d}_{1t}|$ and $|\widehat{d}_{\omega_2 t}/\widehat{d}_{1t}|$, respectively, and thus share a common denominator $\widehat{d}_{1t}$ (itself essentially a weighted average of random-effects estimates $\widehat{\delta}$, which are linear differences between observed data points). The correlations among the mappings are more easily expressed on the logarithmic scale, which replaces the ratios with linear differences, that is, $\log(\widehat{\beta}_{\omega t}) = \log(|\widehat{d}_{\omega t}|) - \log(|\widehat{d}_{1t}|)$. Still, the precise correlation coefficients depend on the relative variances of the estimates $\log(|\widehat{d}_{\omega t}|)$, which depend not only on the network structure but also on the magnitude of $\widehat{d}_{\omega t}$, and thus cannot be specified a priori. However, under the assumption that $\log(|\widehat{d}_{\omega t}|)$ is of equal variance for different

values of ω , the correlation between $\log(\widehat{\beta}_{\omega_1 t})$ and $\log(\widehat{\beta}_{\omega_2 t})$ will be 0.5 on average, and we suggest that this is a reasonable approximation.

We therefore assign a multivariate Normal distribution to the treatment-specific log-mappings, that is, $\log(\beta_t) \sim \text{MVN}(\log(\mathbf{b}), \mathbf{Q})$ where $\mathbf{b} = \{b_2, b_3, \dots, b_\Omega\}$ is the vector of average mappings and $\mathbf{Q}$ is a covariance matrix with diagonal terms equal to the between-treatment mapping variance σ_{map}^2 and off-diagonal terms equal to $0.5 \cdot \sigma_{\text{map}}^2$.

For the “fixed-mapping” model, $\beta_{\omega t}$ simply equals b_ω , the mappings are the same regardless of treatment, and the average treatment effects $d_{\omega t}$ are kept in strict proportion.

As the assumption of proportionality is a strong one, it may be desirable to only map between certain subgroups of outcomes that are especially closely related, rather than between all outcomes simultaneously. We have developed a version of the model that allows for this: the outcomes are partitioned into groups, specified in the data, and within each group G the mapping equation for treatment t and outcome $\omega \in G$ is $d_{\omega t} = \beta_{\omega t} d_{\omega_G^* t}$ where ω_G^* is the reference outcome in group G and $\beta_{\omega t}$ maps the treatment’s effect on outcome ω_G^* to its effect on outcome ω . In this way, mapping relations are assumed to hold between outcomes in the same group but not between outcomes in different groups.

We ran models with the RRMS case study outcomes grouped in the following ways:

- One-group model: all outcomes grouped together
- Two-group model: all efficacy outcomes grouped together, all liver-safety outcomes grouped together
- Three-group model: both relapse outcomes grouped together, both disability progression outcomes grouped together, all liver-safety outcomes grouped together
- For comparison purposes, a model with no mappings was also evaluated (alternatively, this might be thought of as a model with each outcome in its own group).

Note that the groupings apply only to the mappings; they do not impose any restrictions on the between- or within-study correlations. Indeed, although in reality one might expect the mappings and correlations among a set of outcomes to be related by some underlying mechanism, no attempt is made to link the two in the model; the correlation structure is specified independently of the mapping structure and vice versa. For example, the model with no mappings is still able to incorporate correlations between outcomes at the between- and within-study levels.

In theory outcome measures could, if desired, be transformed to different scales for mapping purposes, that is, the mapping equation can be rewritten as $f(d_{\omega t}) = \beta_{\omega t} g(d_{\omega_G^* t})$ for appropriate link functions f and g , but for the purpose of this article we simply assume that mappings apply to the untransformed treatment effects.

3.4 | Extension to a more general correlation structure

The above decomposition (3) of the random effects multivariate normal distribution (2) assumes a universal nonnegative correlation coefficient between all pairs of outcomes. With a slightly modified approach, one can incorporate a broader class of correlation structures, with correlation coefficients that vary in sign and magnitude between pairs of outcomes. This applies equally to the between-study correlation ρ_b and the within-study correlation ρ_w —hereafter in this section we simply use ρ to refer to either coefficient. If ρ is allowed to take vector form over the outcomes $1, \dots, \Omega$, then a similar construction to (3), set out in the Supporting Information, gives a correlation between outcomes ω_1 and ω_2 in the same trial arm equal to

$$\text{sign}(\rho_{j_1}) \text{sign}(\rho_{j_2}) \sqrt{|\rho_{j_1} \rho_{j_2}|}$$

where $\text{sign}(x) = \pm 1$ according to the sign of x .

In other words, in absolute value terms the correlation is the geometric mean of ρ_{j_1} and ρ_{j_2} , and its sign depends on whether the signs of ρ_{j_1} and ρ_{j_2} agree. ρ_j is no longer a correlation coefficient in the strict sense, but it can be thought of as the propensity of outcome j to correlate with other outcomes. Two outcomes with propensities of high magnitude will be highly correlated, two outcomes with propensities of low magnitude will be scarcely correlated, and one of each results in a moderate correlation. This restricts us to a class of correlation structures that are particularly even-handed in the sense that each outcome blindly shares its “propensity to correlate” equally with all other outcomes, weighted only by their own respective propensities; an outcome cannot selectively favor any particular others for correlation. In particular,

if any two outcomes j_1 and j_2 are to be uncorrelated, then we must have $\rho_{j_1} = 0$ or $\rho_{j_2} = 0$, and so at least one of them must be uncorrelated with all other outcomes.

