Received: 14 March 2018 | Revised: 12 December 2018 | Accepted: 11 April 2019

DOI: 10.1002/jrsm.1351

RESEARCH ARTICLE

WILEY Research Synthesis Methods

Modelling time-course relationships with multiple treatments: Model-based network meta-analysis for continuous summary outcomes

Hugo Pedder[1] | Sofia Dias[1] | Margherita Bennetts[2] | Martin Boucher[2] | Nicky J. Welton[1]

[1] Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[2] Pharmacometrics, Pfizer Ltd, Kent, UK

Correspondence
Hugo Pedder, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2PS, UK.
Email: hugo.pedder@bristol.ac.uk

Funding information
Medical Research Council, Grant/Award Numbers: MR/M005232/1 and MR/M005615/1; Pfizer; University of Bristol; National Institute for Health Research Biomedical Centre at the University Hospitals Bristol NHS Foundation Trust; MRC ConDuCT-II Hub for Trials Methodology Research, Grant/Award Number: MR/K025643/1; MRC New Investigator Research Grant, Grant/Award Number: MR/M005232/1; Pfizer Ltd; MRC Methodology Research Programme, Grant/Award Number: MR/M005615/1

Background: Model-based meta-analysis (MBMA) is increasingly used to inform drug-development decisions by synthesising results from multiple studies to estimate treatment, dose-response, and time-course characteristics. Network meta-analysis (NMA) is used in Health Technology Appraisals for simultaneously comparing effects of multiple treatments, to inform reimbursement decisions. Recently, a framework for dose-response model-based network meta-analysis (MBNMA) has been proposed that combines, often nonlinear, MBMA modelling with the statistically robust properties of NMA. Here, we aim to extend this framework to time-course models.

Methods: We propose a Bayesian time-course MBNMA modelling framework for continuous summary outcomes that allows for nonlinear modelling of multiparameter time-course functions, accounts for residual correlation between observations, preserves randomisation by modelling relative effects, and allows for testing of inconsistency between direct and indirect evidence on the time-course parameters. We demonstrate our modelling framework using an illustrative dataset of 23 trials investigating treatments for pain in osteoarthritis.

Results: Of the time-course functions that we explored, the E_{max} model gave the best fit to the data and has biological plausibility. Some simplifying assumptions were needed to identify the ET_{50}, due to few observations at early follow-up times. Treatment estimates were robust to the inclusion of correlations in the likelihood.

Conclusions: Time-course MBNMA provides a statistically robust framework for synthesising evidence on multiple treatments at multiple time points. The use of placebo-controlled studies in drug-development means there is limited potential for inconsistency. The methods can inform drug-development decisions and provide the rigour needed in the reimbursement decision-making process.

1 | INTRODUCTION

In drug development, a key decision is whether to proceed to a further clinical trial and if so, which follow-up outcomes to measure and which comparator agent(s) to include. To inform this decision, it is important to understand the onset and duration of action of not just the agent under development, but also competitor agents. In addition, understanding the time-course of pharmaceutical agents is useful for licencing agencies when making decisions on safety and efficacy of a new agent, and for reimbursement agencies making policy decisions based on manufacturer submissions where different studies and agents have reported results at different follow-up times.

Model-based meta-analysis (MBMA) is a technique for synthesising results of multiple studies, usually randomised controlled trials (RCTs), to gain understanding of the pharmacodynamic profile of different agents during drug development.[1] Key characteristics within this profile are the dose-response and time-course relationships of an agent, and how they compare with competitors. MBMA has been used to understand these relationships across studies.[2-4] Some MBMAs have used indirect evidence to compare several agents simultaneously.[5-7] However, these types of analyses typically model each agent separately, based on pooling individual study arms. Whilst this approach allows the inclusion of nonrandomised and/or single-arm studies and has been proposed in the meta-analysis literature,[8] it breaks the randomisation within RCTs and ignores within-study comparisons, effectively losing the advantages of the RCT design and potentially inducing bias in the resulting estimates.[9]

Network meta-analysis (NMA) allows evidence on multiple treatments to be combined provided they form a connected network of treatment comparisons (where treatment refers to a distinct formulation, such as agent and dose combination).[10-14] NMA pools evidence from RCTs in a way that respects the randomisation in the design of the included studies. It strengthens inference by combining all evidence (both direct and indirect) on the comparisons of interest, under the assumption of consistency—agreement between direct and indirect evidence on the same treatment comparison. When conducting an NMA, it is essential to test for inconsistency where possible and consider results accordingly, as the validity of the resulting estimates rests upon the consistency assumption. Methods have been developed that formally test for inconsistency in NMA,[15,16] something which has not previously been possible when making indirect comparisons in MBMA.

However, recently, MBMA has been extended to incorporate a formal consistency framework by combining ideas from NMA and MBMA in the context of dose-response models with multiple agents.[17] This model-based network meta-analysis (MBNMA) approach respects the randomisation in the included RCTs and allows for formal testing of inconsistency in the network of evidence.

Mawdsley et al[17] described their method for dose-response models with an outcome at a single time-point. In phase II and III trials, there are often multiple follow-up times reported within a study, which represent repeated measures on the same individuals and so will be correlated. This correlation needs to be accounted for either by modelling the time-course,[18-20] or with a multivariate likelihood,[21-24] or both.[2,25] Typically for meta-analysis, only aggregated data are available in published RCTs. This can lead to additional complexities when accounting for correlations between repeated measurements, as the correlation structure may not be known.

Whilst repeated observations over time may be considered discrete observations, from the perspective of drug development, where the focus is on predicting relative efficacy at different time points, it is desirable to estimate a continuous function that describes the relationship between relative effects over time. In this paper, we describe the MBNMA approach with time-course models for repeated observations within the included studies. We also present methods to assess whether the included evidence exhibits inconsistency. Note that in this paper we focus on the time-course relationship only, and so the estimates are obtained for each distinct treatment regimen (agent and dose combination).

The paper is organised as follows. We begin by describing the MBNMA framework to incorporate multiple follow-up times for continuous summary outcomes using models for time-course. We present a general framework, indicating some of the common functional relationships typically seen in practice and accounting for residual correlation between time points. We also outline an approach for model selection and critique and present methods to assess consistency in the time-course setting. We illustrate the methods with an example dataset of treatments for pain in osteoarthritis. Finally, we end with a discussion of the methods in the context of earlier work and indicate areas for further developments.

2 | METHODS

2.1 | Likelihood for continuous summary outcomes

Although individual patient data (IPD) may be available from a manufacturer's own study, it is likely that only aggregate level data from publications are available from other studies. We therefore develop our model at the level of study summaries (eg, mean differences). We assume that for each study we have a continuous summary outcome, such as mean outcome or log-odds of response, $y_{i,k,m}$, together with standard errors, $se_{i,k,m}$, reported for each study i, arm $k = 1, \ldots ,K_i$, and at time point $m = 1, \ldots ,M_i$, where study i has K_i arms and reports at M_i time points and $s_{i,m}$ gives the actual time corresponding to the m^{th} time point in study i. This formulation allows for different studies to report at different times. Typically, $m = 1$ represents a baseline observation at time $s_{i,1} = 0$. The treatment given in study i, arm k, is indicated by $t_{i,k}$.

Because we have repeated measures from the same individuals within each study, the observations may be correlated, which can be captured with a multivariate normal likelihood:

$$\boldsymbol{y}_{i,k} \sim MVN\left(\theta_{i,k}, \boldsymbol{\Sigma}_{i,k}\right) \tag{1}$$

$$\boldsymbol{\Sigma}_{i,k} = \begin{pmatrix} se^2_{i,k,1} & \rho_{i,k,1,2}se_{i,k,1}se_{i,k,2} & \cdots & \rho_{i,k,1,M_i}se_{i,k,1}se_{i,k,M_i} \\ \rho_{i,k,1,2}se_{i,k,1}se_{i,k,2} & se^2_{i,k,2} & \cdots & \rho_{i,k,2,M_i}se_{i,k,2}se_{i,k,M_i} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{i,k,1,M_i}se_{i,k,1}se_{i,k,M_i} & \cdots & \cdots & se^2_{i,k,M_i} \end{pmatrix}$$

where $\boldsymbol{y}_{i,k}$ is a vector of the observed summary measures over time points, $\theta_{i,k}$ is a vector of modelled outcomes, and $\boldsymbol{\Sigma}_{i,k}$ is an $M_i \times M_i$ covariance matrix:
where ρ_{i,k,m_1,m_2} is the within-study correlation between summary measures at time points m_1 and m_2 for study i arm k. In practice, correlations are rarely reported in the literature and will only be available from studies where we have IPD. In addition, the correlations estimated from any available IPD will be at the individual level, which may be different to correlations at the summary level.[26]

One approach to deal with unknown within-study correlations at the summary level, if IPD are available or information on this correlation can be obtained from external data, is to assume that the correlations seen between time points for individual patients are the same as those seen for summary measures, and also that correlations in the study for which correlation information is available also apply in the aggregate data trials.[27,28] However, this might lead to ecological bias as there is no guarantee that correlations at the individual level will be the same as at the aggregate level. An alternative is to estimate within-study correlations based on the aggregate data summaries by giving prior distributions to the ρ_{i,k,m_1,m_2}. This approach allows the possibility of using informative prior distributions based on information gained from external data. In order to identify correlation parameters, some constraints will be required, such as assuming a particular covariance structure. For example, a compound symmetry structure can be assumed, in which a single parameter, ρ, is estimated for the correlation between all time points (assumed to be the same across all studies): $\rho_{i,k,m_1,m_2} = \rho$. Alternatively, it might be more reasonable to assume an autoregressive AR(1) structure in which covariances are dependent on the amount of time between observations where $\rho_{i,k,m_1,m_2} = \rho^{\frac{s_{i,m_2}-s_{i,m_1}}{s_{i,2}-s_{i,1}}}$.

When studies that report mean change from baseline and final values at each time point are included, they can contribute information to ρ_{i,k,m_1,m_2} by modelling separate likelihoods for both pieces of data (see below).[29]

2.1.1 | Mean change from baseline by time

Where aggregated summaries are reported as mean change from baseline (baseline corresponding to m = 1), we have summaries for time points $m = 2, \ldots ,$ M_i defined as $y^{change}_{i,k,m} = \left(y_{i,k,m} - y_{i,k,1}\right)$ and their standard errors $se^{change}_{i,k,m}$. Covariances between mean changes from baseline across time-points m_1 and m_2 (dropping the i,k subscripts for ease of exposition) are

$$Cov\left(\left(y_{m_1} - y_1\right), \left(y_{m_2} - y_1\right)\right) = \\ \rho_{m_1,m_2}se_{m_1}se_{m_2} - \rho_{1,m_1}se_1se_{m_1} - \\ \rho_{1,m_2}se_1se_{m_2} + se^2_1$$

which gives the $(m_1, m_2)^{th}$ element of the covariance matrix for the mean change from baselines, $\boldsymbol{\Sigma}^{change}_{i,k}$.

