



# A Revised Framework to Evaluate the Consistency Assumption Globally in a Network of Interventions

Loukia M. Spineli 

**Background.** The unrelated mean effects (UME) model has been proposed for evaluating the consistency assumption globally in the network of interventions. However, the UME model does not accommodate multiarm trials properly and omits comparisons between nonbaseline interventions in the multiarm trials not investigated in 2-arm trials. **Methods.** We proposed a refinement of the UME model that tackles the limitations mentioned above. We also accompanied the scatterplots on the posterior mean deviance contributions of the trial arms under the network meta-analysis (NMA) and UME models with Bland-Altman plots to detect outlying trials contributing to poor model fit. We applied the refined and original UME models to 2 networks with multiarm trials. **Results.** The original UME model omitted more than 20% of the observed comparisons in both networks. The thorough inspection of the individual data points' deviance contribution using complementary plots in conjunction with the measures of model fit and the estimated between-trial variance indicated that the refined and original UME models revealed possible inconsistency in both examples. **Conclusions.** The refined UME model allows proper accommodation of the multiarm trials and visualization of all observed evidence in complex networks of interventions. Furthermore, considering several complementary plots to investigate deviance helps draw informed conclusions on the possibility of global inconsistency in the network.

## Highlights

- We have refined the unrelated mean effects (UME) model to incorporate multiarm trials properly and to estimate all observed comparisons in complex networks of interventions.
- Forest plots with posterior summaries of all observed comparisons under the network meta-analysis and refined UME model can uncover the consequences of potential inconsistency in the network.
- Using complementary plots to investigate the individual data points' deviance contribution in conjunction with model fit measures and estimated heterogeneity aid in detecting possible inconsistency.

## Keywords

consistency assumption, deviance information criterion, Bland-Altman plot, network meta-analysis

Date received: May 22, 2021; accepted: December 01, 2021

Systematic reviews with network meta-analysis (NMA) have been at the forefront of evidence-based medicine over the past 2 decades.<sup>1</sup> The explosive rate of published

## Corresponding Author

Loukia M. Spineli, Midwifery Research and Education Unit (OE 9210), Hannover Medical School, Carl-Neuberg-Straße 1, 30625, Hannover, Germany; (Spineli.Loukia@mh-hannover.de).

systematic reviews with NMA from several health care fields and a recent comprehensive review on the methodological advances of NMA attest to the reception of this evidence synthesis design from the wide research community.<sup>1,2</sup> NMA has the advantage of providing a hierarchy of interventions for a specific research question to assist the end users of systematic reviews in selecting the best intervention for a condition. The intervention hierarchy results from modeling direct evidence from the relevant clinical trials and indirect evidence for interventions never compared in any clinical trial simultaneously. Consequently, NMA provides coherent evidence for all possible comparisons of interventions under the investigated outcome.<sup>3</sup>

The credibility of the results from NMA strongly depends on the validity of the consistency assumption that underlies this evidence synthesis tool. The consistency assumption dictates the agreement of direct and indirect evidence for any pairwise comparison in a closed loop of interventions (i.e., a path that starts and ends with the same intervention).<sup>4</sup> The evaluation of the consistency assumption includes methods for local and global detection of possible inconsistency.<sup>2,5</sup> The local evaluation is the most prevalent in the published systematic reviews.<sup>6,7</sup> Among the methods for global evaluation, the unrelated mean effects (UME) model, introduced by Dias et al.,<sup>5</sup> is the most frequently applied.<sup>6</sup> The UME model is particularly useful in complex networks, in which the implementation of several statistical tests of inconsistency, such as the loop-specific approach,<sup>8</sup> may become cumbersome, challenging when loops are also informed by multiarm trials and prone to multiplicity issues.

A global evaluation using the UME model is achieved by comparing the Bayesian NMA model with the Bayesian UME model using measures of model fit,<sup>5</sup> such as the deviance information criterion (DIC).<sup>9</sup> The model with the smaller DIC value by 3 or 5 units may be preferred regarding model fit and complexity.<sup>10</sup> If the UME model fits the data better, this is evidence of possible inconsistency in the network.<sup>5</sup> A scatterplot of the posterior mean deviance of the individual data points under the

UME model against the NMA model can reveal the trials with a higher than expected posterior mean deviance.<sup>5</sup> These trials may help identify the loops with possible evidence of inconsistency.<sup>5</sup>

Furthermore, the inspection of the between-trial variance for substantial reductions also offers valuable information on the suitability of the contrasted models. Suppose the between-trial variance estimated from a model such as the UME that does not incorporate the consistency assumption is substantially lower than that estimated from the NMA model. This implies that the estimated between-trial variance had to increase for the NMA model to fit well given the lack of consistent effects, thus suggesting potential inconsistency.

The presence of multiarm trials in the network may challenge the application of the UME model. Suppose a comparison is informed by a multiarm trial alone. In that case, selecting a different baseline intervention for that trial may omit this comparison from the estimation process.<sup>11</sup> This was the case with the network of thrombolytic treatments that Dias et al.<sup>5</sup> considered to illustrate the UME model (figure 3 in Ref 5). A closed-loop of 3 interventions (SK, SK plus t-PA, and Acc t-PA) was informed by a multiarm trial and a 2-arm trial for 1 of the comparisons (SK versus SK plus t-PA). Of the 16 observed comparisons, the UME model estimated 15 treatment effects, inevitably omitting 1 of the comparisons in the multiarm trial (SK plus t-PA versus Acc t-PA). Considering a different baseline intervention for this trial (e.g., SK plus t-PA) would have resulted in the omission of a different comparison (in that case, SK versus Acc t-PA).

