UNIVERSITY *of* York

This is a repository copy of *Methods to Assess Evidence Consistency in Dose-Response Model Based Network Meta-Analysis*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/180525/

Version: Published Version

Article:

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

RESEARCH ARTICLE

Methods to assess evidence consistency in dose-response model based network meta-analysis

Hugo Pedder[1] | Sofia Dias[2] | Martin Boucher[3] | Meg Bennetts[3] |
David Mawdsley[1] | Nicky J. Welton[1]

[1]Population Health Sciences, Canynge Hall, University of Bristol, Bristol, UK

[2]Centre for Reviews and Dissemination, University of York, York, UK

[3]Pharmacometrics, Pfizer Ltd., Sandwich, Kent, UK

Correspondence
Hugo Pedder, Population Health Sciences, Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Rd, Bristol BS8 2PS, UK.
Email: hugo.pedder@bristol.ac.uk

Abstract

Network meta-analysis (NMA) simultaneously estimates multiple relative treatment effects based on evidence that forms a network of treatment comparisons. Heterogeneity in treatment definitions, such as dose, can lead to a violation of the consistency assumption that underpins NMA. Model-based NMA (MBNMA) methods have been proposed that allow functional dose-response relationships to be estimated within an NMA, which avoids lumping different doses together and thereby reduces the likelihood of inconsistency. Dose-response MBNMA relies on appropriate specification of the dose-response relationship as well as consistency of relative effects. In this article we describe methods to check for inconsistency in dose-response MBNMA models. Global and local (node-splitting) tests for inconsistency are described that account for studies with ≥ 3 arms that are typical in dose-finding trials. We show that consistency needs to be assessed with respect to the choice of dose-response function. We illustrate the methods using a network comparing biologics for moderate-to-severe psoriasis. By comparing results from an Emax and an exponential dose-response function we show that failure to correctly characterise the dose-response can introduce apparent inconsistency. The number of comparisons for which node-splitting is possible is also shown to be dependent on the complexity of the selected dose-response function. We highlight that the nature of dose-finding studies, which typically compare multiple doses of the same agent, provide limited scope to assess inconsistency, but these study designs help guard against inconsistency in the first place. We demonstrate the importance of assessing consistency to obtain robust relative effects to inform drug-development and policy decisions.

KEYWORDS

consistency, dose, heterogeneity, meta-analysis, NMA, synthesis

1 | INTRODUCTION

Network meta-analysis (NMA) is a well-established methodology for the combination of results from randomized controlled trials (RCTs) where the evidence forms a connected network of treatment comparisons.[1,2] The popularity of NMA in healthcare has been driven by its ability to simultaneously compare the relative efficacy of multiple treatment options, which is essential to inform health care policy and for the development of clinical guidelines. NMA relies on the consistency assumption, which can be thought of as agreement between the direct and indirect evidence for a particular comparison between two treatments (defined as a contrast).[3] If there is an imbalance in treatment effect modifiers (patient characteristics or other study-level factors that change the relative treatment effects) between the included studies, then this can lead to heterogeneity if these differences are between studies informing the same contrast, or/and inconsistency if there are differences between studies informing different contrasts.[2] Differences in treatment characteristics is one factor that may induce inconsistency and/or heterogeneity. Key characteristics of pharmacological treatments that may vary between studies could be drug formulation, route/frequency of administration and administered dose. It is therefore important to account for these differences in NMA to help avoid inconsistency and heterogeneity.

There are two main approaches to account for dose in NMA. The simplest is to define each agent and dose combination as a distinct treatment in the NMA. However, this may lead to sparse or disconnected evidence networks which yield imprecise estimates, or no estimates at all for some comparisons.[4] An alternative is to model the dose-response relationship with a parametric function. A general model-based NMA (MBNMA) framework has been proposed to fit parametric dose-response functions to combine evidence on multiple agents at multiple doses whilst respecting the randomization in the RCTs.[4] Compared with a standard NMA where each dose and agent combination is considered a distinct treatment, dose-response MBNMA models bring the potential to connect disconnected networks and increase precision of estimates for treatments of interest.[5] Dose-response MBNMA is particularly useful in the context of drug-development to inform decisions as to which dose of an agent to take forwards for further development, where it is crucial to understand not just the dose-response for the compound being developed, but also that of the key competitor drug(s). Whilst accounting for dose helps to reduce inconsistency and heterogeneity that may arise from differences in treatment characteristics, they may not be completely eliminated, and methods are needed to check for it within the dose-response MBNMA framework. A thorough assessment of the consistency assumption is therefore important to increase confidence that the resulting treatment effect estimates provide a firm basis for decision-making, whether for reimbursement decisions, clinical guidelines, or for quantitative drug development.

Standard methods to check for inconsistency in NMA include a simple *z*-test to compare direct and indirect estimates from a single "loop" of evidence (the Bucher method[6]), an extension to multiple independent loops of evidence (repeated Bucher method[7]), fitting models that relax the consistency assumption globally (unrelated mean effects model[3,8]) or locally for each comparison (node-splitting[9]), and cross-validation.[10] Assessing consistency in dose-response MBNMA is complicated because consistency depends on the functional form assumed for the dose-response relationship, and so is inextricably linked to model fit of the dose-response relationship. Furthermore, dose-response MBNMA typically includes many RCTs with three or more arms, which are problematic for standard methods of assessing consistency in NMA.

The aim of this article is to present methods for assessment of evidence consistency in a dose-response MBNMA framework. We begin by introducing a network of evidence comparing different biologics at different doses for the treatment of psoriasis, as an example with which to motivate and illustrate the methods.[11] We then give some background on the dose-response MBNMA modeling framework before describing how methods to assess consistency in NMA can be extended to dose-response MBNMA. We apply the methods to the psoriasis dataset, and end with a discussion and recommendations for practice.

2 | MOTIVATING EXAMPLE: BIOLOGICS FOR PSORIASIS

We motivate the methods with a network of trials comparing the efficacy of biologics for the treatment of moderate to severe psoriasis.[11] The original dataset, which is available within the MBNMAdose R package,[12] included nine different agents (biologics) given at various doses, and placebo. Two agents (Adalimumab and Guselkumab) were only investigated at a single dose. As inclusion of these agents would limit the complexity of different dose-response functions that could be
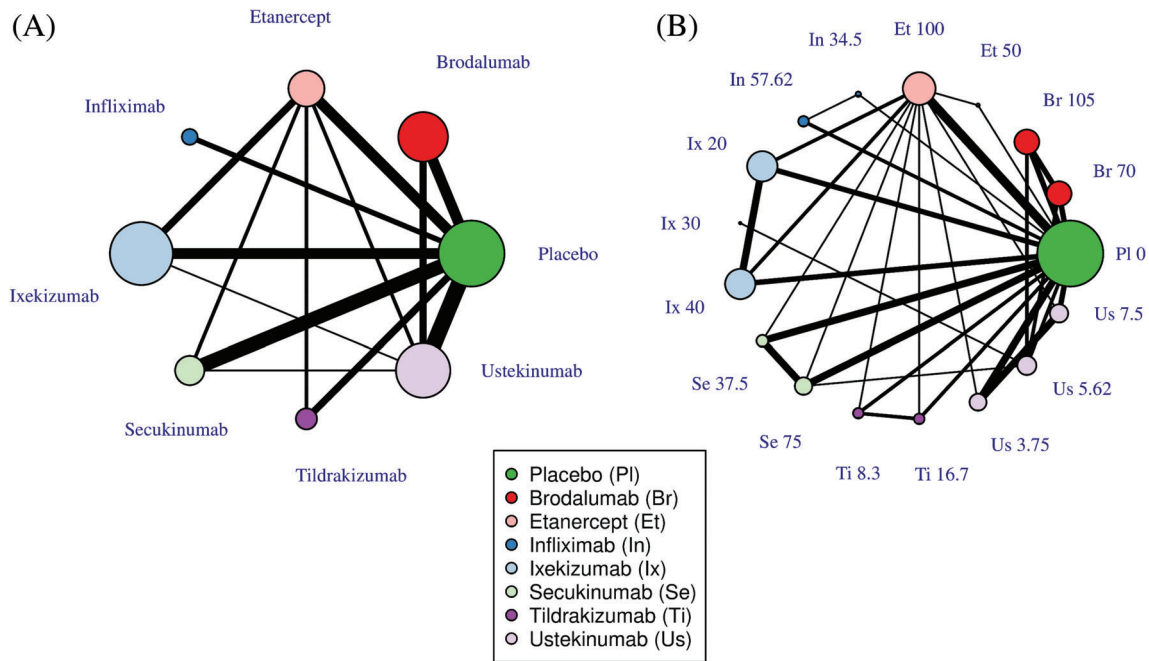
FIGURE 1 Network diagram showing the comparisons made by the RCTs included in the moderate-to-severe psoriasis dataset at (A) agent-level (dose and agent combination) and (B) treatment-level. Where direct trial evidence exists between a pair of agents/treatments they are joined by a line. Each agent/treatment is represented by a node, coloured by agent with agent and dose indicated alongside each node. Width of lines is proportional to the number of studies that make a given agent/treatment comparison. Size of nodes is proportional to the number of participants randomised to a given agent/treatment. Interventions are coded according the first two letters of the agent and dose (mg/week). So Br 70 is Brodalumab at 70 mg/week, and so forth. See Tables S1 and S2 for the full dataset and list of comparisons made in the studies [Colour figure can be viewed at wileyonlinelibrary.com]

fitted, these agents were excluded for the purposes of illustrating consistency checking on a multiparameter dose-response function. The final dataset therefore included seven agents and placebo, investigated in 22 RCTs (Figure S1 and Table S1). Only five of these did not include a placebo arm.

Figure 1A,B show the network of evidence at the agent-level and treatment-level (agent and dose combination) respectively. If we were to ignore dose (agent-level network Figure 1A) then we would be likely to introduce heterogeneity and inconsistency, whereas if we were to consider each treatment (dose-agent combination) as having separate relative effects (agent-by-dose network Figure 1B), then the network becomes very sparse giving imprecise estimates, and only allows comparisons to be made between the specific dose-agent combinations that are connected in the network. Dose-response MBNMA achieves more precision by accounting for dose, however we still need to check for inconsistency.