We can then give a multivariate normal likelihood to the aggregate mean outcomes for all time points:

$$\mathbf{y}^{change}_{i,k} \sim MVN\left(\theta^{change}_{i,k}, \boldsymbol{\Sigma}^{change}_{i,k}\right)$$

where

$$\boldsymbol{\theta}_{i,k}^{change} = \begin{pmatrix} \theta_{i,k,2} - \theta_{i,k,1} \\ \theta_{i,k,3} - \theta_{i,k,1} \\ \vdots \\ \theta_{i,k,M_i} - \theta_{i,k,1} \end{pmatrix}.$$

By writing the model for mean change from baseline in terms of the model for mean outcomes, we can combine studies where some report mean outcomes and some report mean change from baseline by giving each type of data the appropriate likelihood and using a shared-parameter model.[30] If some studies report both mean outcomes and mean change from baseline, then both pieces of data can be included. Modelling both outcomes simultaneously provides sufficient evidence to estimate the correlations, ρ_{i,k,m_1,m_2}.[29]

2.2 | Time-course model

We put the time-course model on the aggregate-level means:

$$\theta_{i,k,m} = f\left(s_{i,m}, \boldsymbol{\lambda}_{i,k}\right)$$

where f defines a functional relationship over time s, and $\boldsymbol{\lambda}_{i,k} = (\lambda_{0,i},\lambda_{1,i,k},\lambda_{2,i,k},\ldots)$ are a set of parameters that describe the relationship in mean outcomes over time. In all time-course models, there will be a "nuisance parameter" $\lambda_{0,i}$ which represents the "intercept" at time, common across arms. Note that for many time-course models the $\lambda_{0,i}$ parameters will cancel out when using change from baseline data. We put our modelling assumptions on the remaining parameters, $\lambda_{1,i,k},\lambda_{2,i,k},\ldots$, leaving the $\lambda_{0,i}$ unconstrained (achieved in a Bayesian analysis by giving independent vague prior distributions to the $\lambda_{0,i}$ parameters).

2.2.1 | Exponential model

One of the most commonly used models is the exponential model, which has intercept $\lambda_{0,i}$, and a single parameter of interest, $\lambda_{1,i,k}$, which represents the rate at which the mean outcome falls over time, assuming a constant rate of growth/decay:

$$\theta_{i,k,m} = \lambda_{0,i} \exp\left(\lambda_{1,i,k} s_{i,k,m}\right). \tag{2}$$

2.2.2 | Linear model

Another model with a single parameter of interest is the linear model:

$$\theta_{i,k,m} = \lambda_{0,i} + \lambda_{1,i,k} s_{i,k,m} \tag{3}$$

where $\lambda_{0,i}$ is the intercept and $\lambda_{1,i,k}$ the fall in mean outcome for a unit increase in time.

2.2.3 | E$_{\mathbf{max}}$ model

A functional form commonly used in pharmacometrics, which has two parameters of interest, is the E$_{\max}$ model:

$$\theta_{i,k,m} = \lambda_{0,i} + \frac{\lambda_{1,i,k} \times s_{i,m}}{\lambda_{2,i,k} + s_{i,m}} \tag{4}$$

where the intercept $\lambda_{0,i}$, often referred to as E$_0$, is the initial outcome at baseline (time = 0), $\lambda_{1,i,k}$, typically referred to as E$_{\max}$, is the maximum possible effect of a treatment relative to baseline, and $\lambda_{2,i,k}$, typically referred to as ET$_{50}$, is the time point at which 50% of the maximum treatment effect has been achieved.

2.2.4 | Piecewise linear model

Piecewise models can allow for considerable flexibility, though they may not so accurately resemble true biological relationships and may not be appropriate when the intention is to predict values close to where the pieces meet (the "knots"). The simplest example of this is a two-piece linear model with a single knot at $s = S$:

$$\theta_{i,k,m} = \begin{cases} \lambda_{0,i} + \lambda_{1,i,k} s_{i,m} & 0 \le s \le S \\ \left(\lambda_{0,i} + \lambda_{1,i,k} S\right) + \lambda_{2,i,k}\left(s_{i,m} - S\right) & s > S \end{cases} \tag{5}$$

where $\lambda_{0,i}$ is the intercept, $\lambda_{1,i,k}$ the change in mean outcome for a unit change in time during time period (0,S), and $\lambda_{2,i,k}$ the change in mean outcome for a unit change in time during time period after S. The intercept for the second piece ($\lambda_{0,i}$+$\lambda_{1,i,k}S$) ensures that the two regression lines meet at the knot.

2.3 | Network meta-analysis model

The NMA model describes the impact of treatments on one or more of the parameters of the time-course model, $\lambda_{1,i,k},\lambda_{2,i,k},\ldots$. If the NMA model is given for a single time-model parameter, $\lambda_{1,i,k}$, we have

$$g\left(\lambda_{1,i,k}\right) = \mu_i + \delta_{i,k}$$

for a given link function g which transforms the outcome to a scale where relative treatment effects may be expected to be additive. μ_i is the time-course model parameter (on the transformed scale) for arm 1 of study i, and $\delta_{i,k}$ the study-specific relative effect for the treatment used in arm k relative to arm 1 of study i.

For example, for an exponential time-course model (Equation 2), it would be natural to put the NMA model on the log-scale:

$$\log\left(\lambda_{1,i,k}\right) = \mu_i + \delta_{i,k}$$

where μ_i is the log growth/decay rate on arm 1, and $\delta_{i,k}$ is the log rate-ratio for treatment arm k compared with treatment arm 1, of study i.

The μ_i are nuisance parameters and given independent vague prior distributions in a Bayesian analysis to allow these to be unconstrained. By treating these as nuisance parameters, the focus of this modelling strategy is on estimating relative treatment effects rather than on characterising the time-course on the reference treatment (eg, placebo effect). In fact, because different studies may have included different control (arm 1) treatments, the μ_i do not have a consistent interpretation across studies.

Treatment effects can be either assumed common ("fixed") or similar/exchangeable ("random") across studies. For the random effects model, study-specific treatment effects are assumed to be normally distributed around a mean treatment effect that adheres to the consistency relationships, with common between-studies variance τ^2 across treatment comparison:

$$\delta_{i,k} \sim N\left(d_{1,t_{i,k}} - d_{1,t_{i,1}}, \tau^2\right). \tag{6}$$

The consistency relationships reflect the comparison made between the treatment $t_{i,k}$ used on arm k and the treatment $t_{i,1}$ used on arm 1 of each study. The fixed effect model is obtained by setting $\tau^2 = 0$. The model estimates "basic parameters" $d_{1,k}$, the pooled mean relative effect for treatment k relative to treatment 1 (the reference treatment for the NMA). The $d_{1,k}$ are each given independent vague normal priors in a Bayesian analysis. All other relative effects for treatment k relative to treatment c, $d_{c,k}$, can then be derived from the consistency relationships[12,30]:

$$d_{c,k} = d_{1,k} - d_{1,c}. \tag{7}$$

Time-course functions with multiple (nonintercept) time-course parameters may have NMA models specified for one or more of these parameters, although a relatively rich dataset is required to estimate NMA models with more than one treatment effect parameter.

Suppose we expect the treatments to influence two parameters of the time-course model, $\lambda_{1,i,k}$ and $\lambda_{2,i,k}$ (for example these could represent E$_{\max}$ and ET$_{50}$). The NMA model proceeds as for a single parameter; however, for a random effects model, we need to allow for correlations between the study-specific treatment effects on the two time-course parameters. Note that the link functions g do not have to be the same for the different parameters:

$$\begin{aligned} g_1\left(\lambda_{1,i,k}\right) &= \mu_{1,i} + \delta_{1,i,k} \\ g_2\left(\lambda_{2,i,k}\right) &= \mu_{2,i} + \delta_{2,i,k} \\ &etc. \end{aligned} \tag{8}$$

The random effects model for $\delta_{1,i,k}$, $\delta_{2,i,k}$, etc needs to be multivariate to allow for correlations between relative effects on the different time-course parameters.

For example, for the E$_{\max}$ model, the E$_{\max}$ parameter $\lambda_{1,i,k}$ can be positive or negative, and so we can put the model on the natural scale, whereas the ET$_{50}$ parameter $\lambda_{2,i,k}$ may only take positive values, and so it makes sense to model this on the log scale, giving

$$\begin{aligned} \lambda_{1,i,k} &= \mu_{1,i} + \delta_{1,i,k} \\ log\left(\lambda_{2,i,k}\right) &= \mu_{2,i} + \delta_{2,i,k} \end{aligned}$$

with a bivariate random effects distribution:

$$\begin{pmatrix} \delta_{1,i,k} \\ \delta_{2,i,k} \end{pmatrix} \sim N\left(\begin{pmatrix} d_{1,1,t_{i,k}} - d_{1,1,t_{i,1}} \\ d_{2,1,t_{i,k}} - d_{2,1,t_{i,1}} \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho_\delta \tau_1 \tau_2 \\ \rho_\delta \tau_1 \tau_2 & \tau_2^2 \end{pmatrix}\right) \tag{9}$$

where all parameters are as before, with an extra subscript to indicate whether they relate to $\lambda_{1,i,k}$ or $\lambda_{2,i,k}$. The correlation between the treatment effects on the two parameters is given by ρ_δ. Different parameterisations are available for the between-studies covariance matrix that may be more computationally stable, such as a Cholesky parameterization or a spherical parameterization.[31] Fixed effect models can be obtained by setting the between-study variance parameters to 0.