The omitted comparisons in the multiarm trials may carry evidence of possible design inconsistency in the network. The design inconsistency indicates disagreement in the treatment effects across different designs (i.e., 2-arm and multiarm trials) for the same comparison. Design inconsistency can be formally investigated using the design-by-treatment interaction model.<sup>12,13</sup> Contrasting the NMA model with the UME model regarding the treatment effect of all observed comparisons in the network offers an informal, exploratory investigation of the design inconsistency.

We aimed to propose a straightforward refinement of the UME model that accommodates the multiarm trials properly and yields treatment effects for all observed comparisons in networks with multiarm trials. Therefore, the proposed refinement allows the scrutiny of all observed evidence in the network to draw informed conclusions about the possibility of consistency. The article has the following structure. We first introduce 2 published systematic reviews with NMA as motivating examples. Then we present the Bayesian random-effects

---

Midwifery Research and Education Unit, Hannover Medical School, Hannover, Germany. The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided entirely by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, grant No. SP 1664/1-3). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

NMA and UME models—the latter as proposed by Dias et al.<sup>5</sup> (called the UME-Dias model). We follow by proposing a straightforward refinement of the Bayesian random-effects UME model (called the refined UME model). We demonstrate the refined UME model using the motivating examples. Finally, we conclude with a discussion of the results, strengths, and limitations of the proposed refinement in complex networks with multiarm trials.

### Motivating Examples

We considered 2 motivating examples: 1) the network of antimanic drugs for the mean change on mania rating scales,<sup>14</sup> and 2) the network of pharmacologic interventions for the exacerbation of chronic obstructive pulmonary disease (COPD).<sup>15</sup> Both networks included multiarm trials. The corresponding articles analyzed these networks in the standardized mean difference (SMD) and log odds ratio (OR) scales. Both outcomes were harmful; therefore, a negative SMD or log OR favored the first intervention in the comparison, and a positive SMD or log OR favored the second intervention. There was information on the number of missing (participant) outcome data in each arm of every trial for both networks. We excluded the missing outcome data from the analysis for illustrative purposes. Namely, we subtracted the number of missing outcome data from the number randomized in each arm of every trial. Methods to properly account for aggregate missing outcome data in NMA have been described elsewhere.<sup>16,17</sup>

### Methods

#### Random-Effects NMA Model with Multiarm Trials

For a network of  $N$  trials comparing different sets of interventions, we have the following information for the investigated outcome in each trial arm. For a binary outcome, we collect the number of events,  $r_{ik}$ , as reported in arm  $k$  of trial  $i$  out of the total randomized participants,  $n_{ik}$ . We assume that  $r_{ik}$  is sampled from a binomial distribution with an underlying probability of an event,  $p_{ik}$ . Then the underlying log odds in arm  $k$  of trial  $i$  is a function of the underlying log odds of the baseline arm,  $u_i$ , and the underlying log OR (the treatment effect),  $\delta_{ik1}$ ,

$$\text{logit}(p_{ik}) = u_i + \delta_{ik1}$$

where  $u_i = \text{logit}(p_{i1})$

For a continuous outcome, we extract the mean outcome,  $y_{ik}$ , as reported in arm  $k$  of trial  $i$  alongside the standard deviation,  $s_{ik}$ , as measured in the total randomized participants,  $n_{ik}$ . We assume a normal distribution

for  $y_{ik}$  with an underlying mean outcome,  $\theta_{ik}$ , and a variance,  $s_{ik}^2$  that is commonly assumed known, even though it has been estimated. Then, the underlying mean in arm  $k$  of trial  $i$  is a function of the underlying mean of the baseline arm,  $v_i$ , and the underlying SMD (the treatment effect),  $\delta_{ik1}$ ,

$$\theta_{ik} = v_i + \delta_{ik1}S_i$$

where  $v_i = \theta_{i1}$  and  $S_i$  is the pooled standard deviation in that trial,

$$S_i = \sqrt{\frac{\sum_{k=1}^{a_i} s_{ik}^2 (n_{ik} - 1)}{\sum_{k=1}^{a_i} (n_{ik} - 1)}}$$

with  $a_i$  being the number of arms in trial  $i$ .

Under the random-effects model,  $\delta_{ik1}$  is assumed to follow a normal distribution with mean  $\mu_{t_{ik}t_{i1}} = \mu_{t_{ikA}} - \mu_{t_{i1A}}$  (the consistency equation) and between-trial variance,  $\tau^2$ , assumed common within the network.<sup>18</sup> With  $t_{ik}$ , we indicate the intervention studied in arm  $k$  of trial  $i$ , and with  $A$ , we indicate the reference intervention of the whole network. The NMA model estimates the treatment effect of  $|T| - 1$  comparisons with the selected reference intervention (i.e.,  $\mu_{t_{ikA}}$  with  $t_{ik} \in T \setminus \{A\}$  and  $T = \{A, B, C, \dots\}$ ) and uses the consistency equation to obtain the remaining possible comparisons.<sup>18</sup>