The outcome of interest is the proportion of patients who achieved ≥75% improvement in psoriasis area and sensitivity index (PASI) at 12 weeks follow-up. The dose for each drug is given in total mg/week to account for different dosing regimens. Two studies reported the dose in mg/kg, so the total dose was calculated by multiplying this by the mean body weight of participants in each arm.[13,14] An additional arm in one of these studies that was not included in the original review was also added to our dataset.[14]

3 | METHODS

3.1 | Dose response model based network meta-analysis (MBNMA)

Mawdsley et al proposed the following framework for dose-response MBNMA.[4] Let $t_{i,k}$ be the treatment given in arm k of study i. Each treatment t is defined according to its dose x_t and agent a_t, where the agent takes values $\{A, B, C, \dots\}$. Placebo is defined as any agent at zero dose, $x_t = 0$, and it is assumed this is equivalent across agents and denoted as $a_t = P$ when $x_t = 0$. However, the inclusion of placebo in the network is not a requirement for the analysis, and these

methods can be used in the absence of studies investigating placebo. For ease of notation we will denote $x_{i,k} = x_{t_{i,k}}$ to indicate the dose of the treatment on arm k of study i. Similarly, we denote $a_{i,k} = a_{t_{i,k}}$. When x and a have two indices they refer to a specific arm and study, and when they have one index this refers to a specific treatment. Each study reports a summary measure for each arm, which, following Dias et al. is modeled using a generalized linear model, where the likelihood is a function of parameter $\theta_{i,k}$ and the MBNMA model is given for an appropriate "linking" function $g\left(\theta_{i,k}\right)$.[2,15] For example, the summary measure may be the number of events $r_{i,k}$ out of $n_{i,k}$ randomized individuals, which has a Binomial likelihood with probability of response parameter, $\theta_{i,k}$, modeled on the logit-scale, so that $g\left(\theta_{i,k}\right) = \log\left(\frac{\theta_{i,k}}{1-\theta_{i,k}}\right)$. Other link functions can be used for other likelihoods/outcome measures.[2,15]

The modeled outcome on arm 1 (the study reference treatment) of each study, μ_i, is treated as a nuisance parameter, with treatment effects $\delta_{i,k}$ estimated relative to this:

$$g\left(\theta_{i,k}\right) = \begin{cases} \mu_i & k = 1 \\ \mu_i + \delta_{i,k} & k \geq 2 \end{cases}.$$

The general random effects model is:

$$\delta_{i,k} \sim N\left(d_{(x_{i,1},a_{i,1}),(x_{i,k},a_{i,k})}, \sigma^2\right), \tag{1}$$

where for any pair of treatments c and k, $d_{(x_c,a_c),(x_k,a_k)}$ is the mean relative effect for treatment k (agent a_k dose x_k) compared with treatment c (agent a_c dose x_c), and σ^2 is the between studies variance in relative effects, assumed to be constant across treatment comparisons. Methods have been proposed to estimate separate between studies variances for different treatment comparisons, but these require more studies than are typically available in a NMA.[16] The common effect model is obtained by setting $\sigma = 0$.

Trials with more than two arms (multiarm trials) are very common in early phase trials that explore dose-response, but they add complications because they provide estimates of multiple relative effects (compared with arm 1) which will be correlated. Equation (1) needs adjusting as follows, to account for correlations in effects for a trial with K arms:

$$\begin{pmatrix} \delta_{i,2} \\ \delta_{i,3} \\ \vdots \\ \delta_{i,K} \end{pmatrix} \sim N_{(K-1)}\left(\begin{pmatrix} d_{(x_{i,1},a_{i,1}),(x_{i,2},a_{i,2})} \\ d_{(x_{i,1},a_{i,1}),(x_{i,3},a_{i,3})} \\ \vdots \\ d_{(x_{i,1},a_{i,1}),(x_{i,K},a_{i,K})} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \cdots & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 & & \sigma^2/2 \\ \vdots & & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \cdots & \sigma^2 \end{pmatrix} \right). \tag{2}$$

We define a dose-response curve $E_0 + f(x, \beta_a)$ for each agent a, governed by agent-specific parameters, β_a, where E_0 corresponds to a placebo effect (Figure 2). Examples of dose-response relationships, $E_0 + f(x, \beta_a)$, include linear, exponential, or Emax models. The choice of an appropriate functional form of dose-response relationship will typically be guided by visual inspection of the data, together with model fit and pharmacological considerations (see model critique and selection). The Emax model is perhaps the most common model used to describe dose-response relationships, and has been shown to be applicable to the majority of compounds and outcome measures.[17,18] The hyperbolic form is given by:

$$f\left(x, E_{\max,a}, ED_{50,a}\right) = \frac{E_{\max,a}\, x}{ED_{50,a} + x}, \tag{3}$$

where $E_{\max,a}$ can be interpreted as the maximum efficacy on the corresponding link scale of agent a relative to placebo, and $ED_{50,a}$ the dose at which half of the $E_{\max,a}$ response has been accrued for agent a. On the logit scale, $E_{\max,a}$ would be the log-odds ratio for agent a relative to placebo. A simpler model with only a single dose-response parameter that is commonly used in pharmacometrics is the exponential model, $f(x, \lambda_a) = \lambda_a(1 - e^{-x})$, where λ_a is the rate on the corresponding link scale of agent a relative to placebo.[17]

In a Bayesian analysis, priors are given to the parameters of the dose-response function (eg, for the Emax model these are: $E_{\max,a}$ and $ED_{50,a}$, noting that $ED_{50,a}$ must be positive), the study-specific baseline parameters μ_i and the between study SD σ.

Dose-response consistency model

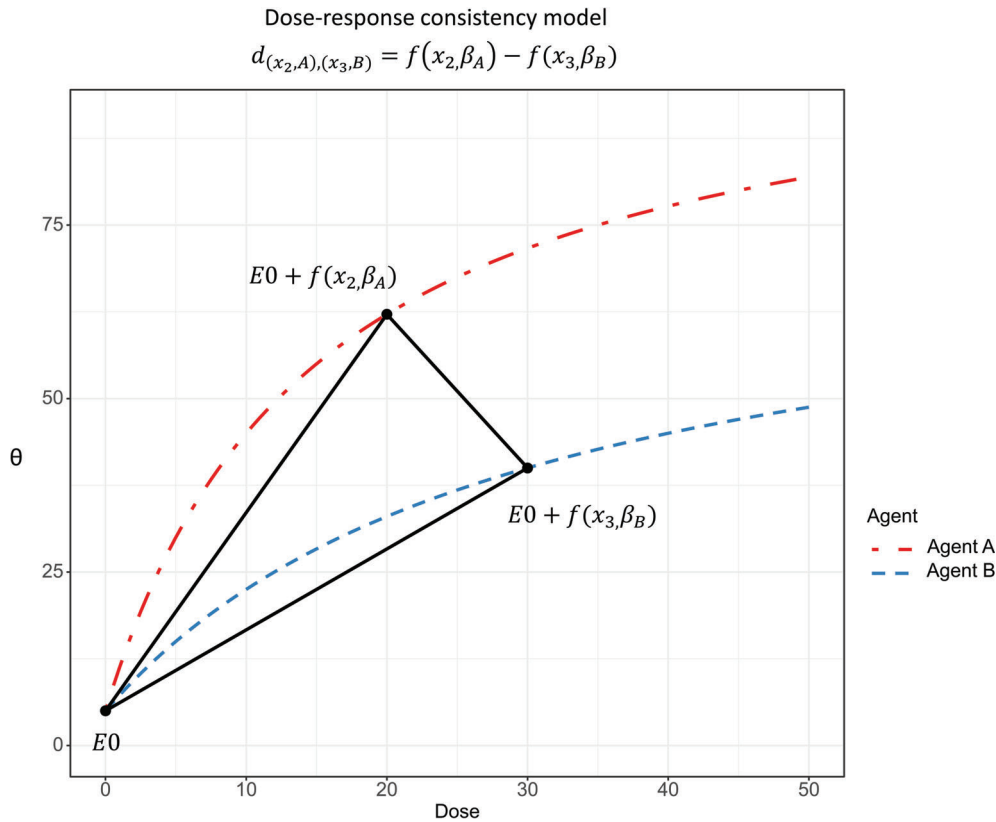$$d_{(x_2,A),(x_3,B)} = f(x_2, \beta_A) - f(x_3, \beta_B)$$

FIGURE 2 Schematic illustration of dose–response functions for agents A and B, illustrating that the comparison of agent A dose x_2 with agent B dose x_3 adheres to the consistency model (Equation (5)) and thus is informed both by direct and indirect evidence. Using these dose–response functions a comparison between any doses of these is possible by extrapolation/interpolation. θ represents the response [Colour figure can be viewed at wileyonlinelibrary.com]

3.1.1 | Consistency equations

At the treatment level we assume consistency of relative effects so that all comparisons can be written in terms of comparisons with the reference treatment 1 (x_1, a_1):

$$d_{(x_c,a_c),(x_k,a_k)} = d_{(x_1,a_1),(x_k,a_k)} - d_{(x_1,a_1),(x_c,a_c)} \text{ for treatments } (x_c, a_c), (x_k, a_k). \tag{4}$$

Equations (1) and (4) define the standard NMA model where each agent and dose combination are treated as a distinct treatment (eg, Figure 1B).

Dose-response MBNMA constrains this model by fitting a functional relationship for the dose-response. For a given dose-response function, $f(x, \beta_a)$, the consistency Equations (4) become (noting that the E_0 terms cancel):

$$d_{(x_c,a_c),(x_k,a_k)} = f\left(x_k, \beta_{a_k}\right) - f\left(x_c, \beta_{a_c}\right) \tag{5}$$

for treatments (x_c, a_c) and (x_k, a_k). Note that consistency will therefore depend on the appropriateness of the dose-response function and may hold for one dose-response function but not another.

The MBNMA model (1) under the treatment level consistency assumptions for a given dose-response function is therefore:

$$\delta_{i,k} \sim N\left(f\left(x_{i,k}, \beta_{a_{i,k}}\right) - f(x_{i,1}, \beta_{a_{i,1}}), \sigma^2\right), \tag{6}$$

where $x_{i,k}$ is the dose on arm k of study i. Equation (6) needs adjusting for multi arm trials, to account for correlations in effects for a trial with K arms, as follows:

$$
\begin{pmatrix} \delta_{i,2} \\ \delta_{i,3} \\ \vdots \\ \delta_{i,K} \end{pmatrix} \sim \mathrm{N}_{(K-1)} \left(\begin{pmatrix} f\left(x_{i,2}, \beta_{a_{i,2}}\right) - f\left(x_{i,1}, \beta_{a_{i,1}}\right) \\ f\left(x_{i,3}, \beta_{a_{i,3}}\right) - f\left(x_{i,1}, \beta_{a_{i,1}}\right) \\ \vdots \\ f\left(x_{i,K}, \beta_{a_{i,K}}\right) - f\left(x_{i,1}, \beta_{a_{i,1}}\right) \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \cdots & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 & & \sigma^2/2 \\ \vdots & & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \cdots & \sigma^2 \end{pmatrix} \right),
\tag{7}
$$

where the correlation between within study relative effects is 0.5 under the homogenous variance model.[1] Setting $\sigma = 0$ in Equations (6) and (7) gives the common effect model.