This extension allows us to incorporate negative correlations between certain outcomes if desired, but note that is not possible for all outcomes to be negatively correlated with one another using this method (unless there are only two outcomes in total); rather, the outcomes are partitioned into two sets with positive intraset correlations and negative interset correlations. We consider this flexible enough for most purposes: the scenario where all pairs of outcomes are negatively correlated seems somewhat improbable.

One advantage of this restricted correlation structure is that it is a sufficient (but not a necessary) condition for the covariance matrix to be positive-definite, which is a fundamental requirement for a multivariate Normal distribution. We offer a proof of this result in the Supporting Information, together with model code that applies this structure to both the within- and between- study correlations.

By assigning vague priors to the correlation propensities this construction allows us to be relatively uninformative about the overall correlation structure. The “LKJ” prior has also been suggested as a vague prior on the correlation structure⁵² but this may be difficult to implement in the BUGS language, especially in arbitrary dimensions.

3.5 | Priors

The prior distributions in general are intended to be noninformative. μ_{ij} and d_{1t} are given Normal priors with mean 0 and variance 1000. The mapping parameters b_ω ($\omega > 1$) are given Normal priors with mean 0 and variance 100 (the lower variance here results in a significantly faster run time and does not appear to have much impact on the posterior distribution), and the absolute value is then taken to ensure $b_\omega > 0$. For the variability parameters, we use the same priors that have been suggested elsewhere,^{27,28} namely, the random effects SD σ is given a uniform prior over the interval (0,10) and the precision of the mappings is given a Gamma (0.005, 0.005) prior restricted to values above 1.

Our main results are based on separate *Uniform* (−0.9, 0.9) priors for the within- and between- study correlation propensity for each outcome. This interval was used because use of the full (−1, 1) interval resulted in the between-study correlations being estimated as either 1 or −1 (a result which has also been noted elsewhere¹³ and it seems implausible that the RRMS outcomes would be so perfectly correlated. We also performed sensitivity analyses where the between- and within-study correlation propensities for each outcome were set equal and assigned a *Uniform* (−0.9, 0.9) prior; and alternatively assuming fixed correlations of 0, 0.3, 0.6, and 0.9 between all outcomes (see the Supporting Information for more details).

3.6 | Rankings

Treatments can be ranked in terms of their effect $d_{\omega t}$ on a given outcome ω . For ranking purposes each outcome is either positive (higher $d_{\omega t}$ is better) or negative (lower $d_{\omega t}$ is better) depending on the context, and this additional information can be supplied in the data. We calculate treatment rankings for each outcome within the model and use the posterior distributions of the rankings to derive the surface under the cumulative ranking curve (SUCRA) statistic,⁵³ which summarizes each treatment’s overall rating for each outcome as a score between 0 and 1 where 0 is a treatment that is always outranked by all others and 1 is a treatment that always outranks all others.

4 | RESULTS

Models were simulated via the Markov Chain Monte Carlo technique using the OpenBUGS software (version 3.2.2 rev 1063). 100 000 iterations were discarded to allow for parameters to converge; and posterior statistics were derived from a further 100 000 iterations using a single Markov chain. Convergence was monitored by inspecting the chain histories and posterior density plots.

A number of variations on the model were fitted-fixed and random effects, fixed and random mappings, and various assumed correlation structures and outcome groupings. Convergence was not achieved in all cases. Here,

we focus on the random effects random mappings model with separate vague priors on each outcome's between- and within-study correlation propensity, and with mappings applied in two groups, as this model resulted in the lowest DIC value out of the models that converged. Results for alternative models are shown in the Supporting Information.

Figure 2 summarizes the posterior distributions of the key parameters and the mean residual deviance for the main model. Figure 3 shows the results from a univariate analysis (with no correlations between outcomes and no mappings) for comparison.

Graphs of the SUCRA statistics by outcome and treatment are shown in Figure 4 for the main model with (i) full allowance for between- and within-study correlations, (ii) between-study correlations only, (iii) no between- or within-study correlations.

The main model fits well, with a mean residual deviance of 131.6, which compares favorably to 152, the number of observations. By contrast, the univariate model with no mappings resulted in a mean residual deviance of 154.8.

If one wishes to see more detail of the underlying rankings, rather than the high-level summary provided by SUCRA, an alternative visualization is suggested in Figure 5, showing for each treatment the proportion of posterior samples in which the treatment was at each ranking level (for a particular outcome, in this case the annualized relapse rate—similar graphs for the other outcomes are available in the Supporting Information).

Comparing Figures 2 and 3 allows us to see the influence that the mappings and correlations have upon the evidence synthesis. The model without mappings is unable to produce data-based estimates for the treatment effects that were not reported in any trials (and hence in this model the treatments cannot all be ranked with regard to every outcome). The models with mappings allow these parameters to be estimated, as shown by the white markers and dashed lines in Figure 2. For the parameters with nonmissing data, there is a noticeable reduction in the posterior SDs when mappings and correlations are used. The mappings also introduce a shift in some of the estimated means, bringing the treatment effect estimates more closely in line with the proportionality assumption. This illustrates the importance of giving careful consideration to the grouping strategy in order to make sure proportionality seems reasonable. Gathering the outcomes into a greater number of groups within which the outcomes are more closely related is the obvious strategy to follow, and this seems borne out by the model fit statistics in some of the alternative models we tried (see the Supporting Information). This approach only worked up to a point, however, as the models with the greatest number of groups (three) were the least likely to achieve convergence, presumably because there were too few observations on which to base the mappings.