Likewise, in a multiarm trial  $i$ , in which the consistency assumption is inherent, we estimate the treatment effects of  $a_i - 1$  comparisons with the baseline intervention of the trial. The  $(a_i - 1) \times 1$  vector of random effects,  $\delta_i = (\delta_{i21}, \delta_{i31}, \dots, \delta_{ia_i1})$ , is assumed to follow a multivariate normal distribution with  $(a_i - 1) \times 1$  mean vector  $\mu = (\mu_{t_{i2A}} - \mu_{t_{i1A}}, \mu_{t_{i3A}} - \mu_{t_{i1A}}, \dots, \mu_{t_{ia_iA}} - \mu_{t_{i1A}})'$  and  $(a_i - 1) \times (a_i - 1)$  variance-covariance matrix  $\Sigma = \tau^2 \begin{pmatrix} 1 & \dots & 0.5 \\ \vdots & \ddots & \vdots \\ 0.5 & \dots & 1 \end{pmatrix}$  (equation (10) in Dias et al.<sup>10</sup>),

which is equivalent to conditional univariate normal distributions of arm  $k > 2$  given the arms from 2 to  $a_i - 1$  (equation (11) in Dias et al.<sup>10</sup>),

$$\delta_{ik1} \mid \begin{pmatrix} \delta_{i21} \\ \vdots \\ \delta_{i(a_i-1)1} \end{pmatrix} \sim N \left( (\mu_{t_{ikA}} - \mu_{t_{i1A}}) + \frac{1}{a_i - 1} \sum_{j=1}^{a_i-1} [\delta_{ij1} - (\mu_{t_{jA}} - \mu_{t_{i1A}})] \right), \frac{k}{2(k-1)} \tau^2 \right).$$

By assuming a common  $\tau^2$  in the whole network, the correlation between any 2 random effects in the multiarm

trial equals 0.5.<sup>19</sup> These random effects refer to contrasts with the baseline arm of the multiarm trial. Choosing a different baseline arm will yield a different vector  $\delta_i$ .

#### *Random-Effects UME Model of Dias and Colleagues*

Contrary to the NMA model, the UME-Dias model does not pose consistency equations. Therefore, the UME-Dias model comprises separate random-effects pairwise meta-analyses for the observed comparisons.<sup>5</sup> The random-effects UME-Dias model also considers a shared between-trial variance across the observed comparisons to borrow strength from comparisons with many trials.<sup>5,11</sup> Hence, the estimation of between-trial variance is greatly improved.<sup>19</sup> Suppose the network includes a total of  $M$  observed comparisons with  $M \leq \binom{|T|}{2}$ .

Then, under the random-effects model, the random effects follow separate normal distributions with mean  $\mu_{t_{ik}t_{i1}}$  ( $t_{ik}, t_{i1} \in T$  and  $t_{ik} \neq t_{i1}$ ) for the observed comparisons and shared between-trial variance,  $\tau^2$ .

In the absence of multiarm trials, the UME-Dias model estimates the treatment effects of all observed comparisons. When the network includes multiarm trials, the UME-Dias model estimates the same vector  $\delta_i$  with the NMA model in multiarm trial  $i$ . However, contrary to the NMA model, the UME model treats the random effects in vector  $\delta_i$  as separate univariate normal distributions (see pp 651–2, appendix in Ref 5),

$$\delta_{ik1} \sim N(\mu_{t_{ik}t_{i1}}, \tau^2) \quad k \geq 2.$$

By making it similar to fitting separate pairwise meta-analyses to the data, the random-effects UME-Dias model retains its simplicity.<sup>11</sup> However, of the  $\binom{a_i}{2}$  possible comparisons in that multiarm trial, the UME-Dias model will not estimate a total of  $\binom{a_i}{2} - (a_i - 1)$  comparisons. These comparisons do not include the baseline arm. If these comparisons are not informed by any 2-arm trial and are not found in the vector  $\delta_i$  of other multiarm trials in the network, they will be omitted completely by the UME-Dias model. In the Bayesian framework, the posterior distribution of  $\mu_{jl}$  for the omitted comparisons will coincide with the prior distribution.<sup>11</sup> Hence, if a noninformative normal prior distribution with zero mean and variance equal to 10,000 is assigned on  $\mu_{jl}$ , the posterior standard deviation of  $\mu_{jl}$  will be approximately 100 for the omitted comparisons.