For example, if study 1 is a 3-arm study comparing placebo and treatments 1 and 2 which are different doses of agent A, $(0, P)\,vs\,(x_1, A)\,vs\,(x_2, A)$, then study 1 estimates:

$$
\begin{pmatrix} \delta_{1,2} \\ \delta_{1,3} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} f\left(x_1, \beta_A\right) \\ f\left(x_2, \beta_A\right) \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 \end{pmatrix} \right),
$$

whereas if study 2 is a 3-arm study comparing treatment 1 a single dose of agent A with treatments 2 and 3 which are two doses of agent B, $(x_1, A)\,vs\,(x_2, B)\,vs\,(x_3, B)$, then study 2 estimates

$$
\begin{pmatrix} \delta_{2,2} \\ \delta_{2,3} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} f\left(x_2, \beta_B\right) - f\left(x_1, \beta_A\right) \\ f\left(x_3, \beta_B\right) - f\left(x_1, \beta_A\right) \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 \end{pmatrix} \right).
$$

3.2 | Model critique and selection

The appropriateness of different dose-response functions and the choice of common or random effect models can be assessed statistically by comparing model fit statistics across competing models. In a Bayesian analysis one would compare the posterior mean residual deviance, Dbar (a measure of fit), the effective number of parameters, pD (a measure of complexity), and the deviance information criteria, DIC, (the sum of the measures of Dbar and pD), preferring smaller values, where differences of at least 3 (or even 5) in Dbar and DIC are considered meaningful.[19] In addition to this, the contribution of each data-point to the posterior mean residual deviance can be plotted against dose as a form of residual analysis to identify any missing structure in the dose-response relationship.[4]

3.3 | Checking the consistency assumption in dose-response MBNMA

In this section we describe how the methods for assessing evidence consistency in NMA can be extended to dose-response MBNMA. Note that the consistency equations under a dose-response MBNMA model are given by Equation (5), which depends on the dose-response function $f(x, \beta_a)$. If each agent and dose combination is treated as a distinct treatment in a standard NMA, then the consistency equations are given by Equation (4) which does not depend on the dose-response relationship. To fully explore the consistency assumption in dose-response MBNMA we need to consider the fit of the dose-response model alongside assessment of consistency.

3.3.1 | Bucher method to assess consistency in dose-response MBNMA

The simplest approach to assess inconsistency is the Bucher method[6] which can be applied to a triangle of evidence. Suppose we have a simple triangle of evidence comparing placebo, $(0, P)$, agent A at dose x_2, (x_2, A), and agent B at dose x_3, (x_3, B) (Figure 3). Because there is only a single dose of each agent, this is a standard NMA without a dose-response model. Then the direct estimate $d^{Dir}_{(x_2,A),(x_3,B)}$, based on head-to-head evidence comparing (x_2, A) and (x_3, B), is compared with the estimate obtained indirectly from $d^{Dir}_{(0,P),(x_2,A)}$ and $d^{Dir}_{(0,P),(x_3,B)}$ via the consistency relation for standard NMA Equation (4), $d^{Ind}_{(x_2,A),(x_3,B)} = d^{Dir}_{(0,P),(x_3,B)} - d^{Dir}_{(0,P),(x_2,A)}$. A test for inconsistency is constructed by forming a z-score for the
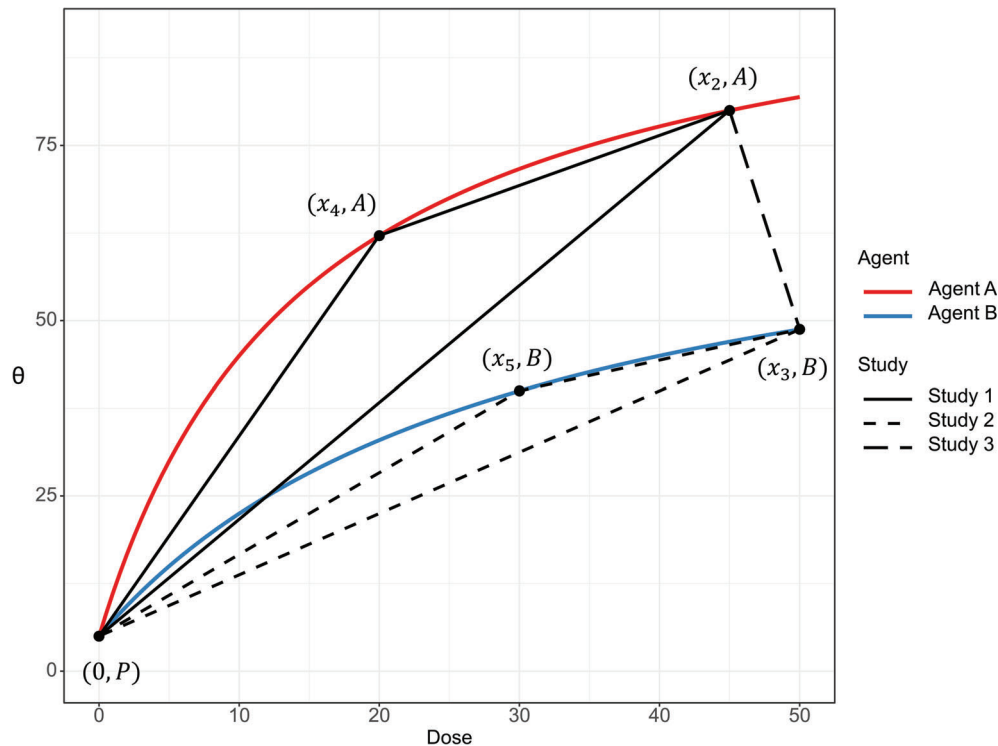
FIGURE 3 An illustrative network and dose–response relationships of agents A and B at various doses, and placebo (P) on which the Bucher test for inconsistency could be used in dose–response MBNMA. Straight black lines connecting points represent comparisons within each study for which direct evidence is available. The network contains at least one study comparing different doses of A and placebo (Study 1), one study comparing different doses of B and placebo (Study 2), and one study comparing (x_2,A) vs (x_3,B) (Study 3). θ represents the response [Colour figure can be viewed at wileyonlinelibrary.com]

difference between the direct and indirect estimates. The method requires that the studies that form the direct comparison are different to those that form the indirect comparison.[2,3] The Bucher method can also be applied to larger loops of evidence (eg, a quadrilateral with four treatments), and can also be extended to the case where there are multiple independent loops of evidence (where each loop is based on different sets of studies with the only exception being a single common edge).[2,3] This allows the construction of a global test for inconsistency across all the independent loops.[7]

In the context of dose-response MBNMA, the Bucher method can also be applied in the situation where we have the same loop of evidence as in Figure 3, but where the (x_2,A) vs $(0,P)$ estimate is based on a dose-response MBNMA for a network of at least one study/studies comparing various doses of agent A and placebo (Study A), and similarly the (x_3,B) vs $(0,P)$ estimate is derived from a dose-response MBNMA for a network of at least one study/studies comparing various doses of agent B and placebo (Study B), and there is a separate study that directly compares (x_2,A) vs (x_3,B) (Study C), as illustrated in Figure 3. Separate dose-response models can be fitted for agents A and B to obtain estimates of $d_{(0,P),(x_2,A)}$ and $d_{(0,P),(x_3,B)}$. Equation (5) then gives the indirect estimate for $d^{Ind}_{(x_2,A),(x_3,B)} = f(x_3,B) - f(x_2,A)$, which can then be compared with the direct estimate using the Bucher method. As with NMA, this method can also be used for multiple independent loops of evidence in this format (where each loop is based on different sets of studies with the only exception being a single common edge).

For applications of the Bucher method with random effects models the between studies variance parameter should be shared across comparisons.[3,8] This can be achieved by fitting a MBNMA model to the indirect evidence simultaneously with a pairwise meta-analysis for the direct evidence, with a shared parameter for the between studies SD. In practice this requires fitting the node-splitting models described below.

The application of the Bucher method is likely to be of limited use in assessing consistency in dose-response MBNMA because networks of evidence are typically complex with many RCTs with three or more arms, such that loops of evidence are not independent.

3.3.2 | Global assessment of inconsistency at treatment-level

In NMA the consistency assumption can be relaxed by fitting an inconsistency model where there is a separate parameter for each comparison based only on the direct evidence, whilst sharing the between-studies SD parameter across comparisons. This is known as the unrelated mean effects (UME) model.[2,3] A global test for inconsistency can be obtained by comparing model fit statistics for the inconsistency and consistency models, and inspecting the estimates of the between-studies SD σ. If σ is much smaller for the inconsistency model, then this indicates evidence of inconsistency (even in absence of differences between model fit statistics).

This UME model can also be fitted in dose-response MBNMA by estimating separate relative effects $d_{(x_c, a_c),(x_k, a_k)}$ for each treatment contrast with a shared between studies SD (Equations (1) and (2)), but without making the additional consistency assumption (Equation (5)).[4]

By comparing Equations (2) and (7) it can be seen that the UME model relaxes both the dose-response relationship and the consistency assumption at the same time. This means that differences in model fit and heterogeneity estimates could be explained by lack of fit of the functional dose-response function $f(x, \beta_a)$, inconsistency of the evidence sources, or both. Furthermore, inconsistency could mask lack-of-fit of the dose-response relationship to the extent that we detect neither lack-of-fit nor inconsistency (this phenomenon has been described in the network meta-regression literature[20]). We therefore recommend that three different models are compared:

1. **Model 1:** The dose-response MBNMA consistency model (Equations (5, 7)), which assumes a dose-response relationship and assumes consistency,
2. **Model 2:** The treatment-level NMA consistency model (Equations (1, 2, 4)) where each treatment (agent and dose combination) is considered a separate treatment. This model does not assume any dose-response relationship, but does assume consistency,
3. **Model 3:** The UME model (Equation (2)) which relaxes both the dose-response relationship and the consistency assumption.