In general, the choice between fixed/random mappings, and the grouping of outcomes, is dependent on context and may be influenced by a number of factors—such as the degree of sparsity in the data, how closely related the outcomes are perceived to be, the estimated SD of mappings, and which outcome(s) from the model will ultimately be used for decision making/modeling purposes (and which are merely there for borrowing strength). More experience with different datasets is needed before any particular model choice and/or grouping strategy can be recommended. One special case is worth highlighting, however: in the fixed-mapping model, the assumption of constant proportionality means that outcomes in the same mapping group give equivalent treatment rankings. This may be desirable if, for example, we would like to combine the treatment effects within a group into a single measure for decision-making purposes: rather than having to come up with an aggregation rule, we simply have to pick an arbitrary outcome within each group and the rankings are equivalent no matter which one we choose. In the three-group fixed-mapping model, for instance, we obtain a ranking for relapse, a ranking for disability progression, and a ranking for liver outcomes, all of them insensitive to the choice of specific outcome measure. If, on the other hand, we want the model to reflect that specific outcomes within a group may occur in different proportions on different treatments, then the random-mapping model must be used instead.

Deciding which treatment is best overall requires subjective judgments as to which form of the model is most appropriate and the relative importance of the included outcomes. The following observations can, however, be made:

- Fingolimod performs extremely well on efficacy outcomes, especially relapses, but is one of the worst performers in terms of liver safety;
- interferon beta-1b performs well on all efficacy measures but poorly on liver safety;
- dimethyl fumarate and glatiramer acetate are both good all-round performers, with dimethyl fumarate looking slightly better in terms of efficacy but glatiramer acetate winning on liver safety (only placebo appears safer);
- intramuscular interferon beta-1a performs well on liver safety but poorly on efficacy.