2.4 | Multi-arm trials

When including multi-arm trials, it is important to account for correlation between relative effects within a trial, as all relative effects will have the same comparator. For a common between-study variance, the correlation between these relative effects will be 0.5.[11] For MBNMAs with a single nonintercept time-course parameter, this can be done either using a multivariate normal distribution to model a vector of random effects,[30] or, for the purposes of writing more generic code, using a conditional univariate distributions formulation for the random effect of arm $k > 2$, given all arms from 2 to $k - 1$[32]:

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

When modelling multiple nonintercept time-course parameters, the correlation between relative effects can be modelled simultaneously to the correlation between the parameters ρ_δ using a multivariate normal distribution on a vector of random effects, $\boldsymbol{\delta_i}$, whose length is equal to the number of parameters multiplied by the number of arms, K_i, in study i. For models with two parameters of interest, this is

$$\boldsymbol{\delta_i}=\begin{pmatrix}\delta_{1,i,k}\\ \vdots\\ \delta_{1,i,K}\\ \delta_{2,i,k}\\ \vdots\\ \delta_{2,i,K}\end{pmatrix}\sim N\left(\begin{pmatrix}d_{1,1,t_{i,k}}-d_{1,1,t_{i,1}}\\ \vdots\\ d_{1,1,t_{i,K}}-d_{1,1,t_{i,1}}\\ d_{2,1,t_{i,k}}-d_{2,1,t_{i,1}}\\ \vdots\\ d_{2,1,t_{i,K}}-d_{2,1,t_{i,1}}\end{pmatrix},\Omega_i\right).$$

The covariance matrix, Ω_i, is a $2K_i \times 2K_i$ matrix:

$$\Omega_i=\begin{pmatrix}\tau_1^2 & \frac{\tau_1^2}{2} & \cdots & 2\rho_\delta\tau_1\tau_2 & \rho_\delta\tau_1\tau_2\\ \frac{\tau_1^2}{2} & \tau_1^2 & \cdots & \rho_\delta\tau_1\tau_2 & 2\rho_\delta\tau_1\tau_2\\ \vdots & \vdots & \ddots & \vdots & \cdots\\ 2\rho_\delta\tau_1\tau_2 & \rho_\delta\tau_1\tau_2 & \cdots & \tau_2^2 & \frac{\tau_2^2}{2}\\ \rho_\delta\tau_1\tau_2 & 2\rho_\delta\tau_1\tau_2 & \cdots & \frac{\tau_2^2}{2} & \tau_2^2\end{pmatrix}.$$

2.5 | Simplifying modelling assumptions

For models with many parameters, there may be insufficient data to be able to estimate all parameters (ie, the parameters may not be identifiable). To aid identifiability when there are two or more non-nuisance parameters, simplifying assumptions can be made to constrain the parameters. In this way, a model can be constrained to be as complex or simple as the data allow, provided there is biological plausibility for any simplification. Note that any shared parameters will be more influenced by studies/treatments with more information (ie, those with more observations within studies). One or more of the following may be considered:

2.5.1 | Fixed effect models

One or more of the modelled parameters could be modelled as a fixed treatment effect, reflecting an assumption of homogeneity where different studies of the same comparison estimate a common effect. So, if there is a fixed effect model on the second parameter, Equation 9 becomes

$$\begin{aligned}\delta_{1,i,k}&\sim N\left(d_{1,1,t_{i,k}}-d_{1,1,t_{i,1}},\tau_1^2\right)\\ \delta_{2,i,k}&=d_{2,1,t_{i,k}}-d_{2,1,t_{i,1}}\end{aligned}$$

and if there is a fixed effect model on both parameters, Equation 9 becomes

$$\begin{aligned}\delta_{1,i,k}&=d_{1,1,t_{i,k}}-d_{1,1,t_{i,1}}\\ \delta_{2,i,k}&=d_{2,1,t_{i,k}}-d_{2,1,t_{i,1}}.\end{aligned}$$

2.5.2 | Class-effect models

Relative treatment effects for one (or more) of the parameters could be assumed to come from a hierarchical model with a common mean, which may depend on class[13,33-35]:

$$d_{2,1,k}\sim\left(D_{2,class},\tau_{class}^2\right)\quad\text{for } k\in\{class\}. \tag{10}$$

For example, in the E_{max} model, it may be that it is reasonable to assume that treatments within the same class might have a similar onset of action (and so have similar, exchangeable treatment effects on ET_{50}) but reach different maximum effects (E_{max}).

An even more constrained model fixes the treatment effects within a class to be equal:

$$d_{2,1,k}=D_{2,class}\quad\text{for } k\in\{class\}. \tag{11}$$

2.5.3 | Constrain the baseline effect

To further aid identifiability of treatment effects, it may be necessary to impose further constrains on the baseline for one (or more) of the time-course parameters, for example with an exchangeable model:

$$\mu_{2,i} \sim N\left(\eta_2, \sigma^2_{\mu_2}\right). \quad (12)$$

This would imply that the reference treatment effects for each study were assumed to be distributed about a single common mean effect, η_2, and would therefore only be suitable in networks for which all included trials have the same reference treatment (eg, placebo).

2.5.4 | Reduce to a single treatment effect

For models with multiple (nonintercept) time-course parameters, a further simplification is to only model treatment effects on one of the time-course parameters. The other parameters are assumed to be treatment independent and modelled on an absolute, rather than relative, scale:

$$g_2\left(\lambda_{2,i,k}\right) = \mu_{2,i,k}.$$

The $\mu_{2,i,k}$ could be left unconstrained or assumed exchangeable for each treatment within a class.

2.6 | Testing for inconsistency

To test whether the consistency assumption (Equation 7) holds, several approaches have been proposed for identifying inconsistency between direct and indirect evidence that arises within a closed "loop" of treatments for which independent sources of information are available.[15,36] It is important to note that available loops of treatments to test for inconsistency will depend on the choice of reference treatment used in the network.[36,37] In addition, the evidence provided by studies with three or more arms is not independent (due to the common reference arm), and within-study relative effects must be internally consistent. Therefore, loops of evidence consisting only of studies with three or more arms will always be consistent.

Furthermore, the choice of reference treatment for a study with three or more arms can affect whether it is possible to test for inconsistency. For example, suppose we have three different studies providing evidence on P (Placebo) vs N (Naproxen 1000 mg/d), P vs C (Celebrex 200 mg/d), and PvsNvsC, respectively. If we take P as the reference for the three-arm trial, then the three-arm trial provides estimates of PvsN and PvsC, so the model only estimates PvsN and PvsC directly (see Figure S3A - Supporting Information). The NvsC effect is derived from the PvsN and PvsC estimates. If, on the other hand, we take N as the reference for the three-arm trial, then the three-arm trial provides estimates of PvsN and NvsC, which together with the two-arm evidence provides independent estimates of PvsN, PvsC, and NvsC, and we can test for inconsistency. We use the convention that we take placebo as reference for all studies that include a placebo arm. In studies without placebo, we use the first drug alphabetically, at its lowest dose.

An unrelated mean effects (UME) model does not include constraints forced by the consistency equations and is equivalent to fitting separate pairwise meta-analyses to each direct comparison whilst sharing treatment-independent parameters across all the comparisons,[16] such as between-study heterogeneity or treatment-independent time-course parameters. The results from this model can then be compared with those from the MBNMA. A better model fit (lower deviance) or lower standard deviations (SD) for exchangeable parameters or random treatment effects would suggest that inconsistency may be present in the network.

A more explicit method for testing inconsistency for specific comparisons is the node-splitting method.[15] This technique involves splitting the evidence for a given comparison within a loop of treatments into "direct" evidence from head-to-head RCT comparisons and "indirect" evidence that arises from the consistency relationships. A Bayesian *P*-value can be calculated for the treatment effects estimated using the direct and indirect evidence, which represents the proportion of the two posterior distributions that overlap.

Note that when performing pairwise meta-analyses to estimate direct evidence, the sharing of parameters across direct comparisons for which limited information is available can make tests for inconsistency conservative, and this should be borne in mind when interpreting them.

For models with multiple time-course parameters (Equation 8), it is important to consider that inconsistency may be present for treatment effects on either *or both* of the time-course parameters.

Donegan et al[38] present inconsistency models to explore consistency on two parameters. We suggest testing for inconsistency on each time-course parameter separately, because if inconsistency is identified in either parameter for a given comparison, then this should be a cause for concern and should prompt further investigation of the included studies to identify the potential cause.

Multiple testing may also be an issue here, as the number of tests in a typical network will be multiplied by the number of time-course parameters in the MBNMA model. However, inconsistency tests are typically underpowered, and we advise erring on the side of caution as it is better to incorrectly identify inconsistency when there is none present than to incorrectly fail to identify inconsistency when true inconsistency is present.

2.7 | Treatment ranking for time-course relationships

In NMA, it is common to calculate ranking probabilities (probability of being first best, second best, etc) for each treatment within a network, as this is an easily interpretable measure for decision-makers to use. In time-course MBNMA, we can rank on any function of the time-course model, which can allow for an extremely flexible decision-making framework. This could include ranking based on any one of multiple time-course parameters or ranking on the predicted response at a desired follow-up time. For time-course functions with multiple parameters, note that the ranking of treatment effects may differ for different time-course parameters. For example, we could have a treatment that ranks highest for ET_{50} indicating that it acts more quickly than other treatments but ranks lowest for E_{max} indicating that the overall response is lower than for other treatments.

For models with less easily interpretable time-course parameters, it may also be beneficial to have an overall ranking that takes into account the full time-course relationship. Calculating the Area Under the Curve (AUC) for the time-course relationship for each treatment using parameters estimated from the model is a pragmatic way of doing this. However, care must be taken when choosing the duration of time-course over which to calculate AUC, as treatment rankings may be sensitive to this choice.

2.8 | Measures of model fit

Models are implemented using a Bayesian approach, and therefore we use the posterior mean of the deviance to compare the goodness-of-fit of the models,[39] where smaller values of deviance are preferred. Model selection is based on the Deviance Information Criterion (DIC) which represents a compromise between model fit and model complexity,[39,40] defined as the sum of the posterior mean deviance (a measure of fit) and the effective number of parameters (a measure of complexity). We use pD calculated using the Kullback-Leibler information divergence as the effective number of parameters.[41] For the selected final model (based on a univariate likelihood), we also report the posterior mean of the residual deviance (defined as the deviance for the model minus the deviance for a saturated model), which can be compared with the number of unconstrained data points to give an overall measure of model fit. Lack of fit is explored by plotting an appropriate posterior summary (median if skewed) of the contribution to the residual deviance for each datapoint against time. Note, we do not compute residual deviances for models with a multivariate likelihood in which the correlation between time points is estimated from the data, as the saturated model is not uniquely defined.

2.9 | Model selection strategy

We propose a step-by-step approach for model selection of time-course relationships in MBNMA, recognising that the available evidence may not be sufficient to be able to estimate some of the more complex, but less restrictive, models.