#### *Refined Random-Effects UME Model*

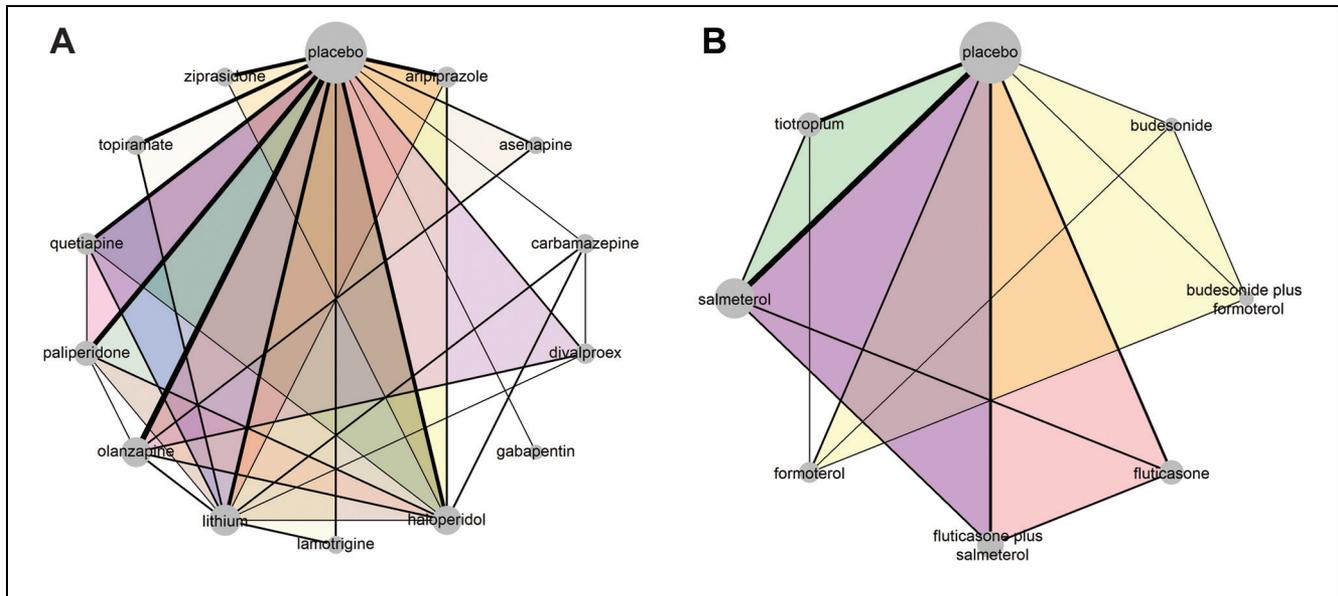
The UME-Dias model does not properly accommodate the multiarm trials. The random effects are inherently correlated in multiarm trials. Hence, as a matter of principle, the conditional univariate normal distributions of the random effects for the multiarm trials (equation (11) in Dias et al.<sup>10</sup>) should be maintained in the UME model. Dias et al.<sup>11</sup> also suggested accounting for the correlated effects in the multiarm trials; however, the authors did not formally implement this model in their book.

In our proposed refinement of the UME-Dias model, we have maintained the conditional univariate normal distributions of the random effects for the multiarm trials. We have also developed an algorithm to automatically detect pairwise comparisons in the multiarm trials that the UME-Dias model would omit. When there is at least 1 omitted comparison, we perform another random-effects NMA in the subset of multiarm trials, and we use the consistency equation to obtain the summary treatment effect of the omitted comparisons. When the subset of multiarm trials forms subnetworks rather than a fully connected network, we perform random-effects NMA with consistency equations in each subnetwork separately. To prevent the multiarm trials from contributing twice to the estimation of  $\tau^2$ , we consider a different  $\tau^2$  for the subset of multiarm trials, indicated by  $\tau_m^2$ . This “companion” model does not share any common parameter with the rest of the model. Therefore, using the multiarm trials twice does not affect the estimation of the parameters outside the companion model.

In the absence of multiarm trials, the refined UME model boils down to the UME-Dias model. In the presence of multiarm trials, comparisons can also be omitted when the fixed-effect model is considered.<sup>5</sup> In that case, the weighting approach proposed by Rucker and Schwarzer<sup>20</sup> can be used to prevent comparison omission. Dias et al.<sup>11</sup> discussed this weighting approach for the fixed-effect UME-Dias model in a Bayesian context.

#### *Model Implementation*

For each network, we applied Bayesian random-effects NMA, the refined UME model, and the UME-Dias model.<sup>5</sup> We considered a normal prior distribution with zero mean and variance 10,000 for the location parameters of the models, and we assigned a half-normal prior distribution with scale parameter 1 on  $\tau$  and  $\tau_m$ . We considered 3 chains of different initial values, 100,000 iterations with 10,000 burn-in and thinning equal to 5 and 20 for the network of antimanic drugs under the NMA and both UME models, respectively, and thinning



**Figure 1** The network of antimanic drugs for the mean change on mania rating scales<sup>14</sup> (plot A). The network of pharmacologic interventions for exacerbating chronic obstructive pulmonary disease<sup>15</sup> (plot B). Each node refers to an intervention, and each link refers to a pairwise comparison. The size of the nodes is proportional to the number of observed comparisons that include that node. The thickness of the edge is proportional to the number of trials that investigated that comparison. The colored intervention loops indicate multiarm trials.

equal to 10 and 20 for the network for COPD under the NMA and both UME models, respectively. We inspected the autocorrelation plots to define the values for the thinning. Convergence was assessed using the Gelman-Rubin convergence diagnostic, where  $\hat{R}$ :  $\hat{R} > 1.1$  indicates a lack of convergence for the corresponding model node.<sup>21</sup> We applied JAGS via the R-package *R2jags* (statistical software R, version 4.1.1) to run the models.<sup>22–24</sup>

We tabulated the posterior median and posterior standard deviation of  $\tau$ , the posterior mean of the residual deviance ( $\bar{D}$ ), the DIC, and the number of effective parameters under all 3 models. A larger DIC by 5 units for the NMA model or  $\bar{D}$  exceeding the number of data points under the NMA model indicates a possible inconsistency in the network. We created scatterplots and Bland-Altman plots on the posterior mean deviance contribution of the individual points to detect possible data points where the compared models have a poor fit. We considered the refined UME model to calculate the bias and limits of agreement in the Bland-Altman plot. We also obtained the leverage plots separately for the NMA and UME models to supplement our observations from the scatterplot and Bland-Altman plots. We used forest plots to illustrate the posterior mean and 95% credible interval of all observed comparisons under the NMA, the refined UME, and UME-Dias models. All figures

were created using the R-package *ggplot2*.<sup>25</sup> For the network plots, we used the R-package *pcnetmeta*.<sup>26</sup> All functions related to the present article can be found as supplementary material (Supplementary Material 1).