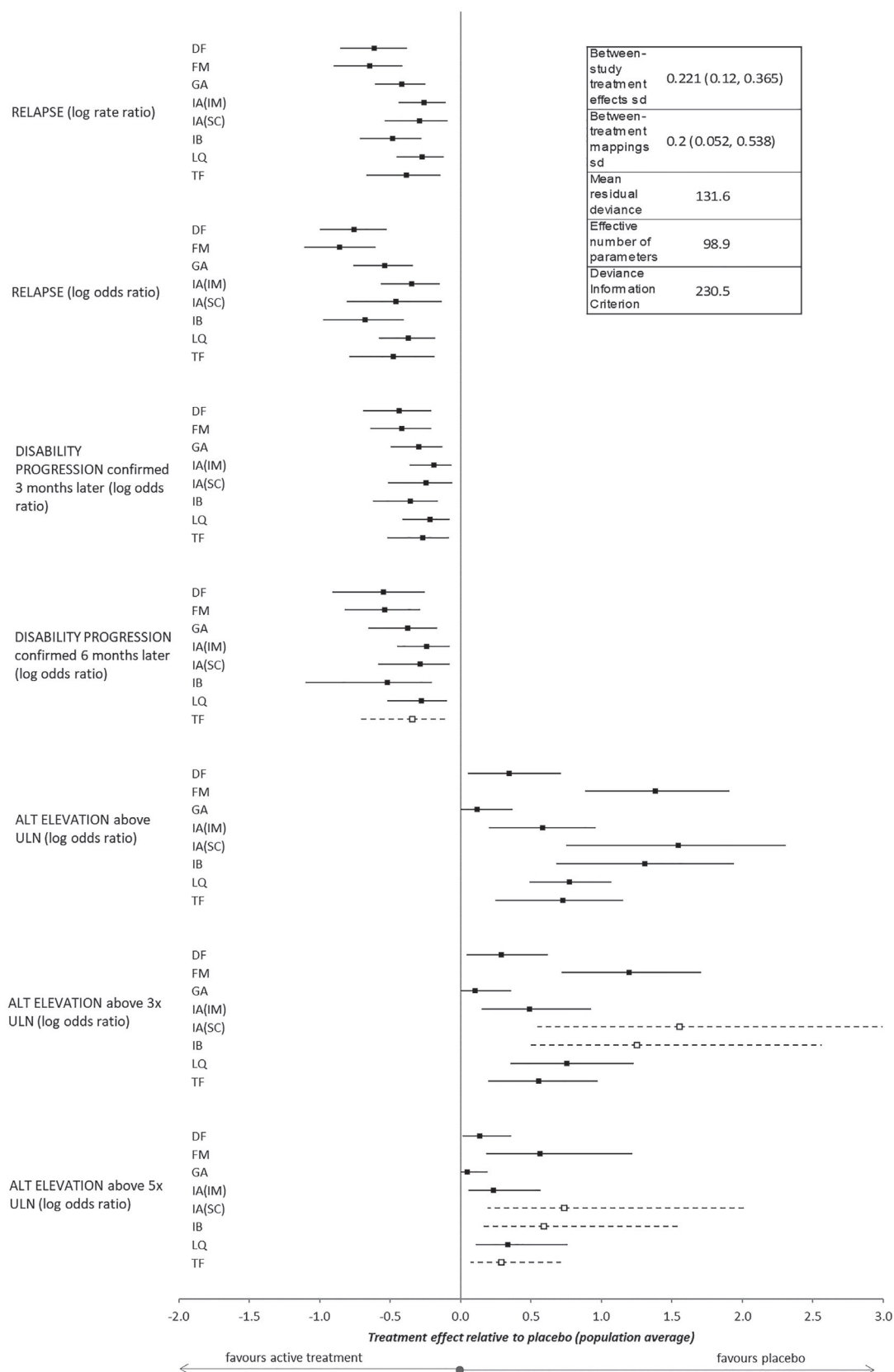


FIGURE 2 Forest plot summarizing the posterior distributions of the population-average treatment effect parameters in the main model (random effects and random mappings in two groups). Markers indicate means and lines indicate 95% credibility intervals. Solid markers and lines are treatment-outcome combinations for which data was available; white markers and dashed lines are estimated by mappings

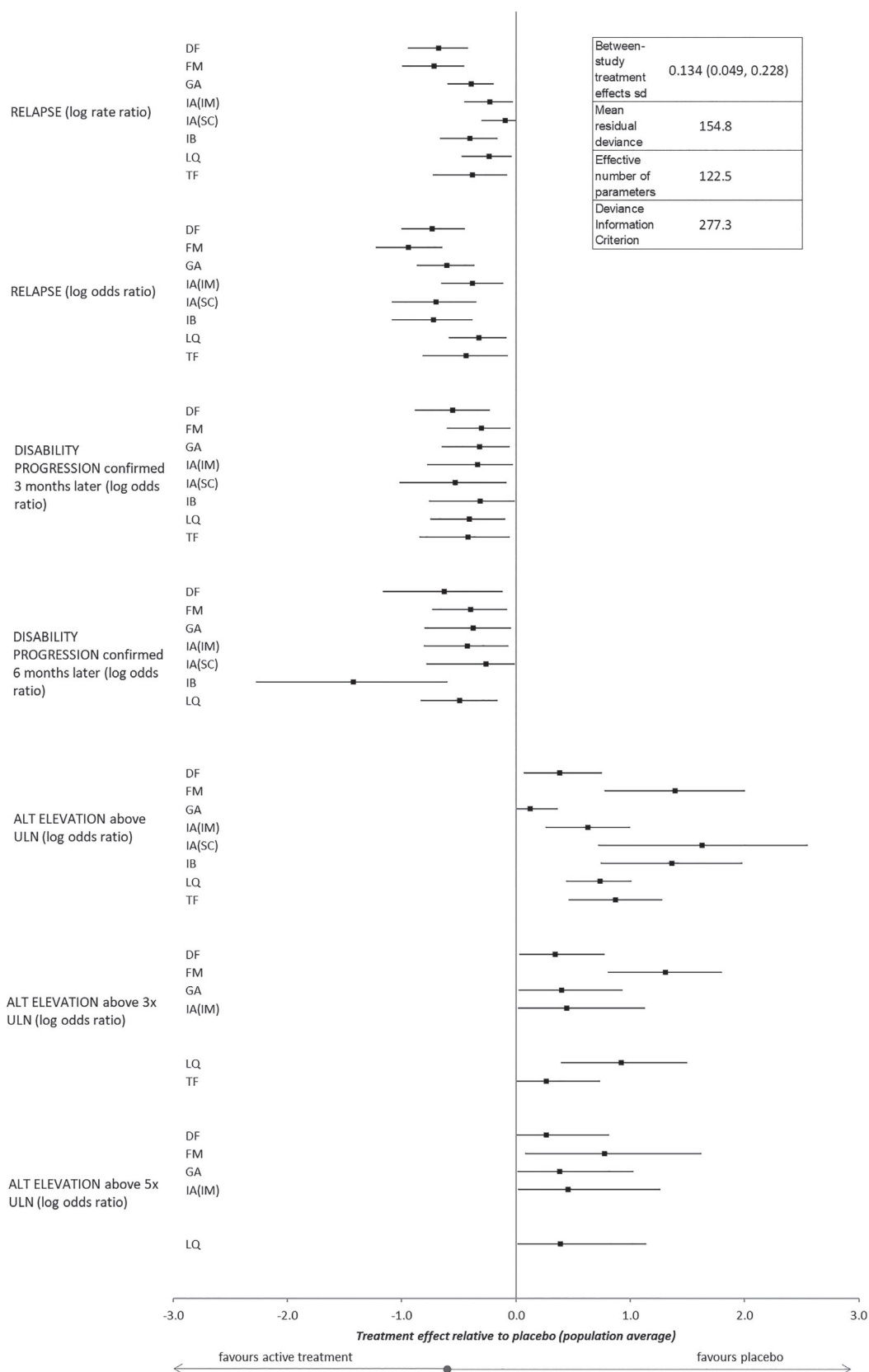


FIGURE 3 Forest plot summarizing the posterior distributions of the population-average treatment effect parameters in a univariate random effects model with no mappings. Markers indicate means and lines indicate 95% credibility intervals