1. Plot study summaries (mean outcome) against time to visually identify potential time-course function candidates and obtain expert opinion to assess their biological plausibility if necessary.
2. Fit candidate time-course models with fixed treatment effects using a univariate likelihood that does not account for correlations over time (univariate models). Use simplifying modelling assumptions, described above, if necessary to estimate the models given the available data.
   For each of these fitted models, plot the posterior median of the contribution of each data-point to the residual deviance against time to check fit and to identify alternative time-course relationships to explore. Compare model fit statistics (posterior mean deviance and DIC) and select a time-course model with adequate balance between fit and complexity (lowest DIC) that also has biological plausibility.
3. For the selected univariate time-course model, fit random treatment effects models (if possible) with the available data. Use model fit statistics and inspection of between studies SD parameters to assess presence of heterogeneity and choose between fixed and random treatment effect models.
4. Check fit of the selected univariate model by comparing posterior mean residual deviance to the total number of data points
5. For the selected time-course and treatment effects model, fit models with multivariate likelihoods that account for correlations over time with different covariance structures. Select between the univariate and multivariate formulations based on estimated correlations and robustness of treatment effects obtained (preferring the simpler models with lower pV). Note the deviance statistics are not directly comparable for models with different likelihoods and so cannot be used for model selection.
6. Check for consistency in final selected model (where possible)

a. Run UME model
b. If suggestive of possible inconsistency, perform node splitting of closed loops

2.10 | Illustrative example—Pain in osteoarthritis

The methodology is illustrated using a dataset of RCTs investigating treatments for pain in patients with osteoarthritis. Pain was measured on the Western Ontario and McMaster Universities Arthritis Index (WOMAC) scale[42] and was recorded at multiple time points up to a maximum of 24 weeks. In order to maintain a consistent imputation method for missing data across studies, only those with last observation carried forward analyses (LOCF) were included, as this was the imputation method reported in the majority of papers. Agents with multiple doses were split to form the network of treatments, meaning that each combination of agent and dose was considered to be a separate treatment. Although SDs were typically available at baseline, they were missing for 269 out of 345 observations and were therefore imputed accounting for changes over time using the method of Boucher.[43] We acknowledge that this is a high proportion of data points for which to impute SD, but our aim here is to illustrate the method rather than to provide clinically useful treatment estimates. We note however that in pharmacometrics, SD is not always reported as weighting is often performed using sample size—in practice, we would always recommend that SDs be reported and measured.

The illustrative dataset consists of 23 RCTs comparing 29 treatments. Each study has a median of 3.5 (range: 2-7) follow-up measurements, and all studies use LOCF imputation for analyses. Figure 1 shows the network of comparisons in the data, and Figure 2 shows mean WOMAC pain in each study arm plotted over time for each treatment. The dataset is freely available in the Supporting Information (OsteoarthritisData.csv).[44]

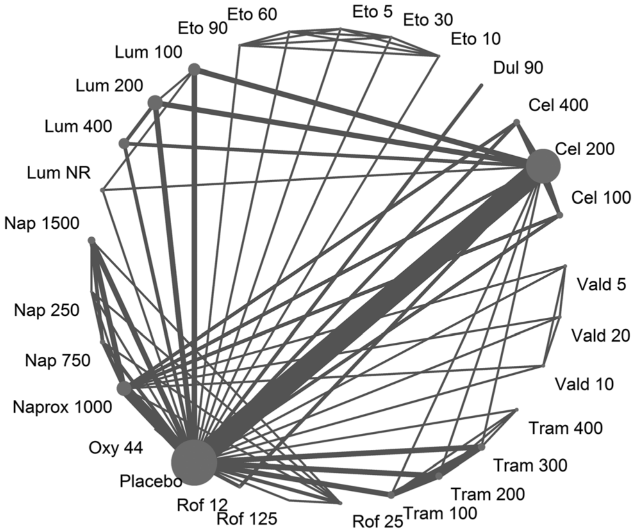

FIGURE 1 Network of treatment comparisons within the MBNMA for the illustrative dataset of 24 RCTs for pain in osteoarthritis. Each treatment is represented by a node. Where direct RCT evidence exists for a particular comparison, the nodes are connected by a line, the thickness of which is proportional to the number of comparisons. All numbers represent doses (total daily dose in mg). *Abbreviations: Cel = Celebrex, Dul = Duloxetine, Eto = Etoricoxib, Lum = Lumiracoxib, Naprox = Naproxcinod, Nap = Naproxen, Oxy = Oxycodone, Rof = Rofecoxib, Tram = Tramadol, Vald = Valdecoxib, NR = Dose not reported* [Colour figure can be viewed at wileyonlinelibrary.com]

2.11 | Implementation

Models were estimated using JAGS 4.2.0 (codes in the Supporting Information). All fixed treatment effect models were run on three independent chains for 20 000 iterations following 80 000 burn-in iterations with a thinning parameter of 10. For random treatment effect models, convergence took more iterations—models were run for 500 000 iterations following 400 000 burn-in iterations with a thinning parameter of 15. Gelman's r statistic[45] and visual inspection of the chains were used to assess convergence. Vague normal prior distributions ($N(0,1000)$) were given to the basic parameters $d_{1,k}$, nuisance parameters μ_i, and treatment effect-independent time-course parameters. For ET_{50} parameters in E_{max} models, it was necessary to ensure that they only took positive values so priors for these were specified on the log-scale. The between-study SD and SDs on exchangeable parameters were given uniform prior distributions ($U(0,5)$). We inspect the posterior for the SDs to ensure they are not being constrained by the prior. We also assessed sensitivity of results to the prior by fitting the same models using half-normal prior distributions ($N(0,6.25)$). In models with a multivariate likelihood, ρ was given a uniform prior distribution ($U(0,1)$) to reflect the belief that outcomes at different time points in the same study are likely to be positively correlated. For bivariate models with two nonintercept parameters, the correlation between these parameters, ρ_δ, was given a uniform prior distribution ($U(-1,1)$). For piecewise time-course models, knot location was selected through trial-and-error by fitting models with different knot values (allowing a minimum of 0.1 weeks between knots in different models) and selecting the value from the best fitting (lowest mean posterior deviance) model.

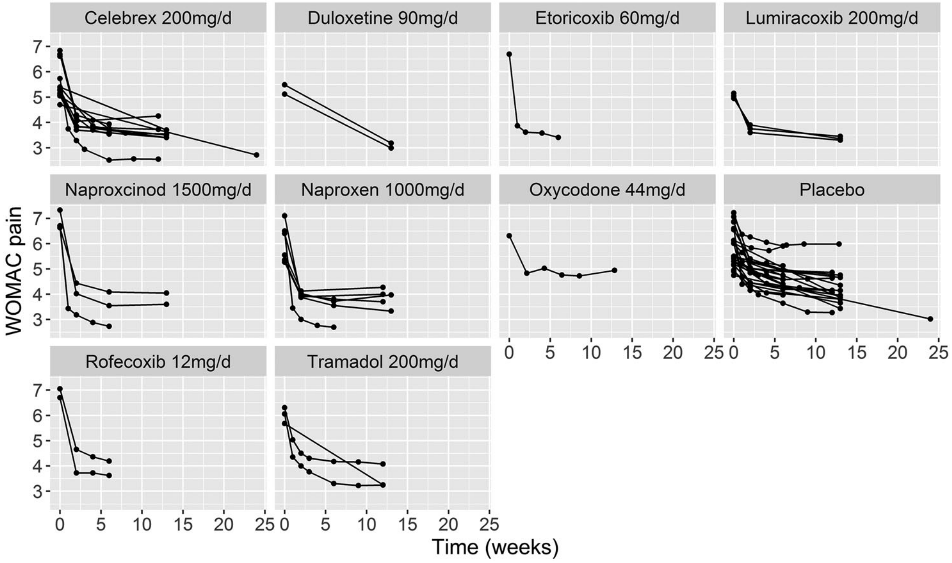

FIGURE 2 Plots of the mean WOMAC pain score for each of the studies in the pain in osteoarthritis dataset showing the most commonly reported dose for each agent, plotted over time

3 | RESULTS

Step 1. Visually inspect the data plotted against time and consider biological plausibility of time-course functions

For all treatments, including placebo, visual inspection of the data shows that WOMAC scores decrease over time in a nonlinear fashion, with a rapid decline in pain during the first 1 to 2 weeks that quickly levels out (Figure 2). This suggests that a simple linear model will not be a good fit for the data, but that exponential, piecewise linear, or E_{max} models may be more suitable. Both exponential and E_{max} models have good biological plausibility and are frequently used for modelling pharmacodynamic properties of drugs. However, the limited number of observations at earlier follow-up times suggests that there may not be enough information to identify parameters that model the rapid decline, such as ET_{50} parameters or parameters in the first piece of piecewise models.

For two treatments (Duloxetine 90 mg/d and Lumiracoxib (NR—dose not reported)), there is only data from studies with baseline measurement and a single follow-up time compared with placebo, which is insufficient information to be able to identify any nonlinear candidate time-course function. We assumed that these treatments followed the same time-course function (though with different time-course parameter values) as the other treatments in the network, and we consider the impact of this assumption in the discussion.

Step 2. Compare time-course models with univariate likelihood and fixed treatment effects

Table 1 shows model fit statistics for linear, exponential, piecewise linear, and E_{max} models (with various simplifying assumptions). As expected, a linear time-course model gave a very poor fit to the data (Equation 3, Figure 3, posterior mean deviance = 6935.2—Table 1). An exponential time-course was also a poor fit (Equation 2, Figure 3, posterior mean deviance = 5856.3—Table 1), as it did not capture the fast rate of decline in WOMAC scores that occurred within the first 2 weeks. A piecewise linear time-course with a knot at week one showed substantially better model fit (Equation 5, Figure 3, posterior mean deviance = −189.3—Table 1) than the linear or exponential models. However, by far, the best fitting time-course appeared to be an E_{max} model (Equation 4, posterior mean deviances less than −441—Table 1). Figure 3 shows that whilst the posterior mean contribution to the deviance displays a pattern for the linear and exponential models (suggesting the time-course is not adequately captured), there is no systematic pattern discernible for the piecewise linear models and best-fitting E_{max} model (see below), and the deviance contributions are much lower for the E_{max} model than the piecewise linear model.