Comparing model fit and σ estimates across these three models allows both the dose-response and consistency assumptions to be checked (Figure 4). Note that the network may be disconnected at the treatment-level, meaning that estimates for disconnected treatments may not be valid. When this occurs, model fit statistics for a treatment-level NMA (Model 2) can still be compared to the dose-response MBNMA consistency model (Model 1), but only relative effects between connected treatments are valid and can be interpreted. There may also be convergence issues in disconnected networks, in which case treatment-level NMAs may have to be fitted separately to different subnetworks.

To further explore potential causes of lack of fit and inconsistency deviance-deviance plots can be useful.[21] The contribution of each data point to the residual deviance can be plotted for (i) the dose-response MBNMA consistency model against the UME model (Model 1 vs Model 3); (ii) the dose-response MBNMA consistency model against the treatment level NMA consistency (Model 1 vs Model 2) and; (iii) the treatment level NMA consistency model against the UME model (Model 2 vs Model 3). These plots can identify which data-points are leading to lack of fit, inconsistency or both. Note however, that this does not necessarily mean that those data points are "wrong". Inconsistency is a property of loops of evidence and identifying inconsistency in a loop does not tell us which piece(s) of evidence are causing the problem, merely that the evidence within that loop does not agree.

3.3.3 | Node-splitting methods for MBNMA

We may wish to explore whether there is inconsistency for specific pairwise contrasts. This could be because the global test identified inconsistency and we wish to know which evidence is contributing to this. Alternatively, attention may be focussed on particular contrasts where there is *a priori* doubt about the consistency assumption, for example due to older studies or differences in study design/populations. In NMA, node-splitting is a method that can be applied to a specific contrast, where the relative effect parameter for that contrast is "split" into two separate parameters estimated by two different sets of evidence.[9] The first is estimated by the direct head-to-head evidence, and the second is estimated indirectly from the NMA model applied to the remaining evidence (ie, excluding the head-to-head evidence for that contrast). The direct and indirect estimates for the contrast can then be compared by computing the difference between them, and model fit can be compared for the "split" model with that from the consistency model.
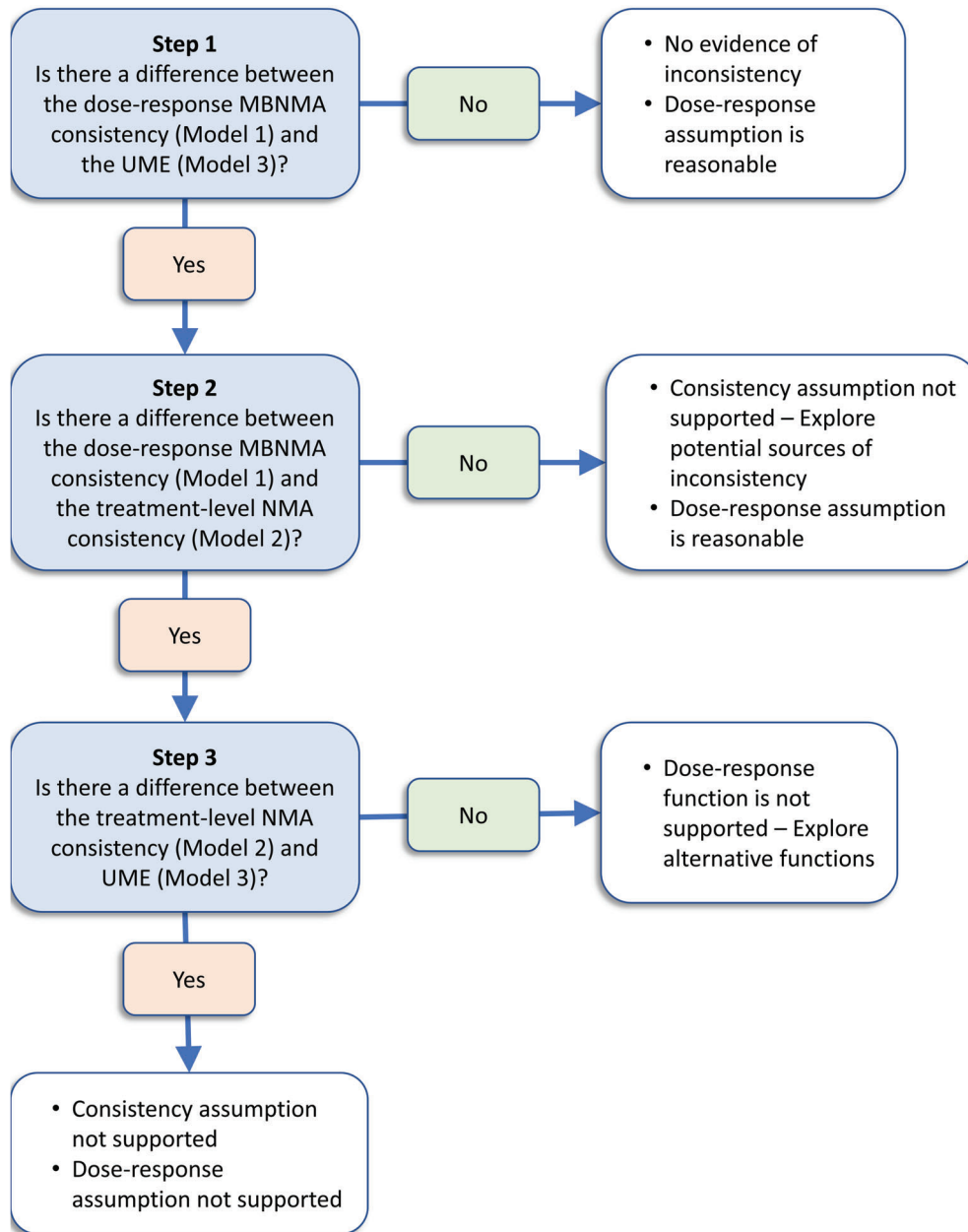
FIGURE 4 Flow chart representing the steps taken to perform a global assessment of inconsistency and to assess the fit of the dose–response function [Colour figure can be viewed at wileyonlinelibrary.com]

Node-splitting for dose-response MBNMA is analogous to that for NMA, except the indirect estimate is derived using the dose-response MBNMA model. For a specific treatment-level contrast, (x_c, a_c) vs (x_k, a_k) we can split the $d_{(x_c,a_c),(x_k,a_k)}$ parameter into two separate parameters, $d^{Dir}_{(x_c,a_c),(x_k,a_k)}$ and $d^{Ind}_{(x_c,a_c),(x_k,a_k)}$ which are estimated by two different sets of evidence. $d^{Dir}_{(x_c,a_c),(x_k,a_k)}$ is estimated from direct head-to-head evidence for (x_c, a_c) vs (x_k, a_k), by replacing $d_{(x_c,a_c),(x_k,a_k)}$ with $d^{Dir}_{(x_c,a_c),(x_k,a_k)}$ in Equation (1), and giving this parameter a prior distribution rather than applying the consistency relations (Equation (5)). $d^{Ind}_{(x_c,a_c),(x_k,a_k)}$ is estimated by applying the consistency relations (Equation (5)) to the estimates obtained from fitting a dose-response MBNMA model to the remaining evidence. The model is fitted to both sets of data simultaneously, allowing us to estimate a common between studies variance parameter, σ^2, which is shared for *all* treatment contrasts, including both $d^{Dir}_{(x_c,a_c),(x_k,a_k)}$ and $d^{Ind}_{(x_c,a_c),(x_k,a_k)}$. The direct and indirect estimates can then be compared by

computing the difference between them, $Diff = d^{Dir}_{(x_c,a_c),(x_k,a_k)} - d^{Ind}_{(x_c,a_c),(x_k,a_k)}$ and obtaining posterior summaries for $Diff$. Model fit measures and heterogeneity can be compared for the "split" model with that from the consistency model. In addition, a Bayesian p-value can be obtained as $2 \times \min\{\Pr(Diff > 0), 1 - \Pr(Diff > 0)\}$.[22]

To illustrate, suppose we were interested in node-splitting for (x_1, A) vs (x_2, B). A 2-arm study, study 1, which compares (x_1, A) vs (x_2, B) estimates:

$$\delta_{1,2} \sim N\left(d^{Dir}_{(x_1,A),(x_2,B)}, \sigma^2\right),$$

whereas a 3-arm study 2 comparing $(0, P) vs (x_3, A) vs (x_4, A)$ contributes to the estimation of the dose-response curve for agent A, estimating:

$$\begin{pmatrix} \delta_{2,2} \\ \delta_{2,3} \end{pmatrix} \sim N_2\left(\begin{pmatrix} f(x_3, \beta_A) \\ f(x_4, \beta_A) \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 \end{pmatrix}\right)$$

and a 3-arm study 3 comparing $(0, P)$ vs (x_5, B) vs (x_6, B) contributes to the estimation of the dose-response curve for agent B, estimating:

$$\begin{pmatrix} \delta_{3,2} \\ \delta_{3,3} \end{pmatrix} \sim N_2\left(\begin{pmatrix} f(x_5, \beta_B) \\ f(x_5, \beta_B) \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 \end{pmatrix}\right).$$

These can be used together to obtain an indirect estimate:

$$d^{Ind}_{(x_1,A),(x_2,B)} = f(x_2, \beta_B) - f(x_1, \beta_A).$$

Node-splitting requires both direct and indirect evidence to exist for a particular treatment-level contrast. The R-packages gemtc[23] and MBNMAdose[12] include functions for finding comparisons in the "treatment-level" network (ie, Figure 1B for psoriasis example) where both direct and indirect evidence exist. However, in dose-response MBNMA there may be additional treatment-level contrasts where indirect estimates can be obtained via the dose-response relationship. This is because the dose-response model can allow indirect comparisons to be made via interpolation/extrapolation (Figure 2).

Due to the nature of phase-II studies exploring dose-response, we expect to be able to obtain indirect estimates for many of the treatment-level contrasts where we have direct evidence. However, the extent to which indirect estimates will be available via the dose-response relationship will depend on the availability of remaining dose-response information once direct evidence has been removed, and on the complexity of the dose-response relationship fitted. Where indirect estimates are not available, then the model will simply predict a very uncertain estimate or may fail to converge. The *inconsistency.loops()* function in MBNMAdose can be used to identify contrasts in which both direct and indirect estimates are available, and to indicate whether it arises from pathways of head-to-head evidence (as for node-splitting in a treatment-level network) or only from the dose-response relationship. It also provides the number of doses of each agent in a comparison on which the indirect dose-response relationship must be estimated, which can help identify the complexity of function that could be used to estimate it.