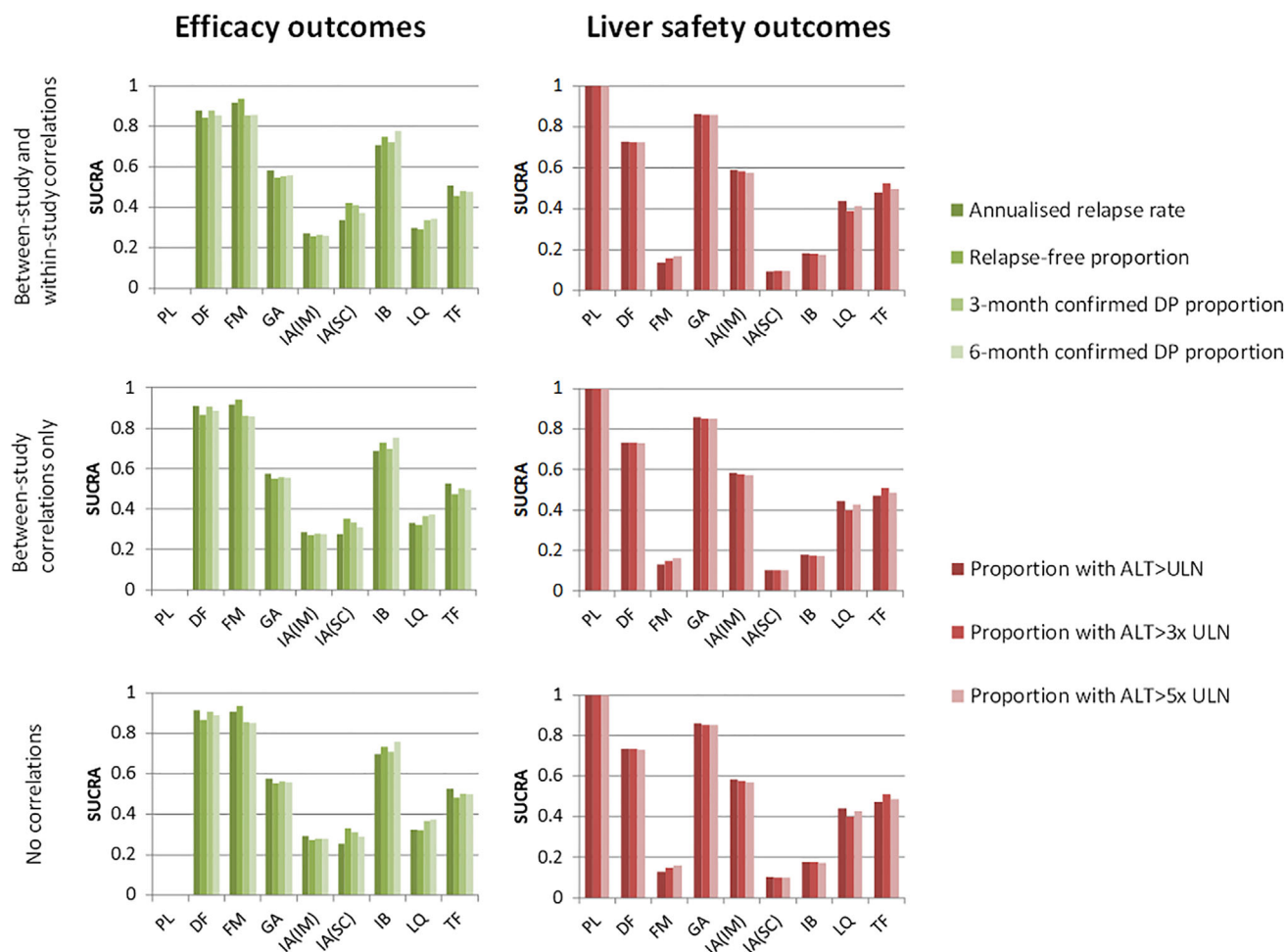


FIGURE 4 SUCRA statistic by treatment and outcome for the main model with and without allowance for outcome correlations. SUCRA, surface under the cumulative ranking curve [Colour figure can be viewed at wileyonlinelibrary.com]

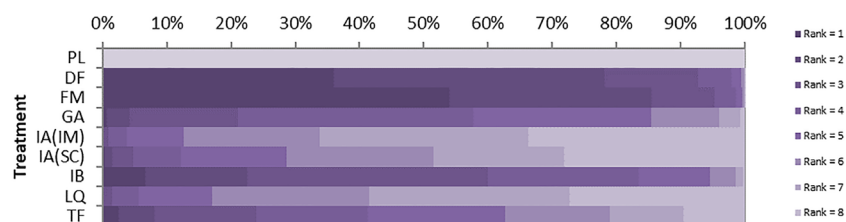


FIGURE 5 Posterior rankings for annualized relapse rate in the main model. The size of the colored bars indicates the proportion of simulations each treatment spent at each ranking position [Colour figure can be viewed at wileyonlinelibrary.com]

The efficacy findings are broadly in line with the results of the Cochrane review³² but there are some differences. Most notably, glatiramer acetate appears less effective in our analysis and dimethyl fumarate appears more effective, even when the analysis is restricted to the Cochrane outcomes and mappings/correlations are not used. The reasons for this are not fully clear but may be a reflection of the differing approaches to nonstandard dosages: we dropped study arms that did not use the normal dosage whereas the Cochrane review pooled all dosages for the same treatment, and this will have impacted on some of the comparisons involving glatiramer acetate and dimethyl fumarate. In terms of safety, our rankings are surprisingly similar to those of the Cochrane review, given that very different outcomes were used: the Cochrane review looked at discontinuation due to any adverse events, whereas we focused on biomarkers for liver damage; the similarity of the results may be due to chance or could perhaps indicate that elevated liver enzymes can be a reasonable proxy for lack of tolerability in a more general sense. It is worth noting, however, that another recent review of RRMS drugs, based on observational data in the postmarketing setting,⁵⁴ found that patient persistence on treatment was highest for fingolimod and other very effective drugs, suggesting that MS patients may value efficacy over safety.