It was not possible to fit an E_{max} model with separate treatment effects on the ET_{50} parameter, due to the limited number of observations at earlier time points in the dataset. We were therefore obliged to make some simplifying assumptions for ET_{50}. We fitted class-effect models

TABLE 1 Model fit statistics for time-course models with univariate likelihood, fitted to the osteoarthritis pain dataset. For exchangeable models, the heterogeneity parameter is reported as standard deviation (SD) = posterior mean SD (95% credible interval)

	Model for $\lambda_{1,i,k}$ (Linear Slope, Exponential Decay, or E_{max})		**Model for $\lambda_{2,i,k}$ (Linear Slope, or ET_{50})**				
Time-course model	**Arm 1 effect, $\mu_{1,i}$**	**Relative treatment effects, $\delta_{1,i,k}$**	**Arm 1 effect, $\mu_{2,i}$**	**Relative treatment effects, $\delta_{2,i,k}$**	**DIC[a]**	**Posterior mean deviance[b]**	**pD[c]**
Linear ($\lambda_{1,i,k}$= slope)	Unconstrained	Fixed effect			7009.1	6935.2	73.9
Exponential ($\lambda_{1,i,k}$= decay rate)	Unconstrained	Fixed effect			5931.8	5856.3	75.5
Piecewise linear ($\lambda_{1,i,k}$= slope period 1, $\lambda_{2,i,k}$= slope period 2, knot = 0.1 wks)	Unconstrained	Fixed effect	Unconstrained	Fixed effect	−69.1	−189.3	120.2
E_{max} model 1 ($\lambda_{1,i,k}$= E_{max}, $\lambda_{2,i,k}$= ET_{50})	Unconstrained	Fixed effect	Exchangeable (Equation 12), SD = 0.53 (0.25, 1.10)	Fixed effect $d_{2,1,k}$ has an exchangeable class effect with an agent-specific mean, SD = 0.11 (0.01, 0.48)	−274.5	−441.2	166.7
E_{max} model 2 ($\lambda_{1,i,k}$= E_{max}, $\lambda_{2,i,k}$= ET_{50})	Unconstrained	Fixed effect	Exchangeable (Equation 12), SD = 0.50 (0.24, 0.97)	Fixed effect $d_{2,1,k}$has a fixed class effect for treatments of same agent	−281.8	−443.1	161.3
E_{max} model 3 ($\lambda_{1,i,k}$= E_{max}, $\lambda_{2,i,k}$= ET_{50})	Unconstrained	Fixed effect	Exchangeable (Equation 12), SD = 0.61 (0.34, 1.10)	Fixed effect $d_{2,1,k}$ has an exchangeable class effect with common mean across all treatments, SD = 0.13 (0.01, 0.48)	−284.3	−444.0	159.6
E_{max} model 4 ($\lambda_{1,i,k}$= E_{max}, $\lambda_{2,i,k}$= ET_{50})	Unconstrained	Fixed effect	Exchangeable (Equation 12), SD = 0.64 (0.39, 1.15)	Fixed effect $d_{2,1,k}$ has a fixed class effect constrained to be equal for all treatments	−289.9	−441.9	152.0
E_{max} model 5 ($\lambda_{1,i,k}$= E_{max}, $\lambda_{2,i,k}$= ET_{50})	Unconstrained	Random effects (Equation 10), SD = 0.09 (0.00, 0.23)	Exchangeable (Equation 12), SD = 0.65 (0.39, 1.16)	Fixed effect $d_{2,1,k}$ has a fixed class effect constrained to be equal for all treatments	−287.9	−448.8	160.8

[a] DIC (= deviance + pD): It is a measure of model fit that penalises complexity.

[b] Deviance (= −2(log-likelihood)): A measure of how closely the fitted values of the model fit the observed data.

[c] pD: The total number of effective parameters in the model, calculated using the Kullback-Leibler information divergence.[41]

with exchangeable ET_{50} treatment effects (Equation 10) with agent-specific means (E_{max} model 1) or with a common mean for all treatments (E_{max} model 3). We also fitted class-effect models with fixed ET_{50} treatment effects (Equation 11) equal to agent-specific values (E_{max} model 2) or with an equal value for all treatments (E_{max} model 4). For all of E_{max} models 1 to 4, we were also obliged to further assume an exchangeable model for the reference treatment effect (placebo in all studies), $\mu_{ET_{50},i}$ (Equation 12), to allow estimation.

Model fit was similar for E_{max} models 1 to 4, but the model that assumed an equal treatment effect for ET_{50} for all treatments compared with placebo, common across studies (E_{max} model 4), had the lowest DIC.

Step 3. Compare random and fixed treatment effect models for selected time-course model

Table 1 shows model fit statistics for a model that is identical to E_{max} model 4, but with a random treatment effects model for E_{max} parameters (Equation 6) (E_{max} model 5). The between-study SD for treatment effects on E_{max} was very low (0.09; 95%CrI: 0.00, 0.23), and although model fit was slightly improved compared with the fixed effects

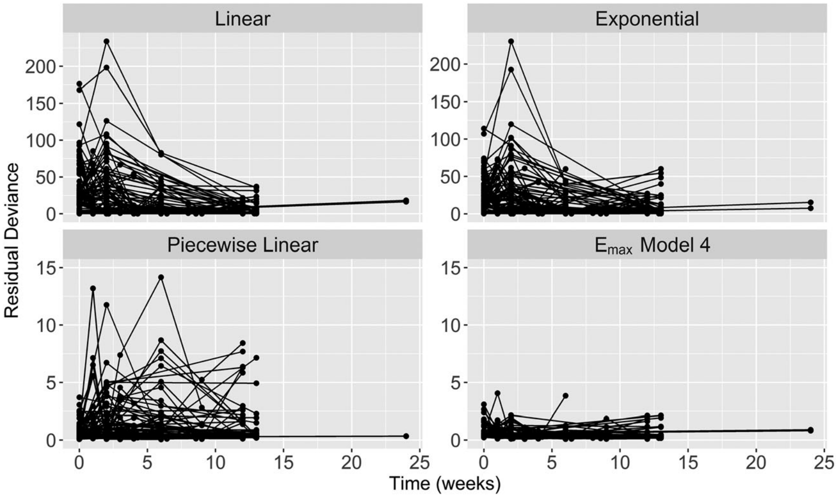

FIGURE 3 Median posterior residual deviance contributions over time from univariate fixed treatment effects models with linear, exponential, piecewise linear, and E_{max} (model 4) time-course relationships in the pain in osteoarthritis dataset. Residual deviances closer to 0 indicate a better fitting model. Nonlinearity in these plots indicates that the effect of time has not been properly accounted for. The scales for residual deviance vary between the upper and lower panels

model (E_{max} model 4), the added complexity resulted in a higher DIC. There was insufficient data to be able to estimate a bivariate random treatment effects model for E_{max} and ET_{50} parameters (Equation 9). Nonetheless, we provide the JAGS code for this model in the Supporting Information. Based on these results, we select the fixed effects E_{max} model 4.

Step 4. Assess overall fit of selected univariate model

The posterior mean residual deviance for E_{max} model 4 was 288.1, which is lower than the number of data points (341), indicating a good fit to the data.

Step 5. Fit the selected time-course model using a multivariate likelihood

Accounting for residual correlation between time points using a multivariate likelihood (Equation 1) for E_{max} model 4 gave an estimated correlation of $\rho = 0.28$ (95%CrI 0.10, 0.41) when using a multivariate compound symmetry covariance structure, and $\rho = 0.50$ (95%CrI 0.19, 0.65) when using a multivariate AR(1) covariance structure (Table 2). Figure S1 (Supporting Information) compares the univariate E_{max} model 4 treatment effects with those from the equivalent multivariate specifications with compound symmetry and autoregressive AR(1) covariance structures. Estimates and their 95% CrIs appear to be reasonably consistent between these models, indicating that accounting for correlation leads to only marginal differences in treatment estimates. Although the differences are very slight, it is interesting to note that use of a multivariate likelihood with compound symmetry covariance structure typically leads to increased precision of treatment estimates compared with the univariate likelihood model, whilst use of a multivariate likelihood with AR(1) covariance structure has more of an effect on the point estimate.

Predictions from the selected model (E_{max} model 4)

Figure 4[46] shows the predicted values from E_{max} model 4 for four illustrative treatments (others are given in Figure S2 - Supporting Information). There are many observations for treatments such as Celebrex 200 mg/d and Naproxen 1000 mg/d, providing rich information on the time-course parameters, whilst for Duloxetine 90 mg/d and Lumiracoxib (NR—dose not reported), the time-course is largely extrapolated and interpolated.

AUC for time-course relationships (E_{max} model 4)

Table 3 shows the median rank and their 95%CrI from E_{max} model 4 for E_{max} treatment effects for each treatment, and the AUC for each treatment calculated over 24 weeks follow-up (the maximum latest follow-up in any of the included studies). As ET_{50} was constrained to be equal across all treatments, the rankings are only dependent on E_{max}, and therefore E_{max} rankings match the AUC rankings. Etoricoxib 60 mg/d was the highest median ranked treatment for both AUC and E_{max}.

Step 6 Test for inconsistency

In the osteoarthritis dataset, all studies included a placebo arm. Within the contrast-based NMA approach, the relative effects within a study are only estimated for each treatment versus the study reference treatment—it is not necessary to estimate relative effects between

TABLE 2 Model fit statistics for the E_{max} model 4 time-course model (see Table 1), comparing univariate and multivariate likelihoods, fitted to the osteoarthritis pain dataset. For the exchangeable baseline parameters, standard deviations (SD) are reported as posterior mean SD (95% credible interval). Correlation is reported as posterior mean (95% credible interval)

	Model for $\lambda_{1,i,k}$ (E_{max})		Model for $\lambda_{2,i,k}$ (ET_{50})					
Time-course model	Arm 1 effect, $\mu_{1,i}$	Relative treatment effects, $\delta_{1,i,k}$	Arm 1 effect, $\mu_{2,i}$	Relative treatment effects, $\delta_{2,i,k}$	Correlation, ρ	DIC[a]	Posterior mean deviance[b]	pD[c]
E_{max} model 4 univariate likelihood	Unconstrained	Fixed effect	Exchangeable (Equation 12), SD = 0.64 (0.39, 1.15)	Fixed effect, $d_{2,1,k}$ equal for all treatments	0	−289.9	−441.9	152.0
E_{max} model 4 multivariate likelihood, compound symmetry	Unconstrained	Fixed effect	Exchangeable (Equation 12), SD = 0.68 (0.43, 1.18)	Fixed effect, $d_{2,1,k}$ equal for all treatments	0.28 (0.10, 0.41)	−266.2	−425.8	159.6
E_{max} model 4 multivariate likelihood, AR(1)	Unconstrained	Fixed effect	Exchangeable (Equation 12), SD = 0.66 (0.41, 1.17)	Fixed effect, $d_{2,1,k}$ equal for all treatments	0.50 (0.19, 0.65)	−278.9	−437.0	158.1

[a] DIC (= deviance + pD): It is a measure of model fit that penalises complexity.