Node-splitting is complicated by the existence of multiarm trials (trials with three or more arms). A K-arm trial provides (K-1) relative effects, but the choice of the "reference" arm 1, determines whether the trial contributes to the direct estimate or not, and this choice is somewhat arbitrary. For example, suppose that we have a 4-arm study i comparing Placebo with agent A at doses x_1 and x_2 and agent B at dose x_3, and we were interested in node-splitting for (x_2, A) vs (x_3, B). If we take Placebo as the reference arm and define three relative effects for each active arm compared to Placebo, then the study does not contribute to the direct estimate for $d^{Dir}_{(x_2,A),(x_3,B)}$ and instead informs the indirect estimate $d^{Ind}_{(x_2,A),(x_3,B)}$. However, if we take the (x_2, A) as the reference arm, then the relative effect for (x_3, B) relative to (x_2, A) contributes to $d^{Dir}_{(x_2,A),(x_3,B)}$, whereas the relative effects for Placebo and (x_1, A) relative to (x_2, A) could contribute to the dose-response model for the indirect estimate. Van Valkenhoef et al. set out a series of rules for how multiarm trials can be used in node-splitting models for NMA.[24] Following these rules and using Placebo as the reference arm, the relative effects for

(x_1, A) and (x_3, B) relative to (x_2, A) in the example above would be discarded since a multiarm trial cannot be inconsistent with itself. However, in dose-response MBNMA there may be inconsistencies within a multiarm trial due to departures from the assumed dose-response relationship. We therefore propose that the following approach be taken for inclusion of multiarm trials in node-splitting for dose-response MBNMA.

1. If the multiarm trial contains the contrast (x_c, a_c) vs (x_k, a_k) that is being split, then reorder the arms with (x_c, a_c) as the reference arm 1 and (x_k, a_k) is arm 2, so that the multiarm trial always contributes to $d^{Dir}_{(x_c, a_c), (x_k, a_k)}$ if possible.

2. Model all the relative effects from the multiarm trial as usual, but replace the relative effect for $\delta_{i,2}$ with $d^{Dir}_{(x_c, a_c), (x_k, a_k)}$

$$
\begin{pmatrix} \delta_{i,2} \\ \delta_{i,3} \\ \vdots \\ \delta_{i,K} \end{pmatrix} \sim N_{(K-1)} \left(\begin{pmatrix} d^{Dir}_{(x_c, a_c), (x_k, a_k)} \\ f\left(x_{i,3}, \beta_{a_{i,3}}\right) - f\left(x_{i,1}, \beta_{a_{i,1}}\right) \\ \vdots \\ f\left(x_{i,K}, \beta_{a_{i,K}}\right) - f\left(x_{i,1}, \beta_{a_{i,1}}\right) \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \cdots & \sigma^2/2 \\ \sigma^2/2 & \sigma^2 & & \sigma^2/2 \\ \vdots & & \ddots & \vdots \\ \sigma^2/2 & \sigma^2/2 & \cdots & \sigma^2 \end{pmatrix} \right).
\tag{8}
$$

The covariance structure reflects the fact that the multiarm trial cannot be inconsistent with itself, the trial contributes to the direct estimate for the contrast that is split, and the remainder of the trial arms contribute to the estimation of the dose-response MBNMA model which is used to form the indirect estimate.

3.4 | Implementation

The models are implemented using the MBNMAdose package version 0.3.1[12] in R version 4.02, which uses Bayesian Markov chain Monte Carlo (MCMC) simulation in JAGS.[25] R code for all analyses can be found on GitHub (https://github.com/hugaped/Dose-response-MBNMA-consistency). We illustrate a global assessment of consistency and methods for node-splitting at the treatment-level in the psoriasis dataset. We focus our analysis on two different dose-response functions, to demonstrate the impacts of a good fitting (Emax) and a badly fitting (exponential) dose-response model. We also include results for a simple linear dose-response model (very badly fitting), as readers will be most familiar with this. For global testing of inconsistency using UME, we also fit an agent-level NMA that lumps different doses of agents together assuming a common effect. Common and random treatment effect models were fitted and compared.

Vague normal prior distributions ($N(0, 1000)$) were given to direct treatment parameters $d^{Dir}_{(x_c, a_c), (x_k, a_k)}$, nuisance parameters μ_i, and dose-response parameter $\lambda_{a_{i,k}}$ for exponential MBNMA models. For Emax MBNMA models, a correlation was modeled between dose-response parameters by assigning them a multivariate normal prior:

$$
\begin{pmatrix} Emax_{a_{i,k}} \\ \log\left(ED50_{a_{i,k}}\right) \end{pmatrix} \sim MVN(0, \Sigma).
\tag{9}
$$

For $ED50_{a_{i,k}}$ parameters it was necessary to ensure that they only took positive values so prior distributions were specified on the log-scale. Σ is the covariance matrix, and its inverse (the precision matrix) was assigned a minimally informative Wishart distribution as a prior, with $2°$ of freedom and a scale matrix $\Omega = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ($\Sigma^{-1} \sim Wishart(\Omega, 2)$). The between-study SD, σ, was given a uniform prior distribution ($U(0, 5)$). Sensitivity of the results to the choice of prior distributions are reported in Supplementary Materials.

Models were run for 50 000 burn-in iterations and monitored for 50 000 iterations on 4 chains. Trace plots, autocorrelation and Brooks–Gelman–Rubin plots were used to assess convergence.[26] Unless otherwise stated, results are presented as posterior medians and 95% credible intervals (95% CrI).

All models were compared using DIC, with a difference of three points or greater being considered as meaningful. DIC was calculated using the Kullback–Leibler divergence for estimation of pD.[27]

TABLE 1 Model fit statistics for all models performed on the psoriasis network

Model	Treatment effect model	Posterior mean residual deviance[a]	pD[b]	DIC[c]	Between-study SD (95% CrI)	Selected (based on DIC)
NMA treatment-level	Common	79.4	37.7	470.4	-	Yes
NMA treatment-level	Random	70.4	43.9	473.2	0.08 (0.003, 0.24)	No
NMA agent-level	Common	288.8	30	672.2	-	No
NMA agent-level	Random	68.2	59.2	486	0.49 (0.37, 0.66)	Yes
MBNMA linear	Common	1318.3	30	1702.3	-	No
MBNMA linear	Random	70.7	68.7	497.9	1.41 (1.12, 1.82)	Yes
MBNMA exponential	Common	285.4	30.8	669.3	-	No
MBNMA exponential	Random	68.4	57.5	484.4	0.49 (0.37, 0.65)	Yes
MBNMA Emax	Common	75.6	35.4	469.7	-	Yes
MBNMA Emax[d]	Random	71.5	41.5	471.7	-	No
UME treatment-level	Common	80.2	43.2	476.7	-	-
UME treatment-level	Random	70.9	47.2	476.7	0.10 (0.002, 0.26)	-

[a] Compared to 70 data points.

[b] pD: The effective number of parameters calculated using the Kullback–Leibler divergence.[27]

[c] DIC: Deviance information criterion = pD + residual deviance (Dbar).

[d] Model did not fully converge for all parameters.

4 | RESULTS

Model fit statistics for all models are reported in Table 1. Heterogeneity was found to be very low when fitting a treatment-level NMA, leading to the selection of a common effect model based on DIC. MCMC chains struggled to converge for between-study SD in the random effect Emax MBNMA, however model fit statistics did converge and indicated that the common effects model gave a lower DIC and model fit was adequate (posterior mean residual deviance 75.6 compared with 70 data points). Lumping doses at the agent-level or using an exponential or linear dose-response resulted in high heterogeneity and the selection of random treatment effect models in all three instances.

Among selected models, DIC was higher in those with higher heterogeneity due to the additional number of effective parameters. Both the MBNMA Emax and treatment-level NMA models had similar DIC which was lower than for other models fitted. The slightly lower residual deviance in the treatment-level NMA model was offset by the fewer parameters required in the Emax MBNMA model. This suggested that the Emax dose-response relationship gave an adequate fit to the data. Predicted probabilities of response for common effects treatment-level NMA, random effects agent-level NMA, random effects exponential MBNMA, random effects linear MBNMA, and common effects Emax MBNMA are shown at different doses for each agent in Figure 5.

Points below the line of equality in the deviance-deviance plots for the common effect treatment-level NMA model against each of the other fitted common effect models indicate lack of fit for the dose-response model assumed. For the Emax MBNMA the deviance-deviance plot showed generally good fit, though there was some lack of fit for studies 15 and 16 that compared different doses of Ustekinumab (Figure 6A). However, they showed poor fit for the exponential MBNMA and the agent-level NMA models, suggesting that these do not explain the dose-response relationship. However, the random effects versions of these models fitted well because the lack of fit is captured as heterogeneity in these models (Figure S2). Deviance-deviance plots are shown separately for common and random effects linear MBNMA models (Figure S3). These showed very similar trend to the exponential MBNMA models.