Figure 4 shows that the overall picture of the SUCRA statistics stays broadly similar with or without the allowance for correlations. There are some changes in the order of treatment preference with respect to specific outcomes, but only in fairly marginal cases (consider, eg, a choice between the two forms of interferon beta-1a and teriflunomide with respect to the relapse rate). Indeed, the treatment effects themselves do not appear to change substantially when correlations are included/excluded (see the Supporting Information), but this may not always be the case in other models or datasets.

Results of the sensitivity analysis to the assumed correlation parameters and mapping structures are shown in the Supporting Information. Broadly speaking, the conclusions based on the SUCRA statistics were very similar for any model using vague priors on the correlation propensities. When fixed positive correlation coefficients were assumed, some of the rankings and treatment effects were more noticeably altered; however, such a correlation structure seems unrealistic in this example given that negative correlations would be expected between some pairs of outcomes (such as the relapse rate and relapse-free proportion).

5 | DISCUSSION

More experience with multivariate NMA models is needed in order to understand their strengths and any associated pitfalls. This is a highly active area of research, with several recent articles proposing multivariate models, as described in the Introduction.

We believe that our model has a number of advantages over others we have seen, with the ability to estimate missing treatment-outcome combinations, a high degree of flexibility in the assumed correlation structure, code that needs minimal adaptation for each dataset, and data specified only in terms of individual outcome variances, rather than covariance arrays. It is generally acknowledged that a common barrier to the use of existing multivariate NMA models is the difficulty of implementation. Our coding approach aims to facilitate applications to other datasets; we recognize, however, that with current software, Bayesian MCMC modeling still requires a good deal of expertise that may restrict its use.

Our approach is not without limitations. The logit scale (for binary outcomes) and log rate scale (for Poisson outcomes) do not permit event rates equal to zero, and any zeroes on the data must be handled via a continuity correction, although this was not an issue with the RRMS outcomes we considered. Our implementation of the model (via decomposition of the multivariate normal distribution) is restricted in terms of the correlation structures it can handle, but the underlying model could in principle be applied to alternative correlation structures if appropriate software routines are written.

The model, in its current form, should only be used for datasets where the treatment effect on the outcome(s) used as the baseline for mappings is nonmissing and nonzero for all treatments. If the baseline outcome for the mappings is missing for any treatment, it must itself be estimated via the mappings and the model becomes unbalanced. If the treatment effect on the baseline outcome is zero for any treatment, all of that treatment's effects will be mapped to zero. It may be possible to further develop the model to avoid these issues by allowing the baseline outcome to vary by treatment according to data availability, and thus to apply the method to datasets with a higher degree of sparsity.

The mapping-based model is not designed for outcomes where the effect size comparing a treatment to the reference treatment ($t = 1$) is zero for some outcomes and nonzero for others, as this is not consistent with a linear scaling between the effects on different outcomes. More experience with different datasets is required to establish how much of a problem this is in practice, but for now we recommend only using mappings between outcomes for which all treatments have a nonzero effect.

Arguably NMAs should only be performed when the data is reasonably consistent; that is, there is minimal conflict between effects estimates that form “loops” in the network diagrams. Techniques for assessing consistency in univariate NMAs do exist^{5,55} and could in principle be extended into the multivariate domain to address consistency of between-outcome mappings as well as between-treatment effects. Although we have not attempted such a formal approach here, examination of two-way plots of the treatment effects provided some reassurance that the mappings were reasonably consistent.

One possible disadvantage of our parameterization (where μ_i represents the outcome values as an average across treatment arms, rather than in an arbitrarily selected first arm) is that summary data for each trial arm is required, rather than contrast-level data comparing each treatment to arm 1. In our experience trials (especially more recent trials) do usually provide data on each arm, but in case this is unavailable we provide code in the Supporting Information for a contrast-data version of the model using the more conventional parameterization. This model has a more complex likelihood that includes additional within-study correlations in trials with three or more arms, as the contrasts observed in arms 2,3,... within a trial are all expressed relative to arm 1 and are therefore not independent.

A limitation of NMAs in general, but perhaps of greater significance for multivariate NMAs, is the implicit assumption that treatments or outcomes that are absent from studies are missing at random. There are reasonable arguments as to why this may not hold in practice, but assessing the validity of this assumption would probably require a much larger set of trials than the one we have used here.

We did not find much impact on the results from correlations, except when the unrealistic assumption of universal positive correlations was imposed, but we would stress that this may be due to any combination of factors such as the particular mix of outcomes in this dataset, the use of mappings, and the correlation structure we used. We plan to carry out further testing of the model with simulation studies to look further into these issues, and to better investigate the model's strengths and limitations in view of its underlying assumptions.