[b] Deviance (= −2(log-likelihood)): A measure of how closely the fitted values of the model fit the observed data.

[c] pD: The total number of effective parameters in the model, calculated using the Kullback-Leibler information divergence.[41]

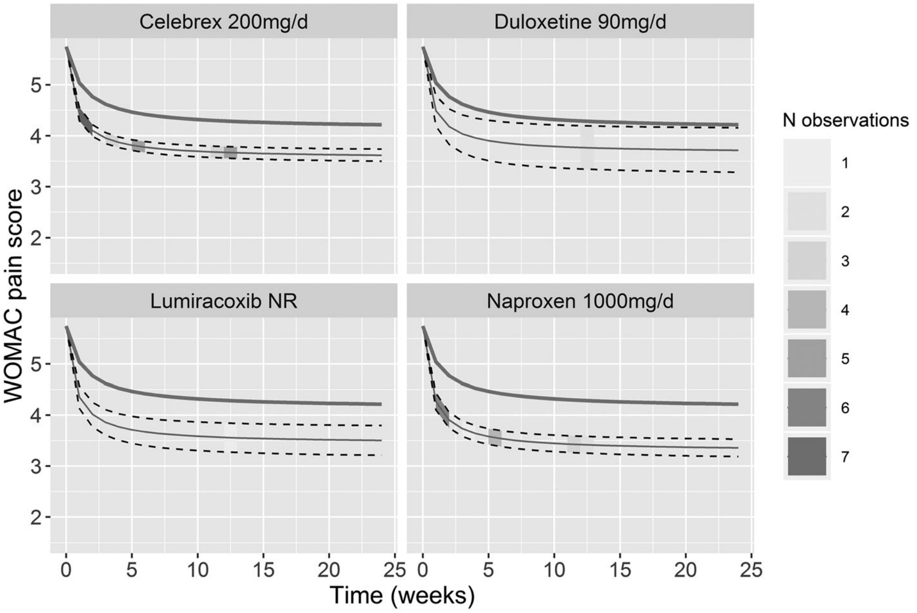

FIGURE 4 Predicted means and 95% CrI from the final model (E_{max} model 4) for the pain in osteoarthritis dataset for Celebrex 200 mg/d, Duloxetine 90 mg/d, Lumiracoxib (dose not reported), and Naproxen 1000 mg/d, plotted over time. The thicker red line indicates the assumed placebo response (calculated from the data). The shading of the 95% CrI indicates observations present in the dataset at each time point [Colour figure can be viewed at wileyonlinelibrary.com]

nonreference treatments within a multi-arm study because these will be defined by the difference between the relative effects for each treatment versus the study reference, as each study must be internally consistent. Therefore, there were no closed loops of treatments in the network that were made up of independent sources of evidence, and as a result it was not possible to test for inconsistency (Figure S3A - Supporting Information).

For illustrative purposes, to create a dataset in which it is possible to test for inconsistency, we expanded our inclusion criteria to all studies irrespective of their method of imputation. This added an extra seven studies

TABLE 3 Median (95%CrI) rankings (1 = best) for AUC and E_{max} treatment effects for E_{max} model 4. Simplifying assumptions on ET_{50} that constrain it to be equal across all treatments mean that the rankings for AUC are identical to the rankings for E_{max} treatment effects

Treatment	Median AUC Rank (95% CrI)	Median E_{max} Rank (95% CrI)
Etoricoxib 60 mg/d	1 (1, 3)	1 (1, 3)
Etoricoxib 90 mg/d	2 (1, 4)	2 (1, 4)
Rofecoxib 125 mg/d	3 (1, 6)	3 (1, 6)
Etoricoxib 30 mg/d	4 (3, 12)	4 (3, 12)
Oxycodone 44 mg/d	5 (1, 25)	5 (1, 25)
Rofecoxib 25 mg/d	6 (4, 15)	6 (4, 15)
Naproxcinod 1500 mg/d	7 (5, 11)	7 (5, 11)
Naproxen 1000 mg/d	10 (6, 14)	10 (6, 14)
Celebrex 400 mg/d	11 (6, 21)	11 (6, 21)
Etoricoxib 10 mg/d	12 (5, 27)	12 (5, 27)
Naproxcinod 750 mg/d	13 (7, 23)	13 (7, 23)
Etoricoxib 5 mg/d	14 (5, 28)	14 (5, 28)
Lumiracoxib (not reported)	14 (7, 24)	14 (7, 24)
Valdecoxib 20 mg/d	15 (6, 25)	15 (6, 25)
Rofecoxib 12 mg/d	16 (7, 25)	16 (7, 25)
Lumiracoxib 100 mg/d	17 (11, 23)	17 (11, 23)
Lumiracoxib 400 mg/d	17 (10, 24)	17 (10, 24)
Tramadol 300 mg/d	17 (8, 24)	17 (8, 24)
Valdecoxib 10 mg/d	17 (7, 26)	17 (7, 26)
Celebrex 200 mg/d	18 (13, 23)	18 (13, 23)
Lumiracoxib 200 mg/d	19 (12, 24)	19 (12, 24)
Valdecoxib 5 mg/d	19 (8, 26)	19 (8, 26)
Tramadol 400 mg/d	20 (8, 27)	20 (8, 27)
Duloxetine 90 mg/d	22 (8, 28)	22 (8, 28)
Celebrex 100 mg/d	25 (17, 27)	25 (17, 27)
Tramadol 200 mg/d	25 (17, 27)	25 (17, 27)
Tramadol 100 mg/d	27 (22, 28)	27 (22, 28)
Placebo 0 mg/d	28 (27, 29)	28 (27, 29)
Naproxcinod 250 mg/d	29 (26, 29)	29 (26, 29)

to create an "augmented dataset" (30 studies in total). The augmented dataset is freely available in the Supporting Information (AugmentedInconsistencyData.csv).[44] One of these additional studies[47] compared Celebrex 200 mg/d, Rofecoxib 25 mg/d, and Naproxen 100 mg/d but not Placebo. This created two loops in the network in which direct and indirect estimates came from independent sources, meaning it was possible to test for inconsistency (Figure S3B - Supporting Information).

In the augmented dataset, results from the UME model were very similar to the MBNMA model (E_{max} model 4). Posterior mean residual deviance was 374.0 for the UME model compared with −370.3 for the MBNMA model, whilst the between-study SD for the reference treatment effect for ET_{50} was almost identical for both UME (0.69 (95%CrI: 0.43, 1.18) and MBNMA models (0.69 (95%CrI: −0.44, 1.18). There is therefore no evidence to invalidate the consistency assumption.

A node splitting model was fitted for the two closed loops of treatments in the network that comprised independent data sources, giving two comparisons on which to node split (Figure S3B - Supporting Information).

For both comparisons, the MBNMA estimate was effectively a weighted average of the direct and indirect estimates, as would be expected, though in both the indirect evidence is more precise and therefore has the greatest influence on the MBNMA result (Figure 5). The

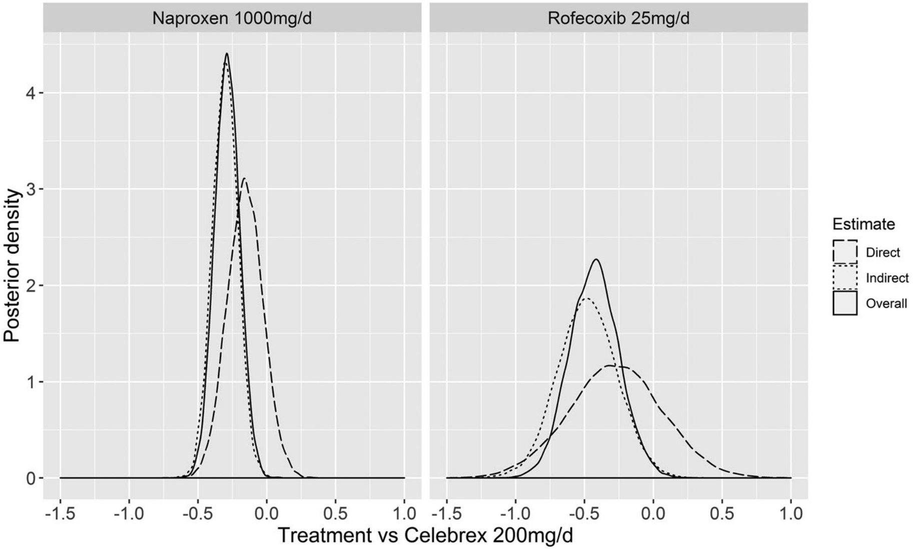

FIGURE 5 Posterior densities for the effect of naproxen (1000 mg/d) versus Celebrex (200 mg/d) and Rofecoxib (25 mg/d) versus Celebrex (200 mg/d) on E_{max} for the direct and indirect evidence arising from node splitting when testing for inconsistency using E_{max} model 4 for the augmented dataset. Bayesian *P*-value of 0.69 and 0.79, respectively, representing the proportion of the densities that overlap

Bayesian *P*-value representing the overlap of the posterior distributions for the direct and indirect evidence was 0.69 for Celebrex 200 mg/d vs Naproxen 1000 mg/d and 0.79 for Celebrex 200 mg/d vs Rofecoxib 25 mg/d indicating no evidence of inconsistency in either loop of treatments.

3.1 | Model checking

For all models for which results are reported, Gelman's r statistic and visual inspection of the chains were indicative of convergence. Convergence diagnostic plots for the final model (E_{max} model 4) and for a nonconverging "ideal" E_{max} model with no simplifying assumptions on ET_{50} are given in the Supporting Information.

Posterior densities were not constrained or strongly influenced by priors. In the final model (E_{max} model 4), the SD for the exchangeable study reference treatment effect was 0.64 (95%CrI: 0.39, 1.15) when using a uniform (*U*(0,5)) prior and 0.64 (95%CrI: 0.39, 1.16) when using a half-normal () prior (Figure S4 - Supporting Information). For the random E_{max} treatment effects model (E_{max} model 5), the between-study SD for random E_{max} treatment effects was 0.09 (95%CrI: 0.00, 0.23) when using either a uniform (*U*(0,5)) prior or a half-normal () prior, suggesting the choice of prior is not influential.