## Results

### Network of Antimanic Drugs

Figure 1A illustrates the network of 13 antimanic drugs and placebo. There were 17 (27%) 3-arm trials in the network. Of the 33 (36%) observed comparisons, 7 (21%) were informed solely by at least 1 multiarm trial and contained no baseline intervention. The UME-Dias model omitted these comparisons. The maximum value of  $\hat{R}$  across the parameters of all 3 models was 1.03, thus indicating convergence. All models yielded a posterior mean of residual deviance that exceeded the total number of 141 trial arms. The NMA model provided the largest  $\bar{D}$ , followed by the UME-Dias and refined UME models ( $\bar{D} = 157.82, 149.73, \text{ and } 146.67$ , respectively; Table 1). The DIC was similar for the NMA and UME-Dias model but lower by 2.54 units for the refined UME model (Table 1). The posterior median and 95% credible interval (CrI) of  $\tau$  were almost identical across the 3 models (Table 2). According to the DIC and the

**Table 1** Measures of Model Assessment for the NMA and UME Models

Model Assessment Measures	Network of Antimanic Drugs			Network for COPD		
	NMA	UME	UME-Dias	NMA	UME	UME-Dias
DIC	264.28	261.54	264.08	92.15	90.47	90.56
$p_D$	106.46	114.87	114.36	35.58	35.62	35.85
$\bar{D}$	157.82	146.67	149.73	56.58	54.85	54.70
Total trial arms	141			50		

COPD, chronic obstructive pulmonary disease;  $\bar{D}$ , posterior mean of the residual deviance; DIC, deviance information criterion; NMA, network meta-analysis;  $p_D$ , number of effective parameters; UME, unrelated mean effects.

estimated  $\tau$ , there was little to choose between a model with and without consistency.

Overall, including the consistency equations yielded similar posterior mean deviance contributions of the trial arms with discounting these equations and almost identical deviance contributions for both UME models (Figure 2A). An exception was 7 trial arms associated with larger deviance under the NMA model compared with the UME models that greatly exceeded the expected deviance contribution equal to 1. Of those trial arms, the arms of the 40th trial yielded a remarkably large posterior mean deviance contribution under all models.

In the Bland-Altman plot, the bias was slightly positive, indicating a tendency for the NMA model to yield slightly larger posterior mean deviance contributions on average than the refined UME model (Figure 2B). The trial arms associated with larger deviance under the NMA model were found outside the upper limit of agreements. These trial arms contributed 1 posterior mean deviance under the UME models (Figure 2A). Two of these points referred to 2-arm trials that were the sole contributors in divalproex versus carbamazepine (the 28th trial) and lithium versus divalproex (the 26th trial; Supplementary Table S1 in Supplementary Material 2). The remaining outlying points referred to haloperidol and lithium compared only in the multiarm trial (the 53rd trial), which included the omitted comparison of paliperidone versus lithium (Supplementary Table S1 in Supplementary Material 2). Supplementary Figure S1 (in Supplementary Material 2) located trials 28, 26, and 53 outside the red parabola (i.e.,  $x^2 + y = 3$ ) of the leverage plot for the NMA model only (plot C) and trial 40 (olanzapine versus lithium) outside the red parabola of the leverage plot for all models. These points contributed to a DIC an amount larger than 3.

Overall, there were similar posterior estimates and sufficient overlapping in the 95% CrIs of SMD for the NMA and either UME model except for the comparisons of divalproex versus carbamazepine and of lithium