4.1 | Global assessment of inconsistency

For global assessment of inconsistency, the posterior mean residual deviances were similar for all selected models and the treatment-level UME model that relaxed the consistency assumption (Table 1). However, whilst there was no evidence of
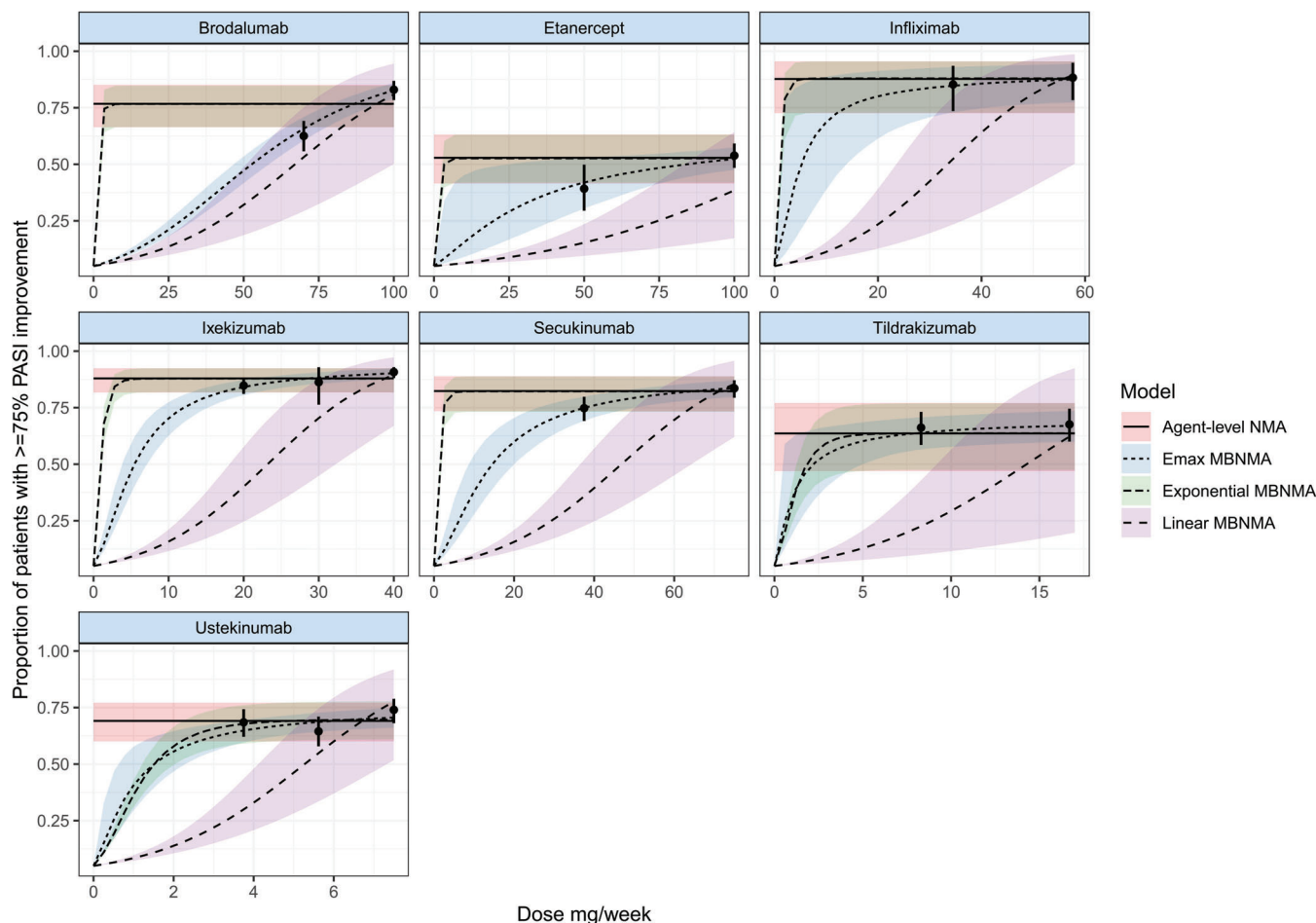
FIGURE 5 Predicted responses for different agents in the network based on results from a common effects treatment-level NMA (thick vertical lines and points), a random effects agent-level NMA (solid horizontal line with red 95% CrI), a common effects Emax MBNMA (dotted curve with blue 95% CrI), a random effects Exponential MBNMA (short dashed curve with green 95% CrI) and a random effects Linear MBNMA (long dashed curve with purple 95% CrI). Predictions were calculated from relative effects applied to a probability of placebo response of 0.05. Predicted responses are shown as posterior medians and 95% CrIs [Colour figure can be viewed at wileyonlinelibrary.com]

inconsistency for the treatment-level NMA and the Emax MBNMA models based on DIC, for the exponential MBNMA and agent-level NMA the DIC were considerably higher than for the UME model, suggestive of inconsistency in these models driven by the high levels of heterogeneity.

Deviance-deviance plots are shown for each model compared to the treatment-level UME (Figure 6B). Whilst the treatment-level NMA model showed no evidence of substantially poorer fit for any data points, most data points in all other models showed considerably higher contributions to the posterior mean deviance compared to the treatment-level UME than when compared to the treatment-level NMA, which may be suggestive of inconsistency (Figure 6A), but may also be due to incorrect specification of the dose-response relationship.

There were two data points in particular in the Emax MBNMA which appeared to have a substantially higher contribution to the posterior mean deviance than in the treatment-level UME. These points correspond to studies comparing placebo with different doses of Ustekinumab (study numbers 15 and 16), which were also identified in Figure 6A as having a poorer fit for the dose-response relationship than other data points.

For the exponential MBNMA and agent-level NMA there were a large number of data points for which deviance contributions were higher when common treatment effects models were compared to a common effects UME model (Figure 6B). However, these differences were no longer discernible when random treatment effects models were compared (Figure S1), since they were masked by high heterogeneity which resulted in lower posterior mean deviances. This could be identified by examining the reduction in between-study SD and DIC for the random effects UME model compared to the random effects exponential MBNMA and treatment-level NMA models (Table 1). For common and random effects
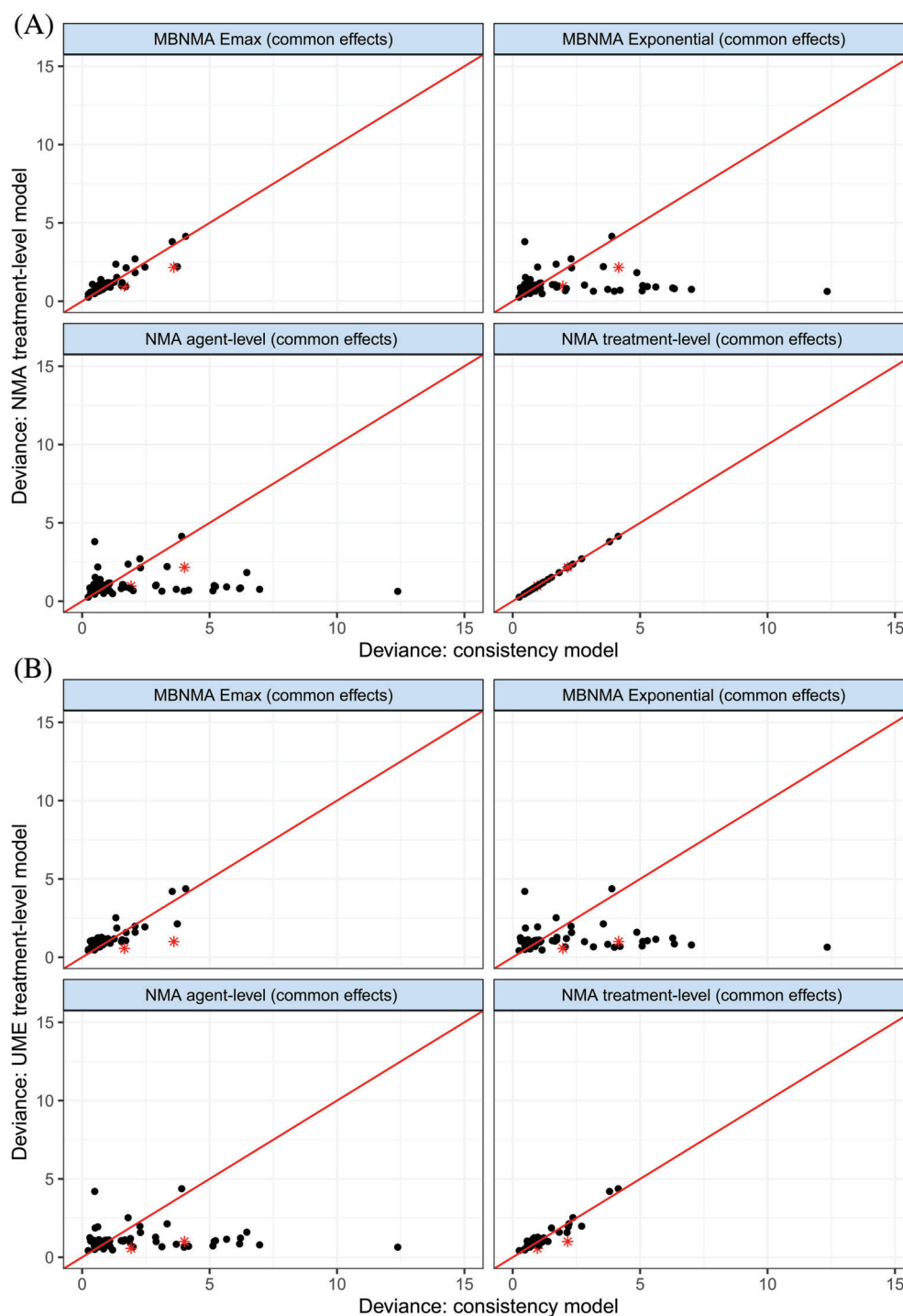
FIGURE 6 Deviance-deviance plots showing the contribution to the posterior mean deviance for each data point from different common treatment effect models compared to treatment-level NMA (A) and UME (B) models. Comparisons in (B) are to UME models. For better visibility points which contribute up to 15 to the posterior mean deviance are plotted, meaning that 10 data points that contribute >15 to the posterior mean deviance have been excluded from MBNMA exponential and NMA agent-level plots. Red stars represent data points described in the text corresponding to arm 1 in studies 15 and 16 that compare different doses of Ustekinumab. The red diagonal line represents the line of equality, at which deviance is equal in both consistency and UME models [Colour figure can be viewed at wileyonlinelibrary.com]

Linear MBNMA models, the trend was similar to that observed for common and random effects exponential MBNMA models, but with more extreme differences in deviance contributions from UME models arising from poorer fit of the dose-response function (Figure S3).

Given that there is poor dose-response fit for these models, it is difficult to further confirm if there is also inconsistency in these models. Node-splitting can be used to provide a local test to further explore the potential inconsistency identified.

4.2 | Node-splitting

In the psoriasis dataset there was direct evidence for 34 different comparisons. However, by parameterising comparisons in each study as relative effects compared to the study reference arm (such that a 3-arm study is parameterised by two relative effects and a study reference treatment effect) this resulted in 20 different comparisons for which direct evidence was available. Both direct and indirect evidence were available for 4 of these 20 comparisons, meaning that node-splitting was possible for these in both MBNMA and NMA (Table S3 and Figure S4). For these comparisons, indirect estimates made use of evidence from both the consistency relationship via a pathway of direct comparisons, and from the dose-response relationship.

For all four comparisons both Emax MBNMA and treatment-level NMA models showed no evidence of inconsistency between direct and indirect evidence (Figure 7). For three of the comparisons (comparisons 1, 2, and 4) direct and indirect estimates from the Emax MBNMA were more similar than in the treatment-level NMA. However, inconsistency was identified in the exponential and linear MBNMAs for the comparison of Ixekizumab 40 mg/wk vs Ixekizumab 20 mg/wk (comparison 1) (Figures 7 and S5). For the other three comparisons (comparisons 2-4) there was no evidence of inconsistency in the exponential or linear MBNMAs.