4 | DISCUSSION

We have presented a method to pool evidence from trials that form a network of comparisons across multiple treatments, reporting continuous summary outcome measures over multiple time points. The method respects the randomisation in RCTs, can include a variety of different functional forms for the time-course, allows for testing of consistency of the evidence, and demonstrates how a multivariate likelihood can be used to account for residual correlation between time points.

In the pain in osteoarthritis dataset, we found that the E_{max} model provided the best fit and allowed for the greatest degree of flexibility, both in the time-course shape and in the specification of various time-course parameters (E_{max} and ET_{50}). The E_{max} model with the lowest DIC (E_{max} model 4) assumed a fixed effect across all nonplacebo treatments on ET_{50} (Equation 12). In this model, we estimated AUC of the time-course function over 24 weeks follow-up and found Etoricoxib 60 mg/d to have the highest ranked AUC.

Whilst WOMAC pain in this dataset was measured as a continuous outcome, MBNMA can also be performed on any data provided it can be *summarised* as a continuous outcome that can be assumed to have a normal likelihood. This therefore allows for binary data (ie, % respondents at each follow-up time) to be analysed using MBNMA if they are summarised as log-odds and are not near the boundary probabilities (0% or 100%).

4.1 | Time-course function

Previous methods for performing NMA on longitudinal data have typically accounted for the dependency between different follow-up times by either modelling an appropriate time-course function[18,19] or using nonparametric approaches to account for responses at each time point.[21,34] To our knowledge, none of the methods for modelling longitudinal continuous data in NMA have described methods for testing inconsistency.

Jansen et al[19] used fractional polynomials to model a nonlinear trend, with an inflated variance for each time point to approximate correlations between observations. These allow for a very flexible shape for modelling the time-course relationship. However, fractional polynomials are difficult to interpret and may not have a clear biological justification, making them less desirable for use in pharmacometric studies, where the objective is to define how the efficacy of a treatment changes over time rather than to simply account for it. For modelling time-course in pharmacometrics, exponential and E_{max} functions are typically used, as these have a biological basis in describing the mechanism of drug action and elimination from the body.[48] Taking this into account, Ding and Fu[18] modelled an exponential shape in NMA and described how their model could be adapted for a sigmoid shape similar to that of an E_{max} function.

Nonparametric models that do not specify a particular time-course relationship have also been proposed, although these do not allow for any interpolation or extrapolation of treatment effects at unmeasured points in time, which makes them less applicable to drug development. Ishak et al[21] used a multivariate likelihood to account for the dependency between different time points. Dakin et al[34] used a saturated model to estimate treatment effects separately for different time bins, although for this method more data are required to reliably estimate the treatment effects for each bin.

One of the strengths of our MBNMA framework is that information on time-course characteristics can be inferred from other treatments or agents by assuming varying degrees of similarity and thus improving identifiability. However, sharing information on time-course parameters across a network assumes exchangeability, and it is important to be aware of the implications of this assumption and to consider whether it holds across different agents or classes of treatments.

4.2 | Correlation between time points

Having selected an appropriate time-course relationship in the pain in osteoarthritis dataset, we found that estimated residual correlation was reasonably high. This was in contrast to our expectation that explicitly modelling the time-course should have generated conditional independence. However, accounting for this correlation had only a slight impact on treatment effect estimates or 95% CrIs. This suggests that for MBNMA where the focus is on summary estimates, whilst accurately characterising the within-study correlation and covariance structure may be important, it is likely to be less critical than accurately characterising the time-course. With regards to the importance of modelling within-study correlation, there is some debate in the literature. A simulation study by Ishak et al[49] suggests that the impact of ignoring within-study correlation on treatment estimates may typically be small, even in cases where there has been no specific modelling of a time-course function. However, Riley[50] has shown that this is only the case when between-study variation is large relative to within-study variation, or when there are complete data with only small differences in the within-study covariance matrices across studies. Ahn and French[2] support this position, demonstrating that ignoring correlation in longitudinal MBMA led to inflated residual variance. We are currently performing a simulation study to further examine the relationship between time-course fit and correlation in MBNMA.

Note that the estimated covariance matrix in the multivariate likelihood will depend on the time-course model fitted, and the strength and importance of correlations between time points are also likely to depend considerably on how close together follow-up measurements are in time, with closer measurements expected to be more strongly correlated.

Previous longitudinal MBMA methods that account for correlated residuals have been developed in a frequentist framework, using NONMEM software to allow for modelling of interarm variability in addition to interstudy variability.[2] The authors used an exponential model for the time-course and also accounted for nonlinear dose-response in their model. Although we follow a Bayesian approach and our model is formulated somewhat differently, our multivariate model with a compound symmetry covariance structure is similar to their method. However, the key difference is that our approach respects randomisation and allows for inconsistency testing. Without these features, the methods are unlikely to meet the requirements of reimbursement agencies.

4.3 | Modelling assumptions

For our selected E_{max} model in the pain in osteoarthritis dataset (E_{max} model 4), assuming a fixed effect across all nonplacebo treatments for ET_{50}, whilst allowing all treatment effects to be different for E_{max} implies that the onset of action is the same for the different treatments relative to placebo, but that treatments differ in the maximal change in outcome achieved. In practice, this might be considered an unusual modelling assumption, as one might expect ET_{50} to differ between active treatments, particularly for those acting via different biological pathways. However, the onset for all these treatments was very rapid, and there were insufficient observations at early time points to reliably estimate this. In fact, none of the included studies report WOMAC scores within a week from baseline. Given that this is later than the estimated ET_{50} (approximately 0.6 weeks), it is not surprising that this parameter is difficult to estimate. For this example, we would therefore caution against making inferences at very early time points. Using informative prior distributions for the $\tau^2_{ET_{50}}$ or for $d_{ET_{50},1,k}$ parameters for which information is sparse may improve estimation. Information from noncomparative pharmacodynamic studies of different agents may be useful to provide support for specific prior distributions. Another approach may be to incorporate information from other treatments in a more biologically plausible manner, such as by using the dose-response relationships between treatments within an agent.[17]

Explicitly modelling the bivariate correlation between E_{max} and ET_{50} may in some circumstances also provide additional information to help identify ET_{50} and reduce the need for such strong simplifying assumptions. This is likely to be the case when correlation between E_{max} and ET_{50} is high. However, in the pain in osteoarthritis example, this still was not sufficient to help identify ET_{50}, even when alternative parameterizations for the covariance matrix were used.[31]

Within time-course MBNMA, it is necessary to assume the same time-course function for the included set of treatments. Whilst the mean responses over time for most treatments supported the use of an E_{max} function, there were only two observations (baseline and one follow-up measurement) for studies comparing Lumiracoxib (NR—dose not reported) vs Placebo and Duloxetine 90 mg/d vs Placebo.

For Lumiracoxib NR, it was reasonable to assume that the time-course function will be similar to other doses of Lumiracoxib for which there are more observations. However, there are no other doses of Duloxetine to make an equivalent assumption, and as the mechanism of Duloxetine is also different to any other agent in the dataset, it may follow a different time-course function. Yet, as this treatment did not contribute any indirect evidence to the rest of the network (which could induce bias in other treatment estimates if modelled appropriately) and there was no evidence to suggest a different time-course function would be applicable for this treatment, we feel it is reasonable to assume a similar time-course function to other treatments provided treatment effect estimates for Duloxetine 90 mg/d are interpreted with caution. This case highlights the importance of understanding the underlying pharmacometrics of treatments in the data, and of dialogue between clinicians, pharmacometricians, and analysts.

An additional assumption made in all the E_{max} models due to the inclusion of Lumiracoxib NR and Duloxetine 90 mg/d was that of an exchangeable placebo (Equation 12), as the lack of multiple follow-up measurements made separate estimation of all three parameters for the time-course (E_0, ET_{50}, and E_{max}) impossible for these comparisons. This is likely to have caused a certain degree of shrinkage and may therefore induce bias in treatment estimates for ET_{50} due to back-propagation of the information on the reference treatment.[9]

4.4 | Inconsistency

In NMA, two approaches are typically used for dealing with longitudinal studies. A single consistent time point may be used for analysis across studies, ignoring evidence from other time points. Alternatively, the final time point from studies with different follow-up times may be "lumped" together to allow for networks to be connected, yet this lumping can often be a source of inconsistency and/or heterogeneity.[16] Whilst MBNMA solves the issue of lumping together studies with different follow-up times or discarding information on multiple time points, the choice of model will affect the presence of inconsistency. We suspect that a poorly fitting time-course model may induce inconsistency. It is therefore important to explore different functional forms and identify a good model *before* testing for inconsistency.

For the purposes of drug development, the potential for inconsistency testing in MBNMA may in fact be relatively rare. The typical design of Phase II trials is multi-arm placebo-controlled, meaning that there are no closed loops of treatments that are not made up of multi-arm trials (as in the illustrative osteoarthritis dataset). As these trials must inherently be internally consistent, this provides no means to test for inconsistency. However, we are still relying on the consistency assumptions to make indirect comparisons, so although in these cases we cannot formally test for inconsistency it is important to consider whether these assumptions are valid.

For the pain in osteoarthritis augmented dataset, we did not find any evidence of inconsistency in the augmented dataset when including non-LOCF studies, and parameter estimates were robust to their inclusion. In practice, we would recommend careful consideration of inclusion criteria to ensure only studies on which the consistency assumption is expected to hold are included.[13]

It is worth noting that a standard NMA performed by "lumping" the latest time point in each of the studies (an approach frequently used but not one that we would recommend) highlights the benefit of performing MBNMA when dealing with different follow-up times. In terms of DIC, a random treatment effects NMA was preferred over a fixed treatment effects NMA, with a nonzero between-study SD (0.25 (95%CrI: 0.17, 0.36)). A Bayesian *P*-value for the node-split of Celebrex 200 mg/d vs Naproxen 1000 mg/d was 0.011 and for Celebrex 200 mg/d vs Rofecoxib 25 mg/d was 0.100 suggesting reasonable evidence of inconsistency in both comparisons that was particularly concerning given that the direct and indirect evidence for both showed opposite directions of effect. Therefore, by accounting for time-course using MBNMA, we have explained heterogeneity and inconsistency that can arise when using standard NMA methodology.