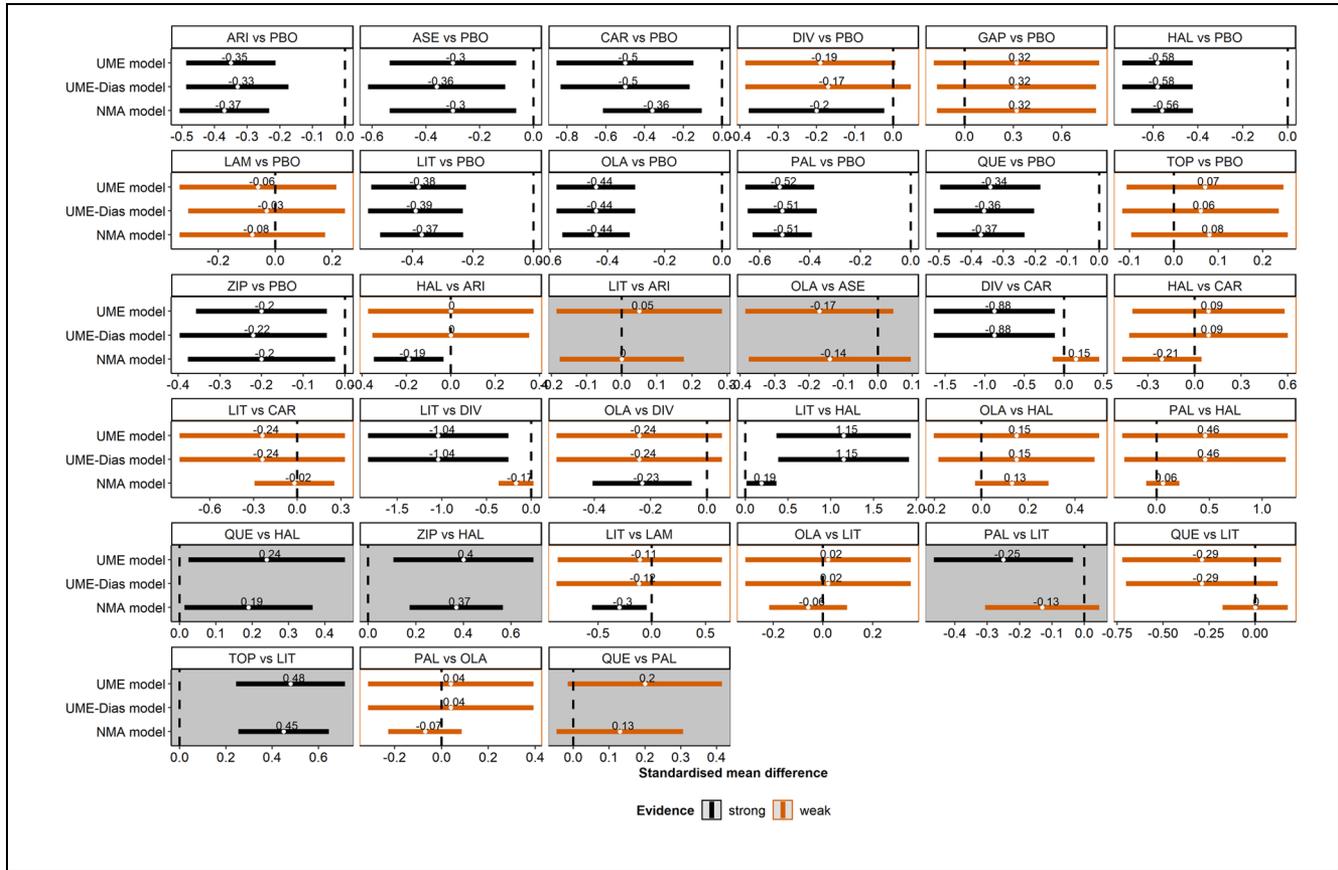
versus divalproex and haloperidol (Figure 3). These comparisons corresponded to trials that contributed to the poor fit of the NMA model (Figure 2; Supplementary Figure S1), thus signaling possible inconsistency in the network

#### Network of Pharmacologic Interventions for COPD

The network of 5 interventions (and their combinations) and placebo for COPD included 5 (24%) multiarm trials: 2 three-arm and 3 four-arm trials (Figure 1B). Of the 15 (54%) observed comparisons, 6 (40%) were omitted from the UME-Dias model. All models converged according to the  $\hat{R}$  diagnostic: the maximum value across the parameters of all models was 1.01. Both UME models yielded a similarly smaller posterior mean of residual deviance ( $\bar{D} \cong 55$ ) than the NMA model that exceeded the total number of 50 trial arms. The DIC indicated little to choose between these models (Table 1). The estimated  $\tau$  was substantial in all models but slightly larger and less precise under the NMA model (Table 2).

Overall, both UME models yielded similar posterior mean deviance contributions (Figure 4A). The scatterplot indicated a poor fit of the NMA model for 2 trial arms that exhibited substantial deviance under the NMA model and deviance close to 1 for both UME models (Figure 4A). These trial arms were flagged as outliers in the Bland-Altman plot as they exceeded the upper limit of agreement (Figure 4B). The 2 points referred to the unique trial that compared formoterol with tiotropium (Supplementary Table S2 in Supplementary Material 2). This trial was found outside the red parabola of the leverage plot for NMA (Supplementary Figure S2, plot C, in Supplementary Material 2) alongside the 2-arm trial 1 (0 events in fluticasone; Supplementary Table S2). Trial 1 also contributed to the poor fit of both UME models (Supplementary Figure S2, plots A and B) as both arms were associated with a substantial posterior mean deviance (Figure 4A).





**Figure 3** A panel of forest plots on the standardized mean difference (SMD) for all observed comparisons under the network meta-analysis (NMA), the refined unrelated mean effects (UME) model, and the UME-Dias model. Results refer to the posterior mean and 95% credible interval of the SMD. Gray panels refer to the omitted comparisons. Red and black indicate weak and strong evidence, respectively. Namely, the corresponding 95% credible interval includes and excludes the null value, respectively.

substantial deviance contribution of the corresponding trial to the poor fit of the NMA model (Figure 4), may suggest possible inconsistency in the network.