A further nine comparisons (comparisons 5-13) made use of indirect evidence that was only available via the dose-response relationship, providing additional potential for node-splitting in MBNMA that was not possible in the treatment-level NMA (Table S3 and Figure S4). For estimating indirect evidence via the dose-response relationship, the number of available doses is critical, as this will limit the complexity of the dose-response function that can be fitted. The number of doses for each agent in the comparison are shown in Table S3. To fit an Emax dose-response function for all agents, we would need at least three doses of each agent available in the indirect evidence. Given that evidence for the direct effect is separated from evidence used to estimate the indirect effect, this reduces the information available to estimate each of the effects. This means that fitting an Emax MBNMA to the indirect evidence is only possible for comparisons 1 to 4 (for which indirect evidence is also available via a pathway of direct comparisons, as discussed above) and additionally for Ustekinumab 5.62 mg/wk vs Ixekizumab 30 mg/wk (comparison 5). For this comparison, an Emax MBNMAs did not show any evidence of inconsistency (Figure 7).

However, if we fit an exponential or linear dose-response function, only two doses for each agent in the indirect evidence are needed to estimate dose-response functions for all agents, which means that we can estimate indirect evidence for all nine indirect dose-response comparisons. Out of the nine comparisons that could be analyzed using a single parameter dose-response relationship, inconsistency was identified in two comparisons (comparisons 6-7) that could be analyzed using an exponential MBNMA and one comparison (comparison 10) that could be analyzed using a linear MBNMA (Figures 7 and S5). Notably, the comparisons in which inconsistency was identified for the different MBNMAs were not the same. For other comparisons there was no evidence of inconsistency.

Overall, this suggested no inconsistency in treatment-level NMA or Emax MBNMA, but that fitting an exponential or linear MBNMA could lead to inconsistency in existing node-splits in which indirect evidence was estimated by both the consistency assumption and dose-response relationship, and that this could also lead to inconsistency in additional comparisons for which indirect evidence was only informed by the dose-response relationship.

5 | DISCUSSION

Ignoring the effects of different doses is a common cause of heterogeneity and inconsistency in NMA. Dose-response MBNMA can reduce heterogeneity and inconsistency by modeling a dose-response relationship. However, it is still important to critique model fit and check consistency assumptions where possible. We have presented several methods to assess model consistency both globally and for specific contrasts.

The methods were illustrated with a dataset of trials comparing different agents for the treatment of moderate-to-severe psoriasis, which had features that are typical in dose-response MBNMA, comprising of (i) phase 2b
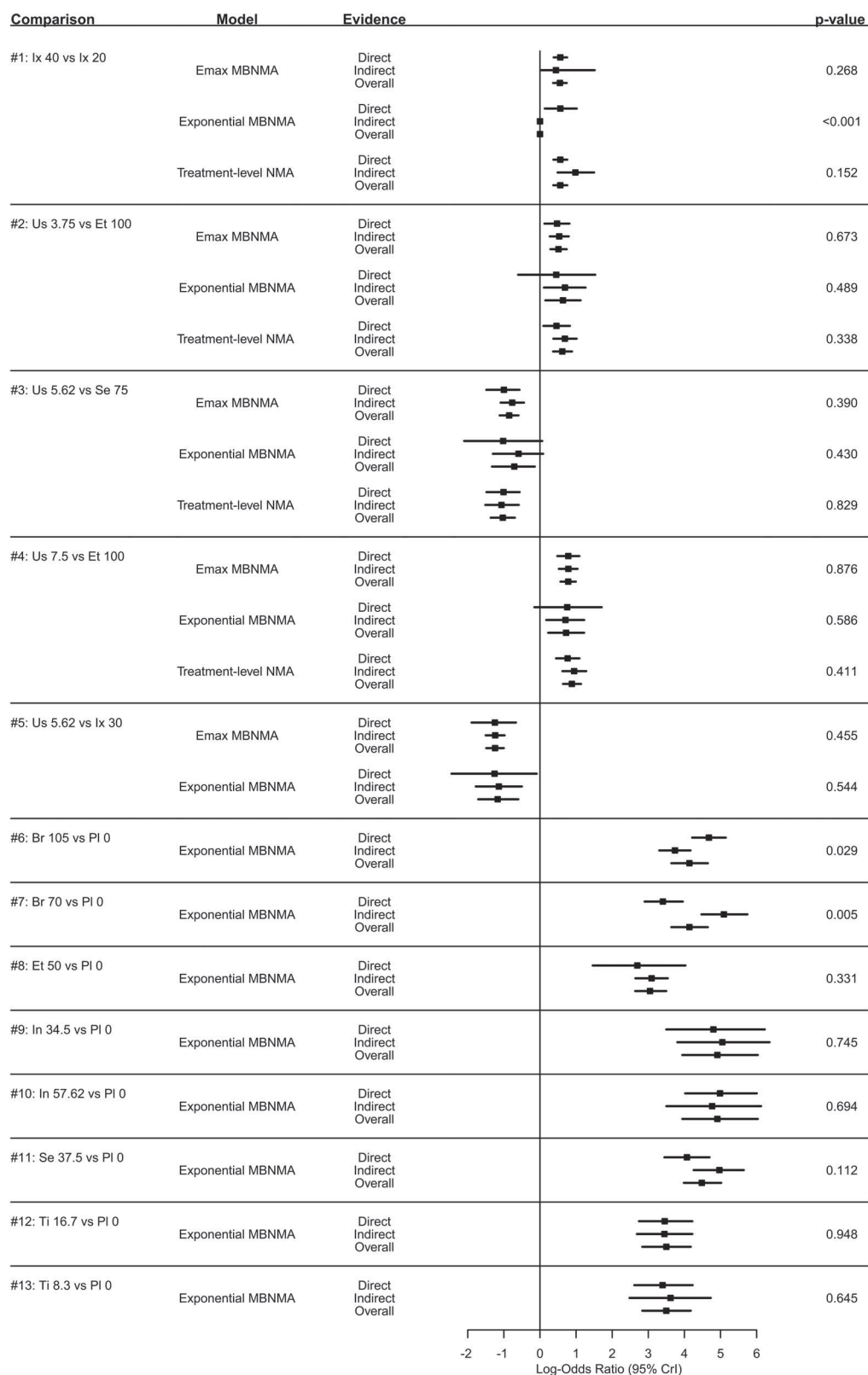
FIGURE 7 Estimates of direct, indirect and overall evidence for each comparison from each model for which node-splitting was possible. Bayesian *p*-values are given to the right of the plot, which show the agreement between direct and indirect estimates. Interventions are coded according the first two letters of the agent and dose (mg/week)

dose-ranging studies where studies have multiple doses of a single agent that are compared with Placebo in multiarm trials, perhaps with a single (licensed) dose of a competitor also included in the study; (ii) phase 3 studies that compare two active agents head to head, each at a single dose. Although in this dataset there were reasonable opportunities for assessing consistency, networks comprised more exclusively of phase 2b dose-ranging studies may have more limited scope for this. The very features of such studies (multiarm trials, comparisons with placebo) are important mechanisms that guard against inconsistency, and so should not be viewed as a problem, but rather as part of the solution to avoid inconsistency in the first place. Only 5 out of 23 studies in the network were 2-arm studies, and the other 18 were multi($\geq$3)-arm studies comparing multiple doses of one or more agents.

We did not find any evidence of inconsistency in the psoriasis dataset when models were used that fitted the data appropriately (treatment-level NMA, Emax MBNMA), but the Emax MBNMA had the advantage that at several doses relative effects were estimated more precisely. However, when an exponential or linear dose-response relationship was used this led to a high level of heterogeneity and the identification of inconsistency within multiple comparisons in the network. Tests for consistency are inextricably linked to model fit and therefore depend on the dose-response relationship assumed. Thorough assessment of model fit for a range of alternative dose-response relationships is therefore recommended, and we advise selection of the best fitting dose-response function prior to comparing common vs random treatment effect models, as the impacts of an inappropriate dose-response function can be masked by fitting random effect models with resulting high heterogeneity.

In the presence of high heterogeneity (such as in exponential MBNMA, linear MBNMA, and agent-level NMA models) a global assessment by comparison with treatment-level NMA and UME models may fail to clearly identify evidence of inconsistency, since inconsistency can instead manifest as high heterogeneity.[3,28] Node-splitting can be a useful tool for highlighting the difference between these forms of heterogeneity by specifically testing for inconsistency on the set of comparisons for which direct and indirect evidence contributions are available.

In MBNMA node-splitting models, indirect estimates can sometimes be obtained via the dose-response relationship even when a pathway of direct evidence is not available (comparisons 5-13) and indirect estimates are not possible for standard NMA. The dose-response relationship can be used in the same way to estimate effects in otherwise disconnected networks.[5] This does not specifically require the inclusion of placebo data, and the methods can be used in the absence of this, although placebo arms typically provide a lot of information to the dose-response relationship, particularly for parameters relating to the efficacy at lower doses (eg, $ED50_a$). In networks with limited evidence at lower doses the indirect estimates via the dose-response relationship may be biased in some circumstances, which may impact the assessment of inconsistency using node-splitting.[5]

Simpler dose-response functions provide more potential to obtain indirect estimates via the dose-response relationship and therefore increase the availability of comparisons on which node-splitting is possible, yet if the true underlying dose-response relationship requires a more complex function, fitting a simpler function may introduce bias. This likely explains the inconsistency identified by node-splitting in several comparisons for exponential and linear MBNMA models. Ensuring that comparisons are only split if the indirect evidence contains sufficient information with which to estimate the same dose-response function as used in the MBNMA consistency analysis limits this bias. In the psoriasis example for our selected Emax MBNMA we would therefore in practice only choose to split comparison 5 (Ustekinumab 5.62 mg/wk vs Ixekizumab 30 mg/wk) in addition to the comparisons for which indirect evidence is available via consistency relationships (comparisons 1-4), given the doses available in the dataset for both agents in this comparison.