4.5 | Limitations

There are a few limitations to the methodology that we seek to investigate further in simulation studies. The first is that the quantity of data required for MBNMA may be significant, particularly for more complex time-course functions. This therefore means that analyses may typically require strong simplifying assumptions within the modelling that are difficult to test. This could relate to the assumption that time-course functions are the same across all treatments within the network, but also to the need to interpolate or extrapolate over the time-course when few observations are available for a particular treatment.

We also do not fully understand the importance of correctly accounting for the correlation between time points, and how failure to do this might affect estimates for different parameters in the model. We believe that the impact of this is likely to depend on the key parameters of interest in the model. Simulation can help to shed light on this issue.

4.6 | Future direction

In future work, we plan to incorporate simultaneous dose-response and time-course modelling into the MBNMA framework,[17] and to develop simulation studies to explore the robustness and data requirements of dose-response, time-course, and methods for assessing inconsistency in MBNMA. We are also developing an R package for MBNMA to facilitate its implementation.

5 | CONCLUSION

MBNMA combines the strengths of both MBMA and NMA, leading to a statistically robust framework for synthesising evidence on multiple treatments at multiple time points whilst preserving randomisation and allowing for assessment of consistency. By unifying these statistical techniques, the methods can provide both the information needed to inform drug-development decisions, and also the rigour required by reimbursement agencies to incorporate valuable evidence from drug development into the decision-making process.

5.1 | Highlights

Within drug development, MBMA is increasingly used to inform drug decisions such as whether to proceed to further clinical trials, and if so, what the design of the study should be. However, these types of analyses typically model each drug separately by pooling individual study arms, which breaks the randomisation and ignores within-study comparisons, effectively losing the advantages of the RCT design and potentially inducing bias in the resulting estimates.

Our MBNMA framework preserves randomisation by modelling relative effects and allows for testing of inconsistency between direct and indirect evidence. This manuscript extends previous methodology on dose-response MBNMA to allow the modelling of nonlinear time-course characteristics, incorporating multiple study time points and accounting for correlation between them.

MBNMA combines techniques from two different disciplines, pharmacometrics and evidence synthesis, thereby acting as a bridge between early phase clinical research and Health Technology Appraisal.

Glossary

Agent = an intervention/compound/drug
Identifiability = the capacity for parameters in a model to be reliably estimated
Treatment = a specific dose and agent combination.

ACKNOWLEDGEMENTS

N.J.W., S.D., and H.P. were supported by a grant from the MRC Methodology Research Programme (ref MR/M005615/1) which included an MRC Industry Collaboration Agreement with Pfizer Ltd. which part-funded H.P. M.B. and M.B. were supported by Pfizer Ltd. S.D. was also supported by MRC New Investigator Research Grant (MR/M005232/1). N.J.W. was also supported by the MRC ConDuCT-II Hub for Trials Methodology Research (MR/K025643/1) and the National Institute for Health Research Biomedical Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol.

CONFLICT OF INTEREST

The author reported no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in "figshare" at https://doi.org/10.6084/m9.figshare.8138045.v1.

ORCID

Hugo Pedder https://orcid.org/0000-0002-7813-3749
Sofia Dias https://orcid.org/0000-0002-2172-0221
Margherita Bennetts https://orcid.org/0000-0001-5198-5274
Nicky J. Welton https://orcid.org/0000-0003-2198-3205

REFERENCES

1. Mould DR. Model-based meta-analysis: an important tool for making quantitative decisions during drug development. *Clin Pharmacol Ther*. 2012;92(3):283-286.
2. Ahn JE, French JL. Longitudinal aggregate data model-based meta-analysis with NONMEM: approaches to handling within treatment arm correlation. *J Pharmacokinet Pharmacodyn*. 2010;37(2):179-201.
3. Mandema JW, Boyd RA, DiCarlo LA. Therapeutic index of anticoagulants for prevention of venous thromboembolism following orthopedic surgery: a dose-response meta-analysis. *Clin Pharmacol Ther*. 2011;90(6):820-827.
4. Mandema JW, Salinger DH, Baumgartner SW, Gibbs MA. A dose-response meta-analysis for quantifying relative efficacy of biologics in rheumatoid arthritis. *Clin Pharmacol Ther*. 2011; 90(6):828-835.
5. Checchio T, Ahadieh S, Gupta P, et al. Quantitative evaluations of time-course and treatment effects of systemic agents for psoriasis: a model-based meta-analysis. *Clin Pharmacol Ther*. 2017;102(6):1006-1016.
6. Demin I, Hamren B, Luttringer O, Pillai G, Jung T. Longitudinal model-based meta-analysis in rheumatoid arthritis: an application toward model-based drug development. *Clin Pharmacol Ther*. 2012;92(3):352-359.
7. Korell J, Martin SW, Karlsson MO, Ribbing J. A model-based longitudinal meta-analysis of FEV1 in randomized COPD trials. *Clin Pharmacol Ther*. 2016;99(3):314-324.
8. Hong H, Chu H, Zhang J, Carlin BP. A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons. *Res Synth Methods*. 2016;7(1):6-22.
9. Dias S, Ades AE. Absolute or relative effects? Arm-based synthesis of trial data. *Res Synth Methods*. 2016;7(1):23-28.
10. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*. 2005;331(7521):897-900.
11. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15(24):2733-2749.
12. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20): 3105-3124.
13. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network meta-analysis for decision making*. Wiley; 2018.
14. Efthimiou O, Debray TP, van Valkenhoef G, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*. 2016;7(3):236-263.
15. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29(7–8):932-944.
16. 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 Mak*. 2013;33(5):641-656.
17. Mawdsley D, Bennetts M, Dias S, Boucher M, Welton NJ. Model-based network meta-analysis: a framework for evidence synthesis of clinical trial data. *CPT Pharmacometrics Syst Pharmacol*. 2016;5(8):393-401.
18. Ding Y, Fu H. Bayesian indirect and mixed treatment comparisons across longitudinal time points. *Stat Med*. 2013;32(15): 2613-2628.
19. Jansen JP, Vieira MC, Cope S. Network meta-analysis of longitudinal data using fractional polynomials. *Stat Med*. 2015;34(15): 2294-2311.
20. Wandel S, Juni P, Tendal B, et al. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. *BMJ*. 2010;341(sep16 2): c4675.
21. Ishak KJ, Platt RW, Joseph L, Hanley JA, Caro JJ. Meta-analysis of longitudinal studies. *Clin Trials*. 2007;4(5):525-539.
22. Jones AP, Riley RD, Williamson PR, Whitehead A. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clin Trials*. 2009;6(1):16-27.
23. Musekiwa A, Manda SO, Mwambi HG, Chen DG. Meta-analysis of effect sizes reported at multiple time points using general linear mixed model. *PLoS ONE*. 2016;11(10): e0164898.
24. Trikalinos TA, Olkin I. Meta-analysis of effect sizes reported at multiple time points: a multivariate approach. *Clin Trials*. 2012;9(5):610-620.
25. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med*. 2007;26(20):3681-3699.
26. 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.
27. 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.
28. Wei Y, Higgins JP. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Stat Med*. 2013;32(7):1191-1205.

29. Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. *Stat Med*. 2005;24(24):3823-3844.
30. 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 Mak*. 2013;33(5):607-617.
31. Pinheiro JC, Bates DM. Unconstrained parameterizations for variance-covariance matrices. *Stat Comput*. 1996;6(3):289-296.
32. Raiffa H, Schlaiffer R. *Applied Statistical Decision Theory*. Wiley Classics Library edn. ed. New York: Wiley Interscience; 1967.
33. Warren FC, Abrams KR, Sutton AJ. Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. *Stat Med*. 2014;33(14):2449-2466.
34. Dakin HA, Welton NJ, Ades AE, Collins S, Orme M, Kelly S. Mixed treatment comparison of repeated measurements of a continuous endpoint: an example using topical treatments for primary open-angle glaucoma and ocular hypertension. *Stat Med*. 2011;30(20):2511-2535.
35. Owen RK, Tincello DG, Keith RA. Network meta-analysis: development of a three-level hierarchical modeling approach incorporating dose-related constraints. *Value Health*. 2015; 18(1):116-126.
36. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc*. 2006;101(474):447-459.
37. van Valkenhoef G, Dias S, Ades AE, Welton NJ. Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. *Res Synth Methods*. 2016;7(1):80-93.
38. Donegan S, Welton NJ, Tudur Smith C, D'Alessandro U, Dias S. Network meta-analysis including treatment by covariate interactions: consistency can vary across covariate values. *Res Synth Methods*. 2017;8(4):485-495.
39. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc B*. 2002; 64(4):583-639.
40. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. The deviance information criterion: 12 years on. *J R Stat Soc B*. 2014; 76(3):485-493.
41. Plummer M. Penalized loss functions for Bayesian model comparison. *Biostatistics*. 2008;9(3):523-539.
42. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15(12):1833-1840.
43. Boucher M. Imputation of missing variance data using non-linear mixed effects modelling to enable an inverse variance weighted meta-analysis of summary-level longitudinal data: a case study. *Pharm Stat*. 2012;11(4):318-324.
44. Boucher M, Pedder H. Dataset containing aggregate data from studies investigating treatments for pain in osteoarthritis. figshare. Dataset. 2019. https://doi.org/10.6084/m9.figshare.8138045.v1
45. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457-472.
46. Donegan S, Dias S, Tudur-Smith C, Marinho V, Welton NJ. Graphs of study contributions and covariate distributions for network meta-regression. *Res Synth Methods*. 2018;9(2):243-260.
47. Sowers J, White W, Pitt B, et al. The effects of cyclooxygenase-2 inhibitors and nonsteroidal anti-inflammatory therapy on 24-hour blood pressure in patients with hypertension, osteoarthritis, and type 2 diabetes mellitus. *Arch Intern Med*. 2005; 165(2):161-168.
48. Ette EI, Williams PJ. *Pharmacometrics: The Science of Quantitative Pharmacology*. John Wiley & Sons; 2007.
49. Ishak KJ, Platt RW, Joseph L, Hanley JA. Impact of approximating or ignoring within-study covariances in multivariate meta-analyses. *Stat Med*. 2008;27(5):670-686.
50. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *J R Stat Soc a*. 2009;172(4):789-811.

SUPPORTING INFORMATION

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

How to cite this article: Pedder H, Dias S, Bennetts M, Boucher M, Welton NJ. Modelling time-course relationships with multiple treatments: Model-based network meta-analysis for continuous summary outcomes. *Res Syn Meth*. 2019;10: 267–286. https://doi.org/10.1002/jrsm.1351