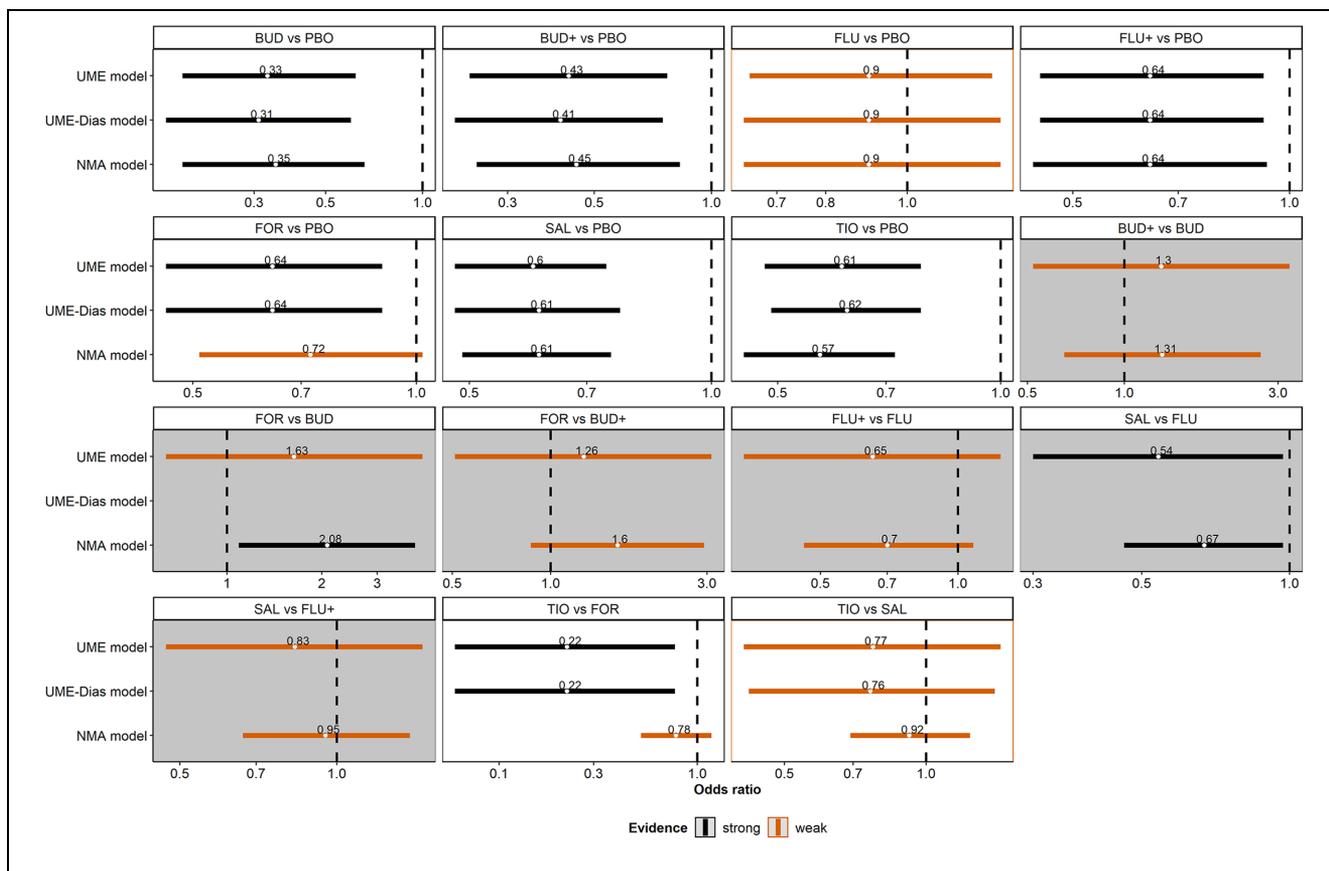
### Discussion

The refined UME model and the UME-Dias model gave the same conclusions regarding possible inconsistency in both examples. Both models elucidated the same data points that contributed to possible inconsistency in the investigated networks. They also yielded similar measures of model fit and almost identical estimated  $\tau$ . However, proper accommodation of the multiarm trials and visualization of all observed comparisons were possible only with the refined UME model, which are the main strengths of this model.

The present study considered a series of complementary plots to thoroughly investigate the individual data

points' deviance contribution, which laid the foundation for signaling possible inconsistency in both examples. Specifically, the scatterplot flagged the data points with higher posterior mean deviance than expected for the NMA model. The Bland-Altman plot complemented the scatterplot by detecting the outlying trial arms among those with substantial deviance contribution as they laid outside the 95% limits of agreement. Furthermore, this plot offers further exploratory insights that are not obvious by looking at the scatterplot. For instance, the lack of randomness in the scattered points and a nonzero bias may indicate a possible mismatch between the direct and NMA evidence that questions the whole evidence base. The leverage plot for NMA revealed that the outlying data points were found outside the red parabola, thus contributing substantially to the DIC (an amount larger than 3) and the model's poor fit. The panel of forest plots pinpointed poor overlapping in the 95% CrI of the treatment effects





**Figure 5** A panel of forest plots on the odds ratio (OR) for all observed comparisons under the network meta-analysis (NMA), refined unrelated mean effects (UME) model, and the UME-Dias model. The results refer to the posterior mean and 95% credible interval of the OR. Gray panels refer to the omitted comparisons. Red and black indicate weak and strong evidence, respectively. Namely, the corresponding 95% credible interval includes and excludes the null value, respectively.

limitation. The choice of parameterization will affect the derived omitted comparisons, because selecting a different baseline intervention for the multiarm trials will derive different omitted comparisons.<sup>5,11</sup> Suppose a connected network comprises an ABC, an AC, and an AB trial. Selecting intervention A to be the baseline arm in the ABC trial will yield the BC comparison as omitted. Selecting intervention B or C as the baseline arm in the ABC trial will not yield omitted comparisons. Suppose the network did not include an AB trial. Selecting intervention C as the baseline arm in the ABC trial will yield the AB as an omitted comparison. However, regardless of the selected parameterization, the refined UME model will estimate all observed comparisons of the network, contrary to the UME-Dias model.