We have also noted that the effects of inconsistency and poor dose-response fit on deviance contributions for individual data points can be aggregated, leading to substantially higher contributions to the posterior mean deviance in a MBNMA model compared to the treatment-level UME. This does not appear to affect the total fit of the model, suggesting that it is balanced by improved fit of other data points. However, it is important to be aware of the causes of this when inspecting deviance-deviance plots and to identify to what degree the difference is caused by inconsistency vs fit of the dose-response curves. Given that there was no evidence of inconsistency in node-splitting for the Emax MBNMA, the issues identified in deviance-deviance plots for studies 15 and 16 that included Ustekinumab was most likely due to poor dose-response fit rather than inconsistency, and it may be that the dose-response relationship for this agent is different than for other biologics. However, given that the dose-response function for Ustekinumab implied by the treatment-level NMA results would be nonmonotonic (Figure 5—the posterior median for the relative effect of Ustekinumab 5.62 mg/wk is lower than at both 3.75 and 7.5 mg/wk), there may be some argument to suggest that the results of the Emax MBNMA at these doses are more biologically plausible.

Whilst we describe methods for identifying inconsistency in dose-response MBNMA, we have not yet suggested steps to take if it is found. Given the link between inconsistency and dose-response fit that we have shown here, our first suggestion would be to ensure that the fit of the dose-response function is reasonable. If this is not the source of inconsistency then other causes can be investigated.[3] Inconsistency can arise from the impact of effect modifiers (variables that influence the treatment effect) that differ between comparisons. Impacts of these effect modifiers may also differ within a comparison, resulting in heterogeneity. One common source of heterogeneity/inconsistency is from "lumping" different doses of agents together, and this is likely to be avoided by explicit modeling of the dose-response relationship using MBNMA. However, other possible variables that could be effect modifiers and may differ between comparisons are age, severity at baseline, and previous treatments.[3] Ensuring strictly defined inclusion/exclusion criteria for a systematic review will help limit heterogeneity/inconsistency, as well as early identification of potential effect modifiers prior to commencing synthesis.

In this study, we have restricted attention to parametric functional forms for the dose-response relationship. Hierarchical models, where different doses are assumed exchangeable, have also been proposed,[29,30] including a hierarchical model that constrains a monotonic relationship with dose.[30] The methods described here can extend naturally to nonparametric dose-response functions and some of these models are implemented in the R package MBNMAdose.[12] However, methods for assessing inconsistency in such models have not yet been fully explored.

Design-by-treatment interaction models have been proposed to assess inconsistency in the NMA literature, where the effect from AB studies is distinguished from those from ABC studies, ACD studies, etc.[31] However, these models do not naturally extend to dose-response models, because dose defines both the study design and the explanatory variable for the dose-response model. Furthermore, there are typically lots of different study designs in dose-response MBNMA (as seen for the psoriasis example, Table S2), making the design-by-treatment interaction models infeasible to estimate in this context.

Cross validation has been proposed to identify potential outliers,[32,33] but can also be used to explore inconsistency in NMA.[10,21,34] The idea is to omit a subset of the data, fit the consistency model (Equations (7) and (8)) to the remaining data, and then compare the predictions from the model with the observed omitted data. This allows the computation of a cross-validation p-value, which represents the probability that the model predicts a value as, or more, extreme than that observed. If *all* the evidence on a specific contrast is excluded, then cross-validation becomes a form of test for inconsistency for that contrast.

Although the idea of cross-validation is very similar to node-splitting, the methods differ in how the comparisons are made between the direct and indirect evidence. In node-splitting *both* direct and indirect estimates of treatment effects are computed and then compared. In cross-validation only an indirect estimate is obtained which is then used to predict the observed data accounting for both sampling error *and* predictive uncertainty. As a consequence, in node-splitting all evidence can be used to estimate the between studies variance jointly, whereas in cross-validation the omitted data do not contribute to the estimate of the between-studies variance. Because the cross-validation method uses the predictive distribution from the random effects model, and the estimate of the between-studies variance is less certain (due to being based on less data), cross-validation is less likely to detect inconsistency than node-splitting. Because there is often little power to detect inconsistency in NMA, and we wish to be confident that we can identify it if it is present, we prefer the node-splitting approach to cross-validation, although this may result in some false positive findings.

Sparsity of evidence may require additional modeling assumptions to be made, such as exchangeable agent effects (within class) for the $Emax_a$ and/or $ED50_a$ parameters, if pharmacologically plausible.[4] Note that such hierarchical models may mask inconsistency which, as mentioned above, can alternatively manifest as heterogeneity. A simulation study to fully explore the ability of the methods to detect inconsistency under a range of modeling assumptions would be useful future work.

6 | CONCLUSIONS

Testing for inconsistency is important when synthesizing direct and indirect evidence to obtain robust relative effects to inform drug-development and policy decisions. We have proposed methods to assess consistency for dose-response MBNMA models and highlight the importance of also considering the fit of the dose-response function as part of this assessment. Poorly fitting dose-response functions are associated with higher heterogeneity and may introduce inconsistency. More complex dose-response functions can allow for better model fit but sufficient data at multiple doses of different agents are required to fit them. This highlights the importance of including phase II dose-finding evidence in

systematic reviews, rather than simply focusing on the licensed doses of different agents. Furthermore, inclusion of phase II multiarm placebo-controlled studies protects against inconsistency because these study designs help guard against inconsistency in the first place.

CONFLICT OF INTEREST

Meg Bennetts and Martin Boucher are employed by Pfizer. David Mawdsley and Hugo Pedder were partially funded by Pfizer for this work. The authors have no other conflicts of interest to declare.

AUTHOR CONTRIBUTIONS

Hugo Pedder: Data curation; formal analysis; investigation; methodology; project administration; software; visualization; writing–original draft preparation (lead); writing–review & editing. **Sofia Dias**: Conceptualization; funding acquisition; methodology; validation; writing–reviewing & editing. **Martin Boucher**: Conceptualization; funding acquisition; methodology; project administration; writing–review & editing. **Meg Bennetts**: Conceptualization; funding acquisition; methodology; project administration; writing–review & editing. **David Mawdsley:** Project administration; writing–reviewing & editing. **Nicky J. Welton:** Conceptualization, funding acquisition (lead); methodology; project administration; supervision; writing–original draft preparation; writing–review & editing.

DATA AVAILABILITY STATEMENT

The final dataset included 7 agents and placebo, investigated in 22 RCTs (Figure S1and Table S1).

ORCID

Hugo Pedder https://orcid.org/0000-0002-7813-3749
Sofia Dias https://orcid.org/0000-0002-2172-0221

REFERENCES

1. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15:2733-2749.
2. Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network Meta-Analysis for Decision Making*. Hoboken, NJ: Wiley; 2018.
3. 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;33:641-656.
4. 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 Pharmacometr Syst Pharmacol*. 2016;5(8):393-401.
5. Pedder H, Dias S, Bennetts M, Boucher M, Welton NJ. Joining the dots: linking disconnected networks of evidence using dose-response model-based network meta-analysis. *Med Decis Making*. 2021;41:194-208.
6. 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.
7. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*. 2005;331:897-900.
8. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. NICE DSU technical support document 4: inconsistency in networks of evidence based on randomised controlled trials; 2011.
9. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*. 2010;29:932-944.
10. Madan J, Stevenson MD, Ades AE, Cooper KL, Whyte S, Akehurst R. Consistency between direct and indirect trial evidence: is direct evidence always more reliable? *Value Health*. 2011;14:953-960.
11. Warren RB, Gooderham M, Burge R, et al. Comparison of cumulative clinical benefits of biologics for the treatment of psoriasis over 16 weeks: results from a network meta-analysis. *J Am Acad Dermatol*. 2020;82(5):1138-1149.
12. MBNMAdose: run dose-response MBNMA models. R package version 0.3.1. [computer program]. CRAN; 2020.

13. Reich K, Nestle FO, Papp K, et al. Infliximab induction and maintenance therapy for moderate-to-severe psoriasis: a phase III, multicentre, double-blind trial. *Lancet*. 2005;366:1367-1374.

14. Menter A, Feldman SR, Weinstein GD, et al. A randomized comparison of continuous vs. intermittent infliximab maintenance regimens over 1 year in the treatment of moderate-to-severe plaque psoriasis. *J Am Acad Dermatol*. 2007;56(1):31.e1-15.

15. 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;33:607-617.

16. Lu G, Ades A. Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics*. 2009;10(4):792-805.

17. Ette EI, Williams PJ. *Pharmacometrics: The Science of Quantitative Pharmacology*. Hoboken, NJ: John Wiley & Sons; 2007.

18. Thomas N, Sweeney K, Somayaji V. Meta-analysis of clinical dose–response in a large drug development portfolio. *Stat Biopharm Res*. 2014;6(4):302-317.

19. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. The deviance information criterion: 12 years on. *J Royal Stat Soc (B)*. 2014;76:485-493.

20. 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.

21. Lu G, Ades A. Assessing evidence consistency in mixed treatment comparisons. *J Am Stat Assoc*. 2006;101:447-459.

22. Presanis AM, Ohlssen D, Spiegelhalter DJ, De Angelis D. Conflict diagnostics in directed acyclic graphs, with applications in bayesian evidence synthesis. *Stat Sci*. 2013;28:376-397.

23. gemtc: network meta-analysis using Bayesian methods. R package. [computer program]. version 081: CRAN; 2016.

24. 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;7:80-93.

25. Plummer M. Bayesian graphical methods using MCMC: package 'rjags'; 2016. Accessed April 4, 2018. https://cran.r-project.org/web/packages/rjags/rjags.pdf.

26. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat*. 1998;7(4):434-455.

27. Plummer Martyn. Penalized loss functions for Bayesian model comparison. *Biostatistics*. 2008;9:(3):523-539. doi:10.1093/biostatistics/kxm049

28. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc*. 2006;101:447-459.

29. Del Giovani C, Vacchi L, Mavridis D, Filippini G, Salanti G. Network meta-analysis models to account for variability in treatment definitions: application to dose effects. *Stat Med*. 2013;32:25-39.

30. 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.

31. Jackson D, Barrett JK, Rice S, White IR, Higgins JPT. A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Stat Med*. 2014;33:3639-3654.

32. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3rd ed. Boca Raton, FL: Chapman & Hall; 2014.

33. Marshall EC, Spiegelhalter DJ. Approximate cross-validatory predictive checks in disease mapping models. *Stat Med*. 2003;22(10):1649-1660.

34. Ades AE, Mavranezouli I, Dias S, Welton NJ, Whittington C, Kendall T. Network meta-analysis with competing risk outcomes. *Value Health*. 2010;13(8):976-983.

35. Plummer M. Penalized loss functions for Bayesian model comparison. *Biostatistics*. 2008;9(3):523-539.

SUPPORTING INFORMATION

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