We have not attempted to reflect the full safety profile of the various treatments in the RRMS case study, as this would have resulted in too cumbersome a dataset for this article. Instead we simply use the liver outcomes for illustrative purposes. We would therefore caution against interpreting our results as a reflection of the overall benefit-risk balance for these drugs. However, to the extent that the data are reliable and consistent, the results for the outcomes we have selected should be meaningful, and the set of efficacy outcomes is reasonably comprehensive.

Finally, although our focus has been on using the model for evidence synthesis purposes, we believe that another possible use is to investigate relationships between outcomes at the between-study level, perhaps in order to establish surrogate measures.

ACKNOWLEDGEMENTS

This work is part of the lead author's PhD research which is funded by an Imperial College PhD Scholarship. The research was supported by the Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London.

DATA AVAILABILITY STATEMENT

The data and model code on which these results are based are available in the Supporting Information.

ORCID

Ed Waddingham  <https://orcid.org/0000-0002-0479-5125>

REFERENCES

1. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR task force on indirect treatment comparisons good research practices: part 1. *Value Health*. 2011;14(4):417-428.
2. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J Clin Epidemiol*. 2014;2(67):138-143.
3. NICE. Guide to the methods of technology appraisal. <https://www.nice.org.uk/process/pmg9>. Published April 4, 2013. Accessed August 24, 2018.
4. Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Stat Med*. 2011;30(20):2481-2498.
5. Riley RD, Jackson D, Salanti G, et al. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ*. 2017;358:j3932.
6. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat Med*. 2007;26(1):78-97.
7. Hughes D, Waddingham E, Mt-Isa S, et al. Recommendations for benefit-risk assessment methodologies and visual representations. *Pharmacoepidemiol Drug Saf*. 2016;25(3):251-262.
8. Mt-Isa S, Hallgreen CE, Wang N, et al. Balancing benefit and risk of medicines: a systematic review and classification of available methodologies. *Pharmacoepidemiol Drug Saf*. 2014;23(7):667-678.
9. Nixon R, Dierig C, Mt-Isa S, et al. A case study using the ProACT-URL and BRAT frameworks for structured benefit risk assessment. *Biom Z*. 2016;58(1):8-27.
10. Waddingham E, Mt-Isa S, Nixon R, Ashby D. A Bayesian approach to probabilistic sensitivity analysis in structured benefit-risk assessment. *Biom Z*. 2016;58(1):28-42.
11. Tovey D. The impact of Cochrane reviews. *Cochrane Database Syst Rev*. 2010;7:ED000007.
12. Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed Core outcomes. *J Health Serv Res Policy*. 2012;17(1):1-2.
13. Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*. 2008;9(1):172-186.
14. Bujkiewicz S, Thompson JR, Sutton AJ, et al. Multivariate meta-analysis of mixed outcomes: a Bayesian approach. *Stat Med*. 2013;32(22):3926-3943.