Another limitation, common to the UME models, is that different parameterizations of the multiarm trials may affect the estimates and possibly the measures of

model fit.<sup>11</sup> Suppose there is also a BC trial in the example above, and we are interested in the BC comparison. Since the multiarm and 2-arm trials inform all comparisons, there are no omitted comparisons. However, selecting a different baseline arm will lead to one comparison being informed solely by the corresponding 2-arm trial. When A is the baseline arm in the ABC trial, the BC trial supports only the BC comparison. When B or C is the baseline arm, both ABC and BC trials contribute to the estimation of BC. The extent to which different parameterizations lead to different conclusions may also depend on the extent of between-trial variance. Different parameterizations may not affect the conclusions if the between-trial variance is low.

Useful clinical decisions can be made based only on models that assume the underlying evidence is consistent to ensure coherent estimates for a proper incremental assessment of benefits and costs. Models like the UME

that synthesize evidence without the consistency assumption are useful for assessing the feasibility of this assumption. When the consistency assumption is not deemed feasible, the whole evidence base should be called into question. Decisions should not be based on results from any syntheses that do not properly account for the reasons for this inconsistency.

## Conclusion

The proposed refinement of the UME model handles multiarm trials properly, and it yields treatment effects for all observed comparisons. A thorough inspection of the deviance contribution of the individual data points in conjunction with visualizing the posterior summaries of all observed comparisons under the NMA and refined UME models can aid our conclusions about possible global inconsistency in the network. In the presence of inconsistency, we should not be making inferences based on any of the models because they do not adequately describe the totality of the evidence available.

## Acknowledgements

We would like to thank 2 anonymous reviewers for their insightful comments that thoroughly improved the article.

## Research Data

Data supporting the findings of this study are available as supplementary material.

## ORCID iD

Loukia M. Spineli  <https://orcid.org/0000-0001-9515-582X>

## Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* website at <http://journals.sagepub.com/home/mdm>.

## References

- Zarin W, Veroniki AA, Nincic V, et al. Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. *BMC Med*. 2017;15(1):3.
- 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):236–63.
- Caldwell DM. An overview of conducting systematic reviews with network meta-analysis. *Syst Rev*. 2014;3:109.
- Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(2):80–97.
- Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making*. 2013;33(5):641–56.
- Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G. Characteristics of networks of interventions: a description of a database of 186 published networks. *PLoS One*. 2014;9(1):e86754.
- Petropoulou M, Nikolakopoulou A, Veroniki AA, et al. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *J Clin Epidemiol*. 2017;82:20–8.
- 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):683–91.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Series B Stat Methodol*. 2002;64:583–639.
- Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making*. 2013;33(5):607–17.
- Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. Checking for inconsistency. In: *Network Meta-Analysis for Decision Making*. Chichester (UK): Wiley; 2018, pp 189–226.
- Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods*. 2012;3(2):98–110.
- White IR, Barrett JK, Jackson D, Higgins JP. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Synth Methods*. 2012;3(2):111–25.
- Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *Lancet*. 2011;378(9799):1306–15.
- Baker WL, Baker EL, Coleman CI. Pharmacologic treatments for chronic obstructive pulmonary disease: a mixed-treatment comparison meta-analysis. *Pharmacotherapy*. 2009;29(8):891–905.
- Spineli LM, Kalyvas C, Papadimitropoulou K. Continuous(ly) missing outcome data in network meta-analysis: a one-stage pattern-mixture model approach. *Stat Methods Med Res*. 2021;30(4):958–75.
- Spineli LM. An empirical comparison of Bayesian modeling strategies for missing binary outcome data in network meta-analysis. *BMC Med Res Methodol*. 2019;19(1):86.
- Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc*. 2006;101:447–59.
- Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15(24):2733–49.

20. Rucker G, Schwarzer G. Reduce dimension or reduce weights? Comparing two approaches to multi-arm studies in network meta-analysis. *Stat Med.* 2014;33(25):4353–69.
21. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci.* 1992;7(4):457–72.
22. Plummer M. *JAGS: Just Another Gibbs Sampler*. Version 4.3.0 user manual. 2017. Available from: [https://people.stat.sc.edu/hansont/stat740/jags\\_user\\_manual.pdf](https://people.stat.sc.edu/hansont/stat740/jags_user_manual.pdf)
23. Su YS, Yajima M. R2jags: Using R to Run 'JAGS'. R package version 0.6-1. 2020. Available from: <https://CRAN.R-project.org/package=R2jags>
24. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna (Austria): R Foundation for Statistical Computing; 2021. Available from <https://www.r-project.org>
25. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2016.
26. Lin L, Zhang J, Hodges JS, Chu H. Performing arm-based network meta-analysis in R with the pcnetmeta Package. *J Stat Softw.* 2017;80:5.
27. Higgins JP. Commentary: heterogeneity in meta-analysis should be expected and appropriately quantified. *Int J Epidemiol.* 2008;37(5):1158–60.