15. Hong H, Carlin BP, Chu H, Shamliyan TA, Wang S, Kane RL. A Bayesian missing data framework for multiple continuous outcome mixed treatment comparisons. <https://effectivehealthcare.ahrq.gov/products/indirect-comparisons/research-2013>. Published 01/10/2013. Accessed May 29, 2020.
16. Dias S, Ades AE. Absolute or relative effects? Arm-based synthesis of trial data. *Res Synth Methods*. 2016;7(1):23-28.
17. Hong H, Chu H, Zhang J, Carlin BP. Rejoinder to the discussion of “a Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons,” by S. Dias and A. E. Ades. *Res Synth Methods*. 2016;7(1):29-33.
18. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50(6):683-691.
19. Hong H, Carlin BP, Shamliyan TA, et al. Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. *Med Decis Making*. 2013;33(5):702-714.
20. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *J R Stat Soc Ser A Stat Soc*. 2009;172(4):789-811.
21. Efthimiou O, Mavridis D, Cipriani A, Leucht S, Bago P, Salanti G. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. *Stat Med*. 2014;33(13):2275-2287.
22. Efthimiou O, Mavridis D, Riley RD, Cipriani A, Salanti G. Joint synthesis of multiple correlated outcomes in networks of interventions. *Biostatistics*. 2015;16(1):84-97.
23. Jackson D, Bujkiewicz S, Law M, Riley RD, White IR. A matrix-based method of moments for fitting multivariate network meta-analysis models with multiple outcomes and random inconsistency effects. *Biometrics*. 2018;74(2):548-556.
24. Madan J, Chen Y-F, Aveyard P, et al. Synthesis of evidence on heterogeneous interventions with multiple outcomes recorded over multiple follow-up times reported inconsistently: a smoking cessation case-study. *J R Stat Soc Ser A Stat Soc*. 2014;177(1):295-314.
25. Welton NJ, Cooper NJ, Ades AE, Lu G, Sutton AJ. Mixed treatment comparison with multiple outcomes reported inconsistently across trials: evaluation of antivirals for treatment of influenza A and B. *Stat Med*. 2008;27(27):5620-5639.
26. Achana FA, Cooper NJ, Bujkiewicz S, et al. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Med Res Methodol*. 2014;14:92-92.
27. Lu G, Kounali D, Ades AE. Simultaneous multioutcome synthesis and mapping of treatment effects to a common scale. *Value Health*. 2014;17(2):280-287.
28. Ades AE, Lu G, Dias S, Mayo-Wilson E, Kounali D. Simultaneous synthesis of treatment effects and mapping to a common scale: an alternative to standardisation. *Res Synth Methods*. 2015;6(1):96-107.
29. Bujkiewicz S, Jackson D, Thompson JR, et al. Bivariate network meta-analysis for surrogate endpoint evaluation. *Stat Med*. 2019;38(18):3322-3341.
30. Rosati G. The prevalence of multiple sclerosis in the world: an update. *Neurol Sci*. 2001;22(2):117-139.
31. Gajofatto A, Benedetti MD. Treatment strategies for multiple sclerosis: when to start, when to change, when to stop? *World J Clin Cases*. 2015;3(7):545-555.
32. Tramacere I, Del Giovane C, Salanti G, D’Amico R, Filippini G, et al. Immunomodulators and immunosuppressants for multiple sclerosis: a network meta-analysis. *Cochrane Database Syst Rev*. 2015;(9):Cd008933.
33. Bornstein MB, Miller A, Slagle S, et al. A pilot trial of cop 1 in exacerbating-relapsing multiple sclerosis. *N Engl J Med*. 1987;317(7):408-414.
34. Cadavid D, Wolansky LJ, Skurnick J, et al. Efficacy of treatment of MS with IFNbeta-1b or glatiramer acetate by monthly brain MRI in the BECOME study. *Neurology*. 2009;23(72):1976-1983.
35. Calabresi PA, Radue E-W, Goodin D, et al. Safety and efficacy of fingolimod in patients with relapsing-relapsing multiple sclerosis (FREEDOMS II): a double-blind, randomised, placebo-controlled, phase 3 trial. *Lancet Neurol*. 2014;13(6):545-556.
36. Comi G, Jeffery D, Kappos L, et al. Placebo-controlled trial of Oral Laquinimod for multiple sclerosis. *N Engl J Med*. 2012;366(11):1000-1009.
37. Durelli L, Verdun E, Barbero P, et al. Every-other-day interferon beta-1b versus once-weekly interferon beta-1a for multiple sclerosis: results of a 2-year prospective randomised multicentre study (INCOMIN). *Lancet*. 2002;359(9316):1453-1460.
38. Ebers GC. Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. *Lancet*. 1998;352(9139):1498-1504.
39. Fox RJ, Miller DH, Phillips JT, et al. Placebo-controlled phase 3 study of Oral BG-12 or Glatiramer in multiple sclerosis. *N Engl J Med*. 2012;367(12):1087-1097.
40. Gold R, Kappos L, Arnold DL, et al. Placebo-controlled phase 3 study of Oral BG-12 for relapsing multiple sclerosis. *N Engl J Med*. 2012;367(12):1098-1107.
41. Jacobs LD, Cookfair DL, Rudick RA, et al. Intramuscular interferon beta-1a for disease progression in relapsing multiple sclerosis. *Ann Neurol*. 1996;39(3):285-294.
42. Johnson KP, Brooks BR, Cohen JA, et al. Copolymer 1 reduces relapse rate and improves disability in relapsing-relapsing multiple sclerosis: results of a phase III multicenter, double-blind, placebo-controlled trial. *Neurology*. 1995;45(7):1268-1276.
43. Kappos L, Radue E-W, O’Connor P, et al. A placebo-controlled trial of Oral Fingolimod in relapsing multiple sclerosis. *N Engl J Med*. 2010;362(5):387-401.
44. Mikol DD, Barkhof F, Chang P, et al. Comparison of subcutaneous interferon beta-1a with glatiramer acetate in patients with relapsing multiple sclerosis (the REBif vs Glatiramer acetate in relapsing MS disease [REGARD] study): a multicentre, randomised, parallel, open-label trial. *Lancet Neurol*. 2008;7(10):903-914.
45. O’Connor P, Filippi M, Arnason B, et al. 250 microg or 500 microg interferon beta-1b versus 20 mg glatiramer acetate in relapsing-relapsing multiple sclerosis: a prospective, randomised, multicentre study. *Lancet Neurol*. 2009;8(10):889-897.

46. O'Connor P, Wolinsky JS, Confavreux C, et al. Randomized trial of Oral Teriflunomide for relapsing multiple sclerosis. *N Engl J Med*. 2011;365(14):1293-1303.
47. Paty DW, IFNB MS. Study group. Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology*. 1993;43(4):655-661.
48. Vollmer TL, Sorensen PS, Selmaj K, et al. A randomized placebo-controlled phase III trial of oral laquinimod for multiple sclerosis. *J Neurol*. 2014;261(4):773-783.
49. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making*. 2013;33(5):607-617.
50. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105-3124.
51. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Series B Stat Methodol*. 2002;64(4):583-639.
52. Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. *J Multivar Anal*. 2009;100(9):1989-2001.
53. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol*. 2011;64(2):163-171.
54. Walker R, Schulz M, Arora B, et al. 057 real world evidence (RWE) on long-term persistence of fingolimod in relapsing-remitting multiple sclerosis (RRMS) in Australia. *J Neurol Neurosurg Psychiatry*. 2018;89(6):A23-A24.
55. 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.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Waddingham E, Matthews PM, Ashby D. Exploiting relationships between outcomes in Bayesian multivariate network meta-analysis with an application to relapsing-remitting multiple sclerosis. *Statistics in Medicine*. 2020;39:3329–3346. <https://doi.org/10.1002/sim.